



**БИБЛИОТЕКА
ТЕХНИЧЕСКОЙ
КИБЕРНЕТИКИ**

Б. С. ФЛЕЙШМАН
**ЭЛЕМЕНТЫ ТЕОРИИ
ПОТЕНЦИАЛЬНОЙ
ЭФФЕКТИВНОСТИ
СЛОЖНЫХ СИСТЕМ**

„СОВЕТСКОЕ РАДИО“

БИБЛИОТЕКА
ТЕХНИЧЕСКОЙ
КИБЕРНЕТИКИ

Б. С. ФЛЕЙШМАН

**ЭЛЕМЕНТЫ ТЕОРИИ
ПОТЕНЦИАЛЬНОЙ
ЭФФЕКТИВНОСТИ
СЛОЖНЫХ СИСТЕМ**

125360



ИЗДАТЕЛЬСТВО
«СОВЕТСКОЕ РАДИО»
МОСКВА—1971

Систематически излагаются известные и новые результаты теории потенциальной эффективности сложных систем, устанавливающей пределы их надежности, помехоустойчивости, управляемости и самоорганизации.

Конструктивные результаты этой теории указывают направление разработкам сложных технических систем, близких к потенциально эффективным. Эти же результаты могут помочь в выявлении структуры и поведения биологических систем.

Изложение сопровождается графиками, облегчающими использование приведенных формул. Книга рассчитана на научных работников, инженеров-системотехников и биологов, а также студентов и аспирантов соответствующих специальностей.

Рис. 21, табл. 3, библи. 54 назв.

Члены редакционного совета

Трапезников В. А. (председатель), Челюсткин А. Б. (зам. председателя), Буселко Н. П., Виленки С. Я., Воронов А. А., Гаазе-Рапопорт М. Г., Дудников Е. Г., Ицкович Э. Л., Круг Г. К., Мамиконов О. Г., Осойков И. О., Пархоменко П. П., Пинскер М. С., Плискин Л. Г., Поспелов Г. С., Райбман Н. С., Самойленко С. И., Таль А. А., Флейшман Б. С., Хургин Я. И., Якобсон Б. М.

БЕНЦИОН СЕМЕНОВИЧ ФЛЕЙШМАН

Элементы теории потенциальной эффективности сложных систем

Редактор *А. Н. Ковалев*

Художественный редактор *З. Е. Вендрова*

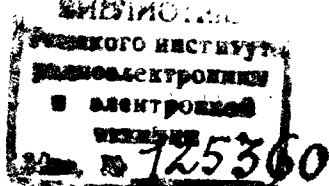
Технический редактор *З. И. Яковлева*

Корректоры *Е. П. Озерецкая, Н. М. Кухтяева*

Сдано в набор 8/VII 1971 г.	Подписано в печать 28/X 1971 г.	T. 17734
Формат 84×108/32	Вумага типографская № 2	Объем 11,76 усл. п. л.
Уч.-изд. л. 12,350	Тираж 6 600 экз.	Зак. 275
Издательство «Советское радио», Москва, Главпочтамт, п/я 693.		Цена 74 коп.

Московская типография № 10 Главполиграфпрома
Комитета по печати при Совете Министров СССР, Москва
Шлюзовая наб., 10.

3-3-14
47-71



Памяти моих родителей

ПРЕДИСЛОВИЕ

Общая теория систем фон Берталанфи по его замыслу должна была быть междисциплинарной наукой, исследующей как материальные, так и абстрактные системы. Кибернетика Н. Винера по его первоначальному замыслу должна была исследовать более конкретную область сложных технических и биологических систем.

Ограниченность возможностей таких рассмотрений приводит к тому, что фактически удается оперировать либо с широкими классами относительно простых моделей, получая общие результаты, либо с узкими классами сложных моделей, получая частные результаты.

В связи с этим теорию сложных систем можно условно разделить на три направления: теорию потенциальной эффективности сложных систем, исследование операций и эвристическое программирование.

Теория потенциальной эффективности сложных систем содержит системную интерпретацию известных результатов теории информации Шеннона, теории игр фон Неймана, а также новые результаты (например, в теории надежности и живучести сложных систем). Исторически первым результатом этого направления была теория потенциальной помехоустойчивости В. А. Котельникова [1]. Так же как теория дискретных автоматов [2], теория потенциальной эффективности сложных систем исследует проблемы оптимального синтеза их простых моделей. Однако в отличие от первой вторая теория носит не структурный, а поведенческий характер. Следует отметить, что в настоящее время наибольших успехов добились в исследовании операций [3], анализирующем модели средней сложности. Существенные результаты в решении задач эвристического программирования [4],

изучающего сложные модели имитации поведения сложных систем, можно ожидать в ближайшем будущем.

Эта книга посвящена теории потенциальной эффективности сложных систем, основной особенностью которой является изучение не того что есть, а того, что может быть [5]. Возможность таких исследований достигается рассмотрением классов простых симметричных моделей систем, определенных небольшим числом параметров, и выделением в них оптимальных систем. Рассмотрение простых моделей реально сложных систем не снижает объективности выводов о предельных возможностях последних, если считать идеальными те качества реальной системы, которые игнорируются в модели.

Основные результаты формулируются в виде простых предельных соотношений (законов), ограничивающих эффективность соответствующих моделей систем. Сами по себе эти законы полезны тем, что исключают попытки решения задач типа «перпетуум мобиле». Конструктивные формулировки этих законов могут быть использованы инженерами и биологами. Первыми — для построения эффективных технических систем, вторыми — для изучения биологических систем, подошедших близко к оптимальному пределу за время эволюции.

В книге используются разрозненные публикации, в том числе и авторские, часть результатов публикуется впервые. При отборе материала автор стремился охватить все известные ему результаты в этой области. Элементарность изложения достигается за счет рассмотрения этих результатов в простейших случаях. Как правило, лишь в этих случаях удается получить простые расчетные формулы.

Книга состоит из введения и семи глав, содержащих подобные определения, формулировки доказываемых предложений, простые доказательства и графики, облегчающие расчеты. Сложные выкладки выведены в приложения для облегчения чтения основного текста.

Выражаю благодарность Б. Р. Левину, В. Д. Зубакову и С. Ю. Рудерману за ценные замечания, учтенные автором при написании книги.

1. Введение

§ 1.1. СИСТЕМЫ

Системой называют совокупность элементов с определенными связями между ними (в отличие от конгломерата). Индивидуальность системы проявляется в ее структуре (относительно постоянной во времени морфологии) и поведении (функционирование во времени), которые взаимно обусловлены. Так, сложности структуры системы, как правило, сопутствует сложность ее поведения. Поэтому можно характеризовать сложность системы сложностью ее поведения, т. е. разнообразием реакций на внешние воздействия.

В порядке возрастания сложности поведения можно различать следующие типы реальных систем: автоматические, решающие, самоорганизующиеся, предвидящие и превращающиеся.

Автоматические системы способны лишь детерминированно реагировать на внешние воздействия. Внутренняя организация их приспособлена к переходу в равновесное состояние при выводе из него. Примерами таких систем могут служить: солнечная система, атом, часы, аппарат поддержания постоянства состава крови и др.

Решающие системы имеют постоянные стохастические критерии различения случайных сигналов и постоянные стохастические реакции на случайные воздействия (им присущ акт решения). Постоянство их структуры поддерживается заменой вышедших из строя элементов. Примерами таких систем могут служить: локатор, рецепторные механизмы организмов и др.

Самоорганизующиеся системы имеют гибкие критерии различения сигналов и гибкие реакции на воздействия, приспособляющиеся к заранее неизвестным сигналам и воздействиям. Устойчивость высших форм

гаких систем обеспечивается постоянным самовоспроизведением (структурно-информационная устойчивость). Примерами таких систем могут служить: модели типа перцептрона, кибернетические игрушки, простейшие организмы и др.

На определенном уровне сложности самоорганизующиеся системы приобретают столь мощную память и устойчивость, что сложность их поведения начинает превосходить сложность воздействий индифферентного внешнего мира. Такого рода системы можно называть *предвидящими*. Помня исходы взаимодействия с внешним миром до данного момента, такая система может «предвидеть» дальнейший ход взаимодействия, полагаясь на повторение прежних ситуаций. Примерами таких систем могут служить высшие животные и человек.

Наконец, можно говорить о системах, не связанных постоянством вещественных носителей своей устойчивой структуры. Такие системы можно назвать *превращающимися*. Проблема устойчивости сводится у них к проблеме помехоустойчивого перекодирования. Примеры таких систем неизвестны.

Важнейшим проявлением сложности поведения можно считать акт решения, т. е. свободный выбор альтернатив, например, с помощью бессознательного случайного механизма. Поэтому системы, не превосходящие по сложности автоматические, включая их самих, будем называть *простыми*, а системы, превосходящие по сложности автоматические, начиная с решающих, будем называть *сложными*.

Большими системами будем называть совокупности разнородных сложных систем со сравнительно слабыми связями между ними, например современные человеко-машинные комплексы.

Формальные математические понятия системы, ее простоты и сложности, структуры, поведения, цели и др. имели бы оправдание, если бы опирающийся на них аксиоматический подход выводил законы теории сложных систем в виде теорем. Отсутствие такого подхода делает, по-видимому, математическую формализацию основных понятий не столь уж актуальной.

Вместе с тем на естественно-научном уровне все эти понятия, как правило, воспринимаются достаточно однозначно. Так, например, отличие простых систем от сложных обнаруживается при следующих рассуждениях.

Низшие формы сложных систем отличаются от высших форм простых систем качественно более резкими проявлениями следующих свойств: открытости (взаимодействие с внешней средой), иерархичности структуры, стохастичности поведения и изменчивости во времени.

В зачаточной форме все эти свойства содержатся и у простых систем. Например, диффузионные процессы на поверхностях любых тел, соприкасающихся с активной средой, иерархическая структура атомов, молекул и тел, стохастичность существования электрона и, наконец, изменчивость тел, связанная с их образованием и старением.

Однако все эти свойства у сложных систем проявляются в качественно новой, более интенсивной форме, что видно, например, из сопоставления простой системы — камня — со сложной системой — собакой. У последней открытость связана с обменом веществ, иерархичность обогащается рангами клеток, тканей и органов, стохастичность связана со сложной психикой поведения, а изменчивость — с ростом, старением, размножением и смертью.

Между простыми неживыми и сложными живыми системами находятся сложные неживые системы, которые образуют как бы переходную ступень между ними. Они обладают рядом отдельных качеств сложных живых систем. Это и позволило обнаружить общие количественные законы, управляющие отдельными качествами сложных систем, при изучении именно сложных неживых технических систем.

§ 1.2. ДВА ПОДХОДА К ПОСТРОЕНИЮ МОДЕЛЕЙ

Реальная сложная система $A^*(t)$, вообще говоря, может быть описана большим числом N^* параметров $\bar{x}(t) = \{x_i(t)\}$ ($i = \overline{1, N^*}$)*, зависящих от времени t (точнее, N^* функциями времени), с $L^* < N^*$ зависимостями между ними

$$G_j[x(t)] = c_j(t) \quad (j = \overline{1, L^*}). \quad (1.1)$$

При заданных параметрах $\bar{c}(t) = \{c_j(t)\}$ система имеет, вообще говоря, $N^* - L^*$ степеней свободы (независимых параметров).

*) Символ $i = \overline{1, N}$ означает $i = 1, 2, \dots, N$.

Моделью $A(t)$ реальной системы $A^*(t)$ называется абстрактное образование, описываемое *лишь* $N < N^*$ параметрами $\tilde{x}(t) = \{x_{ir}(t)\}$ ($r = \overline{1, N}$) с учетом лишь $L < L^*$ зависимостей:

$$G_{j_s}[\tilde{x}(t)] = c_{j_s}(t) \quad (s = \overline{1, L}). \quad (1.2)$$

При заданных параметрах $\tilde{c}(t) = \{c_{j_s}(t)\}$ модель $A(t)$ имеет $N - L$ степеней свободы. Точки $\bar{x}(t)$ и $\tilde{x}(t)$ с ростом t описывают некоторые траектории в N^* -мерном евклидовом пространстве. Чем ближе N к N^* и L к L^* , тем меньше «рассогласование» этих траекторий, измеряемое любой естественной мерой, например максимумом модуля разности соответствующих координат.

Все реальные системы $A^*(t)$ [модели $A(t)$], параметры которых удовлетворяют связям (1.1), (1.2), образуют классы $\mathfrak{A}^*(\mathfrak{A})$.

Пусть имеются некоторые целевые функционалы

$$v^* = F[\bar{x}(t)] \quad \text{и} \quad v = F[\tilde{x}(t)] \quad (1.3)$$

и совокупности параметров $\bar{x}_0(t)$ и $\tilde{x}_0(t)$ обращают их в экстремумы (максимумы или минимумы в зависимости от их смысла) при условиях (1.1) и (1.2) соответственно.

Систему $A^*_0(t)$ и модель $A_0(t)$, соответствующие экстремальным значениям параметров, будем называть *оптимальными*.

Решение экстремальных задач приводит к зависимостям:

$$\bar{x}_0(t) = H^*[\tilde{c}(t)] \quad \text{и} \quad \tilde{x}_0(t) = H[\tilde{c}(t)],$$

из которых следует, что оптимальная система $A^*_0(t)$ и модель $A_0(t)$ при различных параметрах $\tilde{c}(t)$ и $\bar{c}(t)$ имеют L^* и L степеней свободы соответственно.

Имеется два принципиально различающихся подхода в выборе моделей $A(t)$ для реальных систем $A^*(t)$.

Первый (аналитический) подход состоит в приближении модели $A(t)$ к реальной системе $A^*(t)$ за счет приближений N к N^* и L к L^* (анализ реальной системы).

Второй (синтетический) подход рассматривает лишь реальные системы $A^*(t)$, близкие к своим оптимальным

прототипам $A^*_0(t)$. Он начинается с синтеза оптимальных моделей $A_0(t)$, приближающихся к прототипам $A^*_0(t)$ за счет приближения L к L^* и F к F^* .

Аналитический и синтетический подходы используют модели $A(t)$ и $A_0(t)$ с $N - L$ и L степенями свободы, и они тем адекватней действительности, чем ближе N и L к N^* и L^* соответственно. При $L^* \ll N^*$ имеем $L^* \ll N^* - L^*$ и требуемая адекватность выглядит реальнее достижимой для синтетического подхода, чем для аналитического (реальнее возможность преодоления «проклятия многомерности» [6]).

Однако использование второго подхода по сравнению с первым усложнено введением понятия оптимальности.

Исследование операций, как правило, использует аналитический подход в построении моделей сложных систем, теория потенциальной эффективности сложных систем — синтетический.

§ 1.3. ОПТИМАЛЬНОСТЬ В ТЕРМИНАХ (u, v) -ОБМЕНА [7]

В дальнейшем изложении, в отличие от реальных систем: технических (техсистем) и биологических (биосистем), их модели будем называть просто *системами*, а модель среды будем называть просто *средой*.

Система A при фиксации ее цели \underline{A} определяется своей структурой $|A|$ и поведением \bar{A}

$$A = (|A|, \bar{A}).$$

Целью системы является достижение определенного предпочтительного состояния.

Рассматривается открытая система, т. е. система, взаимодействующая со средой B . Вообще говоря, можно рассматривать среду как систему B со своей целью \underline{B} , структурой $|B|$ и поведением \bar{B} :

$$B = (|B|, \bar{B}).$$

Поведение систем A и B , вообще говоря, могут способствовать или препятствовать достижению чужих целей. В первом случае пару систем можно рассматривать как одну систему с общей целью, взаимодействующую со средой. Во втором случае будем говорить о *конфликтной* ситуации. Если не имеет места ни тот, ни другой случай, то будем говорить об *индифферентной* ситуации и лишь в этом случае одну систему по отношению к другой будем называть *средой*. В последнем случае среда

может оказывать непреднамеренное мешающее воздействие на систему. Это воздействие обычно называют шумом.

Функционирование открытых систем можно представить себе как серию обменов некоторыми количествами u расходуемых ресурсов на некоторые количества v потребляемых ресурсов. Будем называть это (u, v) -обменом.

Вещественно-энергетические (u, v) -обмены в индифферентной ситуации описываются законами физики и химии, а в конфликтной ситуации — теорией игр (см. гл. 4). Сигнальные (u, v) -обмены в индифферентной ситуации описываются законами теории информации (см. гл. 3). Дальнейшее изложение касается лишь законов (u, v) -обмена, впервые возникающих на уровне решающих систем и выходящих за рамки физико-химических законов. Поэтому, в частности, здесь не рассматривается принцип обратной связи, возникающий на уровне автоматических систем и основанный на известных законах физики.

Если характеризовать состояние системы величинами пары параметров u и v , то целью системы может являться выгодный (u, v) -обмен. Под последним можно понимать получение как можно большего количества v за фиксированное количество u или, что то же, отдача как можно меньшего количества u за фиксированное количество v . Последнее зависит от величины u , структуры и поведения систем A и B :

$$v = v(u, A, B).$$

Как правило, взаимодействие систем A и B носит стохастический характер и поэтому можно говорить лишь о некоторой вероятности $P(u, v)$ достижения системой A своей цели \underline{A} (выгодного (u, v) -обмена) $P(u, v) = P(\underline{A})$.

В теории потенциальной эффективности сложных систем в качестве целевого функционала удобно рассматривать величину $P(\underline{A}) = P(u, v)$, а не величину v по причинам, о которых будет сказано ниже. Впрочем, в ряде практически интересных случаев, которые будут рассмотрены далее, оптимальные системы, соответствующие этим различным целевым функционалам, совпадают.

Замечательным фактом является то, что для всех известных конкретных моделей решающих систем, приво-

дящих к предельным законам их эффективности, все эти законы могут быть выражены в единой форме через вероятность $P(u, v)$.

Именно при больших значениях u (им соответствуют большие значения v) величина $P(u, v)$ имеет следующее асимптотическое поведение:

$$P(u, v) \approx \begin{cases} 0 & \text{при } v > v_0, \\ 1 & \text{при } v < v_0, \end{cases} \quad (1.4)$$

где фундаментальная константа v_0 определяется следующими соотношениями:

$$v_0 = \begin{cases} v(u, A_0, B) = \max_{A \in \mathfrak{A}} v(u, A, B) & \text{в индифферентной ситуации,} \\ v(u, A_0, B_0) = \max_{A \in \mathfrak{A}} \min_{B \in \mathfrak{B}} v(u, A, B) & \text{в конфликтной ситуации,} \end{cases} \quad (1.5)$$

A_0 и B_0 оптимальные системы из достаточно широких классов \mathfrak{A} и \mathfrak{B} (в случае отсутствия \max и \min берутся \sup и \inf соответственно).

Аналогично при фиксированных u и v определяются оптимальные системы \bar{A}_0 и \bar{B}_0 , вообще говоря, отличные от A_0 и B_0 при оптимальном целевом функционале $P(u, v)$.

Соотношения (1.4) и (1.5) можно рассматривать как общую форму предельного закона (u, v) -обмена для решающей системы. Закон устанавливает пределы ее эффективности (она не может рассчитывать получить величину $v > v_0$, где v_0 определяет ее потенциальную эффективность). Впрочем терминологически удобно эффективностью называть саму вероятность $P(u, v) = P(\underline{A})$, так как в теории надежности она называется надежностью, в теории потенциальной помехоустойчивости — помехоустойчивостью, в то время как величина v_0 связана в теории информации с пропускной способностью, а в теории игр с величиной игры. Как будет показано в следующем параграфе, такое определение имеет и принципиальное оправдание.

Общее асимптотическое поведение (1.4) вероятности $P(u, v)$ для разнообразных моделей, рассмотренных в этой книге, не случайно. Оно связано с одной глубокой математической закономерностью, лежащей в основе асимптотического поведения всех этих моделей. Речь

идет о вероятностных закономерностях так называемых больших уклонений. В простейшей форме эта закономерность состоит в следующем.

Рассмотрим сумму n независимых случайных величин:

$$\mu = \sum_{v=1}^n \varepsilon_v,$$

каждая из которых принимает два значения 1 и 0 с вероятностями p_v и $q_v = 1 - p_v$ соответственно. Тогда вероятность $\mathcal{P}(\mu \leq m) (m = \bar{0}, \bar{n})$ имеет вид

$$\mathcal{P}(\mu \leq m) = \sum_{\sum_{v=1}^n \varepsilon_v \leq m} p_v^{\varepsilon_v} q_v^{1-\varepsilon_v}. \quad (1.6)$$

Имеют место следующие оценки вероятности $\mathcal{P}(\mu \leq m)$ (см. приложение 1):

$$\left. \begin{array}{l} \exp[-\tilde{k}n + O(\ln n)] < \\ \text{при } \theta < \tilde{p}, \\ 1 - \exp[-\tilde{k}n + O(\ln n)] > \\ \text{при } \theta > \tilde{p} \end{array} \right\} \mathcal{P}(\mu \leq m) \left\{ \begin{array}{l} < \exp(-\bar{k}n) \\ \text{при } \theta < \bar{p}, \\ > 1 - \exp(-\bar{k}n) \\ \text{при } \theta > \bar{p}, \end{array} \right. \quad (1.7)$$

где $\theta = m/n$, $\tilde{p} \leq p_v \leq \bar{p} \leq 0,5$, $\tilde{p} = \tilde{p}/(1 - \tilde{p} + \tilde{p})$;

$$\bar{p} = \frac{1}{n} \sum_{v=1}^n p_v, \bar{k} = k(\theta, \bar{p}), \tilde{k} = k(\theta, \tilde{p}) \text{ и} \\ k(\theta, p) = \theta \ln \frac{\theta}{p} + (1 - \theta) \ln \frac{1 - \theta}{1 - p} \geq 0. \quad (1.8)$$

В частности, при постоянных $p_v = p (q_v = q = 1 - p)$, когда вероятность $\mathcal{P}(\mu \leq m)$ имеет вид биномиального распределения,

$$\mathcal{P}(\mu \leq m) = \sum_{r=0}^m C_n^r p^r q^{n-r}, \quad (1.9)$$

при больших значениях n , кроме оценок имеют место асимптотические равенства

$$\left. \begin{array}{l} \exp[-kn + O(\ln p)] \\ \text{при } \theta < p \\ 1 - \exp[-kn + O(\ln p)] \\ \text{при } \theta > p \end{array} \right\} = \mathcal{P}(\mu \leq m) \left\{ \begin{array}{l} < \exp(-kn) \\ \text{при } \theta < p, \\ > 1 - \exp(-kn) \\ \text{при } \theta > p, \end{array} \right. \quad (1.10)$$

где $k = k(\theta, p)$.

В рассматриваемом случае $m = \theta n$ имеет с ростом n порядок $O(n)$. Отклонения случайной величины μ от своего среднего $n\bar{p}$, превосходящие эту величину, имеют экспоненциально убывающую с ростом n вероятность (1.7) и называются большими.

Отклонения случайной величины μ от своего среднего $n\bar{p}$, превосходящие величину $m = O(\sqrt{n})$, оцениваются «хвостами» нормального распределения и их можно называть нормальными.

Большинство физических вероятностных законов имеет дело с нормальными уклонениями.

§ 1.4. ПРОБЛЕМА ЧАСТИ И ЦЕЛОГО

Целостная система A обладает различными качествами, множество которых мы будем обозначать $\mathfrak{X} = \{X\}$, а его элементы X называть X -качествами. Например, будем называть надежность R -качеством, помехоустойчивость — I -качеством, управляемость — C -качеством, а самоорганизацию — L -качеством.

В соответствии с этим общая стратегическая цель системы A — выжить — имеет по X -качествам свои конкретные тактические проявления $A_X (A - \{A_X\})$. При этом в целостной системе $A = (|A|, \bar{A})$ вычлняются подсистемы $A_X = (|A_X|, \bar{A}_X)$, которые мы будем называть X -сечениями системы A .

Фиксируя пару качеств — XU -качество, можно аналогично определить тактическую цель A_{XU} -и XU -сечения $A_{XU} = (|A_{XU}|, \bar{A}_{XU})$ и т. д.

Тактическим целям A_X, A_{XU}, \dots соответствуют (u_X, v_X) , (u_{XU}, v_{XU}) , ... обмены.

Возникает проблема оценки качества целостной системы по качеству ее X - или XU -сечений. Основное преимущество определения целевого функционала в форме эффективности $P(u, v) = P(\underline{A})$ перед другими определениями проявляется при решении этой проблемы. Именно на основании неравенства Буля, например для X -сечений, имеем следующие оценки (см. приложение 2)

$$1 - \sum_{X \in \mathfrak{X}} [1 - P(\underline{A}_X)] \leq P(\underline{A}) \leq \min_{X \in \mathfrak{X}} P(\underline{A}_X). \quad (1.11)$$

Оценки (1.11) повсеместно используются в этой книге.

§ 1.5. ЗАКОНЫ СЛОЖНЫХ СИСТЕМ

По мере возрастания сложности систем у них возникают новые качества, сохраняющиеся для более сложных систем. Поэтому разнообразие качеств системы возрастает с ее сложностью. В настоящее время хорошо изучены общие законы, управляющие автоматическими и более простыми системами (физические законы). Хуже изучены общие законы, управляющие решающими и более сложными системами. Все эти законы вместе с качествами, которых они касаются, как упоминалось выше, сохраняются для систем более сложных по сравнению с теми, где они впервые возникают (рис. 1.1).

Например, упоминавшийся принцип обратной связи (обозначен на рис. 1.1 O -законом), впервые возникнув на уровне автоматических систем, присущ всем сложным системам. То же можно сказать и о X -законах, возникающих впервые на уровне решающих систем (рис. 1.1). Это законы, связанные с их надежностью (R -закон), помехоустойчивостью (I -закон), управляемостью (C -закон) и самоорганизацией (L -закон). Последнее качество в полной мере должно было бы проявиться у самоорганизующихся систем. Однако в зачаточной форме оно могло бы проявиться и у решающих систем.

Указанная выше последовательность рассмотрения отдельных качеств системы в порядке возрастания их активности принята в книге из-за того, что каждое последующее качество имеет смысл только при наличии предыдущих.

В самом деле, если нет устойчивого образования из связанных между собой элементов (качество надежности), то не имеет смысла рассматривать какие-либо другие качества этого образования (системы).

Далее проявляется такое качество системы, как ее помехоустойчивость, связанное с правильной ориентацией системы в среде (правильная информированность о ней). Очевидно, что без наличия последнего качества

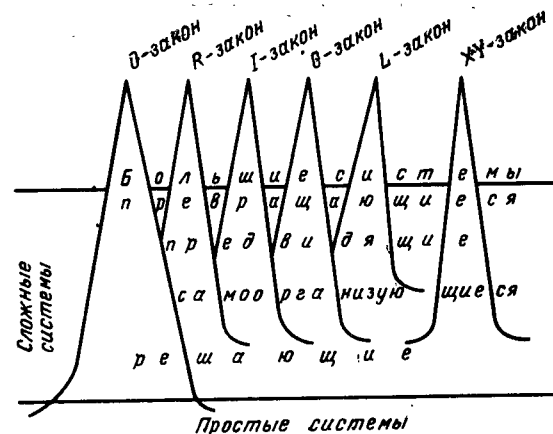


Рис. 1.1. Системы и управляющие ими законы.

любые действия системы (управляемость) неэффективны. И, наконец, такое качество, как самоорганизация, является просто накоплением опыта использования всех перечисленных качеств во времени, его переработки и коррекции.

Внимание к перечисленным качествам сложных систем исторически было обращено не в указанной выше последовательности. Как уже упоминалось, раньше других были найдены предельные законы помехоустойчивости. В этом проявилась авангардная роль техники связи во всей новой технике.

Далее были найдены предельные законы активных действий систем в конфликтных ситуациях (управляемости) на основе успехов теории игр, стимулировавшей военно-экономической проблематикой. И, наконец, в качестве проблемы «номер один» начала фигурировать проблема надежности в связи с возросшей сложностью технических систем.

Первоначально предельные законы сложных систем формулируются как теоремы существования. Формулировки их конструктивных вариантов представляют значительные трудности и до сих пор неизвестны для ряда предельных законов. Известные результаты, связанные с предельными законами для отдельных X-качеств, приводятся в первых четырех главах, следующих за введением. Основанные на них результаты, содержащиеся в двух последних главах, касаются попыток совместного рассмотрения отдельных качеств целостной системы.

В этом отношении особое значение имеет последняя, седьмая глава, посвященная проблеме осуществимости. Эта проблема занимает в рассматриваемой теории не только основное практическое, но и принципиальное теоретическое место.

В самом деле, между теорией потенциальной эффективности сложных систем и математикой существует аналогия. Модели и законы первой имеют характер аксиом и теорем второй соответственно. Однако между ними имеется существенная разница в требованиях, предъявляемых каждой к своим результатам. В то время как к результатам математики предъявляются лишь логические требования и, в первую очередь, требование их логической непротиворечивости, к результатам теории потенциальной эффективности сложных систем предъявляется еще и требование их осуществимости. Последнее требование труднее формализовать, чем, например, требование логической непротиворечивости. Однако на естественно-научном уровне понятие осуществимости можно определить так, что им можно пользоваться в инженерных расчетах, что и иллюстрируется в последней, седьмой главе.

2. Надежность (R-качество)

§ 2.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Устойчивость сложной системы во времени является необходимым качеством, без которого теряют смысл все ее другие проявления. Следует различать два вида устойчивости сложных систем: вещественно-энергетическую и структурно-функциональную.

Первый вид устойчивости связан с постоянством вещественного состава и энергетического баланса сложной

системы. Второй вид устойчивости связан с постоянством структуры системы и постоянством ее реакций на одни и те же внешние воздействия.

В зависимости от сложности системы каждый из указанных видов устойчивости является определяющим. Для автоматических систем, таких, например, как атом или молекула, наблюдается устойчивость первого вида. Целостность атома и особенно его ядра может быть нарушена лишь огромными немеханическими силами. Целостность молекулы может быть нарушена сравнительно меньшими силами, связанными с тепловыми, химическими и механическими воздействиями. Искусственные сооружения (от искусственных молекул до новых материалов и зданий) обладают устойчивостью, связанной с понятием прочности. Этим же качеством обладают машины. Детали современных машин взаимозаменяемы. Поэтому машина уже обладает качеством постоянства структуры при переменном вещественном составе.

Однако в полную меру вторым видом устойчивости обладают решающие системы в их естественном варианте, например простейшие живые клетки.

Клетка в еще большей степени, чем молекула, является «хрупким» образованием по отношению к физико-химическим воздействиям. Однако ее устойчивость по сравнению с «неживыми» молекулярными образованиями обнаруживает новое качество — постоянство структуры при переменном вещественном составе (обмен веществ). В отличие от прочности это качество называют надежностью.

Для более сложных биологических систем, таких, как многоклеточные организмы, клетки играют ту же роль, что и молекулы для клеток.

Здесь основным видом устойчивости является надежность. Клеточный состав организмов является, в основном, переменным. Имеются специальные органы воспроизведения новых и удаления отмирающих клеток. Устойчивой остается лишь структура организма. Следует заметить, что с ростом сложности надежность биологических систем не имеет тенденции падать (клетки, вообще говоря, менее надежны, чем организмы).

В качестве искусственной сложной системы с зачатками «обмена веществ» может служить электронная вычислительная машина с частой заменой отдельных элементов.

Однако с ростом сложности технических систем их надежность падает. Это связано с несовершенством их организации. На определенном уровне сложности автоматических и в полную меру на уровне решающих систем возникает качественно новый вид структурной устойчивости (надежность), связанный с обменом веществ. По мере возрастания сложности биологических систем возрастает их надежность, что указывает на аналогичную возможность и для технических систем.

Последующее изложение касается не прочности, а надежности сложных систем — определяющем виде их устойчивости.

§ 2.2. ИДЕАЛЬНАЯ СИСТЕМА

2.2.1. Отказ элемента и отказ системы. Система A_R определяется набором n ее элементов $A_R = (a_1, \dots, a_v, \dots, a_n)$ и структурой их соединения (взаимодействия). Например, в такой системе, как муравейник, в качестве элементов можно принять отдельных муравьев. В такой системе, как ЭВМ, в качестве элементов можно принять отдельные ее детали.

Каждый элемент системы с точки зрения нормальности функционирования может находиться в различных состояниях, промежуточных между состояниями нормального функционирования и полного выхода из строя.

Рассмотрим два крайних случая: нормальное функционирование (жизнь) и выход из строя (смерть). Выход из строя (смерть) элемента будем называть *отказом*. Отказ происходит в определенный момент времени. Будем упрощенно считать, что переход элемента из состояния жизни в состояние смерти происходит мгновенно.

Итак, формально будем считать, что элементы находятся в каждый данный момент лишь в двух состояниях: жизни или смерти. Причем переход из первого состояния во второе происходит мгновенно, а обратные переходы невозможны.

Выход из строя одного из нескольких элементов, вообще говоря, не влечет за собой выхода из строя всей системы. Имеются жизненно важные и менее важные элементы системы. Кроме того, могут быть параллельно работающие элементы, выполняющие одну и ту же функцию. Выход из строя всех таких элементов, кроме

одного, заметно не нарушает нормальной работы системы. Например, выход из строя одной из двух почек не выводит из строя весь организм. С другой стороны, выход из строя хотя бы одного из последовательно соединенных элементов какого-нибудь технического (или биологического) тракта приводит к выходу из строя всего тракта.

В общем случае естественно считать систему живой в данный момент, если жива хотя бы некоторая заранее заданная минимальная часть ее элементов. В противном случае будем считать ее мертвой (отказ системы) или потенциально мертвой, так как оставшиеся в живых элементы не обеспечивают долговременной жизни системы и последняя постепенно погибает. Таким периодом угасания мы будем пренебрегать.

Множество практически неконтролируемых причин, обуславливающих отказы элементов, приводит к тому, что появление отказов можно считать случайным событием. Экспериментальные данные подтверждают такое допущение.

2.2.2. Вероятностная модель отказа элемента и системы (статическое рассмотрение). Пусть в данный момент τ система A_R состоит из n элементов $\{a_1, \dots, a_v, \dots, a_n\}$ с произвольными, вообще говоря, связями.

Будем считать, что с каждым элементом a_v связана пара взаимно исключающих друг друга событий A_1^v и A_0^v . Событие A_1^v будем считать „благоприятным“ и — связывать с жизнью элемента a_v , событие A_0^v — „неблагоприятным“ и связывать со смертью a_v (отказом). Пусть вероятности событий A_1^v и A_0^v равны p_v и $q_v = 1 - p_v$, соответственно.

Рассмотрим множество R_n всех 2^n возможных состояний $\rho = (i_1, \dots, i_v, \dots, i_n)$ системы A_R , характеризующих состояния жизни и смерти ее элементов ($i_v = 1$, если a_v в состоянии A_1^v , $i_v = 0$, если a_v в состоянии A_0^v).

Разобьем множество R_n на две части \mathcal{E}_1 и $\mathcal{E}_0 = R_n \setminus \mathcal{E}_1$. Первое будем называть множеством жизни системы A_R , а

второе — множеством смерти. По определению будем считать, что система A_R жива в данный момент, если $\rho \in \mathcal{E}_1$, и мертва, если $\rho \in \mathcal{E}_0$.

Будем считать, что состояния ρ системы A_R можно рассматривать как последовательность независимых испытаний с вероятностями исходов $p_v = \mathcal{P}(i_v = 1)$, $q_v = 1 - p_v = \mathcal{P}(i_v = 0)$ ($\mathbf{v} = \overline{1, n}$) каждого v -го испытания.

Тогда вероятность P_v жизни системы A_R в данный момент имеет вид

$$P_v = \mathcal{P}(\rho \in \mathcal{E}_1) = \sum_{\rho \in \mathcal{E}_1} \prod_{v=1}^n p_v^{i_v} q_v^{1-i_v} \quad (2.1)$$

и вероятность смерти (отказа) системы равна

$$Q_v = 1 - P_v.$$

Рассмотрим крайние случаи соотношения (2.1). Пусть множество жизни \mathcal{E}_1 состоит из одного состояния $\mathcal{E}_1 = \rho_1 = \overbrace{(1, 1, \dots, 1)}^n$. Другими словами, рассматривается случай, когда система жива тогда и только тогда, когда живы все ее элементы. Этому случаю соответствует их последовательное соединение.

В этом случае из (2.1) следует, что вероятности P_v и Q_v имеют вид:

$$P_v = \prod_{v=1}^n p_v, \quad Q_v = 1 - \prod_{v=1}^n p_v. \quad (2.2)$$

Пусть теперь множество смерти \mathcal{E}_0 состоит из одного состояния $\mathcal{E}_0 = \rho_0 = \overbrace{(0, 0, \dots, 0)}^n$. Другими словами, рассматривается случай, когда система мертва тогда и только тогда, когда мертвы все ее элементы. Это соответствует их параллельному соединению.

В этом случае из (2.1) следует, что вероятности P_v и Q_v имеют вид:

$$P_v = 1 - \prod_{v=1}^n q_v, \quad Q_v = \prod_{v=1}^n q_v. \quad (2.3)$$

Из соотношений (2.2) и (2.3) видно, что с ростом числа n элементов системы A_R вероятность P_v жизни системы уменьшается в первом случае и возрастает во втором.

Нас интересуют системы с большим числом n элементов. Поэтому естественно рассматривать асимптотический случай больших значений n . При рассмотрении крайних случаев множества жизни \mathcal{E}_1 следует, что с ростом n вероятность P_v может стремиться к нулю и единице. Очевидно, что этим свойством могут обладать и промежуточные множества жизни \mathcal{E}_1 . Естественно классифицировать системы A_R , определяемые своими фиксированными множествами жизни \mathcal{E}_1 , по тому, стремится ли соответствующая вероятность P_v жизни к нулю или единице с ростом числа n элементов.

В первом случае будем называть такие системы *нежизнеспособными*, а во втором — *жизнеспособными*.

В общем случае множества жизни \mathcal{E}_1 могут иметь самую произвольную структуру, по которой трудно судить о жизнеспособности данной системы.

Однако имеется широкий класс систем, в которых правильная структура множеств жизни позволяет судить о жизнеспособности таких систем.

В самом деле, рассмотрим системы, состоящие из n элементов, множество \mathcal{E}_1 жизни которых состоит из состояний, содержащих каждое более чем m живых элементов. Другими словами, в множество \mathcal{E}_1 входят все состояния

$\rho = (i_1, \dots, i_v, \dots, i_n)$ системы, для которых $\sum_{v=1}^n i_v > m$. Со-

ответственно множество смерти \mathcal{E}_0 состоит из всех состояний системы, для каждого из которых $\sum_{v=1}^n i_v \leq m$.

Такого рода системы будем называть *симметричными* системами m -го порядка. Симметричной системой является любой коллектив индивидуумов, отдельные представители которого выполняют совместно однородные действия. При этом минимальное их число для совместного выполнения жизненно важной функции системы равно $m+1$. Каждый индивидуум может иметь свою, отличную от других вероятность смерти q_v . Например, $m=0$ для любой многоклеточной популяции с вегетатив-

ным размножением, если основной функцией системы является выживание.

Если в противоборстве с более сильными индивидуумами требуется по меньшей мере $m+1$ равносильных индивидуумов данной популяции, то мы снова имеем пример симметричной системы, имеющей цель выживание.

Для ряда несимметричных систем симметричные системы могут рассматриваться как приближенные модели. Например, всегда существует некоторое минимальное число членов популяции с половым размножением, обязательно содержащее хотя бы одну самку или самца, для которой не происходит окончательного вырождения. Имеется ряд технических примеров такого же рода, однако здесь мы не будем на них останавливаться.

Для симметричных систем формула (2.1) имеет вид:

$$P_v = \sum_{\sum_{v=1}^n i_v > m} \prod_{v=1}^n p_v q^{1-i_v}, \quad (2.4)$$

Легко видеть, что формулы (2.2) и (2.3) являются частным случаем формулы (2.4) при $m=n-1$ и $m=0$ соответственно.

Для установления жизнеспособности симметричной системы m -го порядка необходимо проанализировать асимптотическое поведение выражения (2.4) при $n \rightarrow \infty$. Для этого достаточно воспользоваться оценкой (1.7).

В самом деле, введем обозначения $\theta = m/n$, $\bar{p} = \frac{1}{n} \sum_{v=1}^n p_v$, $p \leq p_v \leq \tilde{p} < 0,5$ и будем называть θ коэффициентом

смерти системы, а \bar{p} — средней вероятностью жизни элемента. Тогда согласно оценке (1.7) имеем

$$P_v \begin{cases} > 1 - \exp(-\bar{k}n) & \text{при } \theta < \bar{p}, \\ < 1 - \exp[-\tilde{k}n + O(\ln n)] & \text{при } \theta < p, \\ < \exp(-\bar{k}n) & \text{при } \theta > \bar{p}, \end{cases} \quad (2.5)$$

где $\bar{k} = k(\theta, \bar{p})$, $\tilde{k} = k(\theta, p)$ и функция $k(\theta, p)$ определяется

соотношением (1.8):

$$\bar{p} = \frac{1}{n} \sum_{v=1}^n p_v \quad \text{и} \quad \tilde{p} = \frac{\bar{p}}{(1 + \tilde{p} - \bar{p})}. \quad (2.6)$$

Таким образом, симметричная система m -го порядка жизнеспособна, если ее коэффициент смерти θ меньше средней вероятности жизни \bar{p} ее элементов и она нежизнеспособна при $\theta > \bar{p}$.

Ясно, что фактически могут существовать лишь жизнеспособные системы.

2.2.3. Надежность системы. Идеальная система (динамические рассуждения). Итак, ограничимся рассмотрением лишь жизнеспособных систем. Обладая большим, но конечным числом элементов, они имеют близкую к единице «мгновенную» вероятность $P_v(\tau)$ жизни в момент τ . Однако эта вероятность не равна (несколько меньше) единице и если система функционирует длительное время, то дожитие ее до некоторого отдаленного момента может оказаться маловероятным. Для точного анализа этого обстоятельства рассмотрим следующую динамическую модель симметричной жизнеспособной системы переменного порядка $m(\tau)$, которую будем называть просто симметричной.

Пусть в дискретные моменты времени $\tau = 1, 2, \dots, t$, с интервалом, принятым за единицу, система A_R имеет $n(\tau)$ элементов и коэффициент смерти $\theta(\tau) = \frac{m(\tau)}{n(\tau)} < \bar{p}(\tau)$,

где $\bar{p}(\tau)$ — средняя вероятность жизни ее элементов в различные моменты времени τ . Тогда согласно соотношению (2.5) вероятность $P_v(\tau)$ жизни системы в момент τ имеет следующие оценки:

$$P_v(\tau) \begin{cases} > 1 - \exp[-\bar{k}_\tau n(\tau)] & \text{при } \theta(\tau) < \bar{p}(\tau), \\ < 1 - \exp[-\tilde{k}_\tau n(\tau) + O(\ln n(\tau))] & \text{при } \theta(\tau) < p(\tau), \end{cases} \quad (2.7)$$

где $\bar{k}_\tau = k[\theta(\tau), \bar{p}(\tau)]$; $\tilde{k}_\tau = k[\theta(\tau), p(\tau)]$, а величины $\bar{p}(\tau)$, $p(\tau)$ и функция $k(\theta, p)$ определяются соотношениями (2.6) и (1.8) соответственно.

Будем считать, что если система A_R погибла в какой-либо из моментов времени τ , то ничто уже не может возродить ее в последующие моменты времени. То есть будем считать, что для системы, как и для ее элементов, из состояния смерти нельзя перейти в состояние жизни.

Нас интересует вероятность $P_R(t)$ того, что система A_R доживет до некоторого момента t (включительно). Учитывая предыдущее допущение, для этого необходимо, чтобы система была жива во все моменты времени — от первого до t -го включительно. Эту вероятность $P_R(t)$ называют *надежностью* системы A_R . Если схематизировать рассматриваемую модель последовательностью независимых испытаний с вероятностями $P_v(\tau)$ и $Q_v(\tau) = 1 - P_v(\tau)$ жизни и смерти соответственно, то надежность $P_R(t)$ имеет вид:

$$P_R(t) = \prod_{\tau=1}^t P_v(\tau). \quad (2.8)$$

Из соотношения (2.8) ясно, что с ростом времени t надежность системы не увеличивается.

Если учитывать, что с ростом $n(\tau)$ сомножители $P_v(\tau)$ в выражении (2.8) согласно оценке (2.7) стремятся к единице, то очевидны возможности различных предельных значений $P_R(t)$ с ростом t .

Для точного выяснения этого вопроса рассмотрим некоторую *идеальную систему* A_R , обладающую следующими постулируемыми свойствами.

1. *Свойство жизнеспособности.* Система A_R жизнеспособна во все моменты времени. Однако если ее постигла смерть в данный момент времени, то в последующие моменты времени ничто не может возродить ее к жизни.

2. *Свойство неограниченного расширения.* Если система A_R жива в данный момент времени, то в следующий момент времени она может пополниться любым числом элементов.

3. *Физическая ограниченность времени реакции.* О своем состоянии в данный момент система узнает лишь в следующий момент, отстоящий от данного на интервал, принятый за единицу.

4. *Математические ограничения.* Рассматривается симметричная система с независимыми исходами «испытаний» жизни и смерти в каждый данный момент.

Обсудим постулируемые свойства системы.

Первое свойство требует жизнеспособности системы во все моменты времени. Ясно, что если бы система A_R была нежизнеспособна хотя бы для одного момента времени τ , то, дожив до него, система далее имела бы надежность $P_R(t)$, стремящуюся к нулю.

Этот момент можно интерпретировать моментом внезапной смерти системы из-за изменения ее качества (она в этот момент стала нежизнеспособной, для симметричной системы это означает неравенство $\theta(\tau) > \bar{p}(\tau)$).

Мы исключаем такой случай. Система может погибнуть, оставаясь жизнеспособной лишь в силу случайных обстоятельств, так как, например, для симметричной системы всегда имеется хотя и малая, но все же отличная от нуля вероятность $Q_v(\tau) = 1 - P_v(\tau)$ ее гибели и после этого согласно первому свойству ничто не может ее возродить.

Согласно второму свойству среда предоставляет живой системе неограниченное число элементов для их пополнения и это является для системы единственным средством в борьбе за сохранение своей жизни.

Таким образом, природа (внешняя среда) индифферентна к судьбе системы A_R . Она не возрождает ее в случае смерти, но если система жива, то представляет в ее распоряжение неограниченное число элементов для поддержания жизнеспособности. В терминологии, принятой во введении, имеет место индифферентная ситуация между системой A_R и средой B .

Третье свойство идеальной системы избавляет нас от тривиального ее вырождения. В самом деле, если бы система могла мгновенно узнавать о своем состоянии, то, используя второе свойство, она могла бы тут же пополняться нужным ей числом элементов и этим всегда предотвращала бы смертельный исход. Третье свойство исключает такую возможность, устанавливая ограниченность времени реакции системы на свое собственное состояние величиной интервала времени, принятого за единицу. Это ограничение имеет физическое оправдание, так как не существует реальных физических процессов, проходящих с неограниченно большими скоростями (процесс не может происходить со скоростью, например, большей скорости света).

Наконец, четвертое свойство идеальной системы носит математический характер. Оно накладывает на иде-

альную систему некоторые ограничения, при которых возможен математический расчет надежности такой системы.

§ 2.3. ПРЕДЕЛЬНЫЙ ЗАКОН НАДЕЖНОСТИ

2.3.1. Формулировка и доказательство предельного закона надежности. В предыдущем пункте упоминалось, что надежность $P_R(t)$ идеальной симметричной системы с ростом t не может возрастать. Однако предел $\lim_{t \rightarrow \infty} P_R(t)$ может иметь в зависимости от выбора $n(\tau)$ различные значения.

Приступим теперь к выяснению предельного поведения надежности $P_R(t)$ при $t \rightarrow \infty$. Имеет место следующий

Предельный закон надежности [8]. Пусть в момент τ идеальная симметричная система A_R имеет среднюю вероятность жизни элемента $\bar{p}(\tau)$, коэффициент смерти $\theta(\tau)$ и число элементов $n(\tau)$. Тогда вид зависимости числа элементов $n(\tau)$ от момента времени τ определяют следующие оценки надежности $P_R(t)$ системы A_R :

$$P_R(t) \begin{cases} \geq 1 - \frac{1}{a} (1 - t^{-a}), \\ \leq \frac{2O(\ln t)}{1+t}, \\ = \exp(-\lambda t), \end{cases} \quad \text{при } n(\tau) \begin{cases} \geq \frac{1+a}{\bar{k}} \ln(\tau+1), \\ \leq \frac{1}{\tilde{k}} \ln(\tau+1), \\ = n = \text{const} \end{cases} \quad (2.9)$$

соответственно (a — произвольное положительное число), числа \bar{k} и \tilde{k} определены соотношениями

$$\bar{k} = \min_{1 \leq \tau \leq t} \bar{k}_\tau, \quad \tilde{k} = \max_{1 \leq \tau \leq t} \tilde{k}_\tau, \quad (2.10)$$

числа \bar{k}_τ и \tilde{k}_τ имеют вид:

$$\bar{k}_\tau = k[\theta(\tau), \bar{p}(\tau)], \quad \tilde{k}_\tau = k[\theta(\tau), \underline{p}(\tau)],$$

числа $\bar{p}(\tau)$ и $\underline{p}(\tau)$ определены соотношениями (2.6), а константа λ заключена в пределах

$$-\ln[1 - \exp(-\tilde{k}n) + O(\ln n)] \leq \lambda \leq -\ln[1 - \exp(-\bar{k}n)].$$

Доказательство. Оценим выражение надежности (2.8) с учетом соотношений (2.7) и неравенств

$$1 - \sum_{j=1}^n x_j \leq \prod_{j=1}^n (1 - x_j) \leq \exp\left(-\sum_{j=1}^n x_j\right), \quad (2.11)$$

верных для $0 \leq x_j \leq 1$. Нижняя оценка легко доказывается по индукции, верхняя следует из того, что $1 - x_j \leq \exp(-x_j)$. Из (2.8) и (2.11) имеем

$$P_R(t) \begin{cases} \geq 1 - \sum_{\tau=1}^t [1 - P_v(\tau)], \\ \leq \exp\left\{-\sum_{\tau=1}^t [1 - P_v(\tau)]\right\}. \end{cases} \quad (2.12)$$

Используя соотношение (2.7), из соотношения (2.12) имеем

$$R_R(t) \begin{cases} \geq 1 - \sum_{\tau=1}^t \exp[-\bar{k}_\tau n(\tau)], \\ \leq \exp\left\{-\sum_{\tau=1}^t \exp[-\tilde{k}_\tau n(\tau) + O(\ln n(\tau))]\right\}. \end{cases}$$

Положим

$$\bar{k} = \min_{1 \leq \tau \leq t} \bar{k}_\tau \quad \text{и} \quad \tilde{k} = \max_{1 \leq \tau \leq t} \tilde{k}_\tau,$$

тогда из предыдущих оценок получим оценки

$$P_R(t) \begin{cases} \geq 1 - \sum_{\tau=1}^t \exp[-\bar{k}n(\tau)], \\ \leq \exp\left\{-\sum_{\tau=1}^t \exp[-\tilde{k}n(\tau) + O(\ln n(\tau))]\right\}, \end{cases} \quad (2.13)$$

верные при любых значениях $n(\tau)$. Рассмотрим их поведение отдельно при значениях

$$n(\tau) \begin{cases} \geq \frac{1+a}{\bar{k}} \ln(\tau+1), \\ \leq \frac{1+a}{\tilde{k}} \ln(\tau+1) \end{cases} \quad (2.14)$$

соответственно, где a произвольное число,

Тогда, используя (2.13) и (2.14), получим

$$P_R(t) \begin{cases} \leq \exp \left[- \sum_{\tau=1}^t (\tau+1)^{-(1+a)} + O(\ln t) \right], \\ \geq 1 - \sum_{\tau=1}^t (\tau+1)^{-(1+a)}. \end{cases} \quad (2.15)$$

Заметим, что из геометрического представления интеграла площадью следуют известные неравенства:

$$\begin{aligned} \frac{1}{2} [2^{-a} - (1+t)^{-a}] &= \int_2^{t+1} \frac{dx}{x^{1+a}} \leq \sum_{\tau=1}^t (\tau+1)^{-(1+a)} = \\ &= \sum_{\tau=2}^{t+1} \tau^{-(1+a)} \leq \int_1^t \frac{dx}{x^{1+a}} = \frac{1}{a} (1-t^{-a}). \end{aligned} \quad (2.16)$$

Так как предел $\lim_{a \rightarrow 0} \frac{1}{a} [2^{-a} - (1+t)^{-a}] = \ln \frac{1+t}{2}$ и оцениваемая сумма (2.16) монотонно убывают с ростом a , то имеем оценки

$$\ln \frac{1+t}{2} \leq \sum_{\tau=1}^t (1+\tau)^{-(1+a)} \leq \frac{1}{a} (1-t^{-a}), \quad (2.17)$$

левая для значений $a \leq 0$ и правая — для значений $a > 0$. Подставив (2.17) в (2.15), получим первые две оценки (2.9). Далее для случая $n(\tau) = n = \text{const}$ имеем из (2.8), используя оценки (2.7) и введенные величины \underline{k} и \tilde{k}

$$\prod_{\tau=1}^t [1 - \exp(-\underline{k}_\tau n)] \leq P_R(t) \leq \prod_{\tau=1}^t \{1 - \exp[-\tilde{k}_\tau n + O(\ln n)]\}$$

или

$$[1 - \exp(-\underline{k}n)]^t \leq P_R(t) \leq [1 - \exp(-\tilde{k}n) + O(\ln n)]^t. \quad (2.18)$$

Отсюда, обозначая $-\lambda = \frac{\ln P_R(t)}{t}$, видим после логарифмирования оценок (2.18), что величина $-\lambda$ заключена в пределах

$$\ln [1 - \exp(-\underline{k}n)] \leq -\lambda \leq -\ln [1 - \exp(-\tilde{k}n) + O(\ln n)].$$

Таким образом, предельный закон надежности доказан полностью.

Заметим, что при $t \rightarrow \infty$ из предельного закона (2.9) имеем

$$P_R(\infty) = \lim_{t \rightarrow \infty} P_R(t) \begin{cases} \geq 1 - a^{-1} & \text{при } n(\tau) \geq \frac{1+a}{\underline{k}} \ln(\tau+1), \\ = 0 & \text{при } n(\tau) \leq \frac{1}{\tilde{k}} \ln(\tau+1). \end{cases} \quad (2.19)$$

Таким образом, задав достаточно большим значением a величину $1 - a^{-1}$ сколь угодно близкой к единице и определив этим минимальный порядок роста числа элементов $n(\tau)$ со временем $\frac{1+a}{\underline{k}} \ln(\tau+1) = n(\tau)$, мы можем гарантировать значение надежности $P_R(t)$ системы A_R для значений t от 0 до ∞ , не меньшее чем $1 - a^{-1}$. Итак, вопрос о предельном поведении надежности $P_R(t)$ с ростом t выяснен нами полностью.

2.3.2. Обсуждение предельного закона надежности
Предельные соотношения (2.18) делают естественным разбиением совокупности \mathfrak{A} всевозможных идеальных систем A_R на два принципиально различных класса в зависимости от поведения их надежности $P_R(t)$ с ростом t . К классу $\mathfrak{A}_R^{(0)}$ будем относить идеальные системы, у которых с ростом t надежность $P_R(t)$ стремится к нулю. Согласно предельному закону надежности рост числа элементов $n(\tau)$ таких систем с ростом τ не превосходит логарифмического роста $\frac{1}{\tilde{k}} \ln(\tau+1)$, где коэффициент \tilde{k} зависит от вероятностей жизни их элементов. Такого рода системы будем называть *смертными*, а их класс $\mathfrak{A}_R^{(0)}$ — классом смертных систем.

К классу $\mathfrak{A}_R^{(1)}$ будем относить идеальные системы, у которых с ростом t надежность $P_R(t)$, уменьшаясь, стремится к пределу $P_R(\infty) > 0$. Согласно предельному закону надежности с ростом t рост числа элементов $n(t)$ таких

систем не должен быть меньше логарифмического роста $\frac{1+a}{k} \ln(t+1)$, где коэффициент k зависит от вероятностей жизни их элементов.

Выбор достаточно большого значения a согласно (2.19) обеспечивает величине $P_R(\infty)$ значение $1-a^{-1}$, сколь угодно близкое к единице. Таким образом, согласно (2.19) практически такие системы на промежутке времени от $t=1$ до $t=\infty$ не будут подвержены смерти. Такие системы будем называть *бессмертными*, а их класс $\mathfrak{A}_R^{(1)}$ — классом бессмертных систем. Выше было показано, что бессмертными могут быть лишь идеальные системы с возрастающим, по меньшей мере логарифмически со временем, числом элементов (расширяющиеся системы). В частности, любые идеальные системы с ограниченным (не зависящим от времени t) числом элементов n смертны (ограниченные системы).

Можно сказать, что система A_R осуществляет (t, n) -обмен (см. введение). При этом имеет место оптимальный $(t, C_R \ln t)$ -обмен, где константа C_R зависит от вероятности отказов элементов.

Если схематизировать отдельные организмы и их популяции идеальными системами, то первые, будучи ограниченными, оказываются смертными. Вторые, размножаясь со временем, как правило, быстрее, чем по логарифмическому закону, оказываются бессмертными. Оба эти вывода отражают известные биологические факты.

Предельный закон не дает никаких указаний, касающихся конструктивного описания идеальной системы A_R . В нем ничего не говорится о том, как обеспечивается ее жизнеспособность, т. е. как она остается живой по выходе из строя части ее элементов (построение надежных систем из ненадежных элементов), успевает ли она выявить часть разлагающихся элементов до выхода их из строя (методы поиска неисправностей, прогноза и заблаговременной замены), каковы способы ассимиляции ею новых элементов и какова плата за них и т. д. Все эти вопросы остаются открытыми.

Таким образом, предельный закон надежности указывает лишь пределы надежности идеальной системы A_R , достижимые в случае возможности некоторого неизвестного ее жизнеспособного построения. Посмотрим, что может ограничивать возможность такого построения.

В самом деле, для того чтобы быть идеальной, система не должна быть сверхестественной. Всякий раз в интервале времени от τ -го до $\tau+1$ -го моментов времени она должна пересмотреть достаточно большую часть своих элементов и, учтя вышедшие из строя, пополнить их до величины $C_R \ln(\tau+1)$. Возможность распознания качества элементов при наличии мешающих факторов (шумов) ограничены фундаментальными информационными пределами (см. гл. 3). Возможности пополнения элементов, особенно в случае конфликтных ситуаций, также имеют фундаментальные ограничения (см. гл. 4).

Таким образом, после того как доказано существование идеальной системы, как только дело касается ее построения, сразу же мы приходим к необходимости решения проблем, связанных с фундаментальными ограничениями таких качеств системы, как ее помехоустойчивость (информированность) и управляемость, которые будут рассмотрены в следующих главах.

Наконец, остановимся на обсуждении случая конечной системы, состоящей из фиксированного числа элементов $n = \text{const}$. Здесь согласно соотношению (2.9) имеет место известный экспоненциальный закон надежности

$$P_R(t) = \exp(-\lambda t), \quad (2.20)$$

где параметр λ , как легко показать, имеет смысл величины, обратной среднему времени между отказами системы. Этот параметр в теории надежности иногда называют *опасностью отказа* (с его ростом уменьшается среднее время между отказами).

Конструктивные исследования здесь пока ограничиваются рассмотрением конечных систем.

§ 2.4. ИЕРАРХИЧЕСКАЯ СТРУКТУРА ПОИСКА НЕИСПРАВНОСТЕЙ

2.4.1. Физические и функциональные отказы.

Установленный выше предельный закон надежности гарантирует сколь угодно близкую к единице надежность лишь для расширяющихся систем. При этом в нем не содержится каких-либо указаний относительно способов поддержания такой высокой надежности системы, кроме беспредельного пополнения ее элементами. Установление минимального порядка роста числа n элементов системы со временем t для поддержания ее надежности, сколь угодно близкой к единице, — лишь в этом состоит содержа-

ние предельного закона надежности — типичной теоремы существования оптимальной по надежности системы $A_R = (|A_R|, \bar{A}_R)$, осуществляющей (t, n) -обмен, сколь угодно близкий к оптимальному $(t, C_R \ln t)$ -обмену.

Попытки конструктивных исследований в этой области можно начать с более детального изучения механизма поиска неисправных элементов и их пополнения. Вопросы поиска неисправностей носят информационный характер (см. гл. 3). Здесь основные трудности возникают при учете шумов, мешающих различать исправные и неисправные элементы. Вопросы рационального пополнения элементных ресурсов связаны с задачами оптимального управления (см. гл. 4). В самом деле, всякое пополнение системы жизненно важными для нее элементами, имеющимися в ограниченном количестве, может встретить противодействие других заинтересованных в них систем, что приводит к конфликтным ситуациям. Таким образом, кажется, что конструктивная теория здесь не имеет своей специальной проблематики. Однако, как будет показано далее, это не так.

В настоящее время наиболее актуальными являются проблемы продления жизни смертных (конечных) систем.

Дальнейшее изложение целиком посвящено рассмотрению лишь конечных систем A_R , состоящих из конечного числа n элементов. Следует различать две части поведения \bar{A}_R -системы: внешнюю — \bar{A}'_R и внутреннюю — \bar{A}''_R . Первая связана с упоминавшимся уже добыванием элементов (или ресурсов для их создания) и поэтому ее можно включить в виде составной части поведения $\bar{A}_C = \{\bar{A}'_R, \dots\}$ управленческого сечения A_C системы A . Далее внешняя часть поведения \bar{A}'_R не рассматривается. Внутренняя часть поведения A_R связана с локализацией неисправностей и их исправлением. Вместе со структурой $|A_R|$ сечения A_R она будет исследована в дальнейшем изложении. Упоминавшаяся информационная окраска этих вопросов имеет достаточную специфику, чтобы на ее основе построить содержательные модели, имеющие иерархический структурный характер. Поведение в них оптимизируется в поисковом смысле (см. приложение 3).

Реальным прообразом рассматриваемых теоретических моделей могут служить разнообразные подсистемы встроенного контроля и поиска неисправностей совре-

менных сложных технических систем, не говоря уже о развитых биологических контрольных приспособлениях.

Согласно предельному закону надежности (см. (2.9)), конечная, идеальная, симметричная система A_R имеет экспоненциальный закон надежности $P_R(t) = \exp(-\lambda t)$. Далее рассмотрен достаточно широкий класс \mathfrak{A}_R таких систем с различными иерархическими структурами поиска неисправностей, среди которых оптимальной считается система $A_R^{(0)}$, при фиксированном значении t обращающая в максимум величину $P_R(t)$ или, что то же, величину λ^{-1} (среднее время между отказами).

Будем все же в некоторой минимальной степени учитывать внешнее поведение \bar{A}'_R системы A_R , полагая, что ее элементы участвуют в некоторых внешних действиях (работе) без конкретизации этих действий. Далее считается, что время от времени к элементам системы A_R поступают вызовы «на работу» на некоторый отрезок времени. При этом закон поступления и продолжительность таких вызовов далее считается случайным.

Даже такой минимальный учет внешнего функционирования системы приводит к отличному от традиционного определению отказа элемента, принятого в предыдущих параграфах. Далее отказ элемента определяется не как выход его из строя (физический отказ), а как вызов неисправного элемента (функциональный отказ). Биологическим аналогом рассмотренной функциональной модели отказа является процесс внешнего случайного воздействия, требующего функционирования отдельных клеток или органов организма. При этом клетки (органы) независимо от внешних возбуждений могут находиться в активном или пассивном состоянии (спад активности — физический отказ). Функциональный отказ здесь состоит в случайном совпадении внешнего возбуждения с периодом спада активности.

Из-за практической редкости вызовов большинства элементов открываются большие резервы в предупреждении их функциональных отказов непрерывным поиском неисправных элементов с их заменой или исправлением. На функциональное определение отказа [9] автора побудил следующий факт. При демонтаже одной уста-

ревшей ЭВМ были обнаружены некоторые элементы, которые за десятилетний период ее эксплуатации ни разу не использовались.

2.4.2. Обоснование принятых допущений. Рассмотрим независимо и параллельно с функциональной структурой $|A|$ системы A надежность структуру $|A_R|$ ее надежного R -сечения A_R . Основным принципом формирования этой структуры из рабочих элементов системы является выделение таких максимально крупных подсистем, для разладки которых в целом достаточно разладка хотя бы одного из их элементов. Независимая от функционального назначения элементов описанная выше их группировка в подсистемы на все более и более высокие уровни приводит к иерархической структуре $|A_R|$, которая в ряде случаев может обладать следующими свойствами. Во-первых, если объединять в группы зависимые в смысле отказов элементы (в вероятностном смысле), то на определенном высоком иерархическом уровне содержащие их подсистемы можно считать практически выходящими из строя независимо. Во-вторых, несмотря на объединение первичных элементов с вероятностью отказа p в группу из n элементов, вероятность отказа P группы в целом может оказаться все еще малой при достаточной малости величины p (например, если все ее элементы соединены последовательно и имеют независимые отказы, то $P \approx np \ll 1$, если $p \ll 1/n$).

Таким образом, наличие указанных свойств надежной иерархической структуры $|A_R|$ позволяет выделить такой ее уровень, на котором ее подсистемы могут быть приняты за независимые в вероятностном смысле и, вообще говоря, еще относительно редко (с вероятностью $P \ll 1$) выходящие из строя «первичные» элементы. Выбор этого уровня всякий раз является самостоятельной нетривиальной задачей, которая должна учитывать конкретную структуру системы (некоторые соображения на этот счет содержатся в п. 7.3.3).

Далее в принятой модели функционирования элементов они находятся либо в состоянии «отдыха», либо после вызова «на работу» некоторое время функционируют, а затем опять переходят в состояние «отдыха» (например, сердце с регулярным ритмом работы, другие мышцы работают каждый раз случайное время). При этом неисправность у элемента (физический отказ) может возникнуть в любой момент времени. Если физиче-

ский отказ (здесь он считается необратимым без внешнего активного вмешательства) произошел до вызова элемента «на работу», то происходит функциональный отказ в момент вызова элемента на работу».

Если в результате поиска неисправностей наличие физического отказа у данного элемента становится известным системе, то она может устранить неисправность элемента до его вызова «на работу» и этим предотвратит функциональный отказ элемента.

В соответствии со сказанным, нам безразличны физические отказы сами по себе. Они нам страшны лишь как одна из причин функционального отказа (при вызове отказавшего элемента).

2.4.3. Принцип построения иерархической структуры самоконтроля [9]. Локализация неисправностей — одно из отличительных свойств сложной системы в отличие от простых. Принятая модель случайного наступления физических отказов элементов и их вызовов такова, что отказы и вызовы могут наступить в любой момент. Поэтому только непрерывный (постоянный) во времени контроль за каждым из n ненадежных элементов системы может гарантированно избавить от наступления отказа. Но для этого потребовалось бы такое же число n «контрольных» элементов, которые, в свою очередь, требовали бы столько же своих «контрольных» элементов и т. д. При этом возникает так называемая проблема «сторожей сторожей». Конечно, можно требовать повышенной надежности «сторожей» первой ступени, но, имея их столько же (n), что и первичных элементов, мы окажемся в нелогичной ситуации, так как целесообразно сразу же иметь с ними дело, используя их как первичные элементы, и вообще обходиться без «сторожей». Единственным выходом из тупика является отказ от жесткой коммутации рабочих и контрольных элементов, связанной с непрерывным во времени контролем. Вместо этого можно организовать *прерывистый* во времени контроль всей массы n рабочих элементов некоторым, сравнительно меньшим числом m контрольных элементов, имеющих повышенную надежность (за ними уже не ставятся «сторожа», так как в рассматриваемой задаче они считаются практически безотказными).

Такого рода задача может быть сведена к задаче контроля одним контрольным элементом некоторой группы из n рабочих элементов. Поиск неисправности одним

таким элементом естественно рассматривать на некоторых широких классах иерархических структур $|A_R|$ и поисковых процедур \bar{A}_R . Задача состоит в выборе из этих классов систем $A_R = (|A_R|, \bar{A}_R)$ некоторой оптимальной в определенном смысле системы A_R^0 .

Далее будет рассмотрен достаточно широкий класс многопараметрических l -ступенчатых иерархических структур $|A_R|$, схематически изображенных на рис. 2.1. На рисунке вся совокупность n рабочих элементов про-

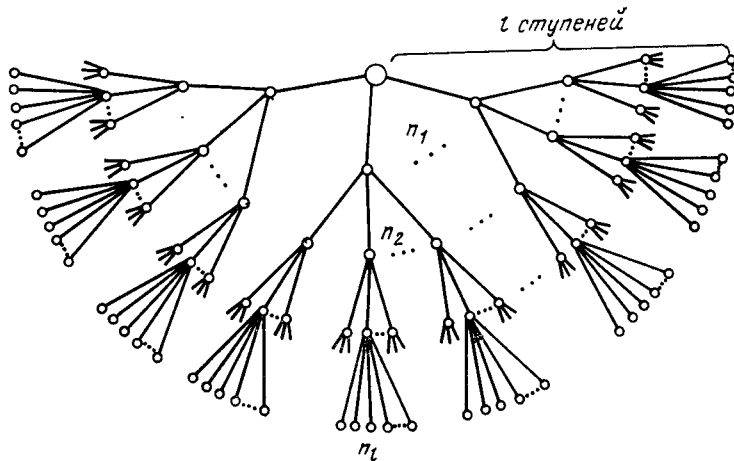


Рис. 2.1. Иерархическая структура поиска неисправностей.

извольным образом разбита на группы по n_i штук каждая (предполагается, что величина n/n_i — целое число). «Отводы» от каждого из элементов группы сходятся в одну точку, в которой помещен индикатор, показывающий наличие или отсутствие неисправности в своей группе элементов. Индикатор предполагается идеально надежным. Далее эти n/n_i индикаторов делятся каким-то образом на группы по n_{i-1} штук индикаторов в каждой группе. «Отводы» от каждого индикатора группы сходятся в одну точку, в которую помещен индикатор (также идеально надежный), показывающий наличие или отсутствие показаний неисправности внутри контролируемой им группы индикаторов (предполагается, что величина $n/n_i n_{i-1}$ — целое число). Новая совокупность ин-

дикаторов в количестве $n/n_i n_{i-1}$ штук снова делится на группы по n_{i-2} штук и т. д.

Определенная таким образом структура $|A_R|$ зависит от l -мерного целочисленного вектора $\bar{n} = (n_1, \dots, n_r, \dots, n_l)$ с компонентами, удовлетворяющими условиям

$$\prod_{r=1}^l n_r = n.$$

Перейдем к описанию алгоритма поиска \bar{A}_R .

В момент появления неисправности одного из рабочих элементов «верхний» (в нашем случае единственный) индикатор показывает неисправность. После этого контрольная аппаратура начинает поочередно просматривать «показания» индикаторов первой ступени в порядке и с продолжительностью каждого просмотра, диктуемыми той или иной поисковой процедурой.

После обнаружения индикатора первой ступени, на котором имеется показание неисправности, начинается просмотр контролируемых им индикаторов второй ступени согласно поисковой процедуре \bar{A}_R , после чего происходит переход к индикаторам третьей ступени и т. д. до l -й ступени, на которой обследуются уже сами рабочие элементы.

После обнаружения неисправного элемента на его замену или восстановление тратится, вообще говоря, некоторое случайное время τ . Такого рода иерархическая l -ступенчатая структура поиска неисправностей в ряде случаев существенно снижает время поиска по сравнению с поиском неисправного элемента непосредственно среди n таких элементов (одноступенчатый случай).

Ставится задача отыскания надежностной системы $A_R = (|A_R|, \bar{A}_R)$ с l -ступенчатой иерархической структурой $|A_R|$, определяемой набором целых чисел $\bar{n} = (n_1, \dots, n_l)$ и поведением \bar{A}_R , определяемым l поисковыми процедурами S_r — порядков просмотра индикаторов r -й ступени ($r = \overline{1, l}$), при которой вероятность Q успеть отыскать и восстановить неисправный элемент до его вызова была бы максимальной. Такую систему A_R^0 мы будем называть оптимальной.

§ 2.5. ОПТИМИЗАЦИЯ ИЕРАРХИЧЕСКОЙ СТРУКТУРЫ ПОИСКА НЕИСПРАВНОСТЕЙ

2.5.1. Вычисление вероятности Q успеть обнаружить неисправность (функциональный отказ). Используем теперь аппарат теории дискретного поиска, развитый в приложении 3 для вычисления вероятности Q . Как уже упоминалось, целочисленные компоненты l -мерного вектора $\bar{n} = (n_1, \dots, n_r, \dots, n_l)$, определяющего структуру $|A_R|$, удовлетворяют условию

$$\prod_{r=1}^l n_r = n, \quad (2.21)$$

где величины n и l будем считать пока заданными.

Рассмотрим дискретные моменты времени $1, 2, \dots$, в которые происходит поиск одного неисправного элемента среди n элементов. Обследование i -го элемента с целью проверки его исправности требует в рассматриваемой нами схеме также дискретного случайного времени τ_i . При этом возможен пропуск исправного элемента с вероятностью β .

Будем считать, что в любой момент времени, независимо от других моментов поисковая процедура может быть прервана с вероятностью $p > 0$ (не прерывается с вероятностью $q = 1 - p$). Такого рода поиск будем называть p -напряженным поиском с β -пропуском или просто p -напряженным поиском.

Если поиск не прерывается до своего успешного окончания (неисправный элемент найден), то будем говорить, что поиск успел окончиться. В противном случае (поиск прерван до отыскания неисправного элемента) будем говорить, что поиск не успел окончиться.

Пусть $\tau_r(1)$ — случайное время 0-напряженного поиска с β_r -пропуском на r -й ступени иерархической структуры $|A_R|$, $\tau_{r i_r}$ — случайное время обследования i_r -го индикатора r -й ступени структуры $|A_R| (i_r = \overline{1, n_r})$, κ_r — случайный номер индикатора r -й ступени структуры $|A_R|$ с сигналом неисправности и τ — случайное время, необходимое для замены или исправления неисправного элемента. Будем искать оптимальную (обращающую в максимум вероятность Q) процедуру поиска $\bar{A}_R = \{S_r\}$ в классе повторяю-

щихся после необнаружения индикатора с сигналом неисправности поисковых процедур $S_r (r = \overline{1, l})$. Возможность таких пропусков связана с $\beta_r > 0$ (отличием вероятностей $1 - \beta_r$ от единицы). Обозначим случайное число таких пропусков на r -й ступени через $\mu_r - 1 (r = \overline{1, l})$ и будем считать его независимым от других рассматриваемых случайных величин.

Обоснование такого ограничения класса поисковых процедур содержится в приложении 3.

Тогда общее время $\tau(1)$ 0-напряженного поиска с пропусками, характеризуемыми набором параметров $\bar{\beta} = \{\beta_r\}$, имеет вид

$$\left. \begin{aligned} \tau(1) &= \sum_{r=1}^l \tau_r(1) + \tau, \\ \tau_r(1) &= \sum_{j=0}^{\mu_r-1} \sum_{i_r=1}^{n_r} \tau_{r i_r}^j + \sum_{i_r=1}^{n_r} \tau_{r i_r}^{\mu_r} \end{aligned} \right\} \quad (2.22)$$

где $\tau_{r i_r}^j$ — случайные величины обследования i_r -го индикатора r -й ступени после $j-1$ -го пропуска (j -го «захода»), одинаково распределенные при различных значениях j и i_r .

Пусть целочисленная случайная величина ξ принимает значения $0, 1, 2, \dots, n \dots$ с вероятностями $p_0, p_1, \dots, \dots, p_n, \dots$

Ее производящей называется сумма

$$g_{\xi}(u) = \sum_{n=0}^{\infty} p_n u^n \left(g_{\xi}(1) = \sum_{n=0}^{\infty} p_n = 1 \right),$$

сходящаяся хотя бы для одного действительного значения $|u| < 1$ (формального аргумента). Подробнее о производящих изложено в приложении 1.

Введем для производящих случайных величин $\tau_{r i_r}, \tau_r(1), \tau(1)$ и τ обозначения $g_{\tau_r}(u), g_{\tau_r(1)}(u), g_{\tau(1)}(u)$ и $g_{\tau}(u)$ соответственно.

Учитывая, что по условию задачи случайные величины $\tau_r(1)$ и τ независимы, будем иметь из соотношения (2.22) для соответствующих производящих функций со-

гласно правилу мультипликативности (см. приложение 1)

$$g_{\tau_r(1)}(u) = \prod_{r=1}^l g_{\tau_r(1)}(u) g_{\tau}(u).$$

Тогда согласно лемме П.3 приложения 3 вероятность $Q = Q(q)$ успеть обнаружить неисправность при p -напряженном поиске неисправности по всей иерархической структуре самоконтроля имеет вид

$$Q = Q(q) = g_{\tau_r(1)}(q) = \prod_{r=1}^l g_{\tau_r(1)}(q) g_{\tau}(q), \quad (2.23)$$

где согласно (П.3.8) и (П.3.10)

$$g_{\tau_r(1)}(u) = \frac{1 - \beta_r}{1 - \beta_r g_{\tau_r}^{n_r}(u)} \sum_{i=1}^{n_r} p_i^r g_{\tau_r}^i(u). \quad (2.24)$$

Рассмотрим случай равновероятной неисправности внутри каждой r -й степени иерархической структуры $|A_R|$, другими словами, случайный номер k_r неисправной группы r -й степени иерархии принимает свои значения i с постоянной вероятностью $p_i^r = 1/n_r$. Тогда исходя из определения производящей

$$g_{\tau_r}(u) = \sum_{i=1}^{n_r} \frac{1}{n_r} u^i = \frac{1}{n_r} \frac{1 - u^{n_r}}{1 - u} u \quad (2.25)$$

и с учетом ограничения (2.21), подставив (2.24) и (2.25) в (2.23), получим

$$Q = Q(q) = Q_l(q, \bar{n}) = \frac{1}{n} \prod_{r=1}^l \frac{(1 - \beta_r) [1 - g_{\tau_r(1)}^{n_r}(q)] g_{\tau_r(1)}(q)}{[1 - g_{\tau_r(1)}(q)] [1 - \beta_r g_{\tau_r(1)}^{n_r}(q)]} g_{\tau}(q). \quad (2.26)$$

Для случая постоянных $\beta_r = \beta$ и $g_{\tau_r(1)}(x) = g_{\tau_1}(x)$ получим

$$Q = Q_l(q, \bar{n}) = \frac{1}{n} \prod_{r=1}^l \frac{(1 - \beta) [1 - g_{\tau_1}^{n_r}(q)] g_{\tau_1}(q)}{[1 - g_{\tau_1}(q)] [1 - \beta g_{\tau_1}^{n_r}(q)]} g_{\tau}(q). \quad (2.27)$$

Найдем экстремальное значение Q по компонентам вектора $\bar{n} = (n_r)$ при условии (2.21), считая их временно не-

прерывными величинами. Для этого используем метод множителей Лагранжа. Составим выражение

$$R = \ln Q + \lambda \left(\ln \prod_{r=1}^l n_r - \ln n \right) = \ln [n/g_{\tau_1}(q)] + \\ + l \ln \{(1 - \beta) g_{\tau_1}(q) / [1 - g_{\tau_1}(q)]\} - \lambda \ln n + \\ + \sum_{r=1}^l \{\ln [(1 - g_{\tau_1}^{n_r}(q)) / [1 - \beta g_{\tau_1}^{n_r}(q)]] + \lambda \ln n_r\}. \quad (2.28)$$

Отсюда условия экстремума имеют вид

$$\frac{\partial R}{\partial n_r} = \frac{d}{dn_r} \ln \frac{1 - g_{\tau_1}^{n_r}(q)}{1 - \beta g_{\tau_1}^{n_r}(q)} + \frac{\lambda}{n_r} = 0 \quad (r = \overline{1, l}). \quad (2.29)$$

Не решая систему (2.29), видим, что величина

$$n_r = n_r(q, \lambda, l)$$

не зависит от r . Поэтому из (2.21) сразу следует, что

$$n_r = n^{1/l}. \quad (2.30)$$

Эти значения n_r приводят к максимуму значения $Q_l(q, \bar{n})$, равному согласно (2.27)

$$Q = Q_l = \min_{\bar{n}} Q_l(q, \bar{n}) = \frac{g_{\tau}(q)}{n} \left\{ \frac{(1 - \beta) (1 - g^{n^{1/l}}) g}{(1 - g) (1 - \beta g^{n^{1/l}})} \right\}^l, \quad (2.31)$$

где $g = g_{\tau_1}(q)$.

В частности, когда время обследования и время ремонта постоянны, принимая их равными единице и $\tau = \text{const}$, получим $g_{\tau_1}(u) = u$, $g_{\tau}(u) = u^{\tau}$.

В этом случае получим из (2.31) соотношение

$$Q_l = \frac{q^{\tau}}{n} \left[\frac{(1 - \beta) (1 - q^{n^{1/l}}) q}{(1 - q) (1 - \beta q^{n^{1/l}})} \right]^l, \quad (2.32)$$

впервые найденное С. Ю. Рудерманом [10] с помощью динамического программирования. Им же было показано, что при $\beta = 0$ максимум величины Q_l по l достигается при бинарной иерархической структуре, когда $n_r = 2$ и $l = \log_2 n$, а также найдены экстремальные значения l для общего случая $\beta > 0$, обращающие Q_l в максимум.

Аналогичные результаты имеют место и в более общем случае Q_l , заданном выражением (2.31) при замене q на $g = g_{\tau_1}(q)$.

2.5.2. Условия вырождения многоступенчатой структуры в одноступенчатую [11]. Покажем прежде всего, что при некоторых условиях максимум Q_l достигается при $l=1$.

Положим $Q_l = C\varphi(l)$, где $\varphi(l)$ — часть Q_l , зависящая от l . Тогда из (2.31) получим

$$\begin{aligned} \frac{1}{\varphi(l)} &= \left[\frac{(1-\beta + \beta - \beta g^{n^{1/l}})(1-g)}{(1-\beta)g(1-g^{n^{1/l}})} \right]^l = \\ &= \left[\frac{\beta \left(\frac{1-\beta}{\beta} + 1 - g^{n^{1/l}} \right) (1-g)}{(1-\beta)g(1-g^{n^{1/l}})} \right]^l = \\ &= \left[\frac{1-g}{g} \left(\frac{1}{1-g^{n^{1/l}}} + \frac{1}{\beta} \right) \right]^l = \left(\frac{B}{1-g^{n^{1/l}}} + \frac{B}{A} \right)^l, \end{aligned} \quad (2.33)$$

где $g = g_{\tau_1}(q)$,

$$B = \frac{1-g}{g} \quad \text{и} \quad A = \frac{1-\beta}{\beta}. \quad (2.34)$$

Так как $0 < g < 1$, то

$$\frac{B}{A} + \frac{B}{1-g^{n^{1/l}}} > \frac{B}{A} + B.$$

Поэтому, если

$$\frac{B}{A} + B \geq 1, \quad (2.35)$$

то

$$\frac{B}{1-g^{n^{1/l}}} + \frac{B}{A} > 1. \quad (2.36)$$

В случае выполнения условия (2.35) для любых l выражение $1/\varphi(l)$ согласно тождеству (2.33) с ростом l возрастает ($\varphi(l)$ — убывает). Таким образом, если выполняется условие (2.35), то максимум $\varphi(l)$ достигается при $l=1$. Преобразуем выражения (2.35) к виду, более удобному для дальнейшего анализа.

Учитывая обозначения (2.34), имеем

$$\frac{B}{A} + B = \frac{(1-g)\beta}{g(1-\beta)} + \frac{1-g}{g} = \frac{1-g}{g} \left(\frac{\beta}{1-\beta} + 1 \right) = \frac{1-g}{g} \frac{1}{1-\beta}. \quad (2.37)$$

Условие (2.35) согласно тождеству (2.37) эквивалентно условию

$$\frac{1-g}{g} \frac{1}{1-\beta} \geq 1.$$

Отсюда

$$g^{-1} - 1 \geq 1 - \beta \quad \text{или} \quad g \leq \frac{1}{2-\beta}. \quad (2.38)$$

Итак, в случае справедливости соотношения (2.38) максимум $\varphi(l)$ (одновременно и максимум Q_l) по l достигается при $l=1$. Это значит, что при выполнении условия (2.38) вообще следует отказаться от многоступенчатой иерархической структуры.

Это обстоятельство имеет следующее интуитивное определение. В самом деле, величина q уменьшается с ростом частоты использования рабочего элемента системой, а величина $g = g_{\tau_1}(q)$ — монотонно растет с ростом q . Если q (а с ним и g) мало ($p=1-q$ — велико), то элемент используется системой весьма часто. В этом случае имеет смысл при поиске начинать осмотр прямо с рабочих элементов, не тратя время на осмотр индикаторов. Вероятность успеть отыскать и восстановить неисправный элемент будет при этом, конечно, весьма мала (рассматривается практически не интересный случай), но она все же здесь будет больше, чем в том случае, когда часть времени тратится на осмотр индикаторов.

Приведенный анализ показывает, что величина $g = \frac{1}{2-\beta}$ в этом смысле является уже достаточно малой. Так

как $0 \leq \beta \leq 1$, то $\frac{1}{2-\beta} \geq \frac{1}{2}$. Это значит, что при $g \leq 1/2$

независимо от значений β наилучшей является одноступенчатая структура $A_R(l=1)$. Условие (2.38) проливает свет и на роль параметра β в рассматриваемой поисковой процедуре.

При $\beta=0$ (рис. 2.2) значение $l=1$ следует брать при $g \leq 1/2$. При увеличении β верхняя граница $g = (2-\beta)^{-1}$,

при которой стоит брать $l=1$, возрастает. В пределе при $\beta=1$ наилучшее значение $l=1$ будет при $g \leq 1$, т.е. при любых g . Таким образом, увеличение вероятности перескока индикатора так же приводит к необходимости уменьшать l , как и уменьшение g . Если β столь велико, что $\frac{1}{2-\beta} \geq g$, то следует брать $l=1$.

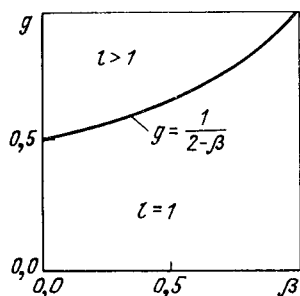


Рис. 2.2. Область параметров g и β , в которой происходит вырождение иерархической структуры в одноступенную ($l=1$).

2.5.3. Оптимизация структуры по числу ступеней [11].

Перейдем теперь к случаю, когда условия $g \leq (2-\beta)^{-1}$ не выполняются и поэтому $\varphi(l)$ не является монотонно убывающей с l функцией.

Заметим, что так как $n_r = n^{1/l} = z$, то

$$l = \frac{\ln n}{\ln z}. \quad (2.39)$$

Поэтому

$$\ln \varphi(l) = \frac{\ln n}{\ln z} \ln \frac{(1-\beta)(1-g^z)g}{(1-\beta g^z)(1-g)} = \Psi(z, g). \quad (2.40)$$

Основной практический интерес представляет случай малых β и $1-g$ (мало p). Действительно, только в этом случае поисковая процедура эффективна (перескок индикатора редок и каждый элемент редко используется системой).

Рассмотрим предельный случай, когда $p \rightarrow 0$, т.е. $g \rightarrow 1$. Тогда

$$\Psi(z, g) = \frac{\ln n}{\ln z} \ln \frac{g - g^{z+1}}{1-g} = \frac{\ln n}{\ln z} \ln (g + g^2 + \dots + g^z). \quad (2.41)$$

Разложим выражение $\Psi(z, g)$ (2.41) в ряд Тейлора в окрестности $g=1$:

$$\begin{aligned} \Psi(z, g) &= \Psi(z, 1) + \left. \frac{\partial \Psi}{\partial g} \right|_{g=1} (g-1) + \frac{1}{2} \left. \frac{\partial^2 \Psi}{\partial g^2} \right|_{g=c} (g-1)^2 = \\ &= \ln n + \left. \frac{\ln n}{\ln z} \frac{1+2g+3g^2+\dots+zg^{z-1}}{g+g^2+g^3+\dots+g^z} \right|_{g=1} (g-1) + \\ &\quad + \frac{1}{2} \Psi''_g(z, c)(g-1)^2, \end{aligned}$$

где величина c удовлетворяет условию $g \leq c \leq 1$. Далее имеем

$$\Psi(z, g) = \ln n + \frac{z+1}{2 \ln z} (g-1) \ln n + \frac{1}{2} \Psi''_g(z, c)(g-1)^2. \quad (2.42)$$

При g , близком к единице ($p \rightarrow 0$), влиянием третьего члена выражения (2.42) по сравнению с двумя первыми членами можно пренебречь.

Таким образом, здесь зависимость $\Psi(z, g)$ от z определяется лишь величиной $(z+1)/\ln z$. Так как $g < 1$, то $\Psi(z, g)$ достигает максимума тогда, когда $(z+1)/\ln z$ достигает минимума. Положим

$$a_z = (z+1)/\ln z. \quad (2.43)$$

Рассмотрим разность

$$\begin{aligned} \Delta_z &= a_{z+1} - a_z = \frac{z+2}{\ln(z+1)} - \frac{z+1}{\ln z} = \\ &= \frac{(z+2) \ln z - (z+1) \ln(z+1)}{(\ln z) \ln(z+1)}. \end{aligned} \quad (2.44)$$

Так как $z > 1$, то знак величины Δ_z совпадает со знаком числителя выражения (2.44).

Исследуем числитель выражения (2.44). Имеем

$$\begin{aligned} (z+2) \ln z - (z+1) \ln(z+1) &= (z+2) \ln z - \\ - (z+1) \ln \left[z \left(1 + \frac{1}{z} \right) \right] &= (z+2) \ln z - (z+1) \ln z - \\ - (z+1) \ln \left(1 + \frac{1}{z} \right) &= \ln z - \ln \left(1 + \frac{1}{z} \right) - \\ - z \ln \left(1 + \frac{1}{z} \right) &= \ln \frac{z}{1+1/z} - \ln \left(1 + \frac{1}{z} \right)^z = \\ &= \ln \frac{z^2}{z+1} - \ln \left(1 + \frac{1}{z} \right)^z. \end{aligned}$$

Последняя разность положительна, если

$$\frac{z^2}{z+1} > \left(1 + \frac{1}{z}\right)^z. \quad (2.45)$$

Так как $e > \left(1 + \frac{1}{z}\right)^z$, то выражение (2.44) положительно, если

$$\frac{z^2}{z+1} \geq e. \quad (2.46)$$

Заметим, что выражение $z^2/(z+1)$ возрастает с z . При $z=2$

$$\frac{z^2}{z+1} = \frac{4}{3} < e = 2,7 \dots$$

При $z=3$

$$\frac{z^2}{z+1} = \frac{9}{4} = 2,25 < e = 2,7 \dots$$

При $z=4$

$$\frac{z^2}{z+1} = \frac{16}{5} = 3,2 > e = 2,7 \dots$$

Итак, при $z=4$ $a_{z+1} - a_z = \Delta_z > 0$, т. е.

$$a_4 < a_5 < a_6 < \dots \quad (2.47)$$

С другой стороны, при $z=3$

$$\begin{aligned} \frac{z^2}{z+1} - \left(1 + \frac{1}{z}\right)^z &= \frac{3^2}{3+1} - \left(1 + \frac{1}{3}\right)^3 = \\ &= \frac{9}{4} - \frac{64}{27} = \frac{243 - 256}{108} = -\frac{13}{108} < 0. \end{aligned}$$

Таким образом,

$$a_3 > a_4. \quad (2.48)$$

Далее при $z=2$

$$\frac{z^2}{z+1} - \left(1 + \frac{1}{z}\right)^z = \frac{4}{3} - \frac{9}{4} = \frac{16-8}{9} = -\frac{2}{9} < 0.$$

Таким образом,

$$a_2 > a_3. \quad (2.49)$$

Собирая неравенства (2.47), (2.48) и (2.49) в одну цепочку, получим

$$a_2 > a_3 > a_4 < a_5 < a_6 < \dots \quad (2.50)$$

Как видно из неравенств (2.50), минимум a_z достигается при $z=4$.

Следовательно, при $\beta \rightarrow 0$ и $g \rightarrow 1$ ($p \rightarrow 0$) наилучшей многоступенчатой иерархической структурой $|A_R|$ является структура с

$$n_1 = \dots = n_r = \dots = n_i^* = z_i^* = 4.$$

2.5.4. Оптимизация структуры по среднему времени поиска неисправностей [11]. Среднее время осмотра индикаторов r -й степени найдем следующим образом. Если сигнал о неисправности поступил на i -й индикатор r -й степени, то для его обнаружения будет затрачено случайное время, равное

$$1 + (\mu - 1)n_r.$$

Здесь μ — случайное число циклов до обнаружения сигнала неисправности. Будем считать, что оно имеет одно и то же (для всех ступеней) геометрическое распределение с вероятностью пропуска, равной β .

Математическое ожидание этого случайного времени обнаружения равно

$$1 + n_r \frac{\beta}{1 - \beta}.$$

Аналогично вычисляется математическое ожидание случайного времени поиска сигнала неисправности, если он поступил на 2 -й индикатор r -й степени. Оно равно

$$2 + n_r \frac{\beta}{1 - \beta}$$

и т. д. и если сигнал неисправности поступил на i -й индикатор r -й степени, то соответствующее математическое ожидание равно

$$i + n_r \frac{\beta}{1 - \beta}.$$

В случае равнонадежности первичных рабочих элементов сигнал о неисправности на индикаторах внутри каждой степени появляется равновероятно. Поэтому вероятность появления сигнала неисправности на i -м индикаторе r -й степени равна $1/n_r$. Итак, безусловное математическое ожидание времени поиска сигнала неисправ-

ности на r -й ступени имеет вид:

$$\sum_{i=1}^{n_r} \frac{1}{n_r} \left(i + n_r \frac{\beta}{1-\beta} \right) = \frac{n_r + 1}{2} + n_r \frac{\beta}{1-\beta}.$$

Полное математическое ожидание t_{cp} случайного времени поиска неисправности по всей l -ступенной иерархической структуре равно

$$t_{cp} = \sum_{r=1}^l \left(\frac{n_r + 1}{2} + n_r \frac{\beta}{1-\beta} \right) = \left(\frac{1}{2} + \frac{\beta}{1-\beta} \right) \sum_{r=1}^l n_r + \frac{1}{2} l. \quad (2.51)$$

Величины $\bar{n} = (n_r)$ и l в соотношении (2.51) должны быть выбраны так, чтобы величина t_{cp} была минимальной. При фиксированном значении l задача сводится к минимизации суммы

$$\sum_{r=1}^l n_r = F, \quad (2.52)$$

при условии

$$\prod_{r=1}^l n_r = n \quad \left(G = \prod_{r=1}^l n_r - n = 0 \right). \quad (2.53)$$

Используя метод множителей Лагранжа (для случая непрерывных величин n_r), получим экстремальные значения из решения уравнений

$$\frac{\partial F}{\partial n_r} + \lambda \frac{\partial G}{\partial n_r} = 1 + \lambda \frac{n}{n_r} = 0. \quad (2.54)$$

Здесь λ — множитель Лагранжа и функции F и G определены соотношениями (2.52) и (2.53).

Из (2.54) получим

$$n_r = -\lambda n \quad (r = \overline{1, l}),$$

отсюда

$$n_1 = \dots = n_l = n^{1/l}. \quad (2.55)$$

Итак, здесь, как и в случае критерия, основанного на вероятности Q_l , количество индикаторов на всех ступенях оптимальной иерархической структуры должно быть постоянным. Минимальное среднее время поиска получим, если подставим соотношения (2.55) в (2.51). Тогда

$$t_{cp \text{ мин}} = \frac{l}{2} + \left(\frac{1}{2} + \frac{\beta}{1-\beta} \right) l n^{1/l}. \quad (2.56)$$

Введя прежнее обозначение $n_r = n^{1/l} = z$, получим $l = (\ln n) / \ln z$. В новых обозначениях равенство (2.56) имеет вид

$$t_{cp \text{ мин}} = \frac{\ln n}{\ln z} \left[\frac{1}{2} + \left(\frac{\beta}{1-\beta} + \frac{1}{2} \right) z \right]. \quad (2.57)$$

При $\beta \rightarrow 0$ (2.57) переходит в соотношение

$$t_{cp \text{ мин}} \rightarrow \frac{z+1}{1} \frac{1}{z} \frac{1}{2} \ln n. \quad (2.58)$$

Из (2.58) следует, что минимум $t_{cp \text{ мин}}$ достигается при минимуме выражения $(z+1)/\ln z$ по z , что имеет место согласно результатам предыдущего пункта при $z=4$. Последний результат означает, что при $\beta \rightarrow 0$ оптимальные структуры по критериям Q_l (при $g \rightarrow 1$) и по $t_{cp \text{ мин}}$ совпадают. Этот результат имеет и интуитивное оправдание. В самом деле, при $g \rightarrow 1$, когда $q \rightarrow 1$, элементы имеют очень редкие вызовы. Если элемент почти не используется в работе системы, то естественно, что вероятность не успеть его отыскать, когда он испортится, будет тем меньше, чем меньше среднее время его поиска.

В общем случае, когда q не слишком близко к единице, оптимальные структуры по разным критериям качества Q_l и $t_{cp \text{ мин}}$ весьма разнятся. Так, при $g \leq 1/2$ оптимальная структура, обеспечивающая минимум вероятности неуспевания Q_l , состоит, как было показано в п. 2.3.2, из одной ступени, т. е. $l=1$ и $z=n$. А в предельном случае

$$l = \frac{\ln n}{\ln z} \text{ и } z = n^{1/l}.$$

На рис. 2.3, а, б, в, г для мгновенного ремонта ($\tau=0$) при различных значениях g , β и n приводятся вычисленные на ЭВМ минимальные значения вероятности неуспевания (отказа) $\min P_{отк} = 1 - Q_l$ согласно соотношению (2.31), достигаемые при оптимальных значениях (l_{opt} , n_r). Приведенные зависимости позволяют проследить неасимптотические (при конечных β) связи указанных величин.

Допущение о мгновенном времени τ восстановления обнаруженного неисправного элемента ($\tau=0$, $g_r(q)=1$) не снижает общности рассмотрений.

Для произвольной случайной величины τ оцениваем здесь и далее значения Q при $\tau=0$ надо помножить на $g_r(q) < 1$.

Итак, из соотношения (2.58), верного при $\beta \rightarrow 0$, имеем для $z=4$:

$$t_{\text{ср мин}} = \frac{1}{2} \frac{4+1}{\ln 4} \ln n = 2,5 \log_4 n = 1,25 \log_2 n.$$

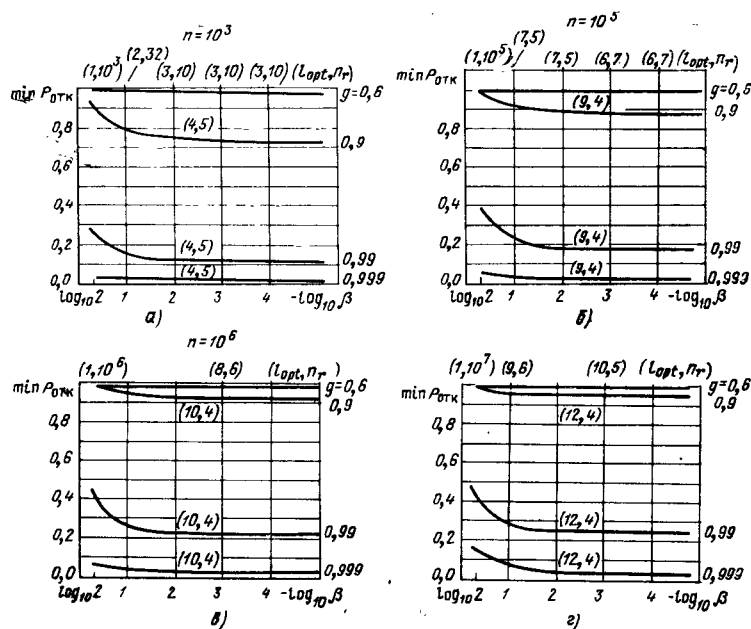


Рис. 2.3. а) Минимальная вероятность $\min P_{\text{отк}}$ отказа при различных значениях g и β и $n=10^3$. Оптимальные значения (l_{opt}, n_r) . б) Минимальная вероятность $\min P_{\text{отк}}$ отказа при различных значениях g и β и $n=10^5$. Оптимальные значения (l_{opt}, n_r) . в) Минимальная вероятность $\min P_{\text{отк}}$ отказа при различных значениях g и β и $n=10^6$. Оптимальные значения (l_{opt}, n_r) . г) Минимальная вероятность $\min P_{\text{отк}}$ отказа при различных значениях g и β и $n=10^7$. Оптимальные значения (l_{opt}, n_r) .

Если же взять неоптимальное в рассматриваемом случае значение $z=2$, то будем иметь существенно большее значение:

$$t_{\text{ср}} = \frac{3}{2} \frac{\ln n}{\ln 2} = 1,5 \log_2 n.$$

2.5.6. Оптимальность двоичной поисковой структуры при $\beta=0$. Рассмотрим теперь случай $\beta=0$, когда пропуск индикатора с сигналом неисправности невозможен. Тогда следует изменить процедуру поиска. Именно, если на r -й ступени уже осмотрены n_r-1 индикаторов и ни на одном из них нет сигнала о неисправности, то сигнал имеется заведомо на n_r -м индикаторе и поэтому не имеет смысла просматривать его и тратить на это единицу времени. В связи с этим случайная величина $\tau(1)$, определяемая соотношением (2.22), должна рассматриваться при $\mu_r=1$, а производящая функция $g_{\tau_r(1)}(u)$, определяемая соотношением (2.24), имеет вид

$$g_{\tau_r(1)}(u) = \sum_{i=1}^{n_r-1} p_i g_{\tau_r}^i(u) + p_{n_r} g_{\tau_r}^{n_r-1}(u).$$

Рассмотрим, как и в предыдущем пункте, случай равновероятности неисправностей и мгновенного восстановления. Тогда величина \tilde{a}_z , соответствующая величине $a_z = (z+1)/\ln z$, как легко показать, имеет вид

$$\tilde{a}_z = \frac{(z+2)(z^2-1)}{z \ln z} = \frac{z^2+z-2}{z \ln z} = a_z + \left(-\frac{2}{z \ln z} \right).$$

Непосредственный просчет значений \tilde{a}_z для $z=2, 3, 4$ показывает, что они монотонно возрастают. Как было показано ранее, начиная с $z=4$, первое из слагаемых последней суммы монотонно возрастает с ростом z . Аналогично ведет себя и второе слагаемое. Поэтому аналогично ведет себя и их сумма \tilde{a}_z . Итак, для случая $\beta=0$ оптимальной по вероятности $Q(q)$ успев обнаружить неисправность является двоичная структура ($z=2$).

Покажем, что эта же структура оказывается оптимальной и по среднему времени $\tilde{t}_{\text{ср}}$ поиска неисправности.

В рассматриваемом случае [см. приложение 3 (П.3.22)] легко показать, что

$$\tilde{t}_{\text{ср}} = \sum_{r=1}^l \frac{1}{n_r} \left(\sum_{i=1}^{n_r-1} i + n_r - 1 \right) = \sum_{r=1}^l \left(\frac{n_r+1}{2} - \frac{1}{n_r} \right).$$

Методом множителей Лагранжа нетрудно показать, что для минимизации $\tilde{t}_{\text{ср}}$ необходимо постоянство значений

$n_1 = \dots = n_l = z$. Тогда $n = z^l$ или $l = \frac{\ln n}{\ln z}$ и

$$\tilde{l}_{\text{оп}} = l \left(\frac{z+1}{2} - \frac{1}{z} \right) = \frac{z^2 + z - 2}{2z \ln z} \ln n = \tilde{a}_z \frac{\ln n}{2}.$$

Таким образом, минимум $\tilde{l}_{\text{оп}}$ совпадает с минимумом \tilde{a}_z , достигаемым при $z=2$. Но стоит только положить $\beta \neq 0$ (хотя и может быть β весьма близко к нулю), как бинарная структура перестает быть оптимальной и при β , близких к нулю, делается скачкообразно четверичной (минуя троичную) ($z=4$, $l=\log_4 n$) вследствие скачкообразного изменения процедуры поиска.

2.5.7. Асимптотическое поведение вероятности Q при $n \rightarrow \infty$. Согласно результатам предыдущих пунктов, при фиксированном значении n максимум величины $Q = Q_l(g, \beta)$, определяемой соотношением (2.31), при различных значениях параметров g и β имеет место для $l = l_{\text{опт}}(n)$.

Последняя величина имеет вид

$$l = l_{\text{опт}}(n) = \begin{cases} \log_2 n & \text{при } \beta = 0, g \rightarrow 1, \\ \log_4 n & \text{при } 0 < \beta < \frac{2g-1}{g}, g \rightarrow 1, \\ 1 & \text{при } \frac{2g-1}{g} \leq \beta \leq 1 \end{cases} \quad (2.59)$$

и приводит к бинарной, четверичной и вырожденной структуре $|A_R|$ соответственно.

В практически интересном случае небольших значений $\beta < (2g-1)/g$, малых $1-g \ll 1$ и больших $n \gg 1$ согласно соотношениям (2.59) имеем

$$l = l_{\text{опт}}(n) = \log_a n, \quad n = a^l, \quad n^{1/l} = a, \quad a = 2; 4. \quad (2.60)$$

Рассмотрим теперь асимптотический случай $n \rightarrow \infty$. Тогда асимптотическое поведение величины Q определяется порядком малости величины

$$1 - g = \frac{c}{l^{1+\varepsilon}} = \frac{c}{(\log_a n)^{1+\varepsilon}}, \quad (2.61)$$

где $c > 0$ и $\varepsilon > -1$ — действительные константы.

Именно, при условии (2.61) и $n \rightarrow \infty$ имеем

$$Q \rightarrow \exp(-\Omega l^{-\varepsilon}) = \begin{cases} \exp(-\Omega l^{\varepsilon}) & \text{при } \varepsilon < 0, \\ \exp(-\Omega) & \text{при } \varepsilon = 0, \\ 1 - \Omega/l^{\varepsilon} + O(l^{-2\varepsilon}) & \text{при } \varepsilon > 0, \end{cases} \quad (2.62)$$

где

$$\Omega = \left(\frac{a+1}{2} - \frac{a\beta}{1-\beta} \right) c > 0 \quad \text{при } \beta < \frac{5}{12} \leq \frac{a+1}{3a} \quad (a=2,4).$$

Для доказательства соотношения (2.62) достаточно положить в (2.31) $g_{\tau}(q) = 1$ и с учетом (2.60) записать его в виде

$$Q = \frac{1}{a^l} \left\{ \frac{(1-\beta) [1 - (1-(1-g))^a] [1 - (1-g)]}{(1-\beta)(1-g) \left[1 - \frac{\beta}{1-\beta} (1 - (1-g))^a \right]} \right\}^l = \left[1 - \left(\frac{a+1}{2} - \frac{a\beta}{1-\beta} \right) (1-g) + O((1-g)^2) \right]^l. \quad (2.63)$$

Подставив сюда выражение (2.61) для величины $1-g$ и переходя к пределу при $l = \log_a n \rightarrow \infty$, получим соотношение (2.62).

Так как при $p \rightarrow 0$ ($g \rightarrow 1$) имеем согласно (П.3.20)

$$g = g_{\tau_1}(g) = 1 - pE\tau_1 + O(p^2),$$

то полученное предельное соотношение (2.62) при условиях (2.61) допускает простую интерпретацию в терминах величин p и $q = 1-p$. В частности, при $\tau_1 = 1$, $g = q = 1-p$ и $1-g = p$.

§ 2.6. НАДЕЖНОСТЬ СИСТЕМЫ В ЦЕЛОМ. ПРИМЕРЫ

2.6.1. Надежность системы с оптимальной структурой поиска неисправностей [9]. Возможность осуществления поиска и устранения неисправностей оказывает влияние на надежность системы в целом.

Для вычисления надежности системы $P_R(T)$ (вероятности безотказной работы во все моменты времени от 1-го до T -го) зададим распределение, характеризующее интенсивность вызовов (в режиме работы) или интенсив-

ность отказов (в режиме готовности) ее первичных элементов.

Это распределение

$$P_T(k) = \mathcal{P}(x=k)$$

случайного числа x вызовов за время T .

Тогда надежность системы $P_R(T)$ совпадает с вероятностью ликвидации неисправностей поисковой процедурой и равна полной вероятности

$$P_R(T) = \sum_{k=0}^T P_T(k) Q^k = g_x(Q), \quad (2.64)$$

где

$$g_x(u) = \sum_{k=0}^T P_T(k) u^k$$

— производящая функция случайной величины g_x , а Q — вероятность единичного успешного поискового акта.

Если $Q \rightarrow 1$, то согласно соотношению (П.3.20)

$$P_R(T) = g_x(Q) = \exp[-(1-Q)E\kappa + O(1-Q)^2]. \quad (2.65)$$

Но, в свою очередь, согласно (П.3.20)

$$Q = g_\tau(q) = \exp(-pE\tau) \approx 1 - pE\tau. \quad (2.66)$$

Поэтому при $E\kappa \gg pE\tau$, подставив соотношение (2.66) в (2.65), получим

$$P_R(T) \approx \exp(-pE\tau E\kappa). \quad (2.67)$$

Для рассматриваемого случая независимых вызовов с вероятностью p вызова элемента имеем

$$E\kappa = Tp. \quad (2.68)$$

Тогда, подставив в (2.67) соотношение (2.68), получим (при $T \gg E\tau \approx (1-Q)/p$) выражение для надежности в виде:

$$P_R(T) \approx \exp[-p(1-Q)T] \approx \exp(-p^2 E\tau T). \quad (2.69)$$

Первое из этих выражений было впервые получено в [10].

2.6.2. Примеры. Рассмотрим некоторые числовые примеры использования кривых рис. 2.2 и 2.3.

Пусть замена (или ремонт) вышедших из строя элементов происходит мгновенно и время обследования τ каждого индикатора (и первичного элемента) имеет геометрическое распределение с параметром π ($0 < \pi < 1$)

$$\mathcal{P}(\tau=t) = (1-\pi)\pi^{t-1}.$$

Тогда производящая и среднее случайной величины τ имеют вид:

$$g_\tau(u) = (1-\pi)u/(1-\pi u) \quad \text{и} \quad E\tau = \frac{1}{1-\pi}, \quad (2.70)$$

а величина параметра g равна

$$g = g_\tau(g) = (1-\pi)q/(1-\pi q), \quad (2.71)$$

где $q=1-p$, причем p — величина вероятности вызова элемента.

Пример 2.1. Пусть $\pi=0,94$ и $p=0,04$. Тогда из соотношения (2.71) получим величину параметра $g=0,6$. Рис. 2.2 показывает, что при $g=0,6$ для значений $\beta \geq 0,33$ нецелесообразно использование многоступенчатых иерархических систем поиска неисправностей ($l > 1$). При $\beta < 0,33$, например при $\beta=0,1$, рис. 2.3 указывают параметры оптимальных многоступенчатых иерархических структур (l_{opt}, n_τ) в зависимости от числа первичных элементов $n=10^3, 10^5, 10^6, 10^7$, равных $(l_{opt}, n_\tau) = (2; 32), (7; 5), (8; 6), (9; 6)$ соответственно.

Однако все они приводят к практически неприемлемым значениям минимальной вероятности отказа $\min P_{отк} > 0,99$. Это указывает на практическую бесперспективность использования рассматриваемой системы самоконтроля.

Пример 2.2. Рассмотрим случай $\pi=0,9$ и $p=0,01$. Здесь из (2.71) получим величину параметра $g=0,99$. При таком значении параметра g видим из рис. 2.2, что для всех значений $\beta \leq 0,99$ имеет смысл использование многоступенчатых иерархических структур поиска неисправностей. Здесь в зависимости от различных значений этого параметра β и исходного параметра n (см. рис. 2.3) имеются разнообразие оптимальные иерархические структуры, приводящие к практически приемлемому диапазону значений $\min P_{отк}$ (для $\beta \leq 0,1$ имеем $\min P_{отк} < 0,3$).

Например, при $n=10^5$ и $\beta=10^{-2}$ имеем $\min P_{отк}=0,2$ (см. рис. 2.3,б).

Чтобы понять, почему в первом и втором примерах получены качественно разные результаты, вычислим соответствующие средние времена $E\tau$ обследования индикаторов [см. (2.70)] и средние времена между вызовами элементов $E\kappa \approx 1/p$. Имеем для примера 2.1

$$E\tau = 1/0,06 = 18 \quad \text{и} \quad E\kappa = 1/p = 25,$$

и для примера 2.2

$$E\tau = 1/0,1 = 10 \quad \text{и} \quad E\kappa = 1/p = 1000.$$

Сопоставление соответствующих величин показывает, в какой мере условия поиска неисправностей в первом примере хуже, чем во втором, что и приводит к качественно лучшим результатам во втором примере по сравнению с первым.

3. Помехоустойчивость [I-качество]

§ 3.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Устойчивость (надежность) сложной системы во времени, как уже отмечалось, является необходимым условием ее целостного функционирования. Но это качество не является характерным для сложных систем. Простые системы и тела тоже обладают устойчивостью (прочностью), необходимой для их целостного существования.

Характерным свойством, отличающим простые системы от сложных, является свойство информированности о внешней среде (ориентации в ней). Это качество сложных систем предполагает, во-первых, различение отдельных объектов внешней среды, несмотря на мешающее воздействие шумов (помехоустойчивость) и, во-вторых, хранение в памяти их абстрактных образов для последующей идентификации (узнавания). В этой главе мы остановимся лишь на предельных возможностях различения отдельных объектов при наличии мешающих факторов (шумов). Предположение об отсутствии последних не реалистично и не приводит к сколько-нибудь содержательным результатам. В гл. 5 будут обсуждаться вопросы, связанные с идентификацией объектов в плане обучения их распознавания.

Авангардная роль техники связи во всей новой технике, как уже отмечалось во введении, привела к открытию предельных законов прежде всего в теории потенциальной помехоустойчивости. Последняя была развита в фундаментальной работе В. А. Котельникова [1]. Он вычислил для канала связи с так называемым гауссовыми шумами максимальную вероятность $P_I(T)$ правильного различения M сигналов за время T . Эту вероятность он назвал помехоустойчивостью. Следуя ему, вероятность $P(A_X)$ достижения цели A_X по любому X -качеству X -сечения A_X сложной системы A мы назвали эффективностью по X -качеству (см. § 1.3).

В этой главе рассматривается решающая система A_I с постоянной структурой и поведением (без предварительного обучения). Если объекты среды представлены сигналами, то говорят, что система A_I имеет целью различение сигналов на фоне шумов.

Снова, как и в предыдущей главе, рассмотрим индифферентную ситуацию, когда система A_I имеет дело со средой B , явно не мешающей системе различать некото-

рые внешние объекты, представленные сигналами. Различные объекты будем называть *сообщениями* и обозначать целыми числами: $1, 2, \dots, u, \dots, M$.

Будем считать, что системе A_I доступна однозначно соотнесенная каждому сообщению u последовательность $x_u = (i_1^u, \dots, i_s^u, \dots, i_t^u)$ t бинарных символов $i_s^u = 0; 1$, которую мы будем называть *входным словом* длины t .

Однозначное отображение φ сообщений u во входные слова x_u

$$x_u = \varphi(u)$$

будем называть *кодом* и отождествлять со структурой $|A_I|$ системы A_I .

Номер s бинарного символа i_s^u будем интерпретировать номером момента дискретного времени. Пусть сообщение u передается входным словом x_u . Мешающее воздействие среды B в простейшем случае может заключаться в искажении в каждый дискретный момент времени s бинарного символа i_s^u входного слова x_u с постоянной для всех моментов $s = \overline{1, t}$ вероятностью p . Другими словами, система A_I воспринимает $i_s^u = 0$ как 1 и $i_s^u = 1$ как 0 с вероятностью p и воспринимает i_s^u правильно с вероятностью $q = 1 - p$. Предполагается, что такого рода искажения происходят в вероятностном смысле независимо друг от друга в различные моменты дискретного времени s .

Принятую модель мешающих воздействий среды B на символы входных слов x_u в теории информации называют *бинарным симметричным каналом* (БСК) с независимыми шумами.

Таким образом, система A_I воспринимает вместо входного слова x_u искаженное шумами выходное слово $y = (j_1, \dots, j_s, \dots, j_t)$ той же длины t , где бинарный символ $j_s = 0; 1$ соответствует искаженному или неискаженному бинарному символу i_s . Целью системы A_I является правильное различение с возможно большей вероятностью P_I возможно большего числа M сигналов, представленных входными словами x_u возможно меньшей длины t . По нашей терминологии (см. введение) цель системы состоит в наиболее выгодном (оптимальном) (M, t) -обмене. Это достигается выбором оптимальной

структуры $|A_I|$ -способа кодирования φ и оптимального поведения A_I -способа декодирования D (указания передававшегося входного слова x_u по выходному слову y)^{*}).

§ 3.2. НЕОПТИМАЛЬНЫЕ СИСТЕМЫ

3.2.1. Посимвольное мажоритарное декодирование.

Естественным способом борьбы с шумами является многократное повторение передачи одного и того же символа с принятием решения по большинству принятых символов (не нарушая общности, при этом предполагается, что $p < 0,5$). В случае $p > 0,5$ решение принимается по меньшинству принятых символов. При $p = 0,5$ имеет место ситуация полной неопределенности, исключающая какое-либо достоверное решение.

В рассматриваемом решении по большинству принятых символов (его иногда называют мажоритарным решением) естественно для исключения неоднозначности повторять символы нечетное число раз $r = 2l + 1$, чтобы принятых нулей и единиц не могло оказаться поровну.

Итак, можно предложить следующую процедуру кодирования $x_u = \varphi(u)$. Каждому из сообщений $u = \overline{1, M}$ сопоставляется последовательность из нулей и единиц длины k , где $k = \log_2 M$ — целое число, что предполагает рассмотрение лишь чисел $M = 2^k$ (так называемое безыбыточное кодирование). Далее каждый нуль и единица такой последовательности повторяются r раз. В результате получается входное слово x_u длины $t = kr$.

Обозначим величину $(\log M)/t = R$. Эта величина важна для дальнейших рассуждений. В нашем случае $R = 1/r$.

Подсчитаем вероятность P_m правильного мажоритарного декодирования при указанном способе кодирования в случае БСК. Декодирование ведется следующим образом. Принятое выходное слово y длины $t = kr$ разбивается на последовательные группы по r символов, каждая из которых соответствует одному входному символу. Далее выносятся решение о том, что передавался тот входной символ, который преобладает в соответствующей группе.

^{*}) Предполагается, что системе A_I известны все M входных слов x_u ($u = \overline{1, M}$), предназначенных для передачи.

Если передавался символ 0, то вероятность P_0 его правильного декодирования имеет вид:

$$P_0 = \sum_{p=(r+1)/2}^r C_r^p q^p p^{r-p} = \sum_{p=0}^{(r-1)/2} C_r^p p^p q^{r-p} = 1 - \sum_{p=(r+1)/2}^r C_r^p p^p q^{r-p} \quad (3.1)$$

Если передавался символ 1, то вероятность P_1 его правильного декодирования, как легко видеть, имеет тот же вид ($P_1 = P_0$).

Таким образом, искомая вероятность P_m правильного декодирования слова x_u длины $t = kr$, состоящего из k групп длины r (необходимо правильно декодировать все k групп), имеет вид:

$$P_m = P_1^k = \left[1 - \sum_{p=(r+1)/2}^r C_r^p p^p q^{r-p} \right]^k. \quad (3.2)$$

Исследуем структуру предлагаемого кодирования для малых значений M . Первое нетривиальное значение $M = 2$ приводит к $k = \log_2 M = 1$. Здесь имеем для кодирования сообщений 0 и 1:

$$x_1 = \underbrace{(0 \ 0 \ \dots \ 0)}_r \quad \text{и} \quad x_2 = \underbrace{(1 \ 1 \ \dots \ 1)}_r$$

соответственно^{*}).

Для случая $M = 4$, когда $k = 2$, имеем для кодирования сообщений 00, 01, 10 и 11:

$$\begin{aligned} x_1 &= (\underbrace{0 \ 0 \ \dots \ 0}_r \ \underbrace{0 \ 0 \ \dots \ 0}_r), \\ x_2 &= (0 \ 0 \ \dots \ 0 \ \ 1 \ 1 \ \dots \ 0), \\ x_3 &= (1 \ 1 \ \dots \ 1 \ \ 0 \ 0 \ \dots \ 0), \\ x_4 &= (1 \ 1 \ \dots \ 1 \ \ 1 \ 1 \ \dots \ 1) \end{aligned}$$

соответственно.

Для случая $M = 8$, когда $k = 3$, имеем для кодирования сообщений 000, 100, 010, 001; 110, 101, 011 и 111:

^{*}) Легко заметить, что для $M > 2$ минимальная разница между входными словами (число различающихся символов на соответствующих местах) не может превосходить половины их длины.

$$\begin{aligned}
x_1 &= (\overbrace{00 \dots 0}^r \quad \overbrace{00 \dots 0}^r \quad \overbrace{00 \dots 0}^r), \\
x_2 &= (11 \dots 1 \quad 00 \dots 0 \quad 00 \dots 0), \\
x_3 &= (00 \dots 0 \quad 11 \dots 1 \quad 00 \dots 0), \\
x_4 &= (00 \dots 0 \quad 00 \dots 0 \quad 11 \dots 1), \\
x_5 &= (11 \dots 1 \quad 11 \dots 1 \quad 00 \dots 0), \\
x_6 &= (11 \dots 1 \quad 00 \dots 0 \quad 11 \dots 1), \\
x_7 &= (00 \dots 0 \quad 11 \dots 1 \quad 11 \dots 1), \\
x_8 &= (11 \dots 1 \quad 11 \dots 1 \quad 11 \dots 1)
\end{aligned}$$

соответственно и т. д.

Из соображений симметрии очевидно, что любой другой способ кодирования, отличный от предлагаемого для случая $M=2$, при том же r может привести лишь к меньшей вероятности правильного декодирования, так как предлагаемый способ приводит к двум максимально различающимся входным словам x_1 и x_2 (они различаются во всех символах на всей своей длине r).

Те же соображения говорят в пользу предлагаемого способа кодирования и для $M=4$. Однако здесь возникает новое качественное отличие от случая $M=2$. В самом деле, здесь в отличие от случая $M=2$ кроме символического декодирования возможно так называемое декодирование в целом (для случая $M=2$ они тождественны). Последний способ декодирования состоит в сопоставлении выходного слова y по всей его длине $l=2r$ со всеми $M=4$ используемыми входными словами x_1 , x_2 , x_3 и x_4 .

При этом принимается решение о том, что передавалось то x_u , которое отличается от y в меньшем числе соответствующих символов или, как принято говорить, имеет минимальное хэмминговское расстояние $d=d(x_u, y)$ от y . Декодирование в целом мы обсудим в следующем пункте. Там будет показано, что и с точки зрения декодирования в целом предлагаемый метод кодирования для $M=4$ все еще не уступает другим по вероятности правильного декодирования, но уже для $M=8$ он становится в этом смысле хуже других способов кодирования.

Исследуем теперь случай больших значений $M=2^k$, когда $k \rightarrow \infty$. Если при этом r остается постоянным

($R=1/r=\text{const}$), то, как следует из соотношения (3.2), вероятность правильного декодирования $P_m \rightarrow 0$.

Рассмотрим теперь случай, когда k остается постоянным, а $r \rightarrow \infty$ ($R=1/r \rightarrow 0$). Легко показать, что в этом случае $P_m \rightarrow 1$. В самом деле, используя соотношение (1.10) при двоичных логарифмах и то, что $\theta = \frac{r+1}{2r} = \frac{1}{2} + \frac{1}{2r} > \frac{1}{2} > p$, имеем

$$\begin{aligned}
& \sum_{r=\frac{r+1}{2}}^r C_r^p p^p q^{r-p} < \\
& < 2^{-r} \left[\left(\frac{1}{2} + \frac{1}{2r} \right) \log \frac{\frac{1}{2} + \frac{1}{2r}}{p} + \left(\frac{1}{2} - \frac{1}{2r} \right) \log \frac{\frac{1}{2} - \frac{1}{2r}}{q} \right] < \\
& < 2^{-r} \left(0,5 \log \frac{0,5}{p} + 0,5 \log \frac{0,5}{q} \right) = 2^{-r} \left(0,5 \log \frac{1}{pq} - 1 \right)^{*}
\end{aligned}$$

Подставив полученную оценку в соотношение (3.2), получим

$$P_m > \left[1 - 2^{-r} \left(\frac{1}{2} \log \frac{1}{pq} - 1 \right) \right]^k > 1 - k \cdot 2^{-r} \left(\frac{1}{2} \log \frac{1}{pq} - 1 \right). \quad (3.3)$$

Из оценки (3.3) следует, что при постоянном k имеем $P_m \rightarrow 1$, если $r \rightarrow \infty$ ($R \rightarrow 0$). Этот же результат имеет место, если k с ростом r растет медленнее, чем по экспоненциальному закону, например по степенному $k=O(r^s)$ при любом $s>0$.

Итак, для того чтобы вероятность правильного символического мажоритарного декодирования P_m стремилась к единице, необходимо стремление $r \rightarrow \infty$ ($R \rightarrow 0$), т. е. бесконечная «растяжка» текста или, как иногда говорят, бесконечная задержка в декодировании или бесконечная избыточность при кодировании. Таковы были представления о возможности борьбы с шумами до фундаментальных исследований Шеннона [12], связанных с идеей приема сигналов в целом.

3.2.2. Хэмминговское декодирование в целом. Хэмминг [13] предложил следующую простую идею асимп-

*) В этой главе все логарифмы (\log) рассматриваются по основанию 2 без специального указания (\log_2).

тотически достоверного декодирования входных слов с конечной задержкой ($R = \text{const}$).

Пусть $M = 2^{tR}$ входных слов длины $t(x_1, \dots, x_u, \dots, \dots, x_v, \dots, x_M)$ выбраны так, что минимальное расстояние Хэмминга $d(x_u, x_v)$ между любой парой из них удовлетворяет соотношению

$$d = \min_{u \neq v} d(x_u, x_v) = 2l + 1,$$

где l — целое число, называемое *корректирующей способностью* кода. Примеры предыдущего пункта показывают, что при $M > 2$ имеем $d \leq t/2$.

Будем считать передававшимся то входное слово x_u , которое имеет минимальное хэмминговское расстояние от принятого выходного слова y :

$$d(x_u, y) = \min_v d(x_v, y).$$

Тогда, как легко видеть, если выходное слово y имеет по отношению к передававшемуся входному слову x_u не более чем l искажений, то оно будет правильно декодировано.

В соответствии с этим вероятность P_x правильного хэмминговского декодирования имеет вид:

$$P_x = \sum_{p=0}^l C_t^p p^p q^{t-p} = 1 - \sum_{p=l+1}^t C_t^p p^p q^{t-p}. \quad (3.4)$$

Учитывая, что для стремления P_x с ростом t к единице нужно требовать, чтобы $p < \theta = l/t$ и так как для $M > 2$ $2l + 1 = d \leq t/2$ ($l/t \leq \frac{1}{4} - \frac{1}{2t} < \frac{1}{4}$), то окончательно

будем иметь для $M > 2$ следующее ограничение на вероятность p искажения для возможности использования хэмминговского декодирования:

$$p < \theta < 1/4. \quad (3.5)$$

При этом вероятность P_x , определяемая (3.4) с использованием (1.10) при двоичных логарифмах, имеет оценку.

$$P_x > 1 - 2^{-t} \left[\left(\theta + \frac{1}{t} \right) \log \frac{\theta + 1/t}{p} + \left(1 - \theta - \frac{1}{t} \right) \log \frac{1 - \theta - 1/t}{q} \right]. \quad (3.6)$$

Оценим теперь число M возможных входных слов x_u , имеющих минимальное хэмминговское расстояние $d = 2l + 1$.

В самом деле, всего имеется $M_0 = 2^t$ всевозможных входных слов длины t . Будем размещать среди них искомые слова x_u с минимальным расстоянием d методом исчерпывания следующим образом. Первое слово x_1 выберем произвольно. Далее удалим из M_0 слов все слова x , отстоящие от x_1 на расстоянии $d(x_1, x) \leq d$.

Тогда всего останется $M_1 = 2^t - \sum_{p=0}^d C_t^p$ слов. Выберем

среди них произвольное слово x_2 и удалим из M_1 все слова x , отстоящие от x_2 на расстоянии $d(x_2, x) \leq d$. Тогда

всего останется $M_2 \geq 2^t - 2 \sum_{p=0}^d C_t^p$ слов. Аналогично на s -м

шаге имеем $M_s \geq 2^t - s \sum_{p=0}^d C_t^p$ оставшихся слов причем знак

«больше», начиная с $s=2$, связан с тем, что на предыдущих шагах может быть уже удалена часть слов, отстоящих от x_s на расстоянии $d(x_s, x) \leq d$, а мы всякий раз удаляем все такие слова.

Такая процедура приведет в конце концов впервые на некотором M -м шаге ($s=M$) к исчерпыванию всех M_0 слов,

т. е. $0 \geq 2^t - M \sum_{p=0}^d C_t^p$. Отсюда следуют нижние оценки числа M :

$$M \geq 2^t \left/ \sum_{p=0}^d C_t^p \right. \quad (3.7)$$

С другой стороны, будем исходить из необходимости обеспечения вероятности правильного декодирования P_x .

Тогда с каждым из M слов x_u должны быть связаны $\sum_{p=0}^l C_t^p$

слов некоторого множества \mathcal{E}_u , отстоящих от x_u на расстоянии, не превосходящем $l = (d-1)/2$ и не имеющих общих слов с аналогичными множествами \mathcal{E}_v для слов x_v ($v \neq u$). Отсюда следует неравенство

$$M \sum_{p=0}^{(d-1)/2} C_t^p \leq 2^t,$$

которое вместе с неравенством (3.7) приводит к извест-

ным оценкам Варшавова — Гильберта [13]:

$$2^t \left/ \sum_{\rho=0}^d C_t^\rho \right. \leq M \leq 2^t \left/ \sum_{\rho=0}^{(d-1)/2} C_t^\rho \right. \quad (3.8)$$

Используя оценки (1.10) при двоичных логарифмах, из соотношений (3.8) будем иметь

$$2^{t[1+2\theta \log 2\theta + (1-2\theta) \log (1-2\theta)]} \leq M \leq 2^{t[1 + (\theta-1/t) \log (\theta-1/t) + (1-\theta+1/t) \log (1-\theta+1/t)]} \quad (3.9)$$

где $\theta = l/t = d/2t - 1/t$.

Известна и другая верхняя оценка числа M , принадлежащая Плоткину [13]:

$$M \leq 2^{t(1-4\theta) + O(\log t)} \quad (3.10)$$

Для ее вывода рассмотрим код из M слов длины t :

$$x_1, \dots, x_u, \dots, x_v, \dots, x_M.$$

Будем получать из слов x_u кода новые слова, складывая их соответствующие символы по модулю два, т. е. полагая $0+0=1+1=0$ и $0+1=1+0=1$. Полученные слова будем символически обозначать так:

$$x_u \dot{+} x_v = x.$$

Рассмотрим такой код из M слов, у которого любая пара слов в сумме дает его же слово $x_u \dot{+} x_v = x_w$. Если подписать его слова друг под другом, то в любом столбце будет одинаковое число нулей и единиц. Такой код, если он существует, назовем идеальным*). Примеры таких кодов были нами уже рассмотрены (см. п. 3.2.1). Найдем верхнюю оценку числа $M = M(t, d)$ слов идеального кода как функцию длины его слов t и их минимального расстояния между собой d .

Для этого заметим, что хэмминговское расстояние обладает свойством

$$d(x_u, x_v) \equiv d(x_u \dot{+} x, x_v \dot{+} x).$$

*) Так, определенный идеальный код охватывает широкий класс бинарных алгебраических кодов [13].

В частности, при $x = x_u$ имеем

$$d(x_u, x_v) \equiv d(0, x_u \dot{+} x_v) \equiv d(0, x_w),$$

где слова $0 = \overbrace{(0 \dots 0)}^t$ и x_w согласно свойствам идеального кода снова принадлежат ему.

Таким образом, вместо изучения расстояний $d(x_u, x_v)$ идеального кода, достаточно рассматривать расстояния $d(0, x_w)$, соответствующие числу единиц в его словах x_w .

Так как в идеальном коде число нулей и единиц одинаковое, то общее число единиц в нем равно $tM/2$ (M предполагается четным числом). При этом нулевое слово 0 можно не учитывать, так как оно не вносит вклада в общее число единиц.

Среднее число единиц \bar{d} ($M-1$ -го слова, отличных от 0 (соответствует среднему расстоянию $M-1$ -го слова до него), имеет значение

$$\bar{d} \leq tM/2(M-1).$$

Ясно, что минимальное расстояние $d < \bar{d}$, т. е. $d \leq tM/2(M-1)$, или при $2d > t$

$$M \leq 2d/(2d-t). \quad (3.11)$$

С другой стороны, у идеального кода с $M(t, d)$ словами в первом столбце половина нулей и единиц. Рассмотрев ту его половину из $M(t, d)/2$ слов, которые имеют на первом месте нули, мы придем к совокупности слов длины $t-1$ с тем же минимальным расстоянием d . Поэтому заведомо $M(t-1, d) \geq M(t, d)/2$. Отсюда $M(t, d) \leq 2M(t-1, d)$. Повторяя эту оценку f раз, получим

$$M(t, d) \leq 2^f M(t-f, d), \quad (3.12)$$

где f — целое число, не большее t . Но если $t = 2d - 1$, т. е. $2d > t$, то имеем указанную выше оценку (3.11). Поэтому полагаем $t - f = 2d - 1$, откуда $f = t - 2d + 1$. Тогда

$$M(t, d) \leq 2^{t-2d+1} \frac{2d}{2d-t} = 2^{t(1-4\theta) + O(\log t)},$$

где $\theta = d/2t$, что и доказывает оценку (3.10).

На рис. 3.1 изображены оценки (3.9) и (3.10) для величины $R = (\log M)/t$ в зависимости от параметра $\theta = d/2t$ для больших значений t , причем заштрихована

область возможных значений величины R . Из рис. 3.1 видно, что оценка Плоткина лучше верхней оценки (3.9) для θ в пределах от 0,156 до 0,5.

В следующем параграфе будет показано, что число M входных слов ограничено лишь верхней оценкой (3.9) и поэтому, не говоря уже о существенном ограничении $p < 1/4$ (вместо $p < 1/2$), метод декодирования Хэмминга

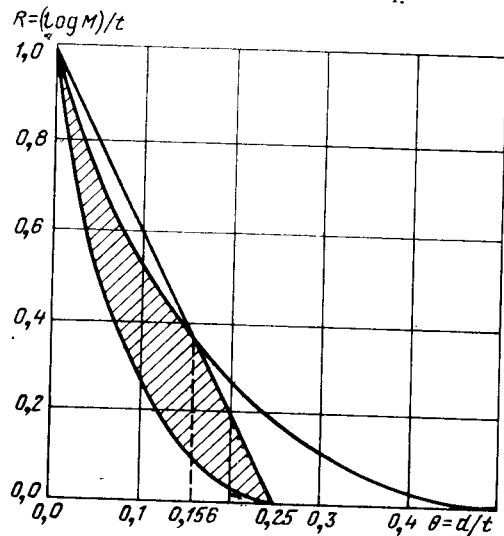


Рис. 3.1. Оценки возможных значений параметра $R = (\log M)/t$ в зависимости от величины параметра $\theta = d/2t$.

по числу M почти достоверно различаемых сигналов не является оптимальным.

Таким образом, системы A_I , использующие в качестве структуры $|A_I|$ и поведения \bar{A}_I указанные методы кодирования и декодирования, являются неоптимальными.

§ 3.3. ПРЕДЕЛЬНЫЙ ЗАКОН ПОМЕХОУСТОЙЧИВОСТИ

3.3.1. Случайное кодирование. Хэмминговское декодирование в целом основывается на оптимальном методе декодирования по минимальному хэмминговскому расстоянию (эквивалентному декодированию по максимуму правдоподобия [14, 15]). Оно приводит к неоптимальным по M результатам из-за несовершенного спо-

соба кодирования, основанного на идее минимального кодового расстояния d .

Докажем существование более совершенных кодов, приводящих к предельно большим числам M , ограниченным лишь верхней оценкой (3.9). Для этого воспользуемся весьма плодотворной идеей Шеннона о так называемом случайном выборе кода.

Начнем с вычисления вероятности правильного декодирования по минимуму хэмминговского расстояния при произвольном, но фиксированном коде

$$|A_I| = \{x_1, \dots, x_u, \dots, x_v, \dots, x_M\},$$

состоящем из M входных слов длины t .

Пусть по БСК передавалось входное слово x_u и принято выходное слово y . Тогда вероятность правильного декодирования имеет вид

$$P(d) = \mathcal{P}(\delta_1 < d), \quad (3.13)$$

где

$$d = d(|A_I|) = \min_{x_v (v \neq u)} d(x_v, y),$$

а случайная величина $\delta_1 = d(x_u, y)$ имеет биномиальное распределение

$$\mathcal{P}(\delta_1 = d) = C_t^d p^d q^{t-d}. \quad (3.14)$$

Пусть теперь символы 0 и 1 входных слов x_u возникают независимо и равновероятно (с вероятностью $1/2$).

Тогда величина $\delta = d(|A_I|)$ и вероятность правильного декодирования $P(\delta)$ являются случайными, причем

$$\mathcal{P}(\delta > d) = \mathcal{P}[d(x_1, y) > d, \dots, d(x_{u-1}, y) > d,$$

$$d(x_{u+1}, y) > d, \dots, d(x_M, y) > d] = \{\mathcal{P}[d(x_1, y) > d]\}^{M-1}.$$

Из предположения независимости и равновероятности символов 0 и 1 входных слов x следует, что для непередававшихся слов $x_v (v \neq u)$ имеем

$$\mathcal{P}[d(x_1, y) > d] = 1 - \sum_{s=0}^d C_t^s \left(\frac{1}{2}\right)^t.$$

Поэтому, подставив последнее соотношение в предыду-

щее, будем иметь

$$\mathcal{P}(\delta > d) = \left[1 - \sum_{s=0}^d C_t^s \left(\frac{1}{2} \right)^t \right]^{M-1}. \quad (3.15)$$

Найдем теперь среднее значение $\bar{P} = EP(\delta)$ случайной вероятности правильного декодирования $P(\delta)$, усредненной по всевозможным случайно выбранным кодам $|A_I|$. Имеем с учетом (3.13)

$$\begin{aligned} \bar{P} = EP(\delta) &= \sum_d \mathcal{P}(\delta = d) P(d) = \sum_d \mathcal{P}(\delta = d) \mathcal{P}(\delta_1 < d) = \\ &= \sum_d \mathcal{P}(\delta > d) \mathcal{P}(\delta_1 = d). \end{aligned} \quad (3.16)$$

Подставим в последнюю сумму (3.16) выражения для сомножителей, определяемые (3.14) и (3.15). В результате получим среднюю вероятность ошибки \bar{P} , усредненную по всем случайно выбранным кодам [14]:

$$\bar{P} = EP(|A_I|) = \sum_{d=0}^{t-1} C_t^d p^d q^{t-d} \left[1 - \sum_{s=0}^d C_t^s \left(\frac{1}{2} \right)^t \right]^{M-1}. \quad (3.17)$$

Верхний предел суммы $t-1$ взят из-за того, что при $\delta_1 = t$ имеет место ошибка декодирования ($d = d(|A_I|)$ не может быть больше t).

3.3.2. Формулировка и доказательство предельного закона помехоустойчивости. Оценка выражения (3.17) средней вероятности \bar{P} правильного декодирования позволяет установить для БСК

Предельный закон помехоустойчивости [12]. Пусть вероятность искажения символов в БСК равна p ($0 < p < 0,5$). Тогда существует такая система A_I (способ кодирования и декодирования), что, в соответствии с видом зависимости числа входных слов $M = 2^{tR}$ от их длины t , вероятность их правильного декодирования $P_I(t)$ имеет следующие оценки:

$$P_I(t) \begin{cases} \geq 1 - 2^{-t(R_{\text{выч}} - R)} & \text{при } 0 < R \leq R_{\text{кр}}, \\ \geq 1 - 2^{-tE_p(R)} & \text{при } R_{\text{кр}} < R < C, \\ < 2^{-tE_p(R)} & \text{при } R > C, \end{cases} \quad (3.18)$$

где $R_{\text{кр}} \leq R_{\text{выч}} \leq C$ имеют вид:

$$\left. \begin{aligned} R_{\text{кр}} &= 1 + \frac{\sqrt{p}}{\sqrt{p} + \sqrt{q}} \log \frac{\sqrt{p}}{\sqrt{p} + \sqrt{q}} + \\ &+ \frac{\sqrt{q}}{\sqrt{p} + \sqrt{q}} \log \frac{\sqrt{q}}{\sqrt{p} + \sqrt{q}}, \\ R_{\text{выч}} &= 1 - \log(1 + 2\sqrt{pq}), \\ C &= 1 + p \log p + q \log q \end{aligned} \right\} \quad (3.19)$$

и

$$E_p(R) = p_R \log \frac{p_R}{p} + (1 - p_R) \log \frac{1 - p_R}{q}, \quad (3.20)$$

причем величина p_R удовлетворяет соотношению

$$R = 1 + p_R \log p_R + (1 - p_R) \log(1 - p_R). \quad (3.21)$$

Доказательство. Начнем с оценки выражения, стоящего в квадратных скобках (3.17). Имеем

$$\begin{aligned} &\left[1 - \sum_{s=0}^d C_t^s \left(\frac{1}{2} \right)^t \right]^{M-1} \geq \\ &\geq \begin{cases} 1 - M 2^{-t} \sum_{s=0}^d C_t^s & \text{при } M 2^{-t} \sum_{s=0}^d C_t^s < 1, \\ 0 & \text{при } M 2^{-t} \sum_{s=0}^d C_t^s > 1. \end{cases} \end{aligned} \quad (3.22)$$

Условия, при которых выполняется каждая из оценок, эквивалентны существованию некоторого значения d_R , такого, что при $d \leq d_R$ имеет место первая оценка, а при $d > d_R$ — вторая.

Подставив оценки (3.22) в (3.17), получим

$$\begin{aligned} \bar{P} &\geq \sum_{d=0}^{d_R} C_t^d p^d q^{t-d} \left(1 - M 2^{-t} \sum_{s=0}^d C_t^s \right) = \\ &= 1 - \sum_{d=0}^{d_R} M 2^{-t} C_t^d p^d q^{t-d} \sum_{s=0}^d C_t^s - \sum_{d=d_R+1}^t C_t^d p^d q^{t-d}. \end{aligned}$$

Используя оценки (1.10) при двоичных логарифмах, получим, полагая $\theta = d/t$,

$$\begin{aligned} \bar{P} &\geq 1 - \\ &- \sum_{d=0}^{d_R} 2^{-t \left[\theta \log \frac{\theta}{p} + (1-\theta) \log \frac{1-\theta}{q} + 1 + \theta \log \theta + (1-\theta) \log (1-\theta) - R \right]} + O(\log t) \\ &- \sum_{d=d_R+1}^t 2^{-t \left[\theta \log \frac{\theta}{p} + (1-\theta) \log \frac{1-\theta}{q} \right]} + O(\log t). \end{aligned} \quad (3.23)$$

Дифференцирование по θ и приравнивание к нулю экспоненты слагаемого первой суммы в выражении (3.23) дает

$$\begin{aligned} & \left[\theta \log \frac{\theta^2}{p} + (1-\theta) \log \frac{(1-\theta)^2}{q} + 1 - R \right]' = \\ & = \log \frac{q}{p} - \log \left(\frac{1-\theta}{\theta} \right)^2 = 0. \end{aligned}$$

Отсюда экстремальное значение $\theta = p_{кр}$, обращающее слагаемое в максимум, имеет вид:

$$p_{кр} = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{q}}. \quad (3.24)$$

Показатель экспоненты при этом значении $\theta = p_{кр}$ имеет вид:

$$1 - \log (\sqrt{p} + \sqrt{q})^2 - R = 1 - \log (1 + 2\sqrt{pq}) - R.$$

Теперь оценим значение величины d_R или, что то же, величины $p_R = d_R/t$, исходя из условий (3.22). Для этого положим

$$M 2^{-t} \sum_{s=0}^d C_t^s = 1. \quad (3.25)$$

Используя асимптотическую оценку (1.10) для двоичных логарифмов при $p = q = 1/2$, получим из (3.25)

$$2^{-t [1 + \theta \log \theta + (1-\theta) \log (1-\theta) - R]} = 1.$$

Отсюда p_R находится из условия

$$R = 1 + p_R \log p_R + (1 - p_R) \log (1 - p_R). \quad (3.26)$$

Заметим, что с ростом R от нуля до $R_{кр} = 1 + p_{кр} \log p_{кр} + (1 - p_{кр}) \log (1 - p_{кр})$ соответствующее значение p_R убывает от 0,5 до $p_{кр}$ и на всем этом участке изменения $R(p_R)$ первая сумма выражения (3.23) экспоненциально преобладает над второй. С ростом R от $R_{кр}$ до величины $C = 1 + p \log p + q \log q$ (p_R убывает от $p_{кр}$ до p) имеет место обратный случай.

Наконец, при $R > C$ ($p_R < p$) вторая сумма выражения (3.23) экспоненциально совпадает с величиной $1 - 2^{-t E_p(R) + O(\log t)}$, что с учетом приводимых далее верхних оценок величины $P_I(t)$ окончательно доказывает верность соотношений (3.18) для величины \bar{P} . Но если эти соотношения верны для среднего значения \bar{P} , то они верны хотя бы для одного из значений случайной величины $P = P_I(t)$, связанной с некоторым конкретным кодом $[A_I]$. Таким образом, доказывается существование кода, обеспечивающего вероятность правильного декодирования $P_I(t)$, которая удовлетворяет соотношениям (3.18). Что и требовалось доказать.

Выясним теперь, в какой мере код $[A_I]$, существование которого было доказано выше, является оптимальным с точки зрения различения максимально большего числа M входных слов минимальной длины t с данной вероятностью P , близкой к единице (оптимальный (M, t) -обмен).

Для этого необходимо дать верхнюю оценку вероятности правильного декодирования $P_I(t)$. Это позволит показать, что при данных t и p величина $M = 2^{tC}$, где $C = 1 + p \log p + q \log q$, является предельно большим числом различимых входных слов.

В самом деле, верхние оценки (3.8) числа M и вероятности $P_I(t)$ приводят к следующим соотношениям:

$$M \leq 2^t \left| \sum_{s=0}^d C_t^s \right|, \quad P_I(t) \leq \sum_{s=0}^d C_t^s p^s q^{t-s}. \quad (3.27)$$

Но при достаточно больших t из (3.27) и (1.10) следует, что лишь при $d > t(p + \varepsilon)$, где $\varepsilon > 0$ произвольно мало ($\theta = d/t > p$), величина $P_I(t) \rightarrow 1$. Итак, имеем в предельном случае сколь угодно малых ε

$$\begin{aligned} M &\leq 2^{t(1 + p \log p + q \log q)}, \\ P_I(t) &\leq 1 - 2^{-t \left[\theta \log \frac{\theta}{p} + (1-\theta) \log \frac{1-\theta}{q} \right]}, \end{aligned}$$

где величина $\theta = d/t$ может быть определена в зависимости от $R = (\log M)/t$ в виде решения уравнения

$$R = 1 + \theta \log \theta + (1 - \theta) \log (1 - \theta).$$

Таким образом, код, существование которого было доказано, приводит к максимально большому числу M и требует при заданном $P_I(t)$ минимальное число t для значений R в интервале $R_{\text{кр}} < R < C$, где нижние и верхние оценки $P_I(t)$ экспоненциально совпадают. Итак, здесь мы близки к оптимальному $(2^{tc}, t)$ -обмену.

Для значений R в интервале $0 < R < R_{\text{кр}}$ эти оценки расходятся. Однако верхняя оценка при определенных значениях t и M может быть достигнута так называемым плотноупакованным кодом. Его построение аналогично построению хэмминговского кода методом исчерпывания.

Однако в отличие от первого у него вместо верхнего предела $(d-1)/2$ в сумме, определяющей $P_I(t)$, стоит d . При нем мы близки к оптимальному $(2^{tc}, t)$ -обмену для всех R ($0 < R < C$).

Прокомментируем полученный результат. Весьма заманчивым было бы использовать метод доказательства предельного закона (случайное кодирование) в качестве алгоритма получения оптимального кода. В самом деле, нами была получена оценка среднего значения \bar{P} отрицательной случайной величины $P = P_I(t)$. Это позволяет дать вероятностную оценку ее значений, используя неравенство Чебышева [14]:

$$\mathcal{P} [1 - P_I(t) \leq u(1 - \bar{P})] \geq 1 - u^{-1},$$

где u — произвольное положительное число.

Отсюда

$$P' = \mathcal{P} [P_I(t) \geq 1 - u(1 - \bar{P})] \geq 1 - u^{-1}. \quad (3.28)$$

Из-за экспоненциального характера величины $1 - \bar{P} = 2^{-\gamma t}$ ($\gamma > 0$) естественно положить $u = 2^{t\delta}$, где $\delta < \gamma$.

Тогда из (3.28) получим

$$P' = \mathcal{P} [P \geq 1 - 2^{-t(\gamma-\delta)}] \geq 1 - 2^{-t\delta}. \quad (3.29)$$

Таким образом, для конкретного случайно выбранного кода $|A_I|$ можно с вероятностью P' , большей чем $1 - 2^{-t\delta}$,

гарантировать, что у него вероятность $P_I(t)$ правильного декодирования будет больше $1 - 2^{-t(\gamma-\delta)}$.

Ясно, что при достаточно больших значениях t можно добиться сколь угодно близости вероятностей $P_I(t)$ и P' к единице. Однако нас интересует вопрос о возможности предельного быстрого стремления $P_I(t)$ к единице, когда $\gamma = E_p(R)$. С помощью оценок Чебышева на этот вопрос ответить нельзя, так как соответствующий показатель экспоненты равен $\gamma - \delta$. Более того, величину δ нельзя делать слишком малой, так как может быть высказан ряд аргументов в пользу задания $\delta > \gamma$. (Выбор «плохого» кода с вероятностью $1 - P'$ может оказаться из-за малости параметра $1 - P_I(t)$ и для «плохого» и для «хорошего» кода неконтролируемым явлением при сравнительно небольшом времени его использования.) Однако наши опасения о ненадежности случайно выбранных кодов, связанные с грубостью оценок Чебышева, оказываются неоправданными. В работе [15] показано, что на самом деле вероятность P' для не слишком близких к единице значений C имеет «сверхэкспоненциальное» стремление к единице ($P' > 1 - 2^{-2At}$, где $A > 0$). Этот результат показывает, что использование случайно выбранного кода практически гарантирует для достаточно больших t получение «хорошего» кода с вероятностью правильного декодирования, близкой к \bar{P} .

Поскольку построить так называемый плотноупакованный код в виде алгоритма $x_u = f(u)$ до сих пор не удается, построение оптимальных кодов случайным выбором является в настоящее время единственно реальной перспективой.

3.3.3. Случай стирающего канала. Рассмотренная выше модель БСК является не единственной простой моделью канала с шумами. Другой простой моделью канала с шумами является модель бинарного стирающего канала (БСтК). Эта модель, так же как и модель БСК, определяется единственным параметром — r -вероятностью стирания входных бинарных символов 0 и 1. В случае отсутствия стираний (с вероятностью $1 - r$) они переходят сами в себя без искажений. Таким образом, на выходе БСтК появляются символы 0, 1 и символ 2, которым мы будем формально обозначать стертые символы. Считается, что их появления являются независимыми событиями в вероятностном смысле.

Установим теперь связь между параметрами r , t , $M=2^{tR}$ и $P_I(t)$ БСтК, которая, как мы увидим далее, аналогична связи между соответствующими параметрами БСК (роль параметра p здесь играет параметр r).

Пусть код $|A_I|$ для БСтК тот же, что и для БСК, и состоит из $M=2^{tR}$ входных слов x_u ($u=1, M$), состоящих из t бинарных символов 0 и 1. Тогда, если передавалось входное слово x_u длины t БСтК производит в нем m стираний (образует m символов 2), в результате чего получим выходное слово y той же длины. Из определения БСтК следует, что случайное число m имеет биномиальное распределение

$$P(m) = C_t^m r^m (1-r)^{t-m}. \quad (3.30)$$

Число $t-m$ нестертых символов (в точности совпадающих с соответствующими символами x_u) имеет то же самое распределение.

Пусть фиксировано m стертых символов 2 в выходном слове y длины t . Тогда передававшееся слово x_u будет совпадать всеми своими $t-m$ нестертыми символами с соответствующими символами y . Если среди $M-1$ непередававшихся слов найдется хотя бы одно, которое на тех же нестертых местах в точности совпадает с y , то будем считать, что имеет место ошибка декодирования. На самом деле возникает неопределенность, но мы обязаны принимать решение и будем считать, что при ее разрешении имеет место ошибка.

Так же как и в случае БСК, будем выбирать бинарные символы M входных слов x_u ($u=1, M$) длины t независимо и равновероятно (с вероятностью $1/2$).

Тогда средняя вероятность $\bar{P}(m)$ правильного декодирования по всем случайно выбранным кодам совпадает с вероятностью случайного (в указанном выше смысле) выбора $2^{tR}-1$ подслов длины $t-m$, из которых ни одно в точности не совпадает с фиксированной нестертой частью y длины $t-m$. Из элементарных вероятностных соображений следует, что

$$\begin{aligned} \bar{P}(m) &= [1 - 2^{-(t-m)}]^{2^{tR}-1} \geq \\ &\geq \begin{cases} 1 - 2^{-t(1 - \frac{t}{m} - R)} & \text{при } R < 1 - \frac{t}{m}, \\ 0 & \text{при } R \geq 1 - \frac{t}{m}. \end{cases} \end{aligned} \quad (3.31)$$

Нами оценена условная средняя вероятность $\bar{P}(m)$ при фиксации m стертых символов. Соответствующая безусловная вероятность \bar{P} с учетом соотношения (3.30) имеет вид

$$\bar{P} = \sum_{m=0}^t C_t^m r^m (1-r)^{t-m} \bar{P}(m). \quad (3.32)$$

Используя оценки (3.31) и замечая, что неравенства $R \geq 1 - m/t$ эквивалентны неравенствам $m \geq t(1-R)$, получим из соотношения (3.32)

$$\begin{aligned} \bar{P} &\geq \sum_{m=0}^{t(1-R)} C_t^m r^m (1-r)^{t-m} \left[1 - 2^{-t(1 - \frac{m}{t} - R)} \right] = \\ &= 1 - \sum_{m=0}^{t(1-R)} C_t^m r^m (1-r)^{t-m} 2^{-t(1 - \frac{m}{t} - R)} - \\ &\quad - \sum_{m=t(1-R)+1}^t C_t^m r^m (1-r)^{t-m}. \end{aligned} \quad (3.33)$$

Полагая $\theta = m/t$ и используя оценки (1.10), получим из (3.33)

$$\begin{aligned} \bar{P} &\geq 1 - \sum_{\theta=0}^{1-R} 2^{-t \left[\theta \log \frac{\theta}{r} + (1-\theta) \log \frac{1-\theta}{1-r} + 1 - \theta - R \right]} - \\ &\quad - \sum_{\theta=1-R+\frac{1}{t}}^1 2^{-t \left[\theta \log \frac{\theta}{r} + (1-\theta) \log \frac{1-\theta}{1-r} \right]}. \end{aligned} \quad (3.34)$$

Найдем экстремум экспоненты слагаемого первой суммы выражения (3.34) по θ . Для этого приравняем нулю соответствующую производную по θ :

$$\begin{aligned} \left[\theta \log \frac{\theta}{r} + (1-\theta) \log \frac{1-\theta}{1-r} + 1 - \theta - R \right]' &= \\ - \log \frac{1-r}{r} - \log \frac{1-\theta}{\theta} - 1 &= 0. \end{aligned}$$

Отсюда экстремальное значение $\theta = \theta_{кр}$ имеет вид $\theta_{кр} = 2r/(1+r)$. При нем слагаемое первой суммы (3.34) обращается в максимум, а соответствующее выражение

В показателе достигает минимума, равного

$$\theta \log \frac{\theta}{r} + (1 - \theta) \log \frac{1 - \theta}{1 - r} + 1 - \theta - R \Big|_{\theta = \frac{2r}{1+r}} =$$

$$= 1 - \log(1 + r) - R. \quad (3.35)$$

Таким образом, первая сумма S_1 выражения (3.34) имеет при больших t асимптотическое значение

$$S_1 = 2^{-t(R_{\text{выч}} - R) + O(\log t)},$$

где

$$R_{\text{выч}} = 1 - \log(1 + r).$$

Заметим, что вторая сумма S_2 выражения (3.34) имеет с учетом оценок (3.31) при больших значениях t асимптотическое значение

$$S_2 = \begin{cases} 2^{-t \left[(1-R) \log \frac{1-R}{r} + R \log \frac{R}{1-r} \right] + O(\log t)} & \text{при } R < 1-r, \\ 1 - 2^{-t \left[(1-R) \log \frac{1-R}{r} + R \log \frac{R}{1-r} \right] + O(\log t)} & \text{при } R \geq 1-r. \end{cases} \quad (3.36)$$

Найдем теперь то значение R , при котором совпадают показатели экспонент обеих сум S_1 и S_2 , быть может, с их производными (как и в случае БСК). Проверим второе требование. В самом деле, производная по R первого показателя равна -1 . Приравняем ей производную второго показателя. Будем иметь

$$\log \frac{r}{1-r} + \log \frac{R}{1-R} = -1 \quad \text{или} \quad \frac{1-R}{R} = \frac{2r}{1-r},$$

отсюда

$$R = R_{\text{кр}} = \frac{1-r}{1+r} = 1 - \theta_{\text{кр}}. \quad (3.37)$$

Подставив вместо R значение $R_{\text{кр}}$ в оба показателя экспонент S_1 и S_2 , легко убеждаемся, что они тождественно совпадают.

Заметим теперь, что с ростом R от нуля до $R_{\text{кр}}$ первая сумма S_1 экспоненциально преобладает над второй суммой S_2 , а с ростом R от $R_{\text{кр}}$ до величины $C = 1 - r$ имеет место обратный случай. Наконец, при $R \geq C$ вто-

рая сумма асимптотически совпадает со вторым выражением в соотношении (3.34). Соображения, аналогичные приведенным для БСК, приводят к асимптотической оценке сверху величины \bar{P} в этом случае первым выражением в соотношении (3.34). Таким образом, нами доказаны следующие оценки для средней вероятности правильного декодирования в случае БСтК:

$$\bar{P} \begin{cases} \geq 1 - 2^{-t(R_{\text{выч}} - R) + O(\log t)} & \text{при } 0 < R \leq R_{\text{кр}}, \\ \geq 1 - 2^{-tE_r(R) + O(\log t)} & \text{при } R_{\text{кр}} < R < C, \\ < 2^{-tE_r(R) - O(\log t)} & \text{при } R > C, \end{cases} \quad (3.38)$$

где $R_{\text{кр}} < R_{\text{выч}} < C$ имеют вид

$$R_{\text{кр}} = \frac{1-r}{1+r}, \quad R_{\text{выч}} = 1 - \log(1 + r), \quad C = 1 - r \quad (3.39)$$

и

$$E_r(R) = (1 - R) \log \frac{1-R}{r} + R \log \frac{R}{1-r}. \quad (3.40)$$

Аналогично случаю БСК доказывается существование кода с вероятностью $P_I(t)$ правильного декодирования, удовлетворяющей соотношениям (3.38), (3.39) и (3.40). Те же соображения, что и в случае БСК, могут быть приведены для оценки возможности использования в качестве кода конкретно выбранного случайного кода. Как и в случае БСК, в случае БСтК имеем следующие верхние оценки для M и $P_I(t)$:

$$M \leq 2^t / 2^m = 2^{t \left(1 - \frac{m}{t} \right)}, \quad P_I(t) \leq \sum_{s=0}^m C_t^{s,r} (1-r)^{t-s}. \quad (3.41)$$

Из (3.41) и (1.10) следует, что лишь при $m > t(r + \varepsilon)$, где $\varepsilon > 0$ произвольно мало ($\theta = \frac{m}{t} > r$), величина $P_I(t) \rightarrow 1$. Итак, имеем в предельном случае сколь угодно малых ε :

$$M \leq 2^{t(1-r)}, \quad P_I(t) \leq 1 - 2^{-t \left[\theta \log \frac{\theta}{r} + (1-\theta) \log \frac{1-\theta}{1-r} \right]},$$

где величина $\theta = m/t$ в рассматриваемом случае совпадает с величиной $1 - R = 1 - (\log M)/t$.

Таким образом, как и в случае БСК, доказано существование кода с максимально большим значением M при заданном $P_I(t)$ (или заданном значении t) в интервале значений $R_{кр} < R < C$. Для значений R в интервале $0 < R < R_{кр}$ оценки $P_I(t)$ для построенного кода уступают оптимальным.

В заключение отметим, что выбор в качестве оптимального поведения системы A_I отбора входных слов по минимальному $d(x_u, y)$ для БСК или совпадающих максимальной своей частью с y для БСтК здесь не обосновывается. Однако в работе [15] этому посвящены более подробные обоснования статистического характера.

§ 3.4. ТЕОРИЯ ИНФОРМАЦИИ. ПРИМЕРЫ

3.4.1. Теория информации Шеннона. Существует другой подход в обосновании решения рассмотренных вопросов. Этот подход основан на аксиоматическом введении характерных величин, встречающихся в показателях экспонент перед величиной t в рассмотренных выражениях M и $P_I(t)$ [12, 14, 16].

Эти выражения имеют энтропийный характер $H = -\sum_i p_i \log p_i$ или в более общем случае $I = -\sum_i p_i \times \times \log q_i$, где $\sum_i p_i = \sum_i q_i = 1$ ($p_i, q_i > 0$).

Величина H энтропии распределения $\{p_i\}$ имеет физический (термодинамический) смысл. Некоторые авторы (см., например, [16]) считают появление величины H в термодинамике и теории информации неслучайным. Однако автору неизвестны какие-либо внутренние связи между этими теориями. Более того, согласно концепции автора (см. введение) их и не должно быть. Для основополагающих теорем Шеннона [12] понятие энтропии имело, по-видимому, наводящее значение.

Основные трудности лежат не в обосновании теории информации, а в сложности практического осуществления ее результатов. Они носят принципиальный характер и связаны с общими так называемыми экспоненциальными трудностями, о которых будет подробно сказано в гл. 7. Здесь же заметим лишь, что они связаны

с необходимостью экспоненциально растущего с ростом t перебора $M = 2^{tR}$ входных слов при указанном оптимальном декодировании, что уже при сравнительно малых значениях величины t делается несущественным.

3.4.2. Примеры. Для возможности инженерного использования развитой теории необходимо вычисление вероятностей правильного декодирования $P_I(t)$ в зависимости от ряда исходных параметров и важнейшего из них параметра $R = (\log M)/t$, называемого в теории информации *скоростью передачи*. Используя формулы (3.19) и

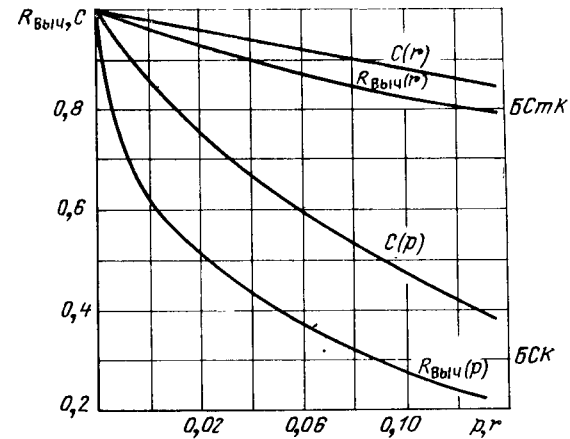


Рис. 3.2. Зависимость констант $C(p)$ и $R_{выч}(p)$ в случае БСК и $C(r)$ и $R_{выч}(r)$ в случае БСтК от параметров p и r соответственно.

(3.39), можно вычислить характерные константы $R_{выч}(p)$, $C(p)$ и $R_{выч}(r)$, $C(r)$ в зависимости от параметров шумов p и r для случаев БСК и БСтК соответственно (рис. 3.2).

Далее можно вычислить характерную функцию $k(\theta, p)$, определяемую соотношением (1.8), которая, помимо самостоятельного значения, полезна для вычислений функций $E(R)$.

Графики семейства функций $k(\theta, p)$ в зависимости от значений параметра p в практически интересном диапазоне помещены на рис. 3.3.

Наконец, для оценки вероятности правильного декодирования при случайном кодировании введем функции

$$E_p(R), E_r(R) = \begin{cases} R_{выч} - R & \text{при } 0 < R \leq R_{кр}, \\ E(R) & \text{при } R_{кр} < R < C \end{cases}$$

для случая БСК и БСтК соответственно. Величины $R_{кр}$, $R_{выч}$, C и функция $E(R)$ для каждого из рассматриваемых случаев определены соответственно соотношениями (3.19), (3.20) и (3.39), (3.40) (см. также рис. 3.2).

Графики семейств функций $E_p(R)$ и $E_r(R)$ в зависимости от параметров p и r помещены на рис. 3.4 и 3.5 соответственно.

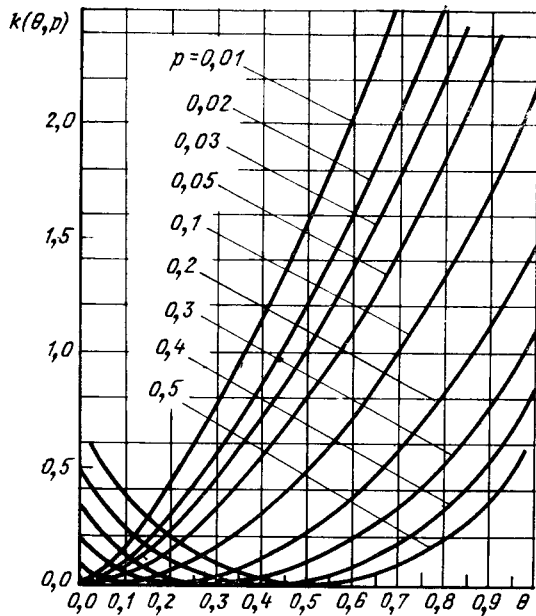


Рис. 3.3. Зависимость функции $k(\theta, p)$ от величины θ при различных значениях p .

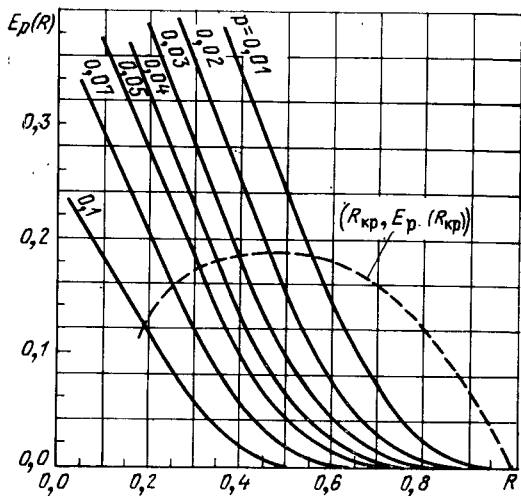


Рис. 3.4. Зависимость функции $E_p(R)$ от величины R при различных значениях p .

Пример 3.1. Пусть уровень шумов в БСК и БСтК определяется значениями параметров $p=0,1$ и $r=0,15$ соответственно. Пусть предельная длина t входных слов ограничена в обоих случаях величиной $t \leq t_0 = 10$. Какое предельно большое число входных слов M_C различимо в каждом из рассматриваемых случаев?

Решение. Из определения фундаментальных констант C имеем

$$M \leq M_C = 2^{t \cdot C}.$$

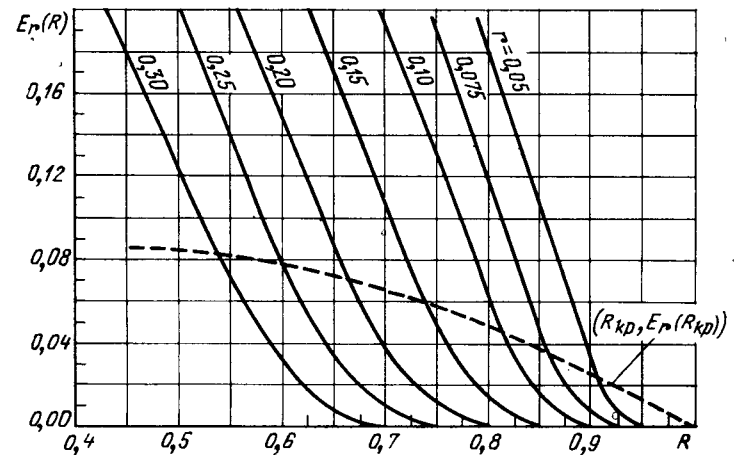


Рис. 3.5. Зависимость функции $E_r(R)$ от величины R при различных значениях r .

Величины $C = C(p) = C(0,1) = 0,53$ и $C = C(r) = C(0,15) = 0,85$ для случаев БСК и БСтК соответственно находим по кривым рис. 3.2. Поэтому для БСК

$$M \leq M_C(p) = 2^{10 \cdot 0,53} = 2^{5,3} \approx 40$$

и для БСтК

$$M \leq M_C(r) = 2^{10 \cdot 0,85} = 2^{8,5} = 366.$$

Пример 3.2. Рассмотрим случай БСК при $p=0,07$ и $R=0,35$. Пусть требуется найти длину t входных слов случайно образованного кода, приводящую к вероятности правильного декодирования:

$$P_I(t) > 1 - 2^{-24}.$$

Решение. По кривым рис. 3.4 находим значение функции $E_{0,07}(0,35) = 0,08$. Используя соотношение (3.18), положим

$$P_I(t) > 1 - 2^{-t E_p(R)} = 1 - 2^{-t \cdot 0,08} = 1 - 2^{-24}.$$

Отсюда $t \cdot 0,08 = 24$ и $t = 300$.

Пример 3.3. Рассмотрим случай БСТК при $r=0,1$ и $R=0,75$. Пусть длина входных слов $t=50$. Требуется найти оценку вероятности правильного декодирования $P_I(t)$ при случайном кодировании. Решение. По кривым рис. 3.5 находим значение функции $E_{0,1}(0,75)=0,13$. Используя (3.38), получим

$$P_I(t) > 1 - 2^{-tE_r(R)} = 1 - 2^{-50 \cdot 0,13} = 1 - 2^{-6,5} \approx 1 - 10^{-2}.$$

4. Управляемость (С-качество)

§ 4.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Управляемость (действенность, активность) является качеством сложной системы, предполагающим ее информированность о внешней среде, так как в противном случае затруднительны целенаправленные действия.

Управляемость системы в отличие от двух изученных выше качеств (надежности и помехоустойчивости) исследуется далее в случае конфликтной ситуации, когда вместо среды фигурирует другая, противоборствующая с первой система. Поэтому здесь все рассуждения приобретают симметричный характер относительно двух систем.

Далее внимание меньше всего концентрируется на конкретных видах проявления активности: механическом движении, выборе определенных действий или других конкретных актов. Основной проблемой является выбор системой оптимального поведения, т. е. одного из возможных набора «абстрактных» действий в условиях незнания ответных действий другой системы.

Для гарантированных заключений здесь приходится исходить из допущения о наиболее неблагоприятных действиях другой системы, что одинаково оправдано как в случае злонамеренного противника, так и в случае индифферентности другой системы, когда мы просто не знаем о ее возможных действиях.

Далее используется следующая модель конфликтной ситуации между системами A_C и B_C . Именно ведется долговременная борьба между системами за обладание жизненно важными для них ресурсами, имеющимися в ограниченном количестве $K = \text{const}$. Это количество K ресурсов до начала борьбы распределено между ними так, что A_C имеет K_A ресурсов, а B_C имеет $K_B = K - K_A$ ресурсов.

Борьба происходит в дискретные моменты времени $1, 2, \dots, t, \dots$, в каждом из которых происходит отдель-

ная схватка за некоторую порцию ресурсов. Схватка определяется следующим образом.

Системы A_C и B_C имеют наборы действий, определяемых как некоторые элементы x и y из абстрактных множеств $X = \{x\}$ и $Y = \{y\}$ (конечных или бесконечных) соответственно. Эти действия системы, не оповещая друг друга, выбирают с помощью, вообще говоря, случайного механизма. Последний определяется вероятностями $p_A(x)$ и $p_B(y)$, так называемыми стратегиями поведения A_C и B_C . Детерминированный выбор является частным случаем случайного, когда вероятности $p_A(x)$ и $p_B(y)$ вырождаются в единицы для некоторых выбираемых значений x_0 и y_0 .

Задается постоянная на все время борьбы функции выигрыша $M(x, y)$, означающая порцию ресурсов, переходящих к A_C , если A_C и B_C выбрали действия x и y соответственно. При этом B_C получает $-M(x, y)$ ресурсов (проигрыш A_C означает выигрыш B_C , и наоборот). Борьба во времени не ограничена и оканчивается истощением первоначального количества ресурсов K_A или $K - K_A$ у одной из систем, что связывается с ее поражением. Истощение ресурсов одной из систем иногда отождествляется с ее гибелью, поэтому можно рассматривать описанную борьбу как борьбу на выживание.

Впервые подобная ситуация была подвергнута вероятностному анализу на модели азартных игр, когда игроки A_C и B_C , доверившись в каждой партии «воле случая», выбирают ходы x и y , в результате чего получают выигрыш или проигрыш величины $M(x, y)$ и $-M(x, y)$ соответственно. Обладая первоначально капиталами K_A и K_B , каждый из них стремится разорить другого и приобрести весь капитал $K = K_A + K_B$. Поэтому описанную борьбу называют еще игрой на разорение.

Постановки задач на оптимум в описанных выше ситуациях появились лишь в послевоенный период и связаны с математической теорией игр [17]. Основные понятия этой теории, необходимые для дальнейшего изложения, будут изложены в следующем параграфе.

§ 4.2. ОДНОШАГОВЫЕ И МНОГОШАГОВЫЕ ИГРЫ

4.2.1. Одношаговые игры. Игрой называют четверку объектов $\{X, Y, M_A(x, y), M_B(x, y)\}$, где $X = \{x\}$ и $Y = \{y\}$ — два конечных или бесконечных множества, эле-

Меню которых называются *чистыми стратегиями*, а $M_A(x, y)$ и $M_B(x, y)$ — действительные функции пары аргументов, называемые *функциями выигрыша* систем A_C и B_C соответственно.

Если множества X и Y конечны или счетны, то функции $M_A(x, y)$ и $M_B(x, y)$ описываются конечными или бесконечными матрицами и поэтому такие игры называют *матричными*. Если множества X и Y несчетные (континуальные), то соответствующие игры называют *непрерывными*.

Если $M_A(x, y) + M_B(x, y) = 0$, то такая игра называется *нульсуммной* (или с нулевой суммой). Эта игра определяется единственной функцией выигрыша $M_A(x, y) = M(x, y)$, так как $M_B(x, y) = -M(x, y)$.

Функции выигрыша интерпретируются, как те выигрыши, которые получают системы при использовании стратегий x и y . В случае нульсуммной игры выигрыш одной системы есть проигрыш другой, и наоборот. Далее мы будем рассматривать лишь нульсуммные игры. В этом случае системы A_C и B_C выбирают свои стратегии x и y так, чтобы иметь

$$\max_{x \in X} \min_{y \in Y} M(x, y) = \underline{c} \quad \text{и} \quad \min_{y \in Y} \max_{x \in X} M(x, y) = \bar{c}, \quad (4.1)$$

причем величины \underline{c} и \bar{c} могут, вообще говоря, не совпадать и в случае отсутствия максимума или минимума берутся соответствующие верхние или нижние грани.

Смысл соотношений (4.1) состоит в следующем. При любой стратегии y системы B_C система A_C получит не меньше величины $\min_{y \in Y} M(x, y)$. Система A_C свободна при

этом выбирать любую стратегию x . Поэтому она может обеспечить себе выигрыш, равный самое меньшее

$$\max_{x \in X} \min_{y \in Y} M(x, y) = \underline{c}.$$

Аналогично система B_C при любой стратегии x системы A_C получит не меньше $\min_{x \in X} [-M(x, y)] = -\max_{x \in X} M(x, y)$, а система A_C получить не больше $\max_{x \in X} M(x, y)$. Поскольку

система B_C свободна в выборе своих стратегий y , то она

может ее выбрать так, чтобы система A_C получила самое большое

$$\min_{y \in Y} \max_{x \in X} M(x, y) = \bar{c}.$$

Как упоминалось выше, величины \underline{c} и \bar{c} могут не совпадать, но всегда удовлетворяют соотношению $\underline{c} \leq \bar{c}$, в силу чего они называются *нижней и верхней величиной* игры соответственно.

В самом деле, для любых x и y

$$\min_{y \in Y} M(x, y) \leq M(x, y) \leq \max_{x \in X} M(x, y).$$

Отсюда

$$\min_{y \in Y} M(x, y) \leq \max_{x \in X} M(x, y).$$

Так как правая часть этого неравенства не зависит от x , то, беря максимум по x от обеих частей, получим

$$\max_{x \in X} \min_{y \in Y} M(x, y) \leq \max_{x \in X} M(x, y).$$

Левая часть этого неравенства не зависит от y . Беря минимум по y от обеих частей, получим искомое соотношение $\underline{c} \leq \bar{c}$.

Если найденное неравенство обращается в равенство

$$\max_{x \in X} \min_{y \in Y} M(x, y) = \min_{y \in Y} \max_{x \in X} M(x, y) = C_c, \quad (4.2)$$

достигаемое при стратегиях x_0, y_0

$$M(x_0, y_0) = C_c,$$

то стратегии (x_0, y_0) , удовлетворяющие уравнению минимакса (4.2), называются *оптимальными*. В этом случае система A_C может получить самое меньшее C_c , используя стратегию x_0 , а система B_C может ей помешать получить больше C_c , используя стратегию y_0 . Итак, системам A_C и B_C ничего не остается лучшего, как выбрать стратегию x_0 и y_0 , поэтому эти стратегии называются *оптимальными*. Они обладают следующими свойствами:

1. Если система A_C выбирает стратегию x_0 , то какую бы стратегию y ни выбрала система B_C , система A_C может получить самое меньшее C_c .

2. Если система B_C выбирает стратегию y_0 , то какую бы стратегию x ни выбрала система A_C , система B_C может получить самое большее C_c .

3. Если бы системы A_C и B_C заранее объявили, что они намерены использовать стратегии x_0 и y_0 , то они не могли бы этим воспользоваться для увеличения своих выигрышей.

Если условия (4.2) удовлетворяются, то говорят, что функция выигрыша $M(x, y)$ имеет *седловую точку*. Значение $M(x_0, y_0) = C_c$ называют *величиной игры*. Вообще говоря, может быть несколько седловых точек, но величина игры во всех них должна быть постоянной. Таким образом, может быть несколько оптимальных стратегий.

Рассмотрим случай отсутствия седловой точки. В этом случае системе A_C гарантирован выигрыш, не меньший чем \underline{c} , но система B_C не может его снизить менее чем до величины $\bar{c} > \underline{c}$. Поэтому система A_C имеет возможность получить выигрыш в интервале значений от \underline{c} до \bar{c} в зависимости от стратегии, выбранной системой B_C . Поэтому в отличие от случая наличия седловой точки здесь для системы A_C существенно знание стратегии y , выбранной системой B_C , а последней существенно держать ее в тайне. Единственным и абсолютно надежным способом обеспечения этого является случайный выбор стратегии y , когда сама система B_C не знает заранее стратегии y , выбирая ее в последний момент. Аналогичные соображения можно высказать и о стратегии x системы A_C .

В этом случае системы A_C и B_C могут выбрать лишь распределения $p_A(x)$ и $p_B(y)$ своих чистых стратегий, которые называются *смешанными стратегиями*. При этом функция выигрыша $M(x, y)$ становится случайной величиной

$$M = \begin{cases} M(x, y), \\ p_A(x) p_B(y) \end{cases}$$

с распределением, равным произведению смешанных стратегий из-за независимости выбора чистых стратегий.

Математическое ожидание и дисперсия случайной функции выигрыша в матричном случае имеют вид:

$$E_{PA PB} M = \sum_{xy} M(x, y) p_A(x) p_B(y)$$

и

$$D_{PA PB} M = \sum_{xy} M^2(x, y) p_A(x) p_B(y) - (E_{PA PB} M)^2^{**}.$$

В непрерывном случае суммы заменяются на соответствующие интегралы.

Нижняя и верхняя величины игры в рассматриваемом случае определяются следующим образом:

$$\max_{PA} \min_{PB} E_{PA PB} M = \underline{c} \quad \text{и} \quad \min_{PB} \max_{PA} E_{PA PB} M = \bar{c}. \quad (4.3)$$

Уравнение минимакса имеет вид

$$\max_{PA} \min_{PB} E_{PA PB} M = \min_{PB} \max_{PA} E_{PA PB} M = C_c. \quad (4.4)$$

Причем смешанные стратегии p_A^0, p_B^0 , для которых

$$E_{PA^0 PB^0} M = C_c,$$

называются *оптимальными*.

Основная теорема фон Неймана о минимаксе [17] утверждает, что величина игры и оптимальные чистые стратегии соответствуют значению функции выигрыша в седловой точке, если она существует. Если она не существует, то всегда имеется величина игры для смешанных стратегий. Это доказано для всех рассматриваемых игр, кроме непрерывных игр с разрывной функцией выигрыша, когда существование величины игры необходимо доказывать для каждой конкретной функции выигрыша.

Заметим, что случай смешанных стратегий обобщает случай чистых стратегий, когда распределения $p_A(x)$ и $p_B(y)$ вырождаются в дельта-функции $p_A(x) = \delta(x - x_A)$ и $p_B(y) = \delta(y - y_B)$, что эквивалентно выбору в матричном случае чистых стратегий x_A и y_B , так как

$$\delta(x - x_A) = \begin{cases} 1, & \text{при } x = x_A, \\ 0, & \text{при } x \neq x_A. \end{cases} \quad (4.5)$$

*) Иногда для простоты обозначений мы будем опускать нижние индексы у выражений E и D . В случае, если величина $EM = 0$, игру называют безобидной.

с соответствующим интегральным определением дельта-функций для непрерывного случая.

Интерпретация решения игры в смешанных стратегиях возможна лишь при рассмотрении длинной последовательности игр со случайно выбираемыми стратегиями x и y , распределенными согласно оптимальным смешанным стратегиям $p_A(x)$ и $p_B(y)$.

Тогда среднее арифметическое функций выигрыша в длинной серии таких игр сходится по вероятности к величине игры C_c .

Заметим, что в ряде случаев игра может происходить лишь однократно (например, приводящая к гибели данной системы). Приведенная интерпретация не годна в таких случаях.

4.2.2. Многошаговые игры. Последовательность приведенных игр можно рассматривать как многошаговую игру, если все игры подчинены некоторой общей цели, как это имеет место в играх на разорение. Далее будут рассматриваться только такие многошаговые игры. Определим их следующим образом. Пусть имеются ограниченные ресурсы K , которые распределены до игры между системами A_C и B_C в количествах K_A и $K_B = K - K_A$ соответственно. Пусть между системами A_C и B_C происходит последовательность t нульсуммных игр рассматривавшегося типа с чистыми стратегиями в i -й игре $x_i \in X$ и $y_i \in Y$, на которых определена одна и та же функция выигрыша $M(x_i, y_i)$. Эту функцию будем называть *локальной* функцией выигрыша. Состояние игры перед t -й игрой характеризуется количеством ресурсов

$$K_{At-1} = K_A + \sum_{i=1}^{t-1} M(x_i, y_i), \quad (4.6)$$

имеющихся у системы A_C после $(t-1)$ -й игры (система B_C имеет при этом $K_{Bt-1} = K - K_{At-1}$ ресурсов). Игра ведется до тех пор, пока у одной из систем не иссякнут ресурсы (она разорится).

Составим из использованных стратегий x_i и y_i t -мерные стратегии $x^t \in X^t$ и $y^t \in Y^t$, где $x^t = (x_1, \dots, x_i, \dots, x_t)$ и $y^t = (y_1, \dots, y_i, \dots, y_t)$. Определим на них *тотальную* функцию выигрыша системы A_C в виде

$$M_t(x^t, y^t, K_A) = P_t(x^t, y^t, K_A) - \frac{1}{2}, \quad (4.7)$$

где $P_t(x^t, y^t, K_A)$ — вероятность разорения системы B_C (выигрыша системы A_C) в последующих после t -й играх (предполагается, что до t -й игры она еще не разорена).

Так как аналогичная вероятность разорения системы A_C (выигрыша системы B_C) равна $1 - P_t(x^t, y^t, K_A)$, то соответствующая тотальная функция выигрыша системы B_C имеет вид

$$1 - P_t(x^t, y^t, K_A) - \frac{1}{2} = \\ = - \left[P_t(x^t, y^t, K_A) - \frac{1}{2} \right] = -M_t(x^t, y^t, K_A),$$

и мы имеем снова нульсуммную многошаговую игру, которую можно рассматривать и как одношаговую игру с двумя t -мерными стратегиями x^t и y^t , на которых определена одна тотальная функция выигрыша $M_t(x^t, y^t, K_A)$. К так определенной игре применима общая теория одношаговых игр. В частности, рассматривая смешанные стратегии $P_{At}(x^t)$ и $P_{Bt}(y^t)$ для случая матричного задания тотальной функции выигрыша $M_t(x^t, y^t, K_A)$, будем иметь для ее математического ожидания

$$EM_t(x^t, y^t, K_A) = \bar{P}(P_{At}, P_{Bt}, K_A) - \frac{1}{2}, \quad (4.8)$$

где

$$\bar{P}(P_{At}, P_{Bt}, K_A) = \\ = \sum_{x^t, y^t} P(x^t, y^t, K_A) P_{At}(x^t) P_{Bt}(y^t) \quad (4.9)$$

средняя вероятность разорения системы B_C с соответствующей интегральной записью в непрерывном случае. Для определения оптимальных смешанных стратегий P_{At} и P_{Bt} будем иметь уравнение минимакса

$$f_t(K_A) = \max_{P_{At}} \min_{P_{Bt}} \{EM_t(x^t, y^t, K_A)\} = \min_{P_{Bt}} \max_{P_{At}} \{...\}. \quad (4.10)$$

При этом величина игры $f_t(K_A)$ зависит лишь от номера шага и начального состояния игры K_A . Выведем теперь функциональное уравнение для определения оптимальных смешанных стратегий P_{At}^0 и P_{Bt}^0 . Из соотношения (4.6) следует, что

$$K_{At} = K_{At-1} + M(x_i, y_i). \quad (4.11)$$

Далее, используя независимость случайных слагаемых в (4.11), запишем вероятность $\bar{P}(P_{At}, P_{Bt}, K_A)$ в виде

$$\bar{P}(P_{At}, P_{Bt}, K_A) = \sum_M \bar{P}(P_{At-1}, P_{Bt-1}, K_A - M) P(M(x, y) = M)$$

или в эквивалентном виде

$$\bar{P}(P_{At}, P_{Bt}, K_A) = \sum_{x, y} \bar{P}[P_{At-1}, P_{Bt-1}, K_A - M(x, y)] p_A(x) p_B(y). \quad (4.12)$$

Вычитая из обеих частей соотношения (4.12) по $1/2$ с учетом (4.8), получим

$$EM_t(x^t, y^t, K_A) = \sum_{xy} EM_{t-1}[x^{t-1}y^{t-1}, K_A - M(x, y)] p_A(x) p_B(y). \quad (4.13)$$

Теперь для отыскания рекуррентных соотношений между величинами игры $f_t(\cdot)$ и $f_{t-1}(\cdot)$ на t -м и $(t-1)$ -м шагах необходимо воспользоваться принципом оптимальности Р. Беллмана [17], который доказывается от противного в сочетании с математической индукцией. Он имеет следующую формулировку для рассматриваемой многошаговой игры.

Каково бы ни было начальное состояние K_A и оптимальные смешанные стратегии P_{At}^0 и P_{Bt}^0 на t -м шаге, на $(t-1)$ -м шаге смешанные стратегии P_{At-1} и P_{Bt-1} должны быть оптимальными $P_{At-1} = P_{At-1}^0$ и $P_{Bt-1} = P_{Bt-1}^0$ относительно соответствующих им „начальных“ состояний $[K_A - M(x, y)]$.

Взяв минимакса от обеих частей соотношения (4.13) с последующим использованием сформулированного принципа Р. Беллмана и обозначений (4.10) приводит к функциональному уравнению для рекуррентного определения оптимальных стратегий P_A^0 и P_B^0 [6]:

$$f_t(K_A) = \max_{PA} \min_{PB} \left\{ \sum_{xy} f_{t-1}[K_A - M(x, y)] p_A(x) p_B(y) \right\} = \min_{PB} \max_{PA} \{ \dots \}. \quad (4.14)$$

Пусть существует предел $\lim_{t \rightarrow \infty} f_t(K) = f(K)$. Тогда, переходя к пределу в обеих частях соотношения (4.14), получим искомое функциональное уравнение в виде

$$f(K_A) = \max_{PA} \min_{PB} \left\{ \sum_{xy} f[K_A - M(x, y)] p_A(x) p_B(y) \right\} = \min_{PB} \max_{PA} \{ \dots \}. \quad (4.15)$$

Пусть $|M(x, y)| \ll K_A$ и функция $f(K)$ имеет отличную от нуля производную $f'(K) \neq 0$. Тогда из приближенного соотношения

$$f[K_A - M(x, y)] \approx f(K_A) - M(x, y) f'(K_A)$$

и уравнения (4.15) легко следует уравнение для приближенного определения оптимальных смешанных стратегий P_A^0 и P_B^0 :

$$\max_{PA} \min_{PB} \left\{ \sum_{xy} M(x, y) p_A(x) p_B(y) \right\} = \min_{PB} \max_{PA} \{ \dots \}, \quad (4.16)$$

не зависящее от вида функции $f(\cdot)$!. Это замечательное свойство многошаговых игр впервые было открыто Р. Беллманом [6].

Прокомментируем его. Основное допущение $|M(x, y)| \ll K_A$ требует малости влияния исхода одной игры $|M(x, y)|$ на величину текущего значения ресурсов K_{At} системы A_C , когда $K_{At} \sim K_A$ (не слишком близко к неудачному для системы A_C концу игры). В этом случае при $f'(K) \neq 0$ общее функциональное уравнение минимакса (4.15) переходит в обычное уравнение минимакса (4.16) для одношаговой игры, не зависящей от вида функции $f(\cdot)$. Таким образом, здесь происходит вырождение многошаговой игры в последовательность независимых друг от друга одношаговых игр, когда на каждом шаге вне зависимости от предыдущих результатов и планов на будущее системы A_C и B_C стремятся получить наибольший выигрыш в каждой данной игре (каждодневный мелочный успех в «серой» без оглядок и перспектив жизни).

§ 4.3. ПРОСТЕЙШИЕ СИММЕТРИЧНЫЕ ИГРЫ

4.3.1. Тринарная функция выигрыша. Основная теорема теории игр о существовании седловой точки и оптимальных стратегий в общем случае не сопровождается какими-либо указаниями о способах их получения, т. е. является типичной неконструктивной теоремой существования.

Только для простейших видов функций выигрыша можно получать явные выражения оптимальных стратегий [18]*. Одной из таких простейших функций выигрыша $M(x, y)$ является тринарная функция двух переменных, принимающая всего три значения 1, a и b , из которых два последних произвольны.

Рассмотрим случай произвольных множеств чистых стратегий $x \in X$ и $y \in Y$, когда

$$M(x, y) = \begin{cases} 1 & \text{при } (x, y) \in G_1, \\ a & \text{при } (x, y) \in G_a, \\ b & \text{при } (x, y) \in G_b, \end{cases} \quad (4.17)$$

где $G_1, G_a, G_b \in X \times Y$ — произвольные множества, такие, что только для G_a его „проекция“ на X и Y совпадают с ними: $\{x: (x, y) \in G_a\} = X$, $\{y: (x, y) \in G_a\} = Y$ и $G_1 \cup G_a \cup G_b = X \times Y$.

Заметим, что функция $M(x, y)$ при $1 > a < b$ или $1 > a < b$ не имеет седловой точки, а при $1 < a < b$ или $1 > a > b$ имеет ее. Вычитая величину a из всех приведенных выше неравенств и деля обе части их на величину $b - a \neq 0$, получим с учетом знака величины $b - a$ условия отсутствия или наличия седловой точки вида

$$\lambda = \frac{1-a}{b-a} \begin{cases} > 0 & \text{— нет седловой точки,} \\ < 0 & \text{— есть седловая точка.} \end{cases} \quad (4.18)$$

В последнем случае к величине игры, равной a , приводит целое множество седловых точек $(x, y) \in G_a$. Далее будет рассматриваться лишь первый случай, когда

* Сводка явных решений игр содержится в издаваемой в этой серии брошюре В. Ф. Крапивина «Теоретико-игровые методы синтеза сложных систем в конфликтных ситуациях». Ему же принадлежит явное решение широкого класса непрерывных игр с функцией выигрыша, зависящей от разности аргументов [18].

оптимальные стратегии ищутся в классе смешанных стратегий $p_A(x)$ и $p_B(y)$, на основе которых вычисляются вероятности

$$\mathcal{P}_1 = \mathcal{P}[(x, y) \in G_1], \\ \mathcal{P}_a = \mathcal{P}[(x, y) \in G_a] \quad \text{и} \quad \mathcal{P}_b = \mathcal{P}[(x, y) \in G_b]. \quad (4.19)$$

и функция выигрыша $M(x, y)$ становится тринарной случайной величиной

$$M = \left\{ \begin{array}{ccc} 1 & a & b \\ \mathcal{P}_1 & \mathcal{P}_a & \mathcal{P}_b \end{array} \right\} (\mathcal{P}_1 + \mathcal{P}_a + \mathcal{P}_b = 1), \quad (4.20)$$

имеющей математическое ожидание

$$EM = \mathcal{P}_1 + a\mathcal{P}_a + b\mathcal{P}_b = \\ = \begin{cases} (1-a)\mathcal{P}_1 + (b-a)\mathcal{P}_b + a & \\ \text{или} & \\ (b-a) \left[\lambda\mathcal{P}_1 + \mathcal{P}_b + \frac{a}{b-a} \right] & \end{cases} \quad (4.21) \\ (4.21')$$

Игра называется *симметричной*, если $M(x, y) = M(y, x)$. Далее рассматриваются лишь симметричные игры. Для возможности дальнейшего исследования симметричных игр с тринарной функцией выигрыша рассмотрим два конкретных случая задания множеств X, Y и G_a ($a = 1, a, b$).

4.3.2. Простейший матричный случай. Пусть $X = \{x_1, x_2\}$, и $Y = \{y_1, y_2\}$ и

$$G_1 = \{(x_1, y_1)\}, G_a = \{(x_1, y_2), (x_2, y_1)\} \quad \text{и} \quad G_b = \{(x_2, y_2)\}. \quad (4.22)$$

Этот случай эквивалентен заданию матрицы игры в виде $\begin{pmatrix} 1 & a \\ a & b \end{pmatrix}$. Причем функция выигрыша становится тринарной случайной величиной (4.20) с вероятностями, согласно определениям (4.19) и (4.22) имеющими вид

$$\mathcal{P}_1 = p_A p_B, \quad \mathcal{P}_a = p_A(1 - p_A) + (1 - p_A)p_B \quad \text{и} \\ \mathcal{P}_b = (1 - p_A)(1 - p_B), \quad (4.23)$$

где $\bar{p}_A = (p_A, 1 - p_A)$ и $\bar{p}_B = (p_B, 1 - p_B)$ ($0 \leq p_A, p_B \leq 1$) — смешанные стратегии в рассматриваемом случае. Из-за

того, что игра симметрична, как легко показать, оптимальные смешанные стратегии достаточно искать в классе совпадающих стратегий

$$\bar{p}_A = \bar{p}_B = \bar{p} = (p, 1 - p). \quad (4.24)$$

Оптимальные стратегии можно найти и без этого упрощения с помощью дифференцирования выражения (4.21) по двум переменным p_A и p_B , от которых оно только и зависит в рассматриваемом случае.

Итак, при выполнении соотношения (4.24) вероятности, определяемые (4.23), будучи подставленными в выражения (4.20) и (4.21), приводят к

$$M = \begin{pmatrix} 1 & a & b \\ p^2 & 2p(1-p) & (1-p)^2 \end{pmatrix}$$

и

$$EM = (1-a)p^2 + (b-a)(1-p)^2 + a. \quad (4.25)$$

Экстремальное значение EM как функции параметра p найдем, приравняв $(EM)'_p$ нулю:

$$2(1-a)p - 2(b-a)(1-p) = 0.$$

Отсюда величина p^0 , дающая седловую точку EM , в чем легко убедиться непосредственной подстановкой, имеет вид

$$p^0 = \frac{b-a}{1-2a+b} = \frac{1}{1+\lambda}. \quad (4.26)$$

Здесь величина λ , определенная соотношением (4.18), при отсутствии седловой точки для чистых стратегий положительна, что приводит согласно (4.26) к существованию седловой точки для смешанных стратегий.

Подставив величину $p = p^0$ в выражение (4.25), получим

$$EM_{p=p^0} = \frac{(1-a)(b-a)^2 + (b-a)(1-a)^2 + a(1-2a+b)^2}{(1-2a+b)^2} = \\ = \frac{(1-a)(b-a)(1-2a+b) + a(1-2a+b)^2}{(1-2a+b)^2} = \frac{b-a^2}{1-2a+b}.$$

Итак, величина игры C_c в рассматриваемом случае имеет вид

$$C_c = \frac{b-a^2}{1-2a+b} = \frac{1+\lambda a}{1+\lambda}. \quad (4.27)$$

4.3.3. Непрерывная игра на m -мерном единичном кубе. Рассмотрим в качестве второго примера непрерывную игру на m -мерном единичном кубе K_m , когда $X = Y = K_m$, причем $x = (x_1, \dots, x_i, \dots, x_m)$, $y = (y_1, \dots, \dots, y_i, \dots, y_m)$, $0 \leq x_i, y_i \leq 1$ ($i = \overline{1, m}$) и между x и y определено евклидово расстояние

$$\rho(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}.$$

Определим множества G_1, G_a и G_b следующим образом:

$$\left. \begin{aligned} G_1 &= \{(x, y) : \rho(x, y) < \varepsilon, x \in K'_m\}, \\ G_a &= \{(x, y) : \rho(x, y) \geq \varepsilon\} \\ \text{и} \\ G_b &= \{(x, y) : \rho(x, y) < \varepsilon, x \in K_m \setminus K'_m\} \end{aligned} \right\} \quad (4.28)$$

где K'_m — произвольное подмножество K_m с объемом, равным половине (интерпретация такого определения содержится в п. 4.3.5).

Введем сокращенные обозначения для m -мерных интегралов:

$$\int_{u < x < v} f(x) dx = \int_u^v \dots \int_u^v f(x_1, \dots, x_m) dx_1, \dots, dx_m,$$

$$\int_{\rho(x, y) < \varepsilon} g(x) dx = \int_{\rho(x, y) < \varepsilon} g(x_1, \dots, x_m) dx_1, \dots, dx_m,$$

а также используем выражение для объема $V_{R, m}$ m -мерной сферы радиуса R :

$$V_{R, m} = \pi^{m/2} R^m / \Gamma(m/2) = S_m R^m, \quad (4.29)$$

где $S_m = \pi^{m/2} / \Gamma(m/2)$ и $\Gamma(m/2)$ — гамма-функция ($\Gamma(1/2) = \pi/2$, $\Gamma(1) = 1, \dots$ и т. д.).

Тогда в соответствии с определениями (4.19) и (4.28), используя теорему о среднем и выражение (4.29), получим

$$\left. \begin{aligned}
 \mathcal{P}_1 &= \int_{x \in K'_m} p_A(x) \left[\int_{p(x,y) < \varepsilon} p_B(y) dy \right] dx = \\
 &= S_m \varepsilon^m \int_{x \in K'_m} p_A(x) p_B(x) dx + o(\varepsilon^m), \\
 \mathcal{P}_b &= \int_{x \in K_m \setminus K'_m} p_A(x) \left[\int_{p(x,y) < \varepsilon} p_B(y) dy \right] dx = \\
 &= S_m \varepsilon^m \int_{x \in K_m \setminus K'_m} p_A(x) p_B(x) dx + o(\varepsilon^m), \\
 \mathcal{P}_a &= 1 - \mathcal{P}_1 - \mathcal{P}_b.
 \end{aligned} \right\} (4.30)$$

Подставив соотношения (4.30) в выражение (4.21'), получим

$$EM = (b - a) \left[S_m \varepsilon^m J_\lambda(p_A, p_B) + o(\varepsilon^m) + \frac{a}{b - a} \right], \quad (4.31)$$

где

$$\begin{aligned}
 J_\lambda(p_A, p_B) &= \lambda \int_{x \in K'_m} p_A(x) p_B(x) dx + \\
 &+ \int_{x \in K_m \setminus K'_m} p_A(x) p_B(x) dx.
 \end{aligned} \quad (4.32)$$

Ясно, что для сохранения зависимости EM от смешанных стратегий p_A и p_B необходимо требовать, чтобы a было пропорционально величине, не большей чем ε^m , т. е. $a = k \varepsilon^m b$ и $\lambda = 1/b + o(\varepsilon^m)$, где k — произвольное отличное от нуля число.

Тогда выражение (4.31) примет вид

$$EM = b [S_m J_\lambda(p_A, p_B) + k] \varepsilon^m + o(\varepsilon^m). \quad (4.33)$$

Теперь вместо решения уравнения минимакса для определения оптимальных стратегий p_A^0 и p_B^0 относительно EM можно решать его относительно $J_\lambda(p_A, p_B)$.

В рассматриваемом непрерывном случае игры функция выигрыша $M(x, y)$ разрывна и поэтому необходим специальный поиск оптимальных стратегий.

Имеем

$$\min_{p_B} J_\lambda(p_A, p_B) = J_\lambda[p_A, \delta(x - \underline{x})],$$

где

$$\underline{x} = \begin{cases} \underline{x}^{(1)}, & \text{если } \lambda p_A(\underline{x}^{(1)}) < p_B(\underline{x}^{(2)}), \\ \underline{x}^{(2)}, & \text{если } \lambda p_A(\underline{x}^{(1)}) \geq p_B(\underline{x}^{(2)}), \end{cases}$$

причем $\underline{x}^{(1)}$ и $\underline{x}^{(2)}$ — значения, в которых $p_A(x)$ достигает минимума на множествах K'_m и $K_m \setminus K'_m$ соответственно.

Далее

$$\max_{p_A} \min_{p_B} J_\lambda(p_A, p_B) = J_\lambda[r_\lambda(x), \delta(x - \underline{x})] = \frac{2\lambda}{1 + \lambda},$$

где

$$r_\lambda(x) = \begin{cases} \frac{2}{1 + \lambda} & \text{для } x \in K'_m, \\ \frac{2\lambda}{1 + \lambda} & \text{для } x \in K_m \setminus K'_m. \end{cases} \quad (4.34)$$

В самом деле, при таком выборе p_A и p_B при любом $x \in K_m$ выражение $J_\lambda \equiv 2\lambda/(1 + \lambda)$.

Аналогично можно показать, что

$$\min_{p_B} \max_{p_A} J_\lambda(p_A, p_B) = J_\lambda[\delta(x - \bar{x}), r_\lambda(x)] = \frac{2\lambda}{1 + \lambda},$$

где

$$\bar{x} = \begin{cases} \bar{x}^{(1)}, & \text{если } \lambda p_A(\bar{x}^{(1)}) > p_B(\bar{x}^{(2)}), \\ \bar{x}^{(2)}, & \text{если } \lambda p_A(\bar{x}^{(1)}) \leq p_B(\bar{x}^{(2)}), \end{cases}$$

причем $\bar{x}^{(1)}$ и $\bar{x}^{(2)}$ — значения, в которых $p_A(x)$ достигает максимума на множествах K'_m и $K_m \setminus K'_m$ соответственно.

Вместе с тем,

$$J_\lambda[r_\lambda(x), r_\lambda(x)] = \lambda \left(\frac{2}{1 + \lambda} \right)^2 \cdot \frac{1}{2} + \left(\frac{2\lambda}{1 + \lambda} \right)^2 \cdot \frac{1}{2} = \frac{2\lambda}{1 + \lambda}$$

и поэтому окончательно имеем решение уравнения минимакса:

$$\max_{p_A} \min_{p_B} J_\lambda(p_A, p_B) = \min_{p_B} \max_{p_A} J_\lambda(p_A, p_B) = J_\lambda[r_\lambda(x), r_\lambda(x)], \quad (4.35)$$

приводящее к оптимальным стратегиям

$$p_A(x) = r_\lambda(x) \quad \text{и} \quad p_B(y) = r_\lambda(y),$$

где $r_\lambda(x)$ определено соотношением (4.34).

Величину рассматриваемой игры получим, если подставим выражение (4.35) в (4.33):

$$\begin{aligned} C_c &= b \left[S_m \frac{2\lambda}{1+\lambda} + k \right] \varepsilon^m + o(\varepsilon^m) = \\ &= \left(S_m \frac{2b}{1+b} + kb \right) \varepsilon^m + o(\varepsilon^m). \end{aligned} \quad (4.36)$$

4.3.4. Частный случай бинарной игры. Заметим, что рассмотренная тринарная игра вырождается в бинарную при $b=1$ (при этом $\lambda=1$).

В этом случае тринарная случайная величина M (4.20) переходит в бинарную:

$$M = \begin{pmatrix} 1 & a \\ 1 - \mathcal{P}_a & \mathcal{P}_a \end{pmatrix},$$

имеющую математическое ожидание

$$H_c = EM = (a - 1) \mathcal{P}_a + 1 \quad (4.37)$$

и дисперсию

$$DM = (a - 1)^2 \mathcal{P}_a (1 - \mathcal{P}_a) = (EM - 1)^2 \frac{1 - \mathcal{P}_a}{\mathcal{P}_a}. \quad (4.38)$$

Легко видеть, что в рассмотренных матричном и непрерывном ($a = k\varepsilon^m$) случаях оптимальные смешанные стратегии являются равномерными распределениями, при которых

$$\mathcal{P}_a = \begin{cases} 1/2 & \text{в матричном случае,} \\ 1 - S_m \varepsilon^m + o(\varepsilon^m) & \text{в непрерывном случае.} \end{cases} \quad (4.39)$$

Подставив \mathcal{P}_a , определяемые (4.39), в выражения (4.37) и (4.38), получим

$$C_c = \begin{cases} (a + 1)/2, \\ (k + S_m) \varepsilon^m + o(\varepsilon^m), \end{cases} \quad (4.40)$$

где k — произвольное, отличное от нуля число и

$$DM = \begin{cases} (a - 1)^2/4, \\ S_m \varepsilon^m + o(\varepsilon^m) \end{cases} \quad (4.41)$$

в матричном и непрерывных случаях соответственно и число S_m определено в соотношении (4.29).

4.3.5. Интерпретация полученных результатов. Выше были рассмотрены два варианта тринарной симметричной игры с нулевой суммой — матричный и непрерывный.

Матричный вариант рассматривается в простейшей бинарной форме, допускающей разнообразные интерпретации во всех случаях, когда каждая из конфликтующих систем располагает лишь двумя возможными действиями.

Непрерывный вариант имеет более специальный характер и допускает меньшее число интерпретаций. Одна из них — это конфликтная ситуация преследования системы B_C системой A_C . Для того чтобы эта ситуация не носила тривиальный характер, следует полагать, что система A_C по сравнению с системой B_C более быстроходна, но менее маневренна. Следует различать два этапа преследования.

Первый этап состоит в сближении систем A_C и B_C . Второй этап состоит в серии попыток системы A_C войти в прямой контакт с системой B_C .

Математическая формализация описанного процесса преследования состоит в следующем. Рассмотрим плоский случай (обобщение на m -мерный случай затруднений не вызывает).

Начнем с первого этапа. Пусть скорости v_A и v_B систем A_C и B_C постоянны ($v_A > v_B$). В силу физических ограничений системы узнают о положениях друг друга лишь спустя интервал времени, принятый за единицу. Рассмотрим дискретные моменты времени, отделенные друг от друга этим интервалом. Тогда в t -й момент системы знают положения друг друга в $(t-1)$ -й момент. Примем, что в t -й момент системы A_C и B_C движутся вдоль прямой $\mathcal{L}(t-1)$, соединяющей их положения в момент $t-1$, имея нормальные отклонения от нее $y_A(t)$ и $y_B(t)$ (рис. 4.1).

При соответствующей нормировке можно добиться того, чтобы

$$-\frac{1}{2} \leq y_A(t), y_B(t) \leq \frac{1}{2}. \quad (4.42)$$

Будем проектировать отрезок $y_A(t)$ на линию $\Pi(t)$, содержащую отрезок $y_B(t)$. Пусть целью системы A_C является попадание своей проекцией A'_C в ε -окрестность системы B_C (см. рис. 4.1). Оправданием такой цели

системы A_C может явиться ее боязнь «проскочить» линию $\Pi(t)$ на значительном удалении от системы.

В соответствии с этим зададим функцию выигрыша в виде

$$M[y_A(t), y_B(t)] = \begin{cases} 1, & \text{если } |y_A(t) - y_B(t)| \leq \epsilon \text{ и } y_A(t)y_A(t-1) \geq 0, \\ b, & \text{если } |y_A(t) - y_B(t)| \leq \epsilon \text{ и } y_A(t)y_A(t-1) < 0, \\ a, & \text{если } |y_A(t) - y_B(t)| > \epsilon, \end{cases} \quad (4.43a)$$

который учитывает перемену знака отклонений $y_A(t)$ системы A_C в соседние моменты времени.

Полагая $a < b < 0$ [$\lambda > 0$, см. (4.18)], мы этим указываем на нежелательность для A_C из-за ее малой маневренности перемены знаков соседних отклонений [см. (4.43a)]. То, что $a < b$, указывает на еще большую нежелательность для системы A_C иметь проекцию A'_C вне

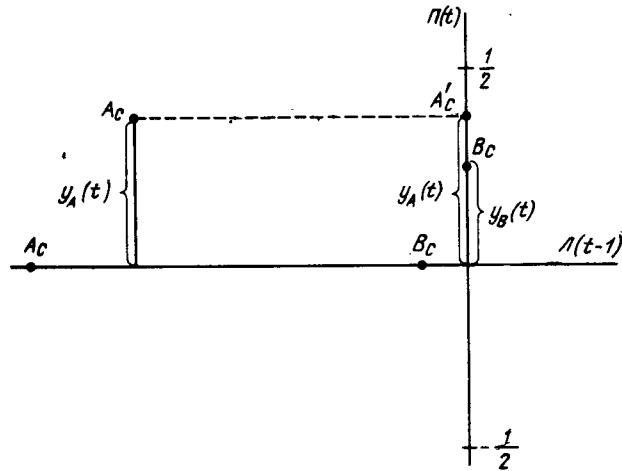


Рис. 4.1. Схема игры преследования.

ϵ -окрестности системы B_C . Если число перемен знаков $y_A(t)$ лимитировано, то, индуцируя их, система B_C имеет шансы уйти от преследования системой A_C .

Итак, картина преследования представляется следующей. Происходит регулярное сближение систем A_C и B_C со скоростью $v = v_A - v_B > 0$. На каждом t -м шаге,

не оповещая друг друга, системы A_C и B_C выбирают знаки и величины $y_A(t)$ и $y_B(t)$ в пределах (4.42).

Таким образом, происходит игра на единичном квадрате с вершинами в точках $(-\frac{1}{2}, -\frac{1}{2})$, $(-\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, -\frac{1}{2})$ и $(\frac{1}{2}, \frac{1}{2})$. Если имело место некоторое значение $y_A(t-1)$, то условие (4.43a) просто всякий раз указывает интервал значений $y_A(t)$ $(-\frac{1}{2}, 0)$ или $(0, \frac{1}{2})$, при котором выпол-

няется первое или второе условие. При этом длина каждого из этих интервалов равна $1/2$. Легко видеть, что функция выигрыша M (4.28) является m -мерным обобщением функции выигрыша (4.43).

Случай $b=1$ ($\lambda=1$), приводящий к бинарной игре, можно интерпретировать вторым этапом преследования — попытками прямых контактов системы A_C с системой B_C . При этом считается, что они находятся в контакте, если система A_C попала в ϵ -окрестность системы B_C . Имеющие место при этом оптимальные стратегии — равномерные распределения — также находят интуитивное оправдание.

§ 4.4. ПРЕДЕЛЬНЫЙ ЗАКОН УПРАВЛЯЕМОСТИ

4.4.1. Формулировка и доказательство предельного закона управляемости. Ряд оценок, основанных на результатах фон Неймана [17] и Беллмана [6], проведенных в предыдущих параграфах, в частности для простейшего случая бинарной симметричной игры, позволяют установить

Предельный закон управляемости [7]. Пусть в дискретные моменты времени $1, 2, \dots, t, \dots$ происходит многошаговая не безобидная игра между системами A_C и B_C с локальной бинарной симметричной функцией выигрыша $M(x, y) = 1$, $a \neq 1$, за обладание ограниченным количеством ресурсов $K = \text{const}$. Пусть до начала игры эти ресурсы распределены между системами A_C и B_C в количествах K_A и K_B ($K_A + K_B = K$), причем

$$|M(x, y)| \leq \min(K_A, K_B).$$

Тогда в зависимости от времени игры t вероятность разорения одной из систем имеет оценки

$$P_c(t) \begin{cases} \geq 1 - 2^{-kt+O(\log t)} & \text{при } K_a < |H_c|t, \\ < 2^{-kt+O(\log t)} & \text{при } K_a > |H_c|t, \end{cases} \quad (4.44)$$

где

$$k = k(\theta_1, p) = \theta_1 \log \frac{\theta_1}{p} + (1 - \theta_1) \log \frac{1 - \theta_1}{1 - p}; \quad (4.45)$$

$$0 \leq \theta_1 = \frac{K_a/t - a}{1 - a} \leq 1, \quad p = 1 - \mathcal{P}_a; \\ a = A, B.$$

При этом для случая $0 < H_c < C_c$ разоряется система B_c ($\alpha = B$), а для случая $C_c < H_c < 0$ разоряется система A_c ($\alpha = A$), где $H_c = EM(x, y)$ при произвольных смешанных стратегиях p_A и p_B , а $C_c = EM(x, y)$ при оптимальных в минимаксном смысле смешанных стратегиях, которые оказываются в рассматриваемом случае равномерными распределениями (величины \mathcal{P}_a и C_c определены соотношениями (4.39) и (4.40) соответственно).

Доказательство. Начнем с доказательства неравенств соотношения (4.44) при $0 < H_c < C_c$, когда оценивается вероятность разорения системы B_c . В самом деле, используя обозначения (4.6), запишем случайное число ресурсов системы B_c после u игр в виде

$$K_{Bu} = K_B + \kappa_{Bu}, \quad (4.46)$$

где

$$\kappa_{Bu} = -\kappa_{Au} = -\sum_{s=1}^u M(x_s, y_s).$$

Определим случайное число τ игр следующим образом. Пусть $\tau = u$, если для $s = \overline{1, u-1}$ имеем

$$K_B + \kappa_{Bs} > 0 \geq K_B + \kappa_{Bu}$$

или эквивалентные им неравенства для $s = \overline{1, u-1}$:

$$\kappa_{As} < K_B \leq \kappa_{Au}.$$

Положим $\kappa_{A0} = 0$ и $K_B > 0$. Тогда, введя несовместимые события

$$A_u = \{\kappa_{As} < K_B \leq \kappa_{Au}, \quad s = \overline{1, u-1}\} (u = \overline{1, t}),$$

можно записать

$$\mathcal{P}(\tau = u) = \mathcal{P}(A_u).$$

Введем события

$$\tilde{A}_u = \{K_B \leq \kappa_{Au}\} (u = \overline{1, t}). \quad (4.47)$$

Тогда легко показать, что $\bigcup_{u=1}^t A_u \supseteq \tilde{A}_t$. Поэтому

$$\mathcal{P}(\tau \leq t) = \mathcal{P}\left(\bigcup_{u=1}^t A_u\right) \geq \mathcal{P}(\tilde{A}_t).$$

С другой стороны, используя очевидные соотношения $A_u \subseteq \tilde{A}_u$, будем иметь

$$\mathcal{P}(\tau \leq t) = \mathcal{P}\left(\bigcup_{u=1}^t A_u\right) = \sum_{u=1}^t \mathcal{P}(A_u) \leq \sum_{u=1}^t \mathcal{P}(\tilde{A}_u).$$

Итак, имеем оценки

$$\mathcal{P}(\tilde{A}_t) \leq \mathcal{P}(\tau \leq t) \leq \sum_{u=1}^t \mathcal{P}(\tilde{A}_u).$$

Но при $H_c > 0$ с ростом u растет и вероятность $\mathcal{P}(\tilde{A}_u)$, поэтому мы лишь усилим оценки, записав

$$\mathcal{P}(\tilde{A}_t) \leq \mathcal{P}(\tau \leq t) \leq t \mathcal{P}(\tilde{A}_t). \quad (4.48)$$

Учитывая представление (4.46) случайной величины κ_{At} и бинарный характер случайной величины $M = \begin{Bmatrix} 1 & a \\ \mathcal{P}_1 & \mathcal{P}_a \end{Bmatrix}$, получим

$$\kappa_{At} = \mu + a(t - \mu) = (1 - a)\mu + at,$$

где μ — случайное число значений $M = 1$ в t играх.

Отсюда

$$\mu = \frac{\kappa_{At}/t - a}{1 - a} t \left(0 \leq \frac{\kappa_{At}/t - a}{1 - a} \leq 1\right)$$

и событие $\tilde{A}_t = \{\kappa_{At} \geq K_B\}$ эквивалентно событию

$$\tilde{A}_t = \{\mu \geq \theta_1(t)\}, \quad \text{где } 0 \leq \theta_1 = \frac{K_B/t - a}{1 - a} \leq 1.$$

Теперь согласно оценкам (1.10) при двоичных логарифмах имеют место неравенства (4.44) в зависимости от неравенств $\theta_1 \lesseqgtr p$, где в рассматриваемом случае $p = \mathcal{P}_1 = 1 - \mathcal{P}_a$.

Но неравенства $\theta_1 \leq 1 - \mathcal{P}_a$ эквивалентны неравенствам $\frac{K_B/t - |a|}{1-a} \leq 1 - \mathcal{P}_a$, которые в свою очередь эквивалентны неравенствам $K_B \leq [(1 - \mathcal{P}_a)(1 - a) + a]t = H_c t$ [см. соотношение (4.37)]. Таким образом, утверждение предельного закона для случая $0 < H_c < C_c$ доказано полностью.

Аналогично оно доказывается и для случая $C_c < H_c < 0$.

4.4.2. Обсуждение предельного закона управляемости.

Условие, при котором был выведен предельный закон $|M(x, y)| \ll \min(K_A, K_B)$, сводит решение многошаговой игры на разорение к решению одношаговой игры. Поэтому вопрос о конструктивности предельного закона управляемости сводится к вопросу о конструктивном решении игры с заданной функцией выигрыша $M(x, y)$. Для простейшей бинарной симметричной функции выигрыша, для которой доказывался предельный закон, этот вопрос решается положительно.

Однако для сколько-нибудь более сложных функций выигрыша решение матричных и непрерывных игр неконструктивно. Таким образом, как правило, предельный закон управляемости является обычной теоремой существования оптимального поведения (смешанных стратегий) систем A_c и B_c .

Фундаментальная константа C_c — величина одношаговой игры — имеет смысл коэффициента линейной функции среднего накопления ресурсов от средней продолжительности игры соответствующей системы. Одновременно эта величина определяет минимальное среднее время, необходимое для выигрыша системой многошаговой игры. Таким образом, по терминологии, принятой во введении, (C_c, t) — обмен между системой A_c и B_c является оптимальным.

§ 4.5. ТЕОРИЯ ИГР. ПРИМЕРЫ

4.5.1. Теория игр. Предыдущее изложение показывает, что теория потенциальной эффективности сложных систем использует игровые модели и сама оказывает «системное» влияние на теорию игр. Из четырех простейших качеств системы: надежности (R), помехоустойчивости (I), управляемости (C) и самоорганизации (L) третье качество — управляемость (C) является просто

системной интерпретацией многошаговой игры на разорение.

Вместе с тем, при рассмотрении смешанных качеств с участием C , таких, как RC -живучесть и IC -управляемость при наличии шумов (см. гл. 6), возникают новые задачи для самой теории игр. При этом ее модели должны быть общи и просты для конструктивности результатов. Это часто приводит к задачам, решение которых находится на грани между тривиальным и невозможным. Примером «слишком конкретных» моделей здесь могут служить модели дифференциальных игр, связанные со спецификой механического движения.

Теория потенциальной эффективности сложных систем вносит структурный элемент в свои игровые модели. Согласно этой теории система $A = (|A|; \bar{A})$ определяется своей структурой $|A|$ и поведением \bar{A} . «Сечение» системы A по ее C -качеству $A_c = (|A_c|, \bar{A}_c)$ приводит к указанной выше многошаговой модели, структурно-аморфной, как и все модели теории игр (оптимизируется лишь поведение \bar{A}_c — стратегия в смысле теории игр). Однако уже «двумерное сечение» $A_{RC} = (|A_{RC}|, \bar{A}_{RC})$ приводит к модели живучести с оптимизацией и по структуре $|A_{RC}|$ и по поведению \bar{B} .

Большое стимулирующее значение для выбора и разумного ограничения моделей сложных систем имеет анализ систем новой техники, биологии, экономики и социологии. В частности, для игровых моделей важны социологические экспериментальные исследования причин и механизма развития и спада конфликтов на разных уровнях, начиная с общих квартир и до международных масштабов. Здесь прежде всего необходимо экспериментальное различие нерегулируемых и регулируемых факторов (на наличие первых указывает, например, постоянство процента писем, отправляемых без адреса).

Специалистам в области «честной» теории игр не следует упускать из вида зарождающуюся теорию игр «без правил», изучающую явления блефов, обманов и тому подобных эффектов. Наиболее точные постановки задач здесь содержатся в теории рефлексивных игр [19]. Последняя имеет ряд эффективных решений в этой области.

Сопоставление классической информационной модели $A_I = (|A_I|, \bar{A}_I)$ Котельникова — Шеннона с рассмотренной моделью $A_c = (|A_c|, \bar{A}_c)$ демонстрирует большую

(все еще не исчерпанную) глубину первой по сравнению со второй. В этом нет ничего удивительного, так как первая оптимизируется и по структуре $|A_I|$ — способу кодирования и по поведению A_I — способу декодирования, а вторая оптимизируется лишь по поведению A_C — стратегии. Структурное углубление игровых моделей встречает большие комбинаторные трудности.

4.5.2. Примеры. Рассмотрим некоторые примеры использования соотношений (4.44) и (4.45) предельного закона управляемости для матричного случая. Здесь в оптимальном случае при величине игры $C_C > 0$ разоряется система B_C , а при $C_C < 0$ — система A_C . Поэтому с учетом соотношения (4.40) имеем для

$$a \begin{cases} > -1 & \text{разорения системы } B_C, \\ < -1 & \text{разорение системы } A_C, \end{cases} \quad (4.49)$$

Далее для оценки близкой к единице вероятности этих событий согласно соотношению (4.44) необходимо иметь

$$\frac{K_\alpha}{t} < |C_c| = \left| \frac{a+1}{2} \right| \begin{cases} < a & \text{при } a > 1, \\ > a & \text{при } a < 1 \end{cases} \quad (4.50)$$

и из соотношения (4.45)

$$0 \leq \theta_1 = \frac{K_\alpha/t - a}{1 - a} \leq 1, \quad (4.50')$$

где $\alpha = A, B$.

Поэтому из соотношения (4.50') имеем

$$a \leq K_\alpha/t \leq 1 \quad \text{при } a < 1 \\ \text{и} \\ 1 \leq K_\alpha/t \leq a \quad \text{при } a > 1. \quad (4.51)$$

Приведенные оценки указывают ограничения на выбор величин K_α и t в зависимости от параметра a .

Пример 4.1. Пусть $a=2$, $K_B=21$ и $t=20$. Оценить вероятность разорения системы B_C .

Решение. Соотношение (4.49) указывает на то, что при $a=2$ разоряется именно система B_C . Вычислив величину $K_B/t=1,05$, видим, что согласно соотношению (4.45) $\theta_1=0,95$.

Так как в рассматриваемом оптимальном случае $p=1-\mathcal{P}_a=0,5$, то согласно графикам рис. 3.3 имеем

$$k=k(\theta_1, p)=k(0,95; 0,5)=0,5$$

и согласно соотношению (4.44)

$$P_C(t) \geq 1 - 2^{-kt} = 1 - 2^{-0,5 \cdot 20} = 1 - 2^{-10},$$

Пример 4.2. Рассчитать, полагая, как и в предыдущем примере, $a=2$, количество ресурсов K_B системы B_C и число шагов t , чтобы при фиксированном значении $\theta_1=0,75$ вероятность разорения системы B_C оценивалась снизу величиной $1-2^{-24}$.

Решение. По графикам рис. 3.3 находим величину

$$k=k(\theta_1, p)=k(0,75; 0,05)=0,12.$$

Далее из соотношения

$$kt=0,12 \cdot t=24$$

находим величину $t=200$.

Используя соотношение (4.50') для выражения K_B через θ_1 , a и t , получим его значение

$$K_B=t[a-(a-1)\theta_1]=200 \cdot 1,25=250.$$

5. Самоорганизация (L-качество)

§ 5.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Самоорганизацию системы A_L можно определить как ее способность менять во времени свою структуру $|A_L|$ и поведение A_L с целью приспособления к изменяющимся условиям среды. Указанное свойство в полной мере проявляется у самоорганизующихся систем (см. введение). Однако в зачаточной форме оно может быть и у решающих систем.

Решающая система определялась как система с постоянной во времени структурой и постоянным, вообще говоря, стохастическим поведением — реакциями на внешние воздействия среды. Постоянство стохастического поведения здесь понимается в том смысле, что распределения вероятностей, определяющие алгоритмы поведения (например, смешанные стратегии), задаются заранее и не меняются в связи с изменяющимися условиями среды. Вместе с тем все известные биологические системы обладают свойством самоорганизации, так что это свойство можно считать определяющим для биологических систем.

В настоящее время рассмотрено много моделей искусственных самоорганизующихся систем как теоретических, так и технически реализованных (например, мышь Шеннона).

Имеется несколько различных теоретических подходов к теории самоорганизации, в частности обучения:

структурный [20] и поведенческий [21] *), однако здесь мы еще весьма далеки от развития этих теорий в плане теории потенциальной эффективности сложных систем. Это связано с тем, что пока в теории самоорганизации не сформулированы такие критерии качества самоорганизующихся систем и широкие классы их моделей, чтобы можно было найти или хотя бы доказать существование оптимальных систем по выбранному критерию (оптимальный синтез).

В настоящее время в большинстве работ заранее предлагаются некоторые «разумные» структуры и поведение самоорганизующейся системы, а затем экспериментально и (или) теоретически оценивается ее эффективность (анализ заданной структуры и поведения).

Нет единства мнения и о законченности теории оптимального поведения аморфных по структуре адаптивных систем [22]. И все же можно выделить узкую область теории самоорганизации, где наряду с общей постановкой задачи имеется частное ее решение. Речь идет о теории различения сигналов на фоне шумов.

§ 5.2. ОБУЧЕНИЕ РАЗЛИЧЕНИЮ СИГНАЛОВ НА ФОНЕ ШУМОВ

Задачу обучения различению сигналов на фоне шумов можно трактовать как естественное обобщение задачи различения сигналов на фоне шумов, рассмотренной в гл. 3.

В самом деле, пусть системе A_L необходимо различить в индифферентной ситуации M входных слов длины t на фоне шумов. Но в отличие от системы A_I в постановке задачи гл. 3 системе A_L не известна вероятностная характеристика шумов, определяемая в случае БСК вероятностью p искажения бинарных символов.

Если задана вероятность p , M входных слов длины

$$x_1, \dots, x_u, \dots, x_v, \dots, x_M \quad (5.1)$$

и выходное слово y той же длины, то этим задано M условных вероятностей $P_u(y)$ появления y на выходе при

* Прекрасным введением в круг рассматриваемых здесь вопросов являются соответствующие разделы книги Б. Р. Левина «Статистическая радиотехника», ч. II, Изд-во «Советское радио», 1969.

x_u на входе, имеющих вид

$$P_u(y) = p^{d_u} (1-p)^{t-d_u},$$

где $d_u = d(x_u, y)$ — хэмминговское расстояние между x_u и y . Постановка задачи в гл. 3 сводилась к определению структуры $|A_I|$ (выбора x -ов — способа кодирования) и поведения \bar{A}_I (декодирования y) системы A_I , при котором сколь угодно достоверно декодируется предельно большое число M входных слов длины t .

В рассматриваемом случае вероятность p неизвестна, а значит неизвестны и M вероятностей $P_u(y)$. Вместо них системе A_L задано $N = \sum_{u=1}^M N_u$ «обучающих» выходных слов

длины t , объединенных в M выборок объемов $N_u (u = \overline{1, M})$:

$$y_{11}, \dots, y_{1N_1}; \dots; y_{u1}, \dots, y_{uN_u}; \dots; y_{M1}, \dots, y_{MN_M}, \quad (5.2)$$

о которых ей заранее известно, что они представляют соответственно M «зашумленных» входных слов (5.1). Требуется найти такую структуру $|A_L|$ [выбор M входных слов (5.1) — способ кодирования] и поведение \bar{A}_L [способ совместной обработки обучающих выходных слов (5.2) и выходного слова y], чтобы система A_L сколь угодно достоверно декодировала (классифицировала) предельно большое число M входных слов длины t .

Ясно, что вторая постановка задачи содержит первую постановку задачи, как частный случай, при котором $N_u \rightarrow \infty (u = \overline{1, M})$. В этом случае M «бесконечных» выборок обучающих выходных слов (5.2) точно определяют M распределений $P_u(y)$, что приводит к постановке первой задачи.

Простота решения частной задачи оправдывает более простую постановку задачи в общем случае. Поэтому можно пытаться решить аморфный по структуре второй вариант задачи, когда не требуется оптимизировать систему A_L по структуре $|A_L|$ (последняя считается заданной и неизменной). Но даже в такой ослабленной постановке решение задачи все еще слишком сложно.

Поэтому можно пытаться решать эту задачу в еще более частной постановке для $M=2$. В последней постановке для случая $t=1$ и так называемых «нормальных» шумов решение этой задачи было получено П. П. Данковым [23]. Оно основывается на результатах математи-

ческой статистики. Здесь этого решения мы не будем обсуждать.

Единственной целью настоящего параграфа являлась точная постановка задачи обучения различению сигналов на фоне шумов, как обобщения задачи без обучения. Это позволило оценить трудности решения указанной задачи.

Из довольно многочисленных результатов в области обучения различению сигналов есть единственный результат [24], носящий по своей форме общий предельный характер законов потенциальной эффективности. Он касается установления принципиальных пределов возможности различения $M=2$ сигналов, представленных на стадии обучения N выходными словами длины t , широким классом систем, обладающих n «степенями свободы». Рассматривая число n как единственный параметр структуры $|A_L|=n$ системы A_L , ставится задача оценки потенциальной возможности еще на стадии обучения произвести разбиение на два класса (дихотомии) всего пространства выходных слов, в том числе и N обучающих выходных слов, имеющих произвольное расположение.

§ 5.3. ПРЕДЕЛЬНОЕ ЧИСЛО ДИХОТОМИЙ

5.3.1. Геометрическая интерпретация и некоторые определения. Будем вести все рассмотрения в t -мерном евклидовом пространстве, интерпретируя множество $Y = \{y_s\}$ ($s = \overline{1, N}$) обучающих слов $y_s = (y_{s1}, \dots, y_{st})$ t -мерными точками в t -мерном евклидовом пространстве R^t .

Ясно, что всевозможных дихотомий множества Y , состоящего из N точек, имеется 2^N , так как для каждой точки независимо от других представляются две возможности (принадлежать к одному или другому классу дихотомии).

При произвольном расположении точек множества Y в пространстве R^t дихотомии могут приводить к весьма «причудливым» их классификациям. Если дихотомии производятся простыми образованиями, такими, например, как $t-1$ -мерные гиперповерхности, то вовсе не очевидно возможности осуществления ими произвольно заданной классификации. Например, для плоского случая $t=2$ и $N=4$ всего имеется $2^4=16$ дихотомий. Однако с помощью $(t-1)$ -мерных (в нашем случае одномерных гиперплоскостей (прямых)) можно произвести лишь 14 дихотомий (рис. 5.1,а), так как имеются две возможные

классификации для каждого из семи возможных разделений множества Y прямыми l_i ($i = \overline{1, 7}$) на две части (за исключением случая коллинеарности каких-либо трех точек из множества Y , рис. 5.1,б). В последнем случае возможно всего лишь $2 \cdot 6 = 12$ дихотомий.

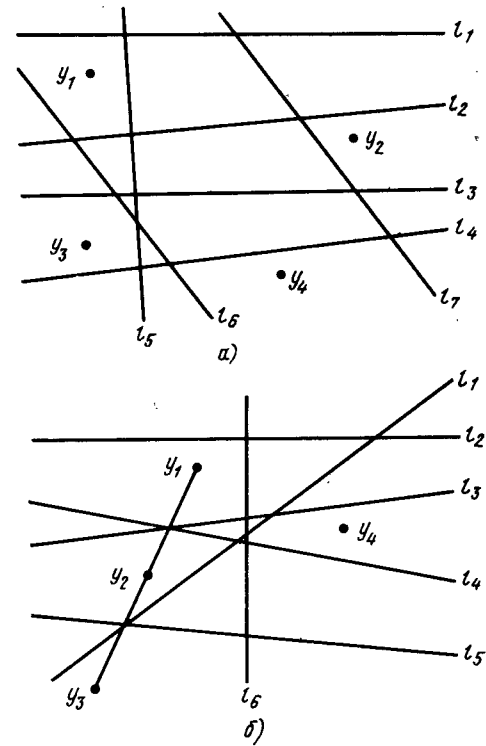


Рис. 5.1. Возможные дихотомии для плоского случая ($t=2$) четырех точек ($N=4$):

а — без коллинеарности каких-либо трех точек; б — при наличии коллинеарности трех точек.

Рассмотренный пример указывает на целесообразность введения следующего определения.

Мы скажем, что N слов множества $Y = \{y_s\} \subset R^t$ имеют расположение *общего типа*, если при $N > t$ ни одно подмножество $Y' \subset Y$, состоящее из $(t+1)$ -го слова, не лежит в $t-1$ -мерной гиперплоскости, а если $N \leq t$, то ни

одна из $(N-2)$ -мерных гиперплоскостей не должна содержать множество Y .

Линейной дихотомией N точек в t -мерном евклидовом пространстве R^t будем называть дихотомию N точек с помощью $(t-1)$ -мерной гиперплоскости. Число таких дихотомий будем обозначать $L(N, t)$. Это число равно удвоенному числу всевозможных разбиений N точек на два подмножества с помощью $(t-1)$ -мерных гиперплоскостей (удвоение связано с возможностью двойной классификации каждого разбиения множества Y на две части).

5.3.2. Вычисление числа $L(N, t)$. Для дальнейшего изложения необходимо вычисление числа $L(N, t)$. Для этого составим конечно-разностное уравнение, определяющее $L(N, t)$. Пусть множество Y' состоит из $(N-1)$ точки с расположением общего типа. Тогда они имеют всего $L(N-1, t)$ линейных дихотомий. На сколько увеличится это число, если добавить еще одну точку y_N так, что у образовавшихся N точек Y сохранится расположение общего типа?

Ясно, что некоторые дихотомии можно получить проведением гиперплоскостей через добавленную точку y_N . Их число обозначим $L_{y_N}(N-1, t)$ из общего числа $L(N-1, t)$. Тогда

$$L(N-1, t) - L_{y_N}(N-1, t)$$

дихотомий нельзя получить, исходя из точки y_N . Каждая из последних дихотомий Y' определяет одну дихотомию в Y (так как и y_N в нее попадает и через y_N гиперплоскость не проходит).

Но каждой дихотомии, полученной проведением гиперплоскости через точку y_N , соответствуют две классификации множества Y (в одной точка y_N принадлежит одному, а в другой — другому классу при незначительном смещении гиперплоскости).

Тогда общее число линейных дихотомий множества точек Y задается выражением

$$L(N, t) = L(N-1, t) - L_{y_N}(N-1, t) + 2L_{y_N}(N-1, t),$$

отсюда

$$L(N, t) = L(N-1, t) + L_{y_N}(N-1, t). \quad (5.3)$$

Покажем теперь, что

$$L_{y_N}(N-1, t) = L(N-1, t-1). \quad (5.4)$$

Для этого проведем через точку y_N и через каждую точку Y' по линии. Всего их будет $N-1$ и все они различны, так как точки в Y имеют расположение общего типа. Выберем некоторую гиперплоскость H , пересекающую все эти линии (она, вообще говоря, выбирается неоднозначно). Обозначим получившееся множество из $(N-1)$ -й точки пересечения через $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_{N-1}\}$.

Ясно, что число классификаций в Y , образуемых $t-1$ -мерными гиперплоскостями, проходящими через точку y_N , равно числу классификаций множества \mathcal{P} , образуемых $t-2$ -мерными гиперплоскостями в гиперплоскости H . Но H является $t-1$ -мерной гиперплоскостью, поэтому число линейных классификаций в ней равно $L(N-1, t-1)$, что и доказывает соотношение (5.4).

Подставив соотношение (5.4) в соотношение (5.3), получим окончательно конечно-разностное уравнение

$$L(N, t) = L(N-1, t) + L(N-1, t-1) \quad (5.5)$$

для определения $L(N, t)$.

Заметим, что

$$L(1, t) = 2 \text{ и } L(N, 1) = 2^N \quad (5.6)$$

и при $t \geq N$

$$L(N, t) = 2^N. \quad (5.7)$$

Таким образом, нам остается найти значение $L(N, t)$ для t в пределах $1 < t < N$.

Перейдем от значений функции $L(N, t)$ от t к ее разностям

$$\Delta L(N, t-1) = L(N, t) - L(N, t-1), \quad (5.8)$$

где $t = \overline{2, N}$, причем при $t > N$ согласно (5.7)

$$\Delta L(N, t) = 0. \quad (5.9)$$

Легко видеть, что конечно-разностное уравнение (5.5) относительно (5.5) переходит в конечно-разностное уравнение

$$\Delta L(N, t-1) = \Delta L(N-1, t-1) + \Delta L(N-1, t-2) \quad (5.10)$$

для $\Delta L(N, t-1)$ при $t = \overline{2, N}$.

Легко проверить, что биномиальные коэффициенты удовлетворяют соотношению

$$C_N^t = C_{N-1}^t + C_{N-1}^{t-1}. \quad (5.11)$$

Поэтому с точностью до постоянного множителя A решение уравнения (5.10) можно искать в виде

$$\Delta L(N, t-1) = AC_{N+\delta}^t. \quad (5.12)$$

Для определения величин A и δ воспользуемся условиями (5.6) и (5.7), полагая $L(N, 0) = 2$, тогда

$$\Delta L(N, 0) = AC_{N+\delta}^1 = L(N, 1) - L(N, 0) = 2N - 2.$$

Отсюда $A(N + \delta) = 2(N - 1)$, что может быть при любых N лишь в случае $A = 2$ и $\delta = -1$.

Итак,

$$\Delta L(N, t-1) = 2C_{N-1}^t. \quad (5.13)$$

Поэтому

$$L(N, t) = \sum_{s=0}^t \Delta L(N, s) = 2 \sum_{s=0}^t C_{N-1}^s. \quad (5.14)$$

Итак [24],

$$L(N, t) = \begin{cases} 2 \sum_{s=0}^t C_{N-1}^s & \text{при } t = \overline{1, N}, \\ 2^N & \text{при } t > N. \end{cases} \quad (5.15)$$

5.3.3. Обобщение линейной дихотомии на общий случай Φ -дихотомии. Рассмотрим вместо гиперплоскостей в t -мерном евклидовом пространстве R^t , определяемых уравнением

$$z = l(\bar{y}) = \tilde{\omega}_1 y_1 + \tilde{\omega}_2 y_2 + \dots + \tilde{\omega}_t y_t + \tilde{\omega}_{t+1}, \quad (5.16)$$

гиперповерхности, определяемые уравнением

$$z = \Phi(\bar{y}) = \omega_1 f_1(\bar{y}) + \omega_2 f_2(\bar{y}) + \dots + \omega_n f_n(\bar{y}) + \omega_{n+1}, \quad (5.17)$$

где $\omega_1, \dots, \omega_{n+1}$ — произвольные постоянные, а $f_1(\bar{y}), \dots, f_n(\bar{y})$ — линейно независимые, действительные, однозначные, не зависящие от ω_i функции векторного аргумента $\bar{y} = (y_1, \dots, y_t)$.

Например, при $n = t$ и $f_i(\bar{y}) = y_i$ и $f_i(\bar{y}) = y_i^2$ имеем линейную и квадратичную (сферическую) функции $f_i(\cdot)$.

В общем случае квадратичной функции $f_i(\cdot)$ $n = t + t(t-1)/2 + t = t(t+3)/2$, так как всего имеем t квадратичных, $t(t-1)/2$ смешанных и t линейных функций $f_i(\bar{y})$.

Линейная независимость n функций $f_i(\bar{y})$ означает непредставимость каждой из них в виде линейной комбинации других. Если n функций $f_i(\bar{y})$ линейно зависимы, то, выражая линейно одну из них через другие, можно было бы уменьшить число слагаемых в выражении (5.17). Так как в рассматриваемом случае этого сделать нельзя из-за их линейной независимости, то их число n является характерным параметром задачи.

Класс функций, определяемых соотношением (5.17), когда набор n функций с указанными выше свойствами

$$F(\bar{y}) = \{f_1(\bar{y}), \dots, f_n(\bar{y})\} \quad (5.18)$$

фиксирован, а набор $n+1$ параметров

$$\omega = \{\omega_1, \dots, \omega_{n+1}\}, \quad (5.19)$$

называемых *весами*, переменный, назовем *классом Φ -функций* с $(n+1)$ -й степенью свободы. Представителей этого класса функций будем называть *Φ -функциями*. Гиперповерхность в t -мерном пространстве R^t , определяемую уравнением (5.17), будем называть *Φ -гиперповерхностью*, а производимую ею дихотомию точек множества Y будем называть *Φ -дихотомией*.

Если не существует Φ -гиперповерхности в пространстве R^t , содержащей $n+1$ точку из Y , то будем говорить, что точки из Y имеют *расположение общего типа* в смысле Φ .

Вычислим число $\Phi(N, t)$ Φ -дихотомий множества Y . Так как каждой точке \bar{y} в t -мерном пространстве R^t соответствует точка $F(\bar{y}) = \{f_1(\bar{y}), \dots, f_n(\bar{y})\}$ в n -мерном пространстве R^n , то множеству Y из N точек, имеющих расположение общего типа в смысле Φ в пространстве R^t , соответствует множество \mathcal{F} из N точек, имеющих расположение общего типа в пространстве R^n . Так как любая линейная дихотомия множества \mathcal{F} в пространстве R^n соответствует Φ -дихотомии множества Y в пространстве R^t , то $\Phi(N, t)$ равно числу $L(N, n)$ линейных дихотомий множества \mathcal{F} из N точек в пространстве R^n .

Таким образом, если точки из Y имеют расположение общего типа в смысле Φ , то число дихотомий множества Y , выполняемое любым классом Φ -функций, равно

$$\Phi(N, n) = L(N, n) = \begin{cases} 2 \sum_{i=0}^n C_{N-1}^i, & \text{если } N > n, \\ 2^N, & \text{если } N \leq n. \end{cases} \quad (5.20)$$

§ 5.4. ПРЕДЕЛЬНЫЙ ЗАКОН КЛАССИФИКАЦИИ. ПРИМЕРЫ

5.4.1. Предельный закон классификации. В общей постановке задачи обучения различению M сигналов на фоне шумов, когда задана обучающая выборка выходных слов объема N , решающая система $A_L = (|A_L|, \bar{A}_L)$ должна иметь некоторую структуру $|A_L|$ и поведение \bar{A}_L . Предыдущие рассуждения можно интерпретировать как попытку анализа предельных возможностей классификации обучающей выборки на этапе обучения. При этом можно считать широкий класс Φ -функций с $(n+1)$ -й степенью свободы элементом структуры $|A_L|$, определяемым одним параметром n . Внутри такого класса каждая Φ -функция определяется набором весов (w_1, \dots, w_{n+1}) . Этот элемент структуры $|A_L|$ будем называть Φ -структурой. Здесь возникает следующая задача, оценивающая предельные возможности обучения системы A_L с Φ -структурой классификации N точек.

В самом деле, пусть в индифферентной ситуации для любых N t -мерных точек из пространства R^t среда *равновероятно* может назначить любую из 2^N возможных их классификаций. Тогда при фиксированном наборе $F(\bar{y}) = \{f_1(\bar{y}), \dots, f_n(\bar{y})\}$ Φ -структуры пробегая Φ -класс при всевозможных значениях весов $\{w_1, \dots, w_{n+1}\}$. Если N точек имеют расположение общего типа в смысле Φ , то Φ -структуры из Φ -класса могут произвести всего $L(N, n)$ классификаций. Поэтому вероятность $P(N, n)$ существования Φ -структуры, т. е. существования такого набора весов $\{w_1, \dots, w_{n+1}\}$, при котором система A_L осуществит выбранную случайно классификацию, равна $P(N, n) = L(N, n) \cdot 2^{-N}$ или с учетом (5.20) имеем:

$$P(N, n) = \begin{cases} \sum_{i=1}^n C_{N-1}^i \left(\frac{1}{2}\right)^{N-1}, & \text{если } n < N, \\ 1, & \text{если } n \geq N. \end{cases} \quad (5.21)$$

Положим $N = n/\theta$. Тогда используя соотношение (1.10), получим из (5.21)

Предельный закон классификации. Вероятность существования Φ -структуры $|A_L|$ системы A_L , определяемой числом степеней свободы n , которая классифицирует N точек с расположением общего типа в t -мерном пространстве, имеет оценки

$$P(N, n) = \begin{cases} > 1 - 2^{-kN} & \text{при } N < 2n, \\ < 2^{-kN} & \text{при } N > 2n, \end{cases} \quad (5.22)$$

где $k = k(\theta; 0,5) = \theta \log \frac{\theta}{0,5} + (1 - \theta) \log \frac{1 - \theta}{0,5} =$

$$= 1 + \theta \log \theta + (1 - \theta) \log (1 - \theta) \quad \text{и} \quad \theta = n/N.$$

Если учесть зависимость величины n от размерности пространства выборок R^t (см. п. 5.3.3), то в линейном и сферическом случаях задания функций $f_i(y)$ будем иметь $N = n/\theta = t/\theta$, а в общем квадратичном случае задания функций $f_i(y)$ $N = n/\theta = t(t+3)/2\theta$. По нашей терминологии (см. введение) здесь мы имеем дело с (N, n) -обменом. Причем $\left(\frac{1}{2} n, n\right)$ -обмен является оптимальным, а соотношение (5.22) устанавливает предельный закон в рассматриваемом случае дихотомий N точек системой A_L с Φ -структурой, имеющей $(n+1)$ -ю степень свободы.

До сих пор не найдено число линейных классификаций N точек на $M > 2$ классов [24]. Очевидно, что без решения этой задачи невозможно обобщение предельного закона (5.22) со случая $M=2$ на случай $M > 2$.

Результаты этой главы, начиная с § 5.3, кроме соотношения (5.22), получены Ковером и Эфроном и изложены ими в недоступных нам изданиях. Здесь эти результаты приведены в том виде, как они изложены в монографии [24].

5.4.2. Примеры. Рассмотрим примеры использования предельного закона классификации, определяемого соотношением (5.22).

Пример 5.1. Пусть $N=100$ точек имеют расположение общего типа. Оценить вероятность $P(N, n)$ их классификации системой A_L с $n+1=91$ степенями свободы.

Решение. Здесь величина $\theta=n/N=0,9$. Отсюда, используя графики рис. 3.3, получим

$$k=k(0,9; 0,5)=0,36.$$

Поэтому, используя соотношение (5.22), имеем оценку

$$P(N, n) > 1 - 2^{-kN} = 1 - 2^{-0,36 \cdot 100} = 1 - 2^{-36}.$$

Пример 5.2. Оценить число степеней свободы n системы A_L , классифицирующей N точек, имеющих расположение общего типа, с вероятностью $P(N, n) \geq 1 - 2^{-9}$ при постоянном отношении $\theta=n/N=0,8$.

Решение. Находим по графикам рис. 3.3 величину

$$k=k(\theta, p)=k(0,8; 0,5)=0,18.$$

Далее находим величину $N=50$ из соотношения $9=Nk=N \cdot 0,18$ и величину $n=\theta N=0,8 \cdot 50=40$.

6. Смешанные RI -, RC - и IC -качества

§ 6.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Во введении уже упоминалась возможность исследования системы A сразу по нескольким, например двум, качествам X и Y . Целесообразность таких рассмотрений объясняется тем, что исследование системы A по отдельным ее качествам предполагает, как уже отмечалось, ее идеальность по другим качествам и поэтому приводит лишь к завышенным оценкам эффективности целостной системы A . Например, при установлении пределов помехоустойчивости (I -качества) последней молчаливо предполагается ее идеальность в смысле надежности (R -качество). При установлении пределов управляемости (C -качества) предполагается ее идеальная надежность (R -качество) и информированность о имеющемся состоянии другой системы (I -качество). Для более реалистических оценок целесообразно совместное рассмотрение этих качеств. Именно при изучении RI -качества рассматривается индифферентная ситуация, когда среда B оказывает мешающее воздействие на RI -сечение A_{RI} внутренними и внешними шумами. Как и в случае I -сечения A_I , цель системы A_{RI} состоит в оптимальном (M, t)-обмене, где M — число различаемых сигналов,

а t — их длительность. Однако здесь оптимизация структуры $[A_{RI}]$ (обобщенное кодирование) состоит во введении избыточности не только в структуру входных слов $[A_I]$, но и в структуру самой системы $[A_R]$. В результате образуются последовательность информационного и вычислительного каналов [25]. В такой общей постановке оптимальный (M, t)-обмен не найден.

При отсутствии внешних шумов оценка оптимального (M, t)-обмена для «вычислительного» канала приведена в работе [25], где решение этой задачи сводится к рассмотрению предельного закона Шеннона (см. гл. 3).

Содержательные рассуждения проблемы передачи информации с учетом надежности кодирующей и декодирующей аппаратуры особенно актуальны в связи с тем, что сложность последней резко возрастает при использовании близких к оптимальным способов передачи информации (подробнее см. следующую главу). Здесь имеются важные конкретные результаты [26]. Однако мы еще весьма далеки от установления предельных законов в этой области*).

Рассмотрение IC -качества связано с исследованием управляемости системы при наличии шумов. IC -сечение A_{IC} в условиях конфликтной ситуации с системой B_{IC} испытывает воздействие шумов, мешающее ей точно фиксировать свое состояние, а также состояние системы B_{IC} .

Возникающие здесь задачи почти совсем не исследовались. В работе [27] решение широкого класса игр при наличии шумов было сведено к решению игр без шумов. Это открывает возможности обобщения результатов, полученных для C -сечения A_C , на более общий случай IC -сечения A_{IC} **).

Наконец, рассмотрение RC -качества RC -сечения A_{RC} связано с его надежностью в случае конфликтной ситуации между ней и системой B . RC -качество системы получило специальное название живучести [28]. Остановимся на этом качестве подробнее, так как для него можно указать предельный закон (u, v)-обмена.

*) Подробное освещение результатов в этой области содержится в издаваемой в этой же серии брошюре Ю. Н. Мельников «Помехоустойчивость информационных систем с учетом надежности».

**) Подробное освещение результатов, полученных в этой области, содержится в издаваемой в этой серии брошюре В. Ф. Крапивина «Теоретико-игровые методы синтеза сложных систем в конфликтных ситуациях».

§ 6.2. ПОНЯТИЕ ЖИВУЧЕСТИ. МОДЕЛЬ СИСТЕМЫ

6.2.1. Понятие живучести. Такое качество, как живучесть системы, было осознано сравнительно недавно. Этому сопутствовали следующие факты. Кроме естественного старения рабочих элементов системы, связанного с их собственным функционированием, они могут подвергнуться внезапному вредному воздействию некоторых активных агентов внешней среды.

Для повышения надежности технических систем в их структуру вводят избыточность дублированием рабочих элементов. При этом поведение системы остается пассивным как по отношению к своим рабочим элементам, так и по отношению к вредным воздействиям внешней среды. Кроме дублирования для повышения надежности системы можно ввести в ее структуру высоконадежные защитные элементы. С их помощью иерархически организованной структуре системы удастся, как правило, заблаговременно обнаружить приближающуюся разладку рабочих элементов и восстановить их (см. гл. 2). При этом поведение системы активно по отношению к своим рабочим элементам, но пассивно к вредным воздействиям внешней среды.

Рассмотрим тот случай, когда защитные элементы выступают как активные внешние агенты системы, подавляющие или нейтрализующие вредные воздействия внешней среды. Таким образом, поведение системы будет активно по отношению к внешней среде и поэтому лишь косвенно касаться рабочих элементов. Учет возможности подавления вредных воздействий внешней среды защитными элементами, а также их уязвимости наряду с рабочими элементами приводит к теоретико-игровым постановкам задач, так как приходится рассматривать случаи наиболее неблагоприятного воздействия внешней среды на систему. Рассмотрение таких ситуаций не обязательно связано с предположением о преднамеренных действиях внешней среды. Как правило, мы просто не знаем заранее о возможных ее воздействиях и вынуждены для гарантированных заключений ориентироваться на наиболее неблагоприятные среди них по отношению к системе. Таким образом, рассматривается некоторое новое качество системы, совмещающее в себе наряду с надежностью (R -качеством) еще и внешнюю активность (C -качество) для ее поддержания (RC -качество).

Это качество системы активно (с помощью соответствующим образом организованной структуры и поведения) противостоять вредным воздействиям внешней среды будем называть коротко *живучестью* системы. В пользу надления этим свойством технических систем говорят многочисленные примеры из биологии.

6.2.2. Описание системы A_{RC} . Будем для краткости называть рабочие элементы системы a -элементами, а ее защитные элементы — R -элементами. Структуру системы A_{RC} , как это принято, будем обозначать $|A_{RC}|$, ее поведение — \bar{A}_{RC} .

Структура $|A_{RC}|$ системы A_{RC} задается следующим образом. Пусть в момент t система A_{RC} имеет $n_a(t)$ и $n_R(t)$ некоторым образом пространственно расположенных a - и R -элементов соответственно. Рассмотрим систему A_{RC} с начальной структурой $|A_{RC}|$, равномерно заполненной элементами.

Это означает, что в момент $t=0$ в любой сфере фиксированного радиуса ϵ , целиком помещающейся внутри системы A_{RC} (ϵ -окрестность любой внутренней точки системы A_{RC}), содержатся постоянные числа a - и R -элементов $\theta n_a(0)$ и $\theta n_R(0)$ соответственно. Здесь $\theta = v/V$, где V и v — объемы, занимаемые системой A_{RC} и сферой радиуса ϵ .

Будем считать, что все элементы системы A_{RC} вне зависимости от их пространственного расположения в ней в одинаковой степени уязвимы и досягаемы для активных агентов внешней среды B . Их мы будем для краткости называть *c -элементами*.

6.2.3. Поведение \bar{B} среды B . Пусть среда B в момент $t=0$ имеет ограниченный и на интервале времени $[0, T]$ взаимодействия с системой A_{RC} непополняемый запас c -элементов численностью $n_c(0)$, который она полностью расходует на систему A_{RC} к моменту $t=T$. Ограниченность и непополняемость запаса c -элементов среды B можно связывать с кратковременностью внезапного контакта системы A_{RC} со средой B . Итак, число c -элементов $n_c(t)$ в момент t с ростом t не возрастает и $n_c(T)=0$.

Пусть интервал $[0, T]$ времени взаимодействия разбит на $k=T/h$ подынтервалов длины $h>0$. Тогда поведение \bar{B} среды B можно описывать двумя наборами чисел $M_a(t)$ и $M_R(t)$ ($M_a(t) + M_R(t) = n_c(t)$, $t=0, h, 2h, \dots, kh$), или двумя наборами их разностей $m_a(t) = M_a(t) - M_a(t+h)$ и $m_R(t) = M_R(t) - M_R(t+h)$, означающих

порции c -элементов, предназначенных средой B соответственно для a - и R -элементов системы A_{RC} в интервалах времени $[t, t+h]$. Итак, обозначим $\bar{B} = \{m_R(t), m_a(t)\}$. Будем считать поведение \bar{B} *ϵ -равномерно воздействующим* своими c -элементами на систему A_{RC} . Это означает, что порции c -элементов $m_a(t)$ и $m_R(t)$ всякий раз для $t=0, h, 2h, \dots, kh$ каждая в отдельности ϵ -равномерно заполняют систему A_{RC} . Происходит как бы поочередное равномерное «обволакивание» элементов системы A_{RC} c -элементами. Последние «впрыскиваются» определенными порциями $m_R(t)$ и $m_a(t)$.

6.2.4. Взаимодействие системы A_{RC} со средой B . Будем предполагать, что в указанных выше симметричных условиях происходит *симметричное* взаимодействие системы A_{RC} со средой B . Это означает, что число a - и R -элементов в любой сфере радиуса ϵ внутри системы A_{RC} в момент $t+h$ зависит лишь от чисел a -элементов, R -элементов и c -элементов в этой сфере в момент t .

Ясно, что в результате такого « ϵ -равномерного» воздействия среды B на систему A_{RC} с ϵ -равномерно заполненной в начале структурой $|A_{RC}|$ последняя из-за симметричного взаимодействия со средой B и в дальнейшем остается ϵ -равномерно заполненной своими элементами. Другими словами, в любой момент t ($0 \leq t \leq T$) в ϵ -окрестности любой внутренней точки системы A_{RC} имеется $\theta n_a(t)$ и $\theta n_R(t)$ a - и R -элементов соответственно.

6.2.5. Поведение \bar{A}_{RC} системы A_{RC} . Будем предполагать, что появление в момент t c -элемента, предназначенного для любого данного элемента системы A_{RC} , вызывает активные защитные действия всех $\theta n_R(t)$ элементов, находящихся в его ϵ -окрестности (пренебрегаем граничным эффектом). Если защитные действия удачны, то этот элемент продолжает функционировать и дальше. Если защитные действия не удачны, то этот элемент выходит из строя. Однако в обоих указанных случаях c -элемент выходит из строя. Предполагается, что система A_{RC} на интервале времени взаимодействия $[0, T]$ не пополняется новыми элементами (отсутствие регенерации). Таким образом, со временем элементы системы A_{RC} лишь «равномерно редеют» и поэтому числа $n_a(t)$ и $n_R(t)$ не возрастают с ростом t . Система A_{RC} выходит из строя, если к моменту $t=T$ имеем $n_a(t) \leq (1-\theta_a)n_a(0)$, т. е. если выходит из строя более чем θ_a -я доля первоначального количества ее рабочих элементов, где θ_a —

некоторая фиксированная величина ($0 \leq \theta_a \leq 1$). Если $n_a(t) > (1-\theta_a)n_a(0)$, то считается, что система A_{RC} продолжает функционировать. Ясно, что чем больше величина $1-\theta_a$, тем меньше система способна противостоять вредным воздействиям среды B . Поэтому естественно назвать величину θ_a *коэффициентом живучести* системы.

6.2.6. Стохастическая постановка задачи. Рассмотрим стохастическую постановку задачи. Для этого зададим основную для дальнейшего изложения вероятность $p[n_R(t)]$ поражения одного элемента системы A_{RC} (a - или R -элемента) в момент t одним c -элементом, если система A_{RC} располагает $n_R(t)$ защитными элементами. Таким образом, на интервале времени $[t, t+h]$ в *среднем* выходит из строя

$$\left. \begin{aligned} n_a(t) - n_a(t+h) &= m_a(t) p[n_R(t)] \\ n_R(t) - n_R(t+h) &= m_R(t) p[n_R(t)] \end{aligned} \right\} \quad (6.1)$$

a - и R -элементов соответственно. Здесь считается, что на интервале времени $[t, t+h]$ действует $n_R(t)$ защитных элементов и лишь в конце него под действием c -элементов часть из них выходит из строя. Определенное выше стохастическое взаимодействие системы A_{RC} со средой B является симметричным лишь в *среднем*. Числа элементов отождествляются с их средними числами. В каждом конкретном случае возможны случайные отклонения от средних. Однако при достаточно больших значениях $m_a(t)$ и $m_R(t)$ этими отклонениями на основании закона больших чисел можно пренебречь, что и делается в дальнейшем изложении, вплоть до § 6.5.

6.2.7. Основные соотношения. Пусть среда B различает a -элементы и R -элементы системы A_{RC} , полностью осведомлена о результатах своих действий и не тратит зря свои c -элементы на элементы A_{RC} , уже вышедшие из строя. Тогда для выведения системы A_{RC} из строя достаточно выполнения соотношения

$$\sum_{t=0}^{T-h} m_a(t) p[n_R(t)] = n_a(0) - n_a(T) = \theta_a n_a(0), \quad (6.2)$$

следующего из (6.1).

Для этого по условию задачи расходуется весь первоначальный запас c -элементов среды B

$$\sum_{t=0}^{T-h} [m_a(t) + m_R(t)] = n_c(0). \quad (6.3)$$

В рассматриваемом случае структура $|A_{RC}|$ системы в момент t зависит от поведения среды $\bar{B} = \{m_a(t), m_R(t)\}$ и начальной структуры $|A_{RC}|$ системы в момент $t=0$. Последняя, будучи ε -равномерно заполненной своими элементами, определяется лишь их начальными числами $|A_{RC}| = \{n_a(0), n_R(0)\}$.

Таким образом, величина $n_c(0)$, определяемая соотношениями (6.1), (6.2) и (6.3), является функцией \bar{B} и $|A_{RC}|$.

$$n_c(0) = F(\bar{B}, |A_{RC}|) = f[m_a(t), m_R(t); n_a(0), n_R(0)].$$

Будем считать оптимальными такие поведение \bar{B}_{opt} и структуру $|A_{RC}|_{opt}$, при которых достигается:

$$\begin{aligned} \max_{|A_{RC}|} \min_{\bar{B}} F(\bar{B}, |A_{RC}|) &= \max_{|A_{RC}|} F(\bar{B}_{opt}, |A_{RC}|) = \\ &= F(\bar{B}_{opt}, |A_{RC}|_{opt}). \end{aligned}$$

§ 6.3. ОПТИМИЗАЦИЯ ПОВЕДЕНИЯ СРЕДЫ

6.3.1. Оптимальность двухшагового поведения среды.

Рассмотрим соотношения (6.1), (6.2) и (6.3) в непрерывном случае, при $h \rightarrow 0$, $T = \text{const}$, когда $k = (T/h) \rightarrow \infty$. Имеем

$$n'_R(t) = -m_R(t) p[n_R(t)], \quad (6.1')$$

$$I_1 = \int_0^T m_a(t) p[n_R(t)] dt \quad (6.2')$$

и

$$I_2 = \int_0^T [m_a(t) + m_R(t)] dt. \quad (6.3')$$

Из дифференциального уравнения (6.1) с разделяющимися переменными для случая $p(x) = \exp(-bx)$, представляющего основной интерес при дальнейшем изложении [см. (6.13)], находим

$$n'_R(t) = -m_R(t) \exp[-bn_R(t)]$$

или

$$\{\exp[bn_R(t)]\} dn_R(t) = -m_R(t) dt.$$

Интегрируя обе части последнего соотношения по t от 0 до t , получим

$$b^{-1} \{\exp[bn_R(t)] - \exp[bn_R(0)]\} = - \int_0^t m_R(\tau) d\tau$$

или

$$\exp\{b[n_R(t) - n_R(0)]\} = 1 - bp[n_R(0)] \int_0^t m_R(\tau) d\tau.$$

Итак,

$$n_R(t) = n_R(0) + b^{-1} \ln \{1 - bp[n_R(0)] \int_0^t m_R(\tau) d\tau\}.$$

Отсюда

$$p[n_R(t)] = p[n_R(0)] \{1 - bp[n_R(0)] \int_0^t m_R(\tau) d\tau\}^{-1}. \quad (6.4')$$

Решим экстремальную задачу $I_2 = \min$ при $I_1 = c = \text{const}$ по всем возможным значениям $m_a(t)$ и $m_R(t)$, рассматриваемым снова как k -мерные векторы. Для этого с помощью метода множителей Лагранжа найдем безусловный экстремум выражения $Q = I_2 + \lambda(I_1 - c)$, заменяя эти интегралы суммами.

Используя соотношения (6.2'), (6.3') и (6.4'), имеем для $0 \leq s \leq T$:

$$\frac{\partial Q}{\partial m_a(s)} = 1 + \lambda p[n_R(s)] = 0,$$

$$\begin{aligned} \frac{\partial I_1}{\partial m_R(s)} &= - \int_s^T m_a(\tau) \frac{p[n_R(0)] \{-bp[n_R(0)]\}}{\left\{1 - bp[n_R(0)] \int_0^t m_R(\sigma) d\sigma\right\}^2} d\tau = \\ &= b \int_s^T m_a(\tau) p^2[n_R(\tau)] d\tau \end{aligned}$$

и

$$\frac{\partial Q}{\partial m_R(s)} = 1 - \lambda b \int_s^T m_a(\tau) p^2[n_R(\tau)] d\tau = 0, \quad (6.5')$$

отсюда при любых $0 < s < t < T$ имеем $p[n_R(s)] = -\lambda^{-1}$. Подставив последнее выражение в (6.2') и (6.5'), должны иметь для экстремальных значений $m_a^*(\tau)$ и $m_R^*(\tau)$:

$$\int_s^T m_a^*(\tau) d\tau = \text{const} \quad \text{и} \quad \int_0^T m_R^*(\tau) d\tau = \text{const} = I_2 + \lambda I_1.$$

Это может быть лишь при

$$m_a^*(\tau) = \delta(T - \tau) m_a \quad \text{и} \quad m_R^*(\tau) = \delta(\tau) m_R,$$

где $\delta(x)$ — «дельта-функция» и m_a и m_R — неотрицательные числа, сумма которых дает величину $I_2 = n_c(0) = m_a + m_R$.

Итак, перераспределение значений $m_a(\tau)$ и $m_R(\tau)$ фактически следует проводить в оптимальном случае лишь в окрестностях начального ($t=0$) и конечного ($t=T$) моментов времени. Поэтому будем ограничиваться рассмотрением двухшаговой задачи.

6.3.2. Решение двухшаговой экстремальной задачи.

Для удобства записи примем интервал времени $h=1$. Тогда для двухшагового случая ($T=k=2$) соотношения (6.3), (6.2) и (6.1) имеют вид:

$$m_a(0) + m_a(1) + m_R(0) + m_R(1) = n_c(0), \quad (6.4)$$

$$m_a(0) p[n_R(0)] + m_a(1) p[n_R(1)] = \theta_a n_a(0), \quad (6.5)$$

где

$$\left. \begin{aligned} n_a(1) &= n_a(0) - m_a(0) p[n_R(0)], \\ n_R(1) &= n_R(0) - m_R(0) p[n_R(0)]. \end{aligned} \right\} \quad (6.6)$$

Для отыскания поведения \bar{B}_{opt} , обращающего величину $n_c(0)$ в минимум, положим прежде всего $m_R(1) \equiv 0$, так как $m_R(1)$ входит лишь в выражение (6.4). Далее, используя (6.5), выразим величину $m_a(1)$ через $m_a(0)$ и подставим ее выражение в соотношение (6.4). Тогда получим

$$m_a(0) \left\{ 1 - \frac{p[n_R(0)]}{p[n_R(0)]} \right\} + m_R(0) + \frac{\theta_a n_a(0)}{p[n_R(0)]} = n_c(0). \quad (6.7)$$

Так как фигурная скобка левой части соотношения (6.7) положительна, то величина $n_c(0)$ с ростом $m_a(0)$ возрастает. Поэтому для обращения $n_c(0)$ в минимум сле-

дует положить $m_a(0) \equiv 0$. Итак, остается найти минимум выражения

$$\begin{aligned} n_c(0) &= m_R(0) + \frac{\theta_a n_a(0)}{p\{n_R(0) - m_R(0) p[n_R(0)]\}} = \\ &= f[m_R(0); n_a(0), n_R(0)] \end{aligned} \quad (6.8)$$

по переменной $m_R(0)$, который может существовать, так как с ростом $m_R(0)$ первое слагаемое возрастает, а второе убывает.

Пусть функция $p(x)$ дифференцируема. Тогда экстремальное значение $m_R(0) = m_R$ может быть найдено из уравнения, получающегося дифференцированием соотношения (6.8) по $m_R(0)$:

$$\frac{p^2[n_R(1)]}{p'[n_R(1)] p[n_R(0)]} = \theta_a n_a(0). \quad (6.9)$$

Оптимальным является поведение \bar{B}_{opt} среды B , при котором

$$\left. \begin{aligned} m_a(0) &\equiv m_R(1) \equiv 0, \quad m_R(0) \equiv m_R \\ \text{и} \\ m_a(1) &= m_a = \frac{\theta_a n_a(0)}{p\{n_R(0) + m_R p[n_R(n_R(0))]\}}, \end{aligned} \right\} \quad (6.9')$$

где величина m_R находится из уравнения (6.9), и минимальное значение

$$n_c = \min n_c(0) = f[m_R; n_a(0), n_R(0)].$$

Полученное оптимальное поведение \bar{B}_{opt} среды B интуитивно можно было бы предвидеть из-за резкого возрастания эффективности s -элементов с понижением защитных возможностей системы A_{RC} . Из-за этого естественно вначале израсходовать $m_R(0)$ s -элементов только на R -элементы, а затем оставшиеся $m_a(1)$ s -элементов только на a -элементы. При этом не ясно лишь, в какой пропорции целесообразно их расходовать.

Рассмотрим подоптимальное поведение \bar{B}_0 среды B , когда

$$m_a(0) \equiv m_R(1) \equiv 0, \quad \text{но} \quad m_R(0) \equiv m_a(1) \equiv \tilde{m}_R.$$

Тогда величина $m_R(0) = \tilde{m}_R$ определяется из уравнения [см. (6.8)]

$$m_R(0) = \frac{\theta_a n_a(0)}{p\{n_R(0) - m_R(0)\} p\{n_R(0)\}}, \quad (6.10)$$

а соответствующее значение $n_c(0) = \tilde{n}_c$ имеет вид:

$$\tilde{n}_c = 2\tilde{m}_R = f[\tilde{m}_R; n_a(0), n_R(0)]. \quad (6.11)$$

Относительный выигрыш $\xi = 1 - \frac{n_c}{\tilde{n}_c}$ в числе c -элементов

при использовании средой B поведения \bar{E}_{opt} вместо \bar{B}_0 имеет вид:

$$\xi = 1 - \frac{f[m_R; n_a(0), n_R(0)]}{f[\tilde{m}_R; n_a(0), n_R(0)]}. \quad (6.12)$$

До сих пор почти нигде не использовалось явное выражение вероятности $p(x)$. Чтобы получить численные результаты, необходимо рассмотреть различные частные ситуации, приводящие к явным зависимостям вероятности $p(x)$ от x .

6.3.3. Случай независимых действий R -элементов. Выше предполагалось, что появление c -элемента вызывает защитные действия всех $\theta n_R(t)$ R -элементов, находящихся в его ε -окрестности.

Рассмотрим случай, когда действия R -элементов происходят независимо друг от друга, причем вероятность вывести из строя c -элемент одним R -элементом равна постоянной величине $D < 1$. Тогда

$$p[n_R(t)] = (1 - D)^{\theta n_R(t)} = \exp[-dn_R(t)], \quad (6.13)$$

где $d = \theta \ln(1 - D)^{-1}$ — «коэффициент эффективности» R -элемента.

В рассматриваемом случае уравнение (6.9) решается в явном виде и для ненулевых компонент \bar{B}_{opt} имеем выражения [см. (6.9')]

$$\left. \begin{aligned} m_R &= d^{-1} \{\exp[dn_R(0)] \ln d \theta_a n_a(0)\}, \\ m_a &= d^{-1} \exp[dn_R(0)]. \end{aligned} \right\} \quad (6.13')$$

Отсюда минимальное число c -элементов

$$\begin{aligned} n_c &= \min_B F(\bar{E}, |A_{RC}|) = f[m_R; n_a(0), n_R(0)] = \\ &= d^{-1} \{\exp[dn_R(0)]\} \ln d \theta_a n_a(0). \end{aligned} \quad (6.14)$$

Логарифмируя соотношение (6.14) и полагая

$$[\ln \ln \theta_a n_a(0)]^{1-\varepsilon} > \ln d, \quad (6.14')$$

где $\varepsilon > 0$, получим

$$\log_{10} n_c \approx dn_R(0) \log_{10} e + \log_{10} \ln \theta_a n_a(0). \quad (6.14'')$$

Уравнение (6.10) для получения значения ненулевых компонент $m_R(0) = m_a(1) = \tilde{m}_R$ подоптимального поведения B среды B здесь оказывается трансцендентным:

$$m_R(0) = \theta_a n_a(0) [\exp dn_R(0)] \exp\{-dp[n_R(0)] m_R(0)\}. \quad (6.15)$$

Логарифмируя выражения (6.15) и пренебрегая величиной порядка $O(\ln \tilde{m}_R / \tilde{m}_R)$, получим приближенное решение уравнения (6.15) в виде

$$\tilde{m}_R \approx d^{-1} [\exp dn_R(0)] [dn_R(0) + \ln \theta_a n_a(0)].$$

Подставив соответствующее значение $\tilde{n}_c = 2\tilde{m}_R = f[\tilde{m}_R; n_a(0), n_R(0)]$ в выражение (6.12), полагая $\ln \theta_a n_a(0) \gg \ln d$ с учетом выражения (6.11), получим приближенное выражение относительного выигрыша

$$\xi = 1 - \frac{n_c}{\tilde{n}_c} \approx 1 - \frac{1}{2} \left[\frac{1 + dn_R(0)}{\ln \theta_a n_a(0)} \right]^{-1}. \quad (6.16)$$

Из соотношения (6.16) следует, что выигрыш в числе элементов при использовании системой B поведения \bar{B}_{opt} вместо \bar{B}_0 всегда более чем двойной. Он существенно возрастает, если

$$dn_R(0) \ll \ln \theta_a n_a(0).$$

§ 6.4. ОПТИМИЗАЦИЯ СТРУКТУРЫ СИСТЕМЫ

6.4.1. Оптимизация структуры системы по числу защитных элементов $n_R(0)$. Пусть число a -элементов фи-

ксийровано из соображений, связанных с работой системы. Число R -элементов $n_R(0)$, выполняющих защитные функции, может быть выбрано лучшим образом при фиксации некоторых физических параметров задачи. В самом деле, пусть система A_{RC} располагает фиксированным количеством E_R некоторого «защитного субстрата» (например, энергии) и может его равномерно распределять между $n_R(0)$ R -элементами так, что на долю каждого приходится

$$E_1 = E_R / n_R(0).$$

Очевидно, что коэффициент эффективности d R -элемента должен расти с ростом E_1 , так как с ростом E_1 должна расти вероятность D . Статистические соображения (см. п. 6.5.3) позволяют принять следующую зависимость D от E_1 :

$$D = 1 - \exp(-FE_1^\alpha), \quad (6.17)$$

где $E_1 = E_R / n_R(0)$, а F и α — некоторые положительные константы, не зависящие от $n_R(0)$.

Используя соотношение (6.17), получим из (6.13)

$$d = \theta \ln(1 - D)^{-1} = G n_R^{-\alpha}(0), \quad (6.18)$$

$$\text{где } G = \theta F E_R^\alpha = d n_R^\alpha = \text{const}. \quad (6.18')$$

Подставив значение d , определяемое выражением (6.18), в (6.14), для n_c получим

$$n_c = \min_{\bar{B}} F(\bar{E}; |A_{RC}|) = \\ = G^{-1} n_R^\alpha(0) [\exp G n_R^{1-\alpha}(0)] [\ln G \theta n_a(0) - \ln n_R^\alpha(0)]. \quad (6.19)$$

Правая часть соотношения (6.19) распадается на произведение возрастающей и убывающей с ростом $n_R(0)$ функций, поэтому для некоторого значения $n_R(0) = n_R$ может существовать максимум. Дифференцируя правую часть (6.19) по $n_R(0)$, получим трансцендентное уравнение для определения n_R :

$$[\ln G \theta n_a(0) - \ln n_R^\alpha(0)] [\alpha n_R^{\alpha-1}(0)] + (1-\alpha) G - \alpha n_R^{\alpha-1}(0) = 0. \quad (6.20)$$

Для больших значений $n_a(0)$ и $\alpha > 0$ приближенное значение n_R можно найти из соотношения (6.20). Оно имеет вид

$$n_R \approx [G \theta n_a(0)]^{1/\alpha}. \quad (6.21)$$

При этом из соотношения (6.18) имеем значение $d \approx [\theta n_a(0)]^{-1}$, которое существенно меньшего порядка, чем значение d , определяемое соотношением (6.14'). Заметим, что при такой структуре системы $|A_{RC}|$ согласно (6.13') для среды оптимальной оказывается вырожденная одношаговая стратегия $m_R \approx 0$ и $m_a \approx n_c$, при которой все c -элементы расходуются на a -элементы системы A_{RC} .

Поэтому у среды может возникнуть «соблазн» в любом случае действовать именно таким способом. Парадокс, возникающий при этой ситуации, состоит в том, что, как правило, система никогда не имеет в точности оптимальную структуру (ее структура подоптимальна). В этом случае среда должна отказаться от указанной вырожденной стратегии.

В частности, для практически важного случая $\alpha = 1$ (см. п. 6.5.3) приближенное выражение (6.21) становится точным. Подставив выражение (6.21) в (6.19), получим

$$n_c = \max_{|A_{RC}|} \min_{\bar{B}} F(B, |A_{RC}|) \approx \\ \approx \theta n_a(0) \exp [G^{1/\alpha} \theta^{(1-\alpha)/\alpha} n_a^{(1-\alpha)/\alpha}(0)], \quad (6.22)$$

причем для $\alpha = 1$ имеет место точное выражение

$$n_c = \max_{|A_{RC}|} \min_{\bar{B}} F(\bar{E}, |A_{RC}|) = \theta n_a(0) \exp G, \quad (6.23)$$

где

$$G = d n_R(0) = \theta F E_R.$$

Итак, для оптимальной структуры $|A_{RC}|_{opt}$ величина $n_R(0)$ при фиксированном значении E_R определяется из соотношения (6.21). Рассмотрим подоптимальную структуру $|A_{RC}|$, при которой берется одно и то же число a - и R -элементов ($n_a(0) = n_R(0)$).

Для этого случая при $\alpha = 1$ имеем из (6.19)

$$\tilde{n}_c = \frac{n_a(0)}{G} (1 + \ln \theta G) \exp G. \quad (6.24)$$

Сопоставление выражений для n_c и \tilde{n}_c (6.23) и (6.24) показывает, что относительный выигрыш среды B в числе c -элементов от использования системой A_{RC} вместо

$[A_{RC}]_{opt}$ подоптимальной структуры имеет вид:

$$\xi' = 1 - \frac{\tilde{n}_c}{n_c} = 1 - \frac{1 + \ln \theta_\alpha G}{\theta_\alpha G}. \quad (6.25)$$

Из соотношения (6.25) видно, что с ростом величины E_R , когда растет G , величина $\xi' \rightarrow 1$, т. е. выигрыш среды B в c -элементах неограниченно растет.

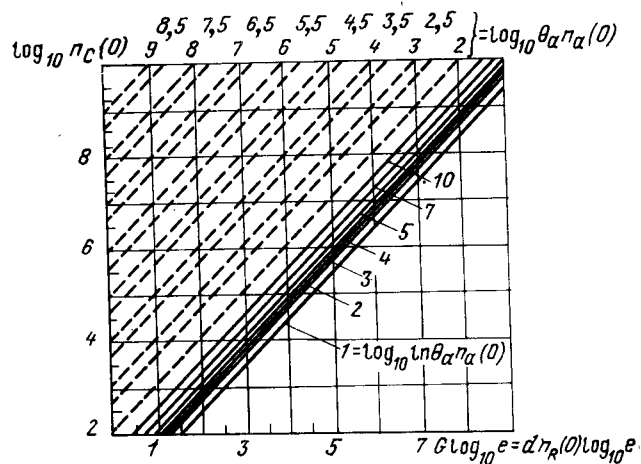


Рис. 6.1. Зависимость среднего числа c -элементов $n_c(0)$ от чисел a -элементов $n_a(0)$ и R -элементов $n_R(0)$ при неоптимальной структуре системы A_{RC} (сплошные линии) и при структуре системы A_{RC} , оптимизированной по числу R -элементов $n_R(0)$ (пунктирные линии).

Если прологарифмировать обе части соотношения (6.23), то с учетом соотношения (6.18') при $\alpha=1$ получим

$$\log_{10} n_c = d n_R(0) \log_{10} e + \log_{10} \theta_\alpha n_a(0). \quad (6.23')$$

Сопоставление выражений (6.14'') и (6.23') величины $\log_{10} n_c$ для оптимального поведения только среды и оптимального поведения среды и оптимальной структуры системы при постоянном значении $d n_R(0) = G = \text{const}$ приведено на рис. 6.1. На нем видно, что оптимизация структуры системы на много порядков увеличивает число n_c c -элементов, необходимых среде для вывода из строя системы, и существенно уменьшает число $n_R(0)$ R -элементов при фиксации прочих параметров.

6.4.2. Дальнейшая оптимизация структуры системы по числу рабочих элементов $n_a(0)$. В предыдущем пункте мы исходили из фиксированного числа $n_a(0)$ жизненно важных a -элементов системы A_{RC} , а также фиксированного значения E_R величины защитного субстрата.

Однако можно выявить связь между этими параметрами системы, исходя из ее целенаправленного развития, предшествующего столкновению со средой B (*). В самом деле, пусть имеется некоторое фиксированное количество E жизненного (творческого) субстрата (например, белка), которое система A_{RC} может равномерно распределять между своими $n_a(0)$ жизненно важными a -элементами, так что каждый из них имеет его в количестве $e = E/n_a(0)$.

Пусть каждый из a -элементов воспроизводит из порции e «первичного» жизненного субстрата наряду с другими «вторичными» производными субстратами также порцию защитного субстрата

$$e_R = f(e) = f[E/n_a(0)].$$

Тогда общее количество защитного субстрата E_R , воспроизводимого системой A_{RC} , имеет вид:

$$E_R = n_a(0) e_R = n_a(0) f[E/n_a(0)] = g[n_a(0)]. \quad (6.26)$$

Естественно считать функцию $g(x)$ убывающей с ростом x . В самом деле, как и в случае защитного субстрата E_R эффект дробления жизненного субстрата понижает общий производительный потенциал системы A_{RC} («в единении силы»). В пользу такого допущения можно привести многочисленные технические, биологические и социальные примеры.

Итак, полагая функцию $g(x)$ дифференцируемой, будем считать, что при $0 \leq x < \infty$:

$$g(x) > 0, \quad g'(x) < 0 \quad \text{и} \quad g(x) \rightarrow 0 \quad \text{при} \quad x \rightarrow \infty. \quad (6.27)$$

Рассмотрим теперь экстремальное значение n_c , определяемое соотношением (6.23) при $\alpha=1$ (рассмотрение более сложного случая $\alpha \neq 1$ качественно не изменит картину).

*) В биологии это соответствует перспективной активности биосистем.

Подставив выражение (6.26) в соотношение (6.23) с учетом соотношения (6.18'), получим экстремальное значение n_c , которое зависит лишь от одной переменной $n_a(0)$:

$$n_c = \theta_a n_a(0) \exp \{ \theta F g [n_a(0)] \}, \quad (6.28)$$

где константы θ и F характеризуют эффективность R -элементов.

Из условий (6.27) следует, что при $n_a(0) \rightarrow \infty$ $g[n_a(0)] \rightarrow 0$, поэтому величина n_c как это следует из соотношения (6.28), имеет асимптоту

$$n_c \xrightarrow{n_a(0) \rightarrow \infty} \theta_a n_a(0). \quad (6.29)$$

Поэтому функция n_c может иметь особенности лишь при конечных значениях $n_a(0)$ в окрестности начального значения $n_a(0) = 1$.

Рассматривая «непрерывный» случай $n_a(0)$, дифференцируя по $n_a(0)$ соотношение (6.28), получим условия экстремума в виде

$$1 + \theta F n_a(0) g' [n_a(0)] = 0. \quad (6.30)$$

Если существует вторая производная $g''(x)$ по x , то условия максимума и минимума получим, взяв вторую производную выражения (6.28) по $n_a(0)$. Тогда

$$g' [n_a(0)] + n_a(0) g'' [n_a(0)] \begin{cases} < 0 & \text{при максимуме } n_c, \\ > 0 & \text{при минимуме } n_c. \end{cases} \quad (6.31)$$

Итак, из условий (6.30) и (6.31) для экстремальных значений $n_a(0) = n_a$ имеем

$$g'(n_a) = \frac{-1}{\theta F n_a} \quad \text{и} \quad g''(n_a) \geq \frac{1}{\theta F n_a^2} \quad (6.32)$$

для максимума и минимума соответственно.

Асимптотическое поведение функции n_c в зависимости от $n_a(0)$ указывает на возможный характер ее особенностей при начальных значениях $n_a(0)$.

Здесь в простейшем случае может вообще не быть экстремальных значений (монотонный рост с приближением сверху к асимптоте). Далее возможен случай наличия двух экстремумов: сначала максимума, потом минимума. Остановимся на рассмотрении этих простейших случаев. Они реализуются, например, гиперболической

функцией $g(x)$ вида

$$g(x) = \frac{\beta}{x + \gamma} \quad \left(g'(x) = -\frac{\beta}{(x + \gamma)^2} \right), \quad (6.33)$$

где β и γ — некоторые положительные константы.

Последние, используя соотношение (6.26) и (6.33) и вычисляя значения $g(1)$ и $g'(1)$, можно представить в следующей форме:

$$\beta = \frac{f^2(E)}{E f'(E) - f(E)} > 0 \quad \text{и} \quad \gamma = \frac{2f(E) - E f'(E)}{E f'(E) - f(E)} > 0, \quad (6.34)$$

Для выполнения условий (6.27) здесь достаточно требование $x > -\gamma$ и $\beta > 0$.

Подставив выражение (6.33) в уравнение (6.30), получим квадратное уравнение

$$[n_a(0) + \gamma]^2 - \theta F \beta [n_a(0) + \gamma] + \theta F \beta \gamma = 0,$$

имеющее в случае $\beta > 0$ и

$$0 < \omega = \frac{4\gamma}{\theta F \beta} < 1 \quad (6.35)$$

два действительных положительных корня

$$n_a^{1,2} = 2\gamma \frac{1 \mp \sqrt{1 - \omega}}{\omega} - \gamma = \theta F \beta \left(\frac{1 \mp \sqrt{1 - \omega}}{2} \right)^2. \quad (6.36)$$

Подставив соотношение (6.36) в (6.33), а затем в (6.28), получим экстремальные значения $n_c^{1,2}$, соответствующие корням (6.36), которые после проверки условий (6.31) оказываются первое — максимумом, а второе — минимумом:

$$n_c^{1,2} = \theta_a n_a^{1,2} \exp \frac{2}{1 \mp \sqrt{1 - \omega}}. \quad (6.37)$$

Для вычислений удобно ввести нормированные значения

$$\begin{aligned} \left(\frac{n_a^{1,2}}{\theta F \beta} \right)^{-1/2} &= \ln n_c^{1,2} - \ln \theta n_a^{1,2} = \psi_{1,2}(\omega) = \\ &= \frac{2(1 \pm \sqrt{1 - \omega})}{\omega} = \frac{2}{1 \mp \sqrt{1 - \omega}} \end{aligned} \quad (6.38)$$

величин $n_a^{1,2}$ и $n_c^{1,2}$, являющиеся функциями $\psi_{1,2}(\omega)$ параметра ω , меняющегося от 0 до 1.

Асимптотические значения функции $\psi_{1,2}(\omega)$ имеют простой вид.

При $\omega \rightarrow 0$ удобно воспользоваться первым представлением $\psi_{1,2}(\omega)$ (6.38):

$$\psi_{1,2}(\omega) = \begin{cases} \frac{4}{\omega} - 1 - \frac{\omega}{4} + O(\omega^2), \\ 1 + \frac{\omega}{4} + O(\omega^2). \end{cases} \quad (6.39)$$

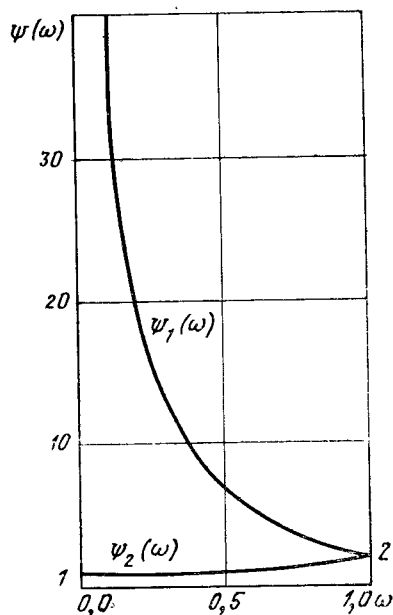


Рис. 6.2. Зависимость функций $\psi_1(\omega)$ и $\psi_2(\omega)$ от параметра ω .

При $\omega \rightarrow 1$ удобно воспользоваться вторым представлением $\psi_{1,2}(\omega)$ (6.38):

$$\psi_{1,2}(\omega) = \begin{cases} 2(1 + \sqrt{1-\omega}) + O(1-\omega), \\ 2(1 - \sqrt{1-\omega}) + O(1-\omega). \end{cases} \quad (6.39')$$

Графики функций $\psi_{1,2}(\omega)$ при значениях параметра ω в пределах $0 \leq \omega \leq 1$ изображены на рис. 6.2.

Для получения зависимости коэффициентов β и γ от параметра E необходимо, как следует из (6.34), задать аналитическую зависимость от него функции $f(E)$.

Положим

$$f(E) = BE^\delta, \quad (6.40)$$

где $B > 0$ и δ ($1 < \delta < 2$) — некоторые константы. Подставив соотношение (6.40) в (6.34) и (6.35), получим

$$\beta = \frac{BE^\delta}{\delta - 1}, \quad \gamma = \frac{2 - \delta}{\delta - 1} \quad (6.41)$$

и

$$\omega = \frac{4(2 - \delta)}{\theta F B E^\delta}. \quad (6.42)$$

Обозначим через $E_{кр}$ то значение параметра E , при котором величина ω обращается в единицу:

$$1 = \frac{4(2 - \delta)}{\theta F B E_{кр}^\delta}. \quad (6.43)$$

Из соотношения (6.43) имеем

$$E_{кр} = \left[\frac{4(2 - \delta)}{\theta F B} \right]^{1/\delta}. \quad (6.44)$$

Поделив соответствующие части соотношений (6.42) и (6.43), получим

$$\omega = (E_{кр}/E)^\delta. \quad (6.45)$$

Таким образом, величина ω меняется в пределах $1 \geq \omega \geq \omega$, когда величина E меняется в пределах $E_{кр} \leq E < \infty$:

Величина n_c , определяемая соотношением (6.28) с учетом соотношений (6.33) и (6.41), имеет вид:

$$n_c = \theta_a n_a(0) \exp \frac{\delta F B E^\delta}{(\delta - 1) \left[n_a(0) + \frac{2 - \delta}{\delta - 1} \right]}. \quad (6.46)$$

На рис. 6.3 изображено семейство кривых $\log_{10} n_c$ в зависимости от $\log_{10} n_a(0)$ при различных значениях $E \geq E_{кр}$ и фиксированных значениях $\theta_a = 0,5$, $\theta F = B = 1$ и $\delta = 1,01$, при которых $E_{кр} \approx 4$.

Прокомментируем полученный результат. Величина E жизненного субстрата, имеющегося у системы A_{RC} , определяет пути, по которым система A_{RC} может пойти для обеспечения своей максимальной живучести. Если $E \leq E_{кр}$, то имеется лишь один путь, состоящий в увеличении живучести за счет увеличения числа элементов все более и более беззащитных. Если $E > E_{кр}$, то имеется еще один путь. Это построение сравнительно не-

большого числа хорошо защищенных элементов (рис. 6.3).

Рассмотрение рис. 6.3 показывает, что возможно резкое увеличение живучести системы с оптимальной структурой (на много порядков), если количество жизненного субстрата E превосходит его критическое значение $E_{кр}$ в несколько раз.

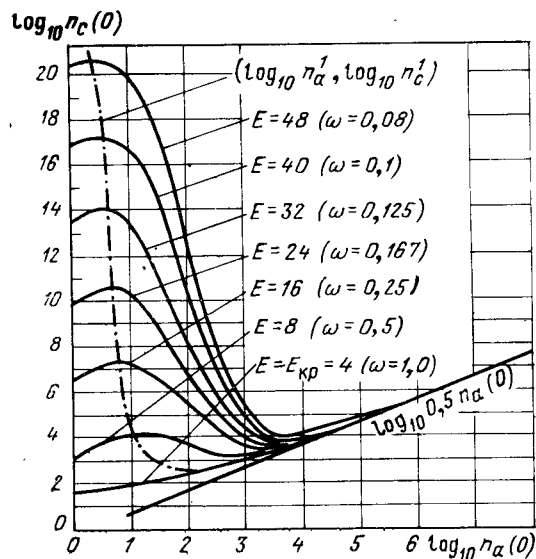


Рис. 6.3. Зависимость среднего числа c -элементов $n_c(0)$, необходимого для уничтожения системы A_{RC} от числа ее a -элементов $n_a(0)$ при различных значениях величины E -жизненно-го субстрата (n_a^1, n_c^1) — оптимальные для системы A_{RC} значения $n_a(0), n_c(0)$.

В природе существуют оба пути обеспечения живучести. Один состоит в безграничном увеличении числа сравнительно незащитных одноклеточных особей популяции для обеспечения живучести последней. Второй путь состоит в образовании немногочисленных популяций, хорошо защищенных многоклеточных особей (например, теплокровных, имеющих белые кровяные тельца; см. далее п. 7.6.3).

Заметим еще, что путь беспредельного увеличения числа элементов системы для обеспечения ее надежности следует и из предельного закона надежности.

§ 6.5. ПРЕДЕЛЬНЫЙ ЗАКОН ЖИВУЧЕСТИ

6.5.1. Уязвимые системы. Прежде всего исключим из рассмотрения тривиальные системы с точки зрения живучести. В самом деле, пусть система имеет $n_a(0)$ a -элементов и для ее уничтожения требуется уничтожить не менее θ_a -й долю их, т. е. $\theta_a n_a(0)$ a -элементов (при этом останется не более $(1-\theta_a)n_a(0)$ a -элементов). Пусть среда имеет всего $n_c(0)$ c -элементов.

Мы скажем, что система A_{RC} уязвимая, если $n_c(0) > \theta_a n_a(0)$.

Если $n_c(0) \leq \theta_a n_a(0)$, то будем говорить, что система A_{RC} неуязвимая. Другими словами, система A_{RC} считается неуязвимой, если, истратив на ее уничтожение все имеющиеся у нее c -элементы и уничтожив такое же число a -элементов системы A_{RC} , которые последняя даже не пытается защищать, среда B все же не добьется ее уничтожения. Рассмотрение таких неуязвимых систем A_{RC} не представляет интереса. Будем рассматривать только уязвимые системы A_{RC} , для которых $n_c(0) > \theta_a n_a(0)$.

Пусть m_a является числом c -элементов, которые среда направляет на уничтожение a -элементов системы A_{RC} .

Ясно, что $m_a \leq n_c(0)$. Поэтому всегда можем иметь такое значение m_a , что $m_a > \theta_a n_a(0)$.

Введем новый коэффициент

$$\theta_{RC} = \frac{\theta_a n_a(0)}{m_a} \quad (6.47)$$

и будем называть его коэффициентом неуязвимости системы A_{RC} .

Из определения уязвимой системы A_{RC} следует, что среда всегда может добиться того, чтобы коэффициент неуязвимости такой системы был меньше единицы ($\theta_{RC} < 1$).

6.5.2. Предельный закон живучести. До сих пор оптимизация структуры A_{RC} и поведения среды B велась по среднему числу элементов. Другими словами, предполагалось, что число элементов столь велико, что имеет место закон больших чисел, т. е. близость значений случайных величин к их средним (пренебрежение отклонениями от них). Такие представления оправданы, если не делать вероятностных утверждений о возможностях ис-

ходов рассматриваемых событий. Однако все предельные законы потенциальной эффективности сложных систем как раз касаются оценок вероятностей больших отклонений случайных величин от своих средних. Поэтому, приступая к установлению предельного закона живучести, приходится вернуться к исходным моделям системы A_{RC} и среды B и заново решать вопросы их оптимизации. Однако мы не будем этого делать в полной мере. Именно будем рассматривать двухшаговую стратегию среды B , оптимальную при рассмотрении в среднем, когда среда из общего числа $n_c(0)$ своих c -элементов на первом шаге расходует m_R их на уничтожение R -элементов системы A_{RC} , а на втором шаге расходует m_a их на уничтожение a -элементов системы A_{RC} . При этом система A_{RC} имеет $n_a(0)$ и $n_R(0)$ a - и R -элементов соответственно. Все перечисленные параметры далее считаются произвольными, но фиксированными и не обязаны быть выбранными оптимальным образом в смысле оптимизации в среднем, как это было сделано выше.

Обозначим через v_R и v_a случайные числа R - и a -элементов системы A_{RC} , уничтоженных средой B на первом и втором шагах соответственно.

Вероятности уничтожения отдельных элементов системы A_{RC} одним c -элементом среды B на первом и втором шагах имеют согласно соотношению (6.13) вид:

$$\text{и } \left. \begin{aligned} p &= p[n_R(0)] = \exp[-dn_R(0)] \\ p_{v_R} &= p[n_R(0) - v_R] = \exp\{-d[n_R(0) - v_R]\} \end{aligned} \right\} \quad (6.48)$$

соответственно.

Ввиду независимости поражений отдельных элементов системы, случайные величины v_R и v_a при фиксированном значении $v_R = n_R$ имеют биномиальные распределения

$$\left. \begin{aligned} \mathcal{P}(v_R = n_R) &= B_{m_R}(n_R, p) = C_{m_R}^{n_R} p^{n_R} q^{m_R - n_R}, \\ \mathcal{P}(v_a = n_a | v_R = n_R) &= B_{m_a}(n_a, p_{n_R}) = \\ &= C_{m_a}^{n_a} p_{n_R}^{n_a} q_{n_R}^{m_a - n_a}, \end{aligned} \right\} \quad (6.49)$$

где $q = 1 - p$ и $q_{n_R} = 1 - p_{n_R}$.

Используя соотношения (6.49), получим безусловное распределение случайной величины v_a в виде

$$\mathcal{P}(v_a = n_a) = \sum_{n_R=0}^{m_R} C_{m_R}^{n_R} p^{n_R} q^{m_R - n_R} C_{m_a}^{n_a} p_{n_R}^{n_a} q_{n_R}^{m_a - n_a}. \quad (6.50)$$

Используя выражение (6.50), можно подсчитать вероятность P_{RC} того, что уязвимая система A_{RC} не будет выведена из строя.

Из определения гибели системы следует, что искомая вероятность $P_{RC} = P_{RC}[n_a(0)]$ имеет вид:

$$P_{RC}[n_a(0)] = \mathcal{P}[v_a < \theta_a n_a(0)] = \sum_{n_a=0}^{\theta_a n_a(0) - 1} \mathcal{P}(v_a = n_a). \quad (6.51)$$

Эту вероятность будем называть *живучестью* системы.

Имеет место

Предельный закон живучести. Пусть уязвимая система A_{RC} имеет коэффициент живучести θ_a и коэффициент неуязвимости θ_{RC} . Тогда ее живучесть $P_{RC}[n_a(0)]$ имеет следующие оценки:

$$P_{RC}[n_a(0)] \left\{ \begin{aligned} &\geq 1 - \exp(-m_a k) \quad \text{при } \theta_{RC} > p^{(m_a)}, \\ &\leq \exp(-m_a k') \quad \text{при } \theta_{RC} < 1 - q^{(m_a)} < p^{(m_a)}, \end{aligned} \right. \quad (6.52)$$

где

$$\theta_{RC} = \theta_a n_a(0) / m_a, \quad k = k(\theta_{RC}, p^{(m_a)}), \quad k' = k(\theta_{RC}, q^{(m_a)}),$$

$$k(x, y) = x \ln \frac{x}{y} + (1 - x) \ln \frac{1 - x}{1 - y},$$

$$\left. \begin{aligned} p^{(m_a)} &= \left[\sum_{n_R=0}^{m_R} p_{n_R}^{m_a} B_{m_R}(n_R, p) \right]^{1/m_a} \\ &= \exp\{-d[n_R(0) - m_R \varphi(dm_a, p)]\}, \end{aligned} \right\} \quad (6.53)$$

$$q^{(m_a)} = \left[\sum_{n_R=0}^{m_R} q_{n_R}^{m_a} B_{m_R}(n_R, p) \right]^{1/m_a}$$

и

$$\varphi(x, p) = \frac{1}{x} \ln [1 + p(\exp x - 1)], \quad (6.54)$$

а величины \dot{a} , p и $B_{m_R}(n_R, p)$ определены соотношениями (6.13), (6.48) и (6.49) соответственно.

Доказательство. Для оценки вероятности $P_{RC}\{n_a(0)\}$, определяемой соотношением (6.51), можно воспользоваться оценками (П.1.24) для случая $l=2$, полагая $a=m_R$, $i=n_R$, $b_i=B_{m_R}(n_R, p)$, $n=m_a$ и $m=\theta_{RC}n_a(0)$, откуда $m/n=\theta_{RC}$.

При этом с учетом (П.1.25) для $l=2$ получим соотношения (6.52) и (6.53), за исключением экспоненциального выражения параметра $p^{(m_a)}$. Чтобы получить последнее, необходимо подставить явное выражение вероятности p_{n_R} , определяемое в соотношениях (6.48), в соответствующую сумму (6.53). С учетом выражения (6.49) будем иметь

$$\sum_{n_R=0}^{m_R} p_{n_R}^{m_a} B_{m_R}(n_R, p) = \sum_{n_R=0}^{m_R} \{\exp[-dm_a n_R(0)]\} \times \\ \times \{\exp[dm_a n_R]\} B_{m_R}(n_R, p) = \{\exp[-dm_a n_R(0)]\} \times \\ \times \{1 + p[\exp(dm_a) - 1]\}^{m_R}.$$

Отсюда после несложного преобразования второго сомножителя последнего соотношения получим экспоненциальное выражение величины $(p^{(m_a)})$, используя функцию $\varphi(x, p)$, определенную соотношением (6.54). Таким образом, предельный закон живучести доказан полностью.

Прежде всего обсудим некоторые аналитические вопросы, связанные с доказанным предельным законом. Следствием использования при его доказательстве оценок (П.1.24), (П.1.25) является неопределенность этих оценок на некотором интервале (см. приложение 1).

В нашем случае интервал неопределенности для коэффициента уязвимости θ_{RC} имеет вид $1 - q^{(m_a)} \leq \theta_{RC} < p^{(m_a)}$. Для того чтобы оценить величину этого интервала, необходимо дать оценку выражения $1 - q^{(m_a)}$. Из (6.53) имеем

$$q^{(m_a)} = \left[\sum_{n_R=0}^{m_R} (1 - p_{n_R})^{m_a} B_{m_R}(n_R, p) \right]^{1/m_a} \leq \\ \leq \left[\sum_{n_R=0}^{m_R} (1 - p_{n_R}) B_{m_R}(n_R, p) \right]^{1/m_a} = (1 - p^{(1)})^{1/m_a}.$$

Таким образом, искомым интервал содержится в интервале $[1 - (1 - p^{(1)})^{1/m_a}, p^{(m_a)}]$. Или, пренебрегая величинами порядка малости $O[(p^{(1)})^2]$ и $O(\ln m_R/m_R)$, с учетом (6.53) имеем интервал

$$[p^{(1)}, p^{(m_a)}],$$

где

$$p^{(1)} = \exp\{-d[n_R(0) - m_R \varphi(d, p)]\} > p = \\ = \exp[-dn_R(0)].$$

Таким образом, полученная оценка заведомо лучше тривиальной.

Из полученного предельного закона следует, что при фиксированном количестве жизненного субстрата E система должна выбрать числа a - и R -элементов $n_a(0)$ и $n_R(0)$ так, чтобы величина коэффициента неуязвимости θ_{RC} была максимальной. Если при этом она больше параметра $p^{(m_a)}$, то живучесть системы $P_{RC}\{n_a(0)\}$ с ростом величины m_a экспоненциально стремится к единице.

В предыдущих параграфах велась минимаксная оптимизация системы и среды по параметру $n_c(0)$ — числа c -элементов среды B , необходимых для уничтожения системы A_{RC} . Эта оптимизация, вообще говоря, может не приводить к оптимальной системе в указанном выше смысле. Решение этих вопросов требует дальнейших исследований.

По нашей терминологии (см. введение) нами был рассмотрен $(n_c(0), n_a(0))$ -обмен между системой и средой. В последнем оптимальном решении мы имеем $(n_c(0), E)$ -обмен.

6.5.3. Заключительные замечания. Защитные R -элементы имеют информационные функции (обнаружение c -элементов) и активные функции (подавление c -элементов).

Эти функции не обязаны выполняться одним и тем же универсальным R -элементом, а могут быть возложены на отдельные специализированные элементы системы A_{RC} . Будем обозначать их I - и c' -элементами соответственно. Если специализированные защитные элементы действуют последовательно и независимо, то вероятность поражения одного элемента системы A_{RC} в момент t од-

ним c -элементом среды B имеет вид

$$p[n_I(t), n_{c'}(t)] = \\ = 1 - [1 - (1 - D_I)^{\theta n_I(t)}][1 - (1 - D_{c'})^{\theta n_{c'}(t)}], \quad (6.55)$$

где $n_I(t)$ и $n_{c'}(t)$ — числа I - и c' -элементов системы A_{RC} в момент t , а D_I и $D_{c'}$ — вероятности обнаружения и поражения c -элемента I - и c' -элементом соответственно.

Для согласованно действующих, эффективных защитных элементов сомножители правой части соотношения (6.55) должны быть равны между собой и близки к единице. Итак, в этом случае имеем

$$n_I(t) \ln(1 - D_I) = n_{c'}(t) \ln(1 - D_{c'}), \quad (6.56)$$

и

$$p[n_I(t), n_{c'}(t)] \approx \\ \approx 2p[n_I(t)] = 2 \exp[-d_I n_I(t)], \quad (6.57)$$

где $d_I = \theta \ln(1 - D_I)^{-1}$.

Если среда B не может нарушить соотношения (6.56) (например, из-за того, что она не различает I - и c' -элементов), то соблюдение условий (6.56) при $t=0$ достаточно для их соблюдения и при любых t в симметричных условиях, определенных выше.

Уточним теперь вид основной вероятности, определяемой (6.57), исходя из статистического смысла вероятности D_I . Последняя имеет смысл вероятности правильного обнаружения и для близких к единице значений, как известно, может быть представлена в форме

$$D_I = 1 - \exp(-FE_I^\alpha),$$

где E_I — энергия сигнала, а F и α — не зависящие от нее константы. Для ряда практически интересных случаев (например, когерентного) $\alpha=1$ [1].

Полученные оптимальные результаты в терминах максимина величины $n_c(0)$ можно переформулировать в терминах минимакса величины $\theta_a n_a(0)$. Именно при фиксированном $n_c(0)$ поведение среды \bar{B} обращает в мак-

симум величину $\theta_a n_a(0)$, а затем последняя обращается в минимум структурной $|A_{RC}|_{opt}$ системы A_{RC} .

Пусть заранее заданы: число $n_c(0)$ c -элементов, которыми располагает среда B , число $n_a(0)$ a -элементов системы A_{RC} , величина θ_a коэффициента живучести и неоптимальным или оптимальным образом введенное число $n_R^*(0)$ c -элементов, при котором система A_{RC} еще выходит из строя. Тогда достаточно ввести соответствующим образом некоторое число $n_R(0) > n_R^*(0)$ R -элементов, чтобы система A_{RC} уже не выходила из строя при любом поведении \bar{B} среды B и число $n_R(0)$ R -элементов является минимальной платой за это.

Полученное решение можно использовать и в конфликтной ситуации, когда вместо среды B фигурирует система B_{RC} со структурой $|B_{RC}|$ и защитным поведением \bar{B}_{RC} , аналогичным структуре $|A_{RC}|$ и защитному поведению \bar{A}_{RC} системы A_{RC} . При этом система A_{RC} оказывает вредное воздействие на систему B_{RC} , аналогичное вредному воздействию \bar{B}_{RC} системы B_{RC} на систему A_{RC} .

Найденное оптимальное поведение \bar{B}_{opt} среды B регулируется не по ходу взаимодействия среды B с системой A_{RC} , как в случае динамического программирования (см. гл. 4), а заранее до такого взаимодействия. Оптимальное поведение среды B по ходу взаимодействия ее с системой A_{RC} может привести и к более тяжелому ущербу для последней. Однако если мы хотим спроектировать живучую систему A_{RC} до взаимодействия ее со средой B , то нам ничего не остается, как использовать приведенные результаты, тем более, что они в среднем описывают всевозможные исходы взаимодействия системы A_{RC} со средой B . Как уже упоминалось выше, при большом числе элементов системы A_{RC} на основе закона больших чисел отклонения (в том числе и неблагоприятные для системы A_{RC}) от средних маловероятны. Выказанные соображения не снижают актуальности расчета живучести системы на основе оценки вероятности выхода ее из строя, применимой при любом (в том числе и малом) числе элементов системы (см. п. 6.5.2).

Приведенные здесь результаты, за исключением предельного закона живучести, были получены в работе [28]. Дальнейшие обобщения этих результатов можно найти в работе [29], где была обобщена приведенная здесь модель живучести на случай конфликтной ситуации между системами A_{RC} и B_{RC} с учетом уязвимости их

c -элементов. В этой работе приводятся общие уравнения, описывающие рассматриваемую модель для любого числа шагов с коэффициентами эффективности защитных элементов, «регулируемыми» во времени. Задача такого регулирования сводится к матричной игре, допускающей численный расчет для ряда конкретных случаев. Получены аналитические условия выживания систем в зависимости от количества их элементов.

6.5.4. Примеры. Рассмотрим некоторые примеры оценки величины числа c -элементов $n_c(0)$ среды B , необходимых для уничтожения системы A_{RC} при различных степенях оптимизации последней.

Пример 6.1. Пусть структура системы A_{RC} оптимизирована по числу a - и R -элементов при заданной величине жизненного субстрата $E=16$. Рассмотрим конкретные значения коэффициентов $\theta_a=0,5$, $\theta F=B=1$ и $\delta=1,01$. Оценить экстремальные значения $n_a^{1,2}$ и $n_c^{1,2}$.

Решение. Из соотношений (6.41), (6.42) и (6.44) вычислим значения параметров

$$\beta \approx 1600, \quad \gamma \approx 100, \quad \omega = 0,25 \quad \text{и} \quad E_{кр} = 4 < E = 16.$$

Далее по графику рис. 6.2 находим значения функции

$$\psi_{1,2}(0,25) \approx 16; 1.$$

Используя соотношения (6.38), находим экстремальные значения

$$n_a^{1,2} \approx (7; 1600)$$

и

$$\ln n_c^{1,2} = (17,2; 7,7)$$

или

$$\log_{10} n_c^{1,2} = (7, 3; 3, 2) \quad (\text{см. рис. 6. 3}).$$

Пример 6.2. Пусть структура системы A_{RC} не оптимизирована по числу a -элементов. При тех же значениях параметров, что и в предыдущем примере, найти число c -элементов $n_c(0)$, необходимых среде для уничтожения системы при экстремальном значении $n_a^2=1600$, если система не оптимизирована и по числу R -элементов.

Решение. По значениям n_a^2 , β и γ вычислим, используя соотношения (6.26) и (6.33), величину $E_R \approx 0,95$.

Далее по величине $\theta F=1$ и E_R найдем, используя соотношение (6.18'), при $\alpha=1$ величину $G=dn_R(0)=\theta F E_R=0,95$, откуда $G \log_{10} e = 0,42$.

Используя графики рис. 6.1 при значениях $G \log_{10} e = 0,42$ и $\log_{10} \theta_a n_a^2 = 2,5$ и $\log_{10} \ln \theta_a n_a^2 = 0,7$, получим

$$\log_{10} n_c(0) = 0,2 \quad \text{и} \quad \log_{10} n_c(0) = 3,2$$

соответственно. При этом последнее значение величины $\log n_c(0) = 3,2$, как это и должно быть, совпадает с полученным в предыдущем примере в качестве минимума $\log n_c^2 = 3,2$.

7. Теория осуществимости

§ 7.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Мы приступаем к изложению важнейшего раздела теории потенциальной эффективности сложных систем — теории осуществимости [9, 15]. Как уже упоминалось во введении, понятие осуществимости имеет принципиальное теоретическое значение, так как оно позволяет отличать модели сложных систем от обычных математических моделей.

Теория осуществимости имеет также и первостепенное практическое значение, так как с ее помощью рассмотренные сравнительно абстрактные модели X - и XU -сечений используются в моделях, где фигурирует реальное физическое время и такие параметры, как быстродействие V [опер/сек] и объем памяти K .

Эти два параметра V и K , вводимые в рассмотрение теорией осуществимости, делают ее как бы промежуточным мостом между теорией сложных систем и практическими приложениями.

Следует сразу заметить, что в отличие от осуществимости понятие технической реализуемости той или другой абстрактной модели далее связывается с принципиальными временными обстоятельствами, например наличием или отсутствием тех или иных известных технических средств.

Самые верхние оценки параметров V и K , приведенные Эшби [30], основываются на принципиальных физических ограничениях. Это константа Бреммермана $V_0 = 1,6 \cdot 10^{47}$ опер/г·сек — максимальное число операций, которое может произвести в секунду один грамм произвольно организованного вещества и число $K_0 = 10^{100}$ атомов во вселенной, ограничивающее объем памяти любой устройства. Ясно, что любые модели сложных систем, реализация или функционирование которых связаны с величинами параметров V и K , выходящими за пределы V_0 и K_0 соответственно, остаются вне рассмотрения теории сложных систем. То же можно сказать о моделях, реализация или функционирование которых требует скоростей v , превышающих скорость света.

Следует заметить, что даже такие грубые ограничения приводят, как это заметил Эшби [30], к резкому сужению круга осуществимых моделей сложных систем.

На самом деле этот круг значительно уже. В свете вышеизложенного те несколько лишних порядков, которые по быстрдействию и доступности памяти имеют ЭВМ по сравнению с человеком, не столь уж заметны. Поэтому ценность ЭВМ в большей мере связана с возможностью освобождения человека от рутинной работы для концентрации своих усилий на постановке и отработке методов решения творческих задач.

Последующее изложение связано с уточнением понятия осуществимости и решением некоторых общих вопросов теории осуществимости, в частности содержательного введения параметров V и K в некоторые общие модели.

§ 7.2. ПОНЯТИЕ ОСУЩЕСТВИМОСТИ. ОСНОВНЫЕ СООТНОШЕНИЯ В ЭКСПОНЕНЦИАЛЬНОМ СЛУЧАЕ

7.2.1. Понятие осуществимости. Рассмотрим системы A и B в условиях конфликтной ситуации, когда они имеют противоположные цели \underline{A} и \underline{B} соответственно.

Это означает, что для фиксированной продолжительности T их взаимодействия вероятности достижения системами A и B своих целей \underline{A} и \underline{B} равны

$$P(T) = \mathcal{P}(\underline{A}) \quad \text{и} \quad Q(T) = 1 - P(T) = \mathcal{P}(\underline{B}). \quad (7.1)$$

соответственно. Таким образом, можно всегда ограничиваться оценками величины $P(T)$ (величина $Q(T)$ дополняет последнюю до единицы).

Пусть из практических соображений назначаются допустимая вероятность $P^{(0)}$ достижения цели \underline{A} системой A и практически приемлемое время $T^{(0)}$, необходимое для этого. Пару значений $(P^{(0)}, T^{(0)})$ будем называть *по-рогами осуществимости*. Основания для конкретного выбора значений $(P^{(0)}, T^{(0)})$ здесь не рассматриваются, так же как и при статистических исследованиях не обсуждаются величины уровней значимости. Очевидна лишь тенденция выбирать значения $P^{(0)}$ близкими к единице, а значения $T^{(0)}$ не слишком большими.

Мы скажем, что достижение цели \underline{A} системой A *осуществимо*, если для некоторого значения T одновременно выполняются два неравенства:

$$P(T) \geq P^{(0)} \quad \text{и} \quad T \leq T^{(0)}. \quad (7.2)$$

Аналогично определяется осуществимость достижения своей цели \underline{B} системой B .

Определение неосуществимости достижения цели более сложно, так как требует предварительного введения нового понятия — оптимальной в целом системы A .

Рассмотрим системы A и B из двух достаточно широких классов \mathfrak{A} и \mathfrak{B} соответственно.

Системы A_1 и B_1 назовем *оптимальными в целом*, если они удовлетворяют минимаксным уравнениям

$$\begin{aligned} \max_{A \in \mathfrak{A}} \min_{B \in \mathfrak{B}} M(A, B, T) &= \min_{B \in \mathfrak{B}} \max_{A \in \mathfrak{A}} M(A, B, T) = \\ &= M(A_1, B_1, T), \end{aligned} \quad (7.3)$$

где

$$M(A, B, T) = P(T) - 1/2, \quad (7.4)$$

и в случае отсутствия \max (\min) берется \sup (\inf). Здесь вычитание $1/2$ из вероятности $P(T)$ производится для возможности формального рассмотрения игры с нулевой суммой между системами A и B (см. гл. 4), так как при этом «функция выигрыша» системы B имеет согласно соотношениям (7.1) и (7.4) вид

$$Q(T) = 1/2 = 1/2 - P(T) = -M(A, B, T).$$

Мы скажем, что достижение цели \underline{A} системой A *неосуществимо*, если для оптимальных в целом систем A_1 и B_1 не выполняется хотя бы одно из неравенств (7.2). Аналогично определяется неосуществимость цели \underline{B} системы B .

В случае индифферентной ситуации в соотношении (7.3) надо опустить знак \min и вместо функции $M(A, B, T)$ мож-

но рассматривать величину $P(T)$ [9]. Далее для простоты мы будем рассматривать именно этот случай. Таким образом, анализ понятия неосуществимости снова возвращает нас к рассмотрению вопроса об оптимальности системы. Основным здесь является проблема оптималь-

ности системы в ее частях и в целом. К этому вопросу мы еще вернемся в дальнейшем изложении. Для возможности дальнейших положительных рассмотрений необходимо задание явного вида зависимости вероятности $P(T)$ от времени T .

7.2.2. Основные соотношения в экспоненциальном случае. С ростом времени T , отпущенным системе

$$A = (A_R, A_I, A_C, A_L)$$

на достижение цели

$$\underline{A} = (\underline{A}_R, \underline{A}_I, \underline{A}_C, \underline{A}_L),$$

отдельные ее сечения A_X ($X=R, I, C, L$) ведут себя по-разному. Эффективность $P_R(T)$ первого R -сечения убывает, а эффективности $P_X(T)$ ($X=I, C, L$) остальных X -сечений возрастают, причем у второго и третьего существенно быстрее, чем у четвертого. В связи с этим ограничимся рассмотрением таких сравнительно небольших значений величины T , когда можно пренебречь влиянием последнего L -сечения, тем более, что зависимость вероятности $P_L(T)$ от времени T еще не выяснена (в гл. 5 приводится зависимость величины P_L от значения лишь одного «структурного» параметра n).

Таким образом, рассмотрим систему

$$A = (A_R, \bar{A}_R),$$

где

$$\bar{A}_R = (A_I, A_C).$$

Вероятность $P(T) = P_A(T)$ с опущенным индексом B в рассматриваемой индифферентной ситуации, а также с учетом особого поведения первого R -сечения удобно записать в следующем виде:

$$P_A(T) = P_R(T) P_{\bar{R}}(T). \quad (7.5)$$

Здесь первый сомножитель означает надежность, т. е. вероятность безотказного функционирования системы A за все время T , а второй сомножитель означает условную вероятность достижения цели \bar{A} системой A за время T при условии ее безотказного функционирования за это же время.

До количественной оценки вероятности $P(T)$, определяемой соотношением (7.5), можно привести ряд качественных соображений относительно ее поведения.

В самом деле, с ростом величины T вероятность $P_R(T)$ безотказного функционирования убывает, а вероятность $P_{\bar{R}}(T)$ — возрастает. В результате вероятность $P_A(T)$, определяемая соотношением (7.5), оказывается произведением убывающей и возрастающей с ростом T функций. Отсюда следует возможность существования максимума $\bar{P} = \max_T P_A(T)$. Качественное поведение функции $P_A(T)$ в зависимости от величины T изображено на рис. 7.1.

Задание порогов осуществимости ($P^{(0)}, T^{(0)}$) приводит к заштрихованной на рис. 7.1 области G плоскости (T, R) . Достижение цели осуществимо тогда и только тогда, когда функция $P_A(T)$ имеет хотя бы одну общую точку с областью G . При этом естественно ориентироваться на такую точку с минимальным значением T .

Приступим теперь к количественной оценке вероятности $P_A(T)$. Для этого используем оценки Буля (П.2.3) для второго сомножителя соотношения (7.5). Получим

$$\min [P_I(T), P_C(T)] \geq P_{\bar{R}}(T) \geq P_I(T) + P_C(T) - 1. \quad (7.6)$$

Допустим теперь, что с ростом величины T вероятности, фигурирующие в оценках (7.6) величины $P_R(T)$, экспоненциально стремятся к единице (это допущение основано на предельных законах, доказанных в предыдущих главах). Тогда в силу оценок (7.6) такого же рода стремление имеет место для величины $P_{\bar{R}}(T)$, т. е.

$$P_{\bar{R}}(T) = 1 - \exp(-f_{\bar{R}} T), \quad (7.7)$$

где величину $f_{\bar{R}}$, не зависящую от T , будем называть *параметром внешней эффективности* системы A .

Оценим теперь первый сомножитель $P_R(T)$ соотношения (7.5). Если исходить из расширяющейся по логарифмическому закону в зависимости от T модели A_R R -сечения (см. гл. 2), то с ростом T величина $P_R(T)$, убывая, может быть сколь угодно близкой к единице. Таким образом, для этой модели практически исключается влияние первого R -сечения A_R на эффективность системы A в целом.

Рассмотрим случай ограниченной модели A_R , когда число ее элементов постоянно. В этом случае при весьма широких предположениях имеет место хорошо известный экспоненциальный закон надежности (см. гл. 2)

$$P_R(T) = \exp(-T/f_R), \quad (7.8)$$

где величину $f_{\bar{R}} = 1/\lambda$ [см. соотношение (2.9)] будем называть *параметром внутренней эффективности системы A*.

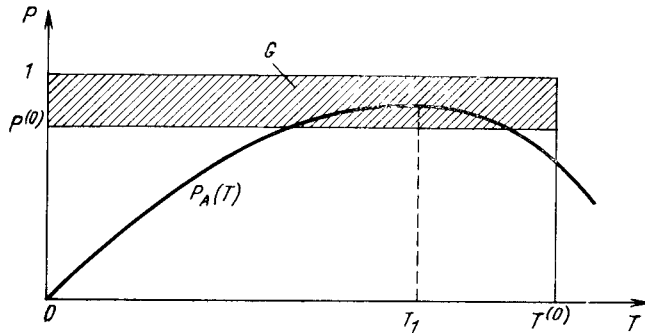


Рис. 7.1. Качественное поведение функций $P_A(T)$ в зависимости от величины T .

Эта величина равна среднему времени между отказами системы A .

В рассматриваемом экспоненциальном случае можно дать явное выражение для величин $\bar{P} = P_A(T_1) = \max P_A(T)$ и T_1 через параметры $f_{\bar{R}}$ и f_R . Для оптимальной системы A_1 эта величина определяет потенциальную эффективность системы A . Приступим к вычислению величины \bar{P} . Согласно соотношениям (7.5), (7.7) и (7.8) вероятность $P_A(T)$ имеет вид:

$$R_A(T) = [1 - \exp(-f_{\bar{R}}T)] \exp(-T/f_R),$$

откуда

$$\begin{aligned} P'_A(T) &= [f_{\bar{R}} - f_R^{-1} (\exp f_{\bar{R}}T - 1)] \exp[-(f_{\bar{R}} + f_R^{-1})T] = \\ &= f_{\bar{R}} \exp[-(f_{\bar{R}} + f_R^{-1})T] - f_R^{-1} P_A(T) \end{aligned}$$

и

$$\begin{aligned} P'_A(T) &= \{-f_{\bar{R}}(f_{\bar{R}} + f_R^{-1}) - f_{\bar{R}}f_R^{-1} + f_{\bar{R}}^{-2} [\exp(f_{\bar{R}}T) - \\ &- 1]\} \exp[-(f_{\bar{R}} + f_R^{-1})T] = -f_{\bar{R}}(f_{\bar{R}} + f_R^{-1}) \exp[-(f_{\bar{R}} + \\ &+ f_R^{-1})T] - f_R^{-1} P_A(T). \end{aligned}$$

Приравнявая $P'_A(T)$ нулю, находим точку $T_1 = T(f_R, f_{\bar{R}})$ экстремума

$$\begin{aligned} T_1^* = T(f_R, f_{\bar{R}}) &= \frac{1}{f_{\bar{R}}} \ln(1 + f_R f_{\bar{R}}) = \\ &= f_R \frac{1}{f_{\bar{R}} f_R} \ln(1 + f_R f_{\bar{R}}), \end{aligned} \quad (7.9)$$

в которой $P_A(T)$ обращается в максимум, равный

$$\begin{aligned} \bar{P} = P_A(T_1) &= \bar{P}(f_R, f_{\bar{R}}) = \\ &= \left(1 - \frac{1}{1 + f_R f_{\bar{R}}}\right) (1 + f_R f_{\bar{R}} - 1/f_R f_{\bar{R}}) = \\ &= f_R f_{\bar{R}} (1 + f_R f_{\bar{R}})^{-(1 + f_R f_{\bar{R}})/f_R f_{\bar{R}}}, \end{aligned} \quad (7.10)$$

а

$$P'_A(T_1) = 0$$

и

$$P''_A(T_1) = -f_{\bar{R}} f_R^{-1} (1 + f_R f_{\bar{R}})^{-1/f_R f_{\bar{R}}} < 0. \quad (7.11)$$

Естественно, введение параметра $f = f_R f_{\bar{R}}$ (его мы будем называть *параметром эффективности системы A*), так как через него легко выразить величины \bar{P} и T_1 .

Из соотношений (7.9), (7.10) и (7.11) имеем

$$\begin{aligned} T_1 = T_1(f) &= f_{\bar{R}}^{-1} \ln(1 + f) = f_R \frac{1}{f} \ln(1 + f), \\ \bar{P} = P_A(T_1) &= \bar{P}(f) = \left(1 - \frac{1}{1 + f}\right) (1 + f)^{-1/f} = \\ &= f(1 + f)^{-(1+f)/f}, \\ P''_A(T_1) &= -\frac{f_{\bar{R}}^2}{f} (1 + f)^{-1/f} = \\ &= -\frac{1}{f^2} (1 + f) \bar{P}(f). \end{aligned} \quad (7.11')$$

Параметр f меняется в пределах $0 \leq f < \infty$, при этом величина $T_1(t)/f_R$ изменяется в пределах $1 \geq T_1(f)/f_R \geq 0$, а величина $\bar{P}(f)$ изменяется в пределах $0 \leq \bar{P}(f) \leq 1$.

На рис. 7.2 приведены графики зависимости предельного значения вероятности $\bar{P} = \bar{P}(f)$ достижения цели системой A и идущего на это времени $T_1(f)/f_R$, нормированного средним временем между отказами f_R , в зависимости от величины параметра эффективности системы f .

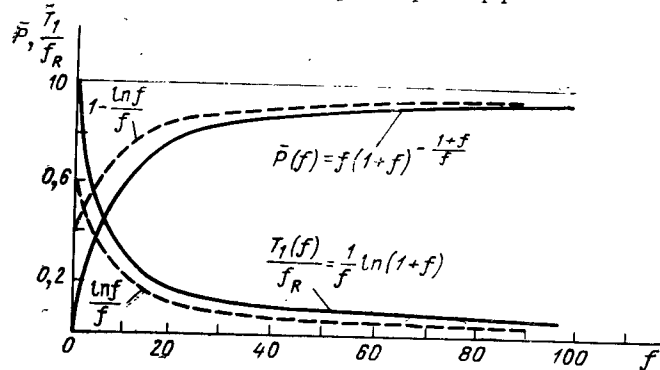


Рис. 7.2. Зависимость предельного значения вероятности $\bar{P} = \bar{P}(f)$ достижения цели системы и идущего на это времени $T_1(f)/f_R$ нормированного средним временем между отказами f_R системы A от параметра эффективности f .

системы A . Асимптотические формулы при $f \rightarrow 0$ и $f \rightarrow \infty$ таковы:

$$T_1 = T_1(f) = \begin{cases} f_R \left(1 - \frac{f}{2}\right) & \text{при } f \rightarrow 0, \\ f_R \frac{\ln f}{f} & \text{при } f \rightarrow \infty, \end{cases} \quad (7.12)$$

$$\bar{P} = \bar{P}(f) = \begin{cases} fe^{-1} + o(f) & \text{при } f \rightarrow 0, \\ 1 - \frac{\ln f}{f} + o(f^{-1}) & \text{при } f \rightarrow \infty. \end{cases} \quad (7.13)$$

Зная величины $P_A(T_1)$ и $P''_A(T_1)$ [см. (7.11')], можно вычислить значения $P_A(T)$ в окрестности точки T_1 по формуле $P_A(T) \approx P_A(T_1) + (T - T_1)^2 P''_A(T_1)/2 =$

$$= \bar{P}(f) \left\{ 1 - \left[\frac{T}{f_R} - \frac{1}{f} \ln(1+f) \right]^2 (1+f)/2 \right\}. \quad (7.14)$$

Так как основное практическое значение имеют значения вероятности $\bar{P}(f)$, близкой к единице, на рис. 7.3, а, б, в

приведены ее значения, близкие к единице, вычисленные по асимптотической формуле (7.13), при $f \rightarrow \infty$, начиная с $f=150$, когда асимптотическая формула начинает давать почти полное совпадение с точной. Заметим, что согласно асимптотической формуле (7.12) при $f \rightarrow \infty$ значение T_1/f_R просто вычисляется вычитанием значения

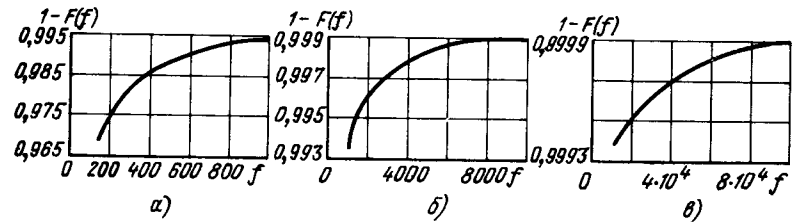


Рис. 7.3. Зависимость значений вероятности $\bar{P}(f) = 1 - F(f)$, близких к единице от величины f :

а) $f \leq 1000$; б) $1000 \leq f \leq 10\,000$ и в) $10\,000 \leq f \leq (\sim) 100\,000$.

$\bar{P}(f)$ из единицы. Итак, в этом случае имеем

$$1 - \bar{P} \approx \frac{T_1}{f_R} \approx F(f) = \frac{\ln f}{f}. \quad (7.15)$$

При заданных порогах осуществимости ($P^{(0)}, T^{(0)}$), полагая $\bar{P} = P^{(0)}$ и $T_1 = T^{(0)}$, получим из соотношений (7.15) выражения для необходимых параметров эффективности:

$$f = F^{-1}(1 - P^{(0)}), \quad f_R = T^{(0)} / (1 - P^{(0)}) \quad \text{и} \quad f_{\bar{R}} = f / f_R. \quad (7.16)$$

7.2.3. Примеры. Рассмотрим несколько числовых примеров.

Пример 7.1. Пусть среднее время между отказами системы A $f_R = 1$ сутки $= 24 \cdot 60 \cdot 60 \text{ сек} \approx 8 \cdot 10^5 \text{ сек}$. Пусть для осуществления некоторой цели A система имеет параметр внешней эффективности $f_{\bar{R}} = 0,01 \text{ сек}^{-1}$. Пороги осуществимости назначены $T^{(0)} = 15 \text{ мин}$ и $P^{(0)} = 0,995$.

Осуществимо ли достижение цели A ?

Решение. Подсчитаем параметр эффективности системы A $f = f_{\bar{R}} \cdot f_R = 10^{-2} \cdot 8 \cdot 10^5 = 8 \cdot 10^3$. Далее по графику рис. 7.3, б находим по $f = 8 \cdot 10^3$ значение $\bar{P}(8 \cdot 10^3) = 0,9989$. Далее, по $\bar{P}(8 \cdot 10^3)$ находим величину $T_1(f) = f_R [1 - \bar{P}(8 \cdot 10^3)] = 8 \cdot 10^5 \cdot 10^{-3} = 8 \cdot 10^2 \text{ сек} = 13 \text{ мин } 20 \text{ сек}$. Таким образом, в нашем случае $\bar{P}(f) \geq P^{(0)}$ и $T_1(f) < T^{(0)}$ и цель A осуществима.

Пример 7.2. Каково должно быть среднее время f_R безотказной работы системы A и параметр внешней эффективности $f_{\bar{R}}$ ее функ-

ционирования, чтобы цель, поставленная перед ней, была осуществлена при порогах осуществимости $T^{(0)} = 50 \text{ сек}$ и $P^{(0)} = 0,99$.

Решение. Определим по графику рис. 7.3, а значение f , приводящее к $\bar{P}(f) \geq P^{(0)} = 0,99$. Оно оказывается равным $f = f_R \bar{f}_R \geq \geq 650$. Одновременно ему соответствует величина $T_1(f) = f_R [1 - \bar{P}(f)] = f_R \cdot 0,01 \leq 50 \text{ сек}$. Откуда $f_R \leq 5000 \text{ сек} = 1 \text{ час } 23 \text{ мин } 20 \text{ сек}$ и $\bar{f}_R \geq 650/5000 = 0,13 \text{ сек}^{-1}$.

Подчеркнем, что полученные результаты, дающие простую зависимость максимальной вероятности осуществления цели $\bar{P}(f)$ и соответствующего времени осуществления цели $T_1(f)$ от параметра $f = f_R \bar{f}_R$ эффективности системы A , нацеливают дальнейшие исследования зависимостей сомножителей f : f_R и \bar{f}_R от таких, например, параметров как быстродействие $V[\text{опер/сек}]$ и объем памяти K системы A .

Выше уже упоминалось о том, что задание порогов осуществимости ($P^{(0)}$, $T^{(0)}$) выходит за рамки развиваемой теории. Однако порядок этих величин можно себе представить в зависимости от «ответственности» решаемых задач. В связи с этим рассмотрим три примера таких задач с нарастающей ответственностью. При этом удобно иметь выражение привычных интервалов времени в секундах:

Секунды	Минуты	Часы	Сутки	Месяцы	Годы	Столетия	Миллионы лет
10 сек	10^2 сек	10^4 сек	10^6 сек	10^8 сек	10^9 сек	10^{10} сек	10^{15} сек

Пример 7.3. Рассмотрим повседневный технический процесс, срывы которого не связаны с риском для жизни людей. Здесь можно задать пороги осуществимости $(P^{(0)}, T^{(0)}) = (1-10^{-3}, 10 \text{ сек})$. Каковы величины f_R и \bar{f}_R необходимы для осуществления рассматриваемого процесса?

Решение. Используя соотношения (7.16), получим
 $f_R = 10^4 \text{ сек}$, $\bar{f}_R = 1 \text{ сек}^{-1}$.

Пример 7.4. Рассматривается технический процесс, срывы которого связаны с риском для жизни людей. Пороги осуществимости здесь задаются на уровне ежедневного риска уличных катастроф в городе с многомиллионным населением $(P^{(0)}, T^{(0)}) = (1-10^{-6}, 10^5 \text{ сек})$. Каковы соответствующие значения f_R и \bar{f}_R ?

Решение. Используя соотношения (7.16), получим
 $f_R = 10^{11} \text{ сек}$, $\bar{f}_R = 10^{-3,75} \text{ сек}^{-1}$.

Пример 7.5. Рассмотрим глобальный процесс, срыв которого связан с риском для существования всего человечества. Пороги осуществимости задаются здесь на уровне ежегодного ожидания космической катастрофы Земли $(P^{(0)}, T^{(0)}) = (1-10^{-15}, 10^9 \text{ сек})$. Каковы соответствующие значения f_R и \bar{f}_R ?

Решение. Используя соотношения (7.16), получим
 $f_R = 10^{24} \text{ сек}$, $\bar{f}_R = 10^{-7} \text{ сек}^{-1}$.

§ 7.3. ОЦЕНКА ЭФФЕКТИВНОСТИ СИСТЕМЫ ПО ЭФФЕКТИВНОСТИ ЕЕ X-СЕЧЕНИЙ. ПРАВИЛО СРЕДНЕГО ГАРМОНИЧЕСКОГО

7.3.1. Проблема части и целого. В § 1.4 введения уже рассматривались подходы к решению этой проблемы на основе неравенства Буля. Именно, пусть имеется конечный набор $\mathfrak{X} = \{X\}$ из m качеств системы $A = \{A_X\}$. При этом стратегическая цель системы $\underline{A} = \{\underline{A}_X\}$ распадается на m тактических подцелей \underline{A}_X ее отдельных X-сечений A_X .

Тогда эффективность системы $\mathcal{P}(\underline{A})$ оценивается через эффективность $\mathcal{P}(\underline{A}_X)$ ее отдельных X-сечений согласно формулам Буля (П.2.3)

$$1 - \sum_{X \in \mathfrak{X}} [1 - \mathcal{P}(\underline{A}_X)] \leq \mathcal{P}(\underline{A}) \leq \min_{X \in \mathfrak{X}} \mathcal{P}(\underline{A}_X). \quad (7.17)$$

Ослабим оценку (7.17), полагая

$$1 - m \max_{X \in \mathfrak{X}} [1 - \mathcal{P}(\underline{A}_X)] \leq \mathcal{P}(\underline{A}) \leq \min_{X \in \mathfrak{X}} \mathcal{P}(\underline{A}_X). \quad (7.18)$$

Из результатов предыдущих глав следует, что рассматривавшиеся там величины $\mathcal{P}(\underline{A}_X)$ при $X \neq R$ имеют экспоненциальный с ростом времени T порядок стремления к единице [см. соотношения (7.1) и (7.7)]. Если величина m с ростом T не имеет экспоненциального роста, то оценки (7.18) переходят в асимптотические с ростом T оценки

$$1 - \max_{X \in \mathfrak{X}} [1 - \mathcal{P}(\underline{A}_X)] \lesssim \mathcal{P}(\underline{A}) \leq \min_{X \in \mathfrak{X}} \mathcal{P}(\underline{A}_X),$$

откуда имеем

$$\min_{X \in \mathfrak{X}} \mathcal{P}(\underline{A}_X) \leq \mathcal{P}(\underline{A}) \leq \min_{X \in \mathfrak{X}} \mathcal{P}(\underline{A}_X)$$

и

$$\mathcal{P}(\underline{A}) \approx \min_{X \in \mathfrak{X}} \mathcal{P}(\underline{A}_X), \quad (7.19)$$

т. е. в экспоненциальном случае эффективность системы A в целом асимптотически при больших T совпадает с минимальной эффективностью своего X-сечения. Таким образом, эффективность системы в целом определя-

ется эффективностью ее слабого звена («узкого места»).

Соотношение (7.19) указывает на целесообразность назначения экспоненциально совпадающих эффективностей $\mathcal{P}(A_x)$ отдельных X -сечений системы A :

$$\mathcal{P}(A_x) \approx \mathcal{P}(A) (X \in \mathfrak{X}). \quad (7.20)$$

В самом деле, задание их экспоненциально различными неощутимо для экспоненциального значения величины $\mathcal{P}(A)$, так как согласно (7.19) ее порядок определяется максимальным порядком величины $1 - \mathcal{P}(A_x)$.

В свете вышеизложенного найденные в предыдущих главах предельные законы эффективности X -сечений можно рассматривать как законы, устанавливающие минимальную временную или элементную избыточность в качестве платы за «девятки после запятой» у величин $\mathcal{P}(A_x)$.

Соотношение (7.20) можно трактовать как требование к X -сечениям иметь эффективность с одним и тем же количеством девяток после запятой. Это избавляет от излишнего введения избыточности в X -сечениях, не приводящего к какому-нибудь окончательному эффекту. Другими словами, соотношение (7.20) определяет систему A с «равнопрочными» частями A_x без «узких мест». Соображения «равнопрочности» могут быть использованы также и для получения так называемого правила среднего гармонического [7].

7.3.2. Правило среднего гармонического. Пусть для достижения некоторой цели A системе A необходимо на интервале времени $[0, T]$ последовательно предпринять l успешных независимых действий, каждое продолжительностью t_λ ($\sum_{\lambda=1}^l t_\lambda = T$). При этом каждое из таких действий приводит к своей эффективности, имеющей параметр внешней эффективности $f_{R\lambda}$ ($\lambda = \overline{1, l}$).

Тогда в экспоненциальном случае зависимости эффективности от времени, при больших значениях T и неэкспоненциальной зависимости величины l от T общий параметр эффективности $f_{\overline{R}}$ системы достигает своего

наибольшего значения

$$f_{\overline{R}} \approx \Gamma_{\overline{R}} / l \quad (7.21)$$

при выборе значений

$$t_\lambda = T \Gamma_{\overline{R}}^{-1} / f_{R\lambda} \quad (\text{заданы } f_{\overline{R}} \text{ и } T) \quad (7.22)$$

или

$$f_{R\lambda} = T f_{\overline{R}} / t_\lambda \quad (\text{заданы } t_\lambda \text{ и } f_{\overline{R}}), \quad (7.22')$$

где

$$\Gamma_{\overline{R}} = l \left/ \sum_{\lambda=1}^l f_{R\lambda}^{-1} \right.$$

— среднее гармоническое коэффициентов внешней эффективности $f_{R\lambda}$.

В самом деле, по условию имеем

$$\begin{aligned} P(A) &= \prod_{\lambda=1}^l [1 - \exp(-f_{R\lambda} t_\lambda)] \approx \\ &\approx 1 - \sum_{\lambda=1}^l \exp(-f_{R\lambda} t_\lambda). \end{aligned} \quad (7.23)$$

Последнее выражение при больших значениях t_λ и неэкспоненциальной зависимости величины l от T асимптотически определяется одним слагаемым с минимальным значением произведения $f_{R\lambda} t_\lambda$. Поэтому, чтобы излишне не завышать величины этих произведений, следует их выбрать равными друг другу, т. е. положить

$$f_{R\lambda} t_\lambda = C = \text{const}. \quad (7.24)$$

Из соотношения (7.24) следует, что

$$t_\lambda = C / f_{R\lambda}. \quad (7.25)$$

Для определения константы C просуммируем обе части (7.25) по λ . Получим

$$T = \sum_{\lambda=1}^l t_\lambda = C \sum_{\lambda=1}^l f_{R\lambda}^{-1}.$$

Отсюда

$$C = T \left| \sum_{\lambda=1}^l \bar{f}_{R\lambda}^{-1} \right| = T \Gamma_{\bar{R}} / l$$

и

$$t_{\lambda} = T \Gamma_{\bar{R}} / l \bar{f}_{R\lambda},$$

а если заданы $\bar{f}_{\bar{R}}$ и t_{λ} , то $\bar{f}_{R\lambda} t_{\lambda} = \bar{f}_{\bar{R}} T$, откуда

$$\bar{f}_{\bar{R}} = \Gamma_{\bar{R}} / l, \quad \bar{f}_{R\lambda} = T \bar{f}_{R\lambda} / t_{\lambda} = T \Gamma_{\bar{R}} / l t_{\lambda}$$

и согласно (7.23)

$$P(\underline{A}) \approx 1 - \exp(-T \Gamma_{\bar{R}} / l). \quad (7.26)$$

Итак, $\bar{f}_{\bar{R}} = \Gamma_{\bar{R}} / l$, и таким образом соотношения (7.21), (7.22) и (7.22') доказаны полностью. Эти соотношения можно назвать правилом среднего гармонического для внешней эффективности. Аналогичное правило имеет место и для параметра внутренней эффективности.

Будем говорить, что система A_R стареет в l раз быстрее системы A'_R , если параметр внутренней эффективности \bar{f}_R первой системы в l раз меньше параметра внутренней эффективности \bar{f}'_R второй системы ($\bar{f}_R = \bar{f}'_R / l$).

Другими словами, так называемая опасность отказа первой системы \bar{f}_R^{-1} в l раз больше, чем опасность отказа $(\bar{f}'_R)^{-1}$ второй системы.

Пусть на интервале времени $[0, T]$ система A параллельно (одновременно) осуществляет $l-1$ независимых внешних действий, каждое из которых в отдельности может привести к отказу системы A . Причем надежность системы всякий раз при этом определяется параметром внутренней эффективности $\bar{f}_{R\lambda}$ ($\lambda=2, l$). Пусть, кроме того, надежность системы A в пассивном состоянии (независимо от активных состояний) определяется параметром внутренней эффективности \bar{f}_{R1} .

Тогда общий параметр внутренней эффективности системы \bar{f}_R имеет вид:

$$\bar{f}_R = (\Gamma_R / l)^{-1}, \quad (7.27)$$

где

$$\Gamma_R = l \left| \sum_{\lambda=1}^l \bar{f}_{R\lambda}^{-1} \right| \quad (7.28)$$

— среднее гармоническое коэффициентов внутренней эффективности $\bar{f}_{R\lambda}$, а надежность системы A имеет вид:

$$P_R(T) = \exp(-T \Gamma_R / l). \quad (7.29)$$

В самом деле, по условию с учетом (7.28) имеем

$$R_R(T) = \prod_{\lambda=1}^l \exp(-T / \bar{f}_{R\lambda}) = \exp(-T \Gamma_R / l), \quad (7.30)$$

что доказывает одновременно соотношения (7.27) и (7.29).

Другими словами, в указанной выше ситуации система A стареет в l раз быстрее системы A' , имеющей параметр внутренней эффективности

$$\bar{f}'_R = \Gamma_R^{-1}.$$

7.3.3. Использование правила среднего гармонического для совместной оптимизации системы на нескольких иерархических уровнях. Изучим теперь связь между подсистемами трех иерархических рангов: нулевого (элементов a_i), первого (системы A) и второго (системы \mathfrak{A}).

Последняя мыслится как совокупность s подсистем $A^{(\sigma)}$ ($\sigma = \overline{1, s}$), рассматриваемых как ее «элементы»:

$$\mathfrak{A} = \{A^{(1)}, A^{(2)}, \dots, A^{(\sigma)}, \dots, A^{(s)}\}.$$

Зафиксируем число n первичных элементов a_i . Пусть они распределены между подсистемами $A^{(\sigma)}$ так, что каждая из них содержит n_{σ} элементов. Тогда

$$\sum_{\sigma=1}^s n_{\sigma} = n. \quad (7.31)$$

Пусть каждая из подсистем является симметричной порядка m_{σ} системой $A_{m_{\sigma}}^{(\sigma)}$ со средней вероятностью жизни элемента $p^{(\sigma)}$ и коэффициентом смерти $\theta_{\sigma} = m_{\sigma} / n_{\sigma}$ (см. п. 2.2.2). Рассмотрим характеристическую функцию δ_{σ} на множестве состояний системы $A_{m_{\sigma}}^{(\sigma)}$, равную единице на множестве ε_1^{σ} ее жизни и нулю на множестве ε_0^{σ} ее смерти.

Будем характеризовать состояния систем $A_{m_{\sigma}}^{(\sigma)}$ их характеристическими функциями δ_{σ} . Тогда состояния системы \mathfrak{A}

второго ранга характеризуются всевозможными 2^s наборами

$$\{(\delta_1, \delta_2, \dots, \delta_\sigma, \dots, \delta_s)\}.$$

Пусть подсистемы $A_{m_\sigma}^{(\sigma)}$ независимы в вероятностном смысле и для каждой из них подсчитана вероятность жизни $P_v^{(\sigma)}$ по формуле (2.1). Тогда можно определить симметричную порядка l систему \mathfrak{A}_l второго ранга, так же как определялась система A_m первого ранга. Ее множество жизни имеет вид

$$\left\{(\delta_1, \delta_2, \dots, \delta_\sigma, \dots, \delta_s) : \sum_{\sigma=1}^s \delta_\sigma > l\right\}.$$

Рассмотрим тот случай, когда все подсистемы $A_{m_\sigma}^{(\sigma)}$ без исключения важны для существования системы \mathfrak{A}_l . Другими словами, рассмотрим систему \mathfrak{A}_{s-1} , множество жизни которой состоит из единственного состояния $(\underbrace{1, 1, \dots, 1}_s)$. Вероятность жизни системы \mathfrak{A}_{s-1} согласно (2.2) и (2.5) имеет вид

$$P_v = \prod_{\sigma=1}^s P_v^{(\sigma)} \geq \prod_{\sigma=1}^s [1 - \exp(-n_\sigma k_\sigma)], \quad (7.32)$$

где

$$k_\sigma = \theta_\sigma \ln \frac{\theta_\sigma}{p^{(\sigma)}} + (1 - \theta_\sigma) \ln \frac{1 - \theta_\sigma}{1 - p^{(\sigma)}}. \quad (7.33)$$

Пусть индивидуальность каждой из подсистем $A_{m_\sigma}^{(\sigma)}$ характеризуется фиксированными значениями средней вероятности жизни ее элемента $p^{(\sigma)}$ и коэффициентом смерти $\theta_\sigma = m_\sigma/n_\sigma$, а ее порядок m_σ и число элементов n_σ переменны. При этом соблюдается условие (7.31) постоянства общего числа элементов n .

Естественно искать такое экстремальное распределение чисел элементов $\{n_\sigma^0\}$ по системам, при котором нижняя оценка (7.32) достигает максимума. Точное решение этой задачи методом множителей Лагранжа при больших значениях n_σ (а лишь этот случай и интересен) асимптотически

совпадает с решением, получающимся при значениях n_σ , обеспечивающих постоянство множителей в оценке (7.32). Последнее, как легко показать, достигается при

$$n_\sigma = n \Gamma_s / k_\sigma s, \quad (7.34)$$

и при этом из соотношения (7.32) имеем

$$P_v \geq [1 - \exp(-n \Gamma_s / s)]^s, \quad (7.35)$$

где

$$\Gamma_s = s \left/ \sum_{\sigma=1}^s k_\sigma^{-1} \right. \quad (7.36)$$

является *средним гармоническим* коэффициентов k_σ . Этот результат является следствием первого варианта более общего правила среднего гармонического при интерпретации времени числами элементов (см. п. 7.3.2). В частном случае постоянных вероятностей гибели элементов $p_\sigma = p$, когда при больших значениях n_σ неравенство (7.32) переходит в асимптотическое равенство, этот результат содержится в работе Пирса [31]. Им же была дана полезная стоимостная интерпретация чисел элементов n_σ и рекомендации о рациональном последовательном повышении жизнеспособности системы \mathfrak{A}_{s-1} .

Интересна биологическая интерпретация соотношения (7.34), которое удобно переписать в виде

$$\frac{n_\sigma}{n} = \frac{\Gamma_s}{s k_\sigma}. \quad (7.37)$$

В самом деле, если интерпретировать элементы клетками, подсистемы $A^{(\sigma)}$ — органами, а систему \mathfrak{A}_{s-1} — организмом, то при росте последнего за счет числа клеток для максимальной жизнеспособности организма в целом должны соблюдаться пропорции (n_σ/n) , определяющие размеры органов с учетом их жизнеспособности $(\Gamma_s/s k_\sigma)$ согласно соотношению (7.37). Важно отметить, что использование формулы Буля приводит к тому же результату без требования вероятностной независимости состояний жизни и смерти подсистем $A_{m_\sigma}^{(\sigma)}$ ($\sigma = \overline{1, s}$).

Интересно исследовать вопрос о целесообразности той или иной мелкости дробления системы \mathfrak{A} на s подсистем $A^{(\sigma)}$ с точки зрения жизнеспособности системы \mathfrak{A} в смысле определения п. 2.2.2. Для этого рассмотрим случай системы \mathfrak{A}_{s-1} с постоянными вероятностями гибели элементов, когда неравенство (7.32) переходит с ростом n_s в асимптотическое равенство [см. соотношение (1.10)]

$$P_v = \left\{ 1 - \exp \left[-n \frac{\Gamma_s}{s} + O \left(\ln \frac{n}{s} \right) \right] \right\}^s. \quad (7.38)$$

В этом случае имеет место

Предельный закон жизнеспособности. Пусть число элементов n системы \mathfrak{A}_{s-1} растет вместе с числом s ее подсистем так, что

$$n = \frac{1 + \varepsilon}{\Gamma_s} s \ln s - \frac{\ln C}{\Gamma_s} s, \quad (7.39)$$

где $C > 0$ и ε — произвольные действительные числа. Тогда вероятность жизни системы \mathfrak{A}_{s-1} имеет при больших значениях $s \rightarrow \infty$ следующие оценки:

$$P_v \approx \exp(-Cs^{-\varepsilon}) \begin{cases} = \exp(-Cs^{\varepsilon|}) & \text{при } \varepsilon \begin{cases} < 0 \\ > 0 \end{cases} \\ > 1 - Cs^{-\varepsilon} & \end{cases} \quad (7.40)$$

соответственно. Таким образом, система \mathfrak{A}_{s-1} — жизнеспособна (нежизнеспособна), если с ростом s порядок роста n больше (меньше) величины $s \ln s / \Gamma_s$.

Доказательство. Положим в (7.38)

$$\exp(-n\Gamma_s/s) = Cs^{-(1+\varepsilon)},$$

где $C > 0$ и ε — произвольные действительные числа. Отсюда

$$n = \frac{1 + \varepsilon}{\Gamma_s} s \ln s - \frac{\ln C}{\Gamma_s} s,$$

и

$$P_v \approx [1 - Cs^{-(1+\varepsilon)}]_{s \rightarrow \infty}^s \approx \exp(-Cs^{-\varepsilon}).$$

Последний предел в зависимости от знака величины имеет две различные оценки, указанные в соотношении (7.40), что и требовалось доказать.

Из соотношений (7.37), (7.39) и (7.40) следует, что для жизнеспособности системы \mathfrak{A}_{s-1} необходимо выполнение неравенства

$$n_s > \frac{\ln s}{k_s} \quad (\sigma = \overline{1, s}). \quad (7.41)$$

Таким образом, для жизнеспособности системы \mathfrak{A}_{s-1} , состоящей из n элементов, предельно мелкое ее дробление на s подсистем должно быть таким, чтобы последние имели не менее чем логарифмический порядок s числа своих элементов [см. соотношение (7.41)].

§ 7.4. УЧЕТ ОГРАНИЧЕННОСТИ БЫСТРОДЕЙСТВИЯ СИСТЕМЫ

7.4.1. Выражение параметра $f_{\bar{r}}$ через быстродействие системы A . Пусть перед системой A , имеющей быстродействие V операций/сек^{*}, поставлена цель распознавания неизвестного фактора a . В общем случае из-за наличия шумов системе A доступно лишь стохастическое следствие x фактора a (причины). Другими словами, задана условная вероятность $p_a(x)$ (плотность, если x непрерывно) следствия x при фиксации причины a . Известно статистическое решение поставленной задачи методами оптимальной оценки параметров [32], когда $x = (x_1, \dots, x_n)$ является выборкой объема n , а a , вообще говоря, — k -мерный действительный параметр. Это решение основывается на методе максимума правдоподобия и ограничивается указанием оптимального способа обработки выборки x для решения задачи.

Ниже приводится кибернетическое решение той же задачи, учитывающее возможность использования системы A с большим, но ограниченным быстродействием V . Предлагаемое решение целесообразно использовать, например, в тех случаях, когда ограниченность объема выборки n не позволяет получить требуемую точность оценки параметра (широки доверительные интервалы). По сути дела далее в этом параграфе приводится решение задачи Эшби [5] относительно рационального перебора большого числа вариантов. Существенным оказывается сочетание такого перебора со статистическим анализом.

^{*} Операции системы в общем случае могут занимать несколько тактов ее работы.

Главные затруднения в решении поставленных задач связаны с многомерностью параметра $a = (a_1, \dots, a_i, \dots, \dots, a_k)$. В самом деле, пусть заданы практически достоверные конечные диапазоны изменения компонент $a(a_i \leq a_i \leq \bar{a}_i)$, найденные, например, упоминавшимися методами оценки параметров. Пусть из-за ограниченной разрешающей способности системы нельзя различать значений a_i в интервале длины Δa_i . Тогда каждая компонента a_i определяется дискретным набором $N_i = (\bar{a}_i - a_i) / \Delta a_i$ своих значений, а вектор a определяется

$$N = \prod_{i=1}^k N_i \text{ своими значениями.}$$

Очевидно, что с ростом k величина $N \geq [\min_{1 \leq i \leq k} N_i]^k$ экспоненциально возрастает и это существенно затрудняет упоминающийся выше перебор (так называемые *экспоненциальные трудности*). Основное значение для последнего изложения имеет статистическое различительное свойство. Для его формулировки зафиксируем i первых и $k-i$ последних компонент $a^{(i)} = (a_1, \dots, a_i)$ и $a^{(i)} = (a_{i+1}, \dots, a_k)$ вектора a соответственно, а также априорную вероятность $p(a)$ [по ней легко найти вероятности $p(a_{(i)})$ и $p(a^{(i)})$].

Пусть заданы вероятности:

$$p_0(x) = p(x) = \sum_a p_a(x) p(a),$$

$$p_i(x) = p_{a_{(i)}}(x) = \sum_{a^{(i)}} p_a(x) \frac{p(a)}{p(a^{(i)})} \quad (i = \overline{1, k-1}), \quad (7.42)$$

$$p_k(x) = p_a(x).$$

Легко показать, что если $a_{(i)} = a^*_{(i)}$ является истинным значением параметра $a_{(i)}$, то функция правдоподобия имеет вид:

$$\begin{aligned} \mathcal{P}(x/a_{(i)}) &= \\ &= \begin{cases} p_i(x), & \text{если } a_{(i)} = a^*_{(i)}, \\ \frac{p_0(x) - p(a^*_{(i)}) p_i(x)}{1 - p(a^*_{(i)})}, & \text{если } a_{(i)} \neq a^*_{(i)} \quad (i = \overline{1, k}). \end{cases} \end{aligned} \quad (7.43)$$

В соответствии с соотношением (7.43) и известными результатами математической статистики [32] оптимальный выбор между двумя гипотезами $H_1^{(i)}(a_{(i)} = a^*_{(i)})$ и $H_0^{(i)}(a_{(i)} \neq a^*_{(i)})$ можно основывать на любой монотонной функции отношения правдоподобия

$$L_i = [1 - p(a^*_{(i)})] / [(L'_i)^{-1} - p(a^*_{(i)})],$$

где $L'_i = p_i(x) / p_0(x)$, в частности на $\ln L'_i$.

Те же соотношения имеют место, если вместо полной выборки $x = (x_1, \dots, x_n)$ объема n рассмотреть любую ее часть $x_{(i)}$. Ее объем удобно обозначить n_{i-1} ($i = \overline{1, k}$). При этом будем полагать $x_{(0)} = x$ и $n_0 = n$.

При заданном пороге X_i принимается гипотеза $H_1^{(i)}$ или $H_0^{(i)}$ в зависимости от неравенств $L'_i \geq X_i n_{i-1}$ или $L'_i < X_i n_{i-1}$ соответственно. Вероятность первого неравенства, если имеет место гипотеза $H_0^{(i)}$, обозначим α_i (вероятность ошибки 1-го рода). Вероятность второго неравенства, если имеет место гипотеза $H_1^{(i)}$, обозначим β_i (вероятность ошибки 2-го рода).

При больших значениях объемов выборки n_{i-1} имеют место следующие асимптотические соотношения [14, 15, 32, 33]:

$$\alpha_i \approx \exp(-k_i^0 n_{i-1}), \quad \beta_i \approx \exp(-k_i^1 n_{i-1}), \quad (7.44)$$

где $k_i^0 = k_i^0(X_i)$ и $k_i^1 = k_i^1(X_i)$ — функции, вид которых зависит от соответствующих распределений (7.43). В простейшем биномиальном случае этих распределений указанные функции имеют вид

$$k_i^0 = k(\theta_i, p_0) \quad \text{и} \quad k_i^1 = k(\theta_i, p_1),$$

где θ_i является порогом хэмминговского расстояния $\theta_i = d_i/n$, а p_0 и p_1 — соответствующие параметры биномиальных распределений (см. гл. 3). Широко используемая в предыдущих главах функция $k(\theta, p)$ имеет вид

$$k(\theta, p) = \theta \ln \frac{\theta}{p} + (1 - \theta) \ln \frac{1 - \theta}{1 - p}.$$

Перейдем теперь к описанию процедуры кибернетической оценки параметра a . Она состоит в одновременном отборе

компонент истинного варианта параметра $a = a^*$. Это достигается одновременной работой k последовательных подсистем A_i ($i = \overline{1, k}$) выбора между двумя гипотезами $H_1^{(i)}$ и $H_0^{(i)}$ ($i = \overline{1, k}$) (рис. 7.4). Каждая из них осуществляет N_i таких выборов с вероятностями ошибок α_i и β_i по выборке объема n_{i-1} , основываясь в каждом выборе на одном из N_i отношений правдоподобия L'_i . Так как

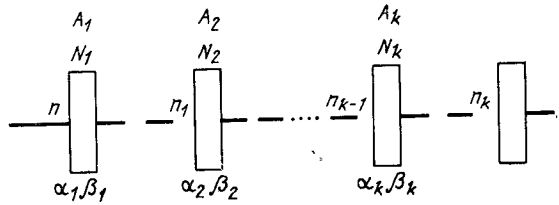


Рис. 7.4. Схема одновременной работы k подсистем перебора вариантов со статистическим отбором (кибернетическая оценка параметра).

усреднение (7.42) может привести лишь к «сближению» гипотез, то имеем

$$k_1^1 \leq k_2^1 \leq \dots \leq k_i^1 \leq \dots \leq k_k^1. \quad (7.45)$$

Кроме того, ясно, что возможен следующий выбор объемов выборок:

$$n = n_0 \geq n_1 \geq n_2 \geq \dots \geq n_i \geq \dots \geq n_{k-1}. \quad (7.46)$$

Одновременная работа всех k подсистем отбора достигается согласованием потоков вариантов значений параметров, проходящих через них. Так что между поступлением вариантов с A_{i-1} -й подсистемы на A_i -ю последняя успевает перебрать все N_i вариантов компоненты a_i . С учетом визуального быстрействия $v \ll V$ отбора с просмотром выборки объема $n_k \leq n_{k-1}$ в конце процедуры и заданного времени T на осуществление всей процедуры указанное согласование в среднем обеспечивается соотношением

$$\begin{aligned} VT &= nN_1 = \alpha_1 n_1 N_1 N_2 = \alpha_1 \alpha_2 n_2 N_1 N_2 N_3 = \\ &= \dots = \alpha_1 \dots \alpha_{k-1} n_{k-1} N_1 \dots N_k = \alpha_1 \dots \alpha_k N_1 \dots N_k \frac{V}{v}, \end{aligned} \quad (7.47)$$

откуда

$$\begin{aligned} n &= \frac{V}{N_1} T, \quad \alpha_1 = \frac{1}{N_2} \frac{n}{n_1}, \quad \alpha_2 = \frac{1}{N_3} \frac{n_1}{n_2}, \dots, \\ \alpha_{k-1} &= \frac{1}{N_k} \frac{n_{k-2}}{n_{k-1}}, \quad \alpha_k = \frac{v}{V} \frac{n_{k-1}}{n_k}. \end{aligned} \quad (7.48)$$

Таким образом, вероятности ошибок 1-го рода здесь используются лишь для согласования потоков вариантов. Заметим, что соотношения (7.47) согласуют лишь средние числа вариантов. Однако постоянная задержка и накопление после каждого из устройств анализа достаточно большого числа возникающих через случайные промежутки времени вариантов с последующей передачей их на следующее устройство через постоянные интервалы (равные средним промежуткам времени между поступлениями вариантов) позволяет «стабилизировать» случайную картину на основе центральной предельной теоремы. Подробные расчеты в этом направлении были проведены в работе [34].

Оценим теперь вероятность $P_{\bar{R}}(T)$ осуществления процедуры кибернетической оценки системой A за время T . В рассмотренном варианте успешное ее окончание эквивалентно прохождению истинного варианта a^* значения параметра через k последовательных подсистем. Вероятность прохождения истинных значений $a^{*(i)}$ через i -ю подсистему A_i равна $1 - \beta_i$. Ввиду того что соответствующие решения выносятся по частично общим выборочным значениям при различных i (т. е. испытания зависимы), то здесь для оценки $P_{\bar{R}}(T)$ снизу воспользуемся неравенством Буля (см. приложение 2):

$$P_{\bar{R}}(T) \geq \sum_{i=1}^k (1 - \beta_i) - (k - 1) = 1 - \sum_{i=1}^k \beta_i. \quad (7.49)$$

Пусть k не растет экспоненциально с ростом n_k . Тогда из (7.44) и (7.49) имеем оценку

$$P_{\bar{R}}(T) \geq 1 - k \max_{1 \leq i \leq k} \beta_i \approx 1 - \exp(-\min_{1 \leq i \leq k} k_i n_{i-1}). \quad (7.50)$$

Из (7.50) следует, что объемы n_i ($i = \overline{1, k}$) промежуточных выборок (параметр $n = n_0$ фиксирован) нужно выбрать так, чтобы

$$k_1^1 n = k_2^1 n_1 = \dots = k_k^1 n_{k-1} = \text{const}. \quad (7.51)$$

Соотношениям (7.47) и (7.51) всегда можно удовлетворить, так как всего имеется $2k$ уравнений с $2k$ неизвестными $n_i, X_i (i = \overline{1, k})$ величины k^1 являются функциями порогов $X_i (i = \overline{1, k})$. Из соотношений (7.48) и (7.50) следует оценка

$$P_{\bar{R}}(T) \geq 1 - \exp(-k^1 n) = 1 - \exp\left(-k^1 \frac{V}{N_1} T\right). \quad (7.52)$$

Можно показать, что последовательный во времени анализ компонент a приводит к аналогичной вероятности $\tilde{P}_{\bar{R}}(T)$, имеющей оценку

$$\tilde{P}_{\bar{R}}(T) \geq 1 - \exp\left(-k_n^1 \frac{V}{N} T\right). \quad (7.53)$$

Если использовать оценки (7.52) и (7.53), когда их правые части близки к единице (тогда они более точны, см. приложение 2), то получим условия целесообразности использования одновременного во времени анализа компонент в виде

$$\frac{k_1^1}{N_1} > \frac{k_n^1}{N}. \quad (7.54)$$

Соотношение (7.54), как правило, имеет место, так как несмотря на то, что $k_1^1 \leq k_n^1$, всегда $N \gg N_1$.

Итак, величина параметра $f_{\bar{R}}$ внешней эффективности в задаче кибернетической оценки параметра a , принимающего N значений, с помощью системы с быстрым действием V может достигать величины

$$f_{\bar{R}} = \frac{k_1^1 V}{N_1}. \quad (7.55)$$

7.4.2. Параллельное декодирование. Изложенные здесь идеи с некоторыми модификациями легли в основу метода параллельного декодирования [35]. Этот метод был разработан в связи с необходимостью преодоления упоминавшихся выше экспоненциальных трудностей, возникающих при попытке осуществления оптимального декодирования Шеннона (см. гл. 3). В самом деле, проводя декодирование, например, по минимуму хэмминговского расстояния $d = d(x, y)$, необходимо провести 2^{iR} сравнений, число которых с ростом длины

слов t экспоненциально возрастает. Впервые здесь экспоненциальные трудности, связанные со средним числом вычислений, были преодолены Розенкратом в предложенной им процедуре последовательного декодирования [36]. Однако ни его процедура, ни последующие ее модификации не привели к преодолению экспоненциального роста с длиной слов t объема необходимой буферной памяти. Кроме того, лишь при $R < R_{\text{выч}} < C$ средний объем вычислений не имеет экспоненциального роста. Эти обстоятельства являются органическим дефектом использованных здесь методов кодирования.

Основанное на принципиально других методах кодирования, параллельное декодирование не преодолевает экспоненциальные трудности ни по среднему числу вычислений, ни по объему памяти. Однако оно для всего допустимого диапазона значений $0 < R < C$ приводит к экспоненциальным зависимостям указанных величин от t с коэффициентами при t , наименьшими по сравнению с аналогичными коэффициентами у других процедур, в том числе и последовательного декодирования. Конкретные значения средних объемов вычислений и памяти для реализации параллельного декодирования в ряде случаев оказываются практически приемлемыми.

§ 7.5. УЧЕТ ОГРАНИЧЕННОСТИ ОБЪЕМА ПАМЯТИ СИСТЕМЫ

7.5.1. Выражение параметра f_R через объем памяти K системы A . При анализе внешнего (\bar{A}^*) и внутреннего (\bar{A}^{**}) поведения системы A возникает следующая задача. Имеется N элементов (например, N элементов «разрешения» среды B или N частей самой системы A), являющихся одновременно действующими источниками статистических данных. Выдвигаются некоторые предположения о том или ином вероятностном характере этих элементов (например, шумовом или сигнальном). Требуется по статистическим данным в некотором смысле (см. далее) оптимальным образом распознать характер анализируемых элементов, имея ограниченный объем памяти системы, содержащий $K = \text{const}$ ячеек (память объема K).

Перейдем к более точной постановке задачи. Пусть каждый i -й элемент характеризуется случайной функцией $\xi_i(t)$ (измеряемой характеристикой), зависящей

от дискретных моментов времени $t=0, h, 2h, \dots$, взятых через интервал времени h , соответствующий такту работы системы A . Если $K \geq N$, то можно подключить каждую случайную функцию $\xi_i(t)$ к определенной ячейке памяти системы A и в ней, накопив достаточно данных, вынести статистическое решение о вероятностном характере каждой $\xi_i(t)$ известными методами математической статистики. Однако, как правило, $K \ll N$ и невозможен непрерывный анализ сразу всех $\xi_i(t)$. Поэтому естествен периодический через такт h просмотр всех N элементов с выделением из них за каждый просмотр части «подозрительных» (подозреваемых в сигнальном характере) для вынесения окончательного статистического решения в соответствующих ячейках памяти. Далее принят простейший критерий подозрительности, состоящий в превышении $\xi_i(t)$ некоторого порога X_i ($\xi_i(t) \geq X_i$), естественный в простейшей постановке статистического различения элементов шумового (гипотеза H_0) и сигнального (гипотеза H_1) характера. Ставится задача оптимального выбора порогов X_i и параметров статистических процедур, приводящих к минимальному среднему времени $E\tau$, идущему на выделение сигнального элемента, и практически исключающих переполнение ограниченного объема K памяти системы A . Последние два требования оказываются противоречивыми, что будет видно из следующих рассуждений.

Пусть плотность вероятности характеристики $\xi_i(t)$ i -го элемента в зависимости от его шумового или сигнального характера имеет вид $f_\varepsilon^{(i)}(x) = f_{a_\varepsilon^{(i)}}^{(i)}(x)$, где $\varepsilon = 0$ в шумовом и $\varepsilon = 1$ в сигнальном случае, а $a_0^{(i)} \neq a_1^{(i)}$ — параметры соответствующих плотностей вероятностей.

Из определения подозрительности i -го элемента следует, что условные вероятности того, что i -й элемент окажется подозрительным, имеют вид

$$p_\varepsilon^{(i)} = \int_{X_i}^{\infty} f_\varepsilon^{(i)}(x) dx \quad (7.56)$$

при условии, что $\varepsilon = 0$ и $\varepsilon = 1$.

Пусть шумовые элементы имеют в момент t индексы $(i_1, \dots, i_{N_0}) = \mathcal{G}_0^{(t)}$, а сигнальные индексы $(i_{N_0+1}, \dots, i_N) = \mathcal{G}_1^{(t)}$.

Тогда, как легко показать, среднее значение случайного числа $\varkappa(t)$ вызовов в момент t равно

$$E \varkappa(t) = \sum_{i \in \mathcal{G}_0^{(t)}} p_0^{(i)} + \sum_{i \in \mathcal{G}_1^{(t)}} p_1^{(i)}. \quad (7.57)$$

Вместе с тем случайное число ρ_i просмотров всех элементов до того, как i -ый элемент окажется подозрительным, имеет распределение

$$\mathcal{P}(\rho_i = r) = (1 - p_1^{(i)})^{r-1} p_1^{(i)} \approx p_1^{(i)} \exp(-p_1^{(i)} r)$$

со средним

$$E \rho_i = \frac{1 - p_1^{(i)}}{p_1^{(i)}} \approx \frac{1}{p_1^{(i)}}. \quad (7.57')$$

Соотношения (7.57) и (7.57') обнаруживают противоположные требования к назначению порогов. В самом деле, увеличением X_i , приводящим к уменьшению $p_\varepsilon^{(i)}$, можно добиться согласно (7.56) столь малого значения $E \varkappa(t) \ll K$, что вероятность переполнения памяти может быть сделана сколь угодно малой. Но при этом согласно (7.57') будет возрастать среднее время (задержка) обнаружения сигнального элемента. В следующем пункте решается задача оптимального выбора порогов. Сейчас мы оценим загрузку памяти и вероятность ее переполнения при произвольных, но фиксированных значениях X_i (или $p_\varepsilon^{(i)}$). Превышение величиной $\xi_i(t)$ порога X_i можно интерпретировать вызовом системы A i -м элементом.

Пусть заданы функции распределения случайных времен v_i обслуживания каждого из вызовов (время вынесения окончательного решения о сигнальном или шумовом характере i -го элемента) $\mathcal{P}(v_i < n) = F^{(i)}(n)$.

Требуется найти распределение случайного числа $\mu(t)$ ячеек памяти системы в момент t .

Это число имеет следующее представление

$$\mu(t) = \sum_{s=1}^t \sum_{i \in R_s} \sum_{j=1}^{\varepsilon_i(s)} \varepsilon_{ij}(s, t), \quad (7.58)$$

где R_s — совокупность индексов элементов, не фиксированных в ячейках памяти системы A в момент s , $\varepsilon_i(s)$ — случайная функция, определенная на индексах R_s , равная нулю, если элемент не подозрителен, или равная

единице в противном случае, $\varepsilon_{ij}(s, t)$ — случайная функция, определенная на подозрительном элементе, возникшем в момент s , равная нулю, если он обслужен (фиксирован в своей ячейке) до момента t , или равная нулю в противном случае, $\varepsilon_{i0}(s, t)$ — случайная функция, тождественно равная нулю.

Упростим соотношение (7.58). В самом деле, число индексов в множестве R_s , как легко видеть, равно $N - \mu(s-1)$ (ясно, что в начальный момент $\mu(1) = 0$ и число индексов в R_1 равно N). Далее рассматривается случай малой вероятности переполнения памяти, для чего необходимо иметь с большой вероятностью соотношение $\mu(t) < K \ll N$. Поэтому вместо (7.58) можно рассматривать соотношение

$$\mu(t) = \sum_{s=1}^t \sum_{i=1}^N \sum_{j=0}^{\varepsilon_i(s)} \varepsilon_{ij}(s, t), \quad (7.59)$$

где под знаком суммы стоят независимые между собой и от $\varepsilon_i(s)$ случайные величины $\varepsilon_{ij}(s, t)$. Как правило, на практике основная масса элементов имеет шумовой характер ($N \approx N_0$). Поэтому будем вести расчет загрузки памяти для случая, когда все элементы являются шумовыми.

Таким образом, бинарные случайные величины $\varepsilon_i(s)$ и $\varepsilon_{ij}(s)$ имеют вид:

$$\varepsilon_i(s) = \begin{cases} 0 & 1 \\ 1 - p_0^{(i)} & p_0^{(i)} \end{cases}, \quad \varepsilon_{i0}(s, t) \equiv 0,$$

$$\varepsilon_{i1}(s, t) = \begin{cases} 0 & 1 \\ F^{(i)}(t-s) & 1 - F^{(i)}(t-s) \end{cases}.$$

С соответствующими производящими (см. приложение 1)

$$\left. \begin{aligned} g_{\varepsilon_i(s)}(u) &= 1 + p_0^{(i)}(u-1), \\ g_{\varepsilon_{i0}(s,t)}(u) &\equiv 1, \\ g_{\varepsilon_{i1}(s,t)}(u) &= 1 + [1 - F^{(i)}(t-s)](u-1). \end{aligned} \right\} \quad (7.59')$$

Найдем производящую функцию $g_{\mu(t)}(u)$ случайной величины $\mu(t)$. Для этого заметим, что производящая функция $g_{si}(u)$ внутренней суммы представления (7.59)

согласно свойству итераций производящих (см. П.1.42) и с учетом соотношений (7.59') имеет вид

$$\begin{aligned} g_{si}(u) &= g_{\varepsilon_i(s)} [g_{\varepsilon_{ij}(s,t)}(u)] = \\ &= (1 - p_0^{(i)}) g_{\varepsilon_{i0}(s,t)}(u) + p_0^{(i)} g_{\varepsilon_{i1}(s,t)}(u) = \\ &= 1 + p_0^{(i)} [1 - F^{(i)}(t-s)](u-1). \end{aligned} \quad (7.60)$$

Используя независимость внутренних сумм представления (7.59) между собой, свойство мультипликативности производящих (см. П.1.42) и выражение (7.60), получим для производящей $g_{\mu(t)}(u)$ выражение, следующее из представления (7.59):

$$g_{\mu(t)}(u) = \prod_{s=1}^t \prod_{i=1}^N \{1 + p_0^{(i)} [1 - F^{(i)}(t-s)](u-1)\}. \quad (7.61)$$

Тогда из (7.61), используя (П.1.56) и (П.1.12), получим

$$E_{\mu}(t) = g'_{\mu(t)}(1) = \sum_{i=1}^N p_0^{(i)} \sum_{s=1}^t [1 - F^{(i)}(t-s)]. \quad (7.62)$$

$$\begin{aligned} D_{\mu}(t) &= g''_{\mu(t)}(1) + g'_{\mu(t)}(1) - [g'_{\mu(t)}(1)]^2 = E_{\mu}(t) - \\ &- \sum_{i=1}^N (p_0^{(i)})^2 \sum_{s=1}^t [1 - F^{(i)}(t-s)] \leq E_{\mu}(t) \end{aligned} \quad (7.63)$$

и

$$g_{\mu(t)}(u) \leq \left[1 + \frac{E_{\mu}(t)}{Nt}(u-1)\right]^{Nt}.$$

Теперь на основании (П.1.24) найдем вероятность $\mathcal{P}[\mu(t) > K]$ переполнения памяти объема K системы A в момент t . Имеем в случае $K > E_{\mu}(t)$:

$$\mathcal{P}[\mu(t) > K] < \exp(-Ntk), \quad (7.64)$$

где

$$\begin{aligned} k &= k \left(\frac{K}{Nt}, \frac{E_{\mu}(t)}{Nt} \right) = \frac{K}{Nt} \ln \frac{K}{E_{\mu}(t)} + \\ &+ \left(\frac{Nt - K}{Nt} \right) \ln \frac{Nt - K}{Nt - E_{\mu}(t)}. \end{aligned} \quad (7.65)$$

При фиксированных значениях K и N , когда $t \rightarrow \infty$, используя соотношение (7.65), получим

$$Ntk = \Psi[K, E_{\mu}(t)] + O(t^{-1}) \approx \Psi[K, E_{\mu}(t)], \quad (7.66)$$

$$\Phi[K, E] = K \ln \frac{E}{K} - [K - E]. \quad (7.67)$$

Подставив выражение (7.66) в (7.64), получим

$$\mathcal{P}[\mu(t) > K] \leq \exp\{-\Phi[K, E_{\mu}(t)]\}. \quad (7.68)$$

Заметим, что с ростом t (при $t \rightarrow \infty$)

$$\lim_{t \rightarrow \infty} \sum_{s=1}^t [1 - F(s)] = \sum_{s=1}^{\infty} [1 - F(s)] = E_0 \nu_2$$

$$= \sum_{s=1}^{\infty} s \Delta F(s) = E_0 \nu_2, \quad (7.69)$$

где $E_0 \nu_2$ — среднее время обнаружения шумового характера шумового элемента.

Поэтому с учетом (7.62) и (7.69) заключаем, что величина $E_{\mu}(t)$ с ростом t , возрастая, стремится к пределу

$$E_{\mu}(t) > \lim_{t \rightarrow \infty} E_{\mu}(t) = \sum_{\nu=1}^N p_{(\nu)}^0 E_0 \nu_2 = A. \quad (7.70)$$

Но функция $\Psi[K, E_{\mu}(t)]$ с ростом $E_{\mu}(t)$ убывает, поэтому усилим оценку (7.68), полагая

$$\mathcal{P}[\mu(t) > K] > \exp\{-\Phi(K, A)\}. \quad (7.71)$$

Вычислим теперь вероятность $P^{*R}(T)$ того, что ни в один момент времени $t = 1, \dots, T$ не произойдет переполнение памяти

$$P^{*R}(T) = \mathcal{P}[\mu(t) \leq K, \quad t = 1, \dots, T].$$

Ввиду того, что случайные величины $\mu(t)$ — зависимость при различных значениях t , здесь приходится использовать неравенство Буля (см. приложение 2). Тогда с учетом соотношения (7.71) будем иметь

$$P^{*R}(T) \geq 1 - \sum_{t=1}^T \mathcal{P}[\mu(t) > K] \geq 1 - T \exp\{-\Phi(K, A)\}.$$

Для значений T , имеющих практический интерес, когда $T \exp\{-\Phi(K, A)\} \gg 1$, последняя оценка имеет вид:

$$P^{*R}(T) \approx \exp(-T/f_{R^*}), \quad (7.72)$$

$$f_{R^*} = \exp\Phi(K, A). \quad (7.73)$$

Пусть переполнение памяти системы A хотя бы в один момент из интервала времени $[1, T]$ приводит к расстройству ее работы на этом интервале, эквивалентному *ее выходу из строя* и не зависит от надежности отказов системы. С другой стороны, пусть в пассивном состоянии система A имеет надежность (безотказность) работы на интервале времени $[1, T]$

$$P^{**R}(T) = \exp(-T/f_{R^{**}}). \quad (7.74)$$

Тогда в соответствии со вторым вариантом правила среднего гармонического [см. соотношение (7.27) и (7.29)] имеем в рассматриваемом случае $l=2$ для надежности $P_R(T)$ системы A в активном состоянии выражение

$$P^R(T) = P^{*R}(T) P^{**R}(T) \approx \exp(-T/f_R), \quad (7.75)$$

где

$$f_R = \frac{f_{R^*} + f_{R^{**}}}{f_{R^*} f_{R^{**}}} = \frac{f_{R^*} + f_{R^{**}}}{1}, \quad (7.76)$$

и величина f_R выражается через объем памяти K системы A соотношениями (7.73) и (7.67).

Важность соотношений (7.55) и только что полученного соотношения (7.76) состоит в том, что с их помощью параметры внешней и внутренней эффективности системы f_R и f_{R^*} ставятся в зависимость от быстроты действия V и объема памяти K системы соответственно. Эти соотношения были получены в работе [9]. В частном случае одинаково распределенных случайных функций $\xi_i(t)$ задача о разгрузке памяти системы была решена в работе [37]. В работе [38] было проведено моделирование самоанализа вычислительной машины загрузки своей памяти.

7.5.2. Оптимальный выбор порогов X_i . Присутствие теперь к постановке задачи оптимального выбора порогов X_i ($i=1, N$) при распознавании характера (шумового или сигнального) N элементов системой A с ограниченным объемом памяти K .

Прежде всего заметим, что при больших значениях A и K можно получить простое асимптотическое выражение

ние функции $\Psi(K, \Lambda)$, определяемой соотношением (7.67). В самом деле, полагая $K = \Lambda + u\sqrt{\Lambda}$, где u — некоторая положительная константа, а $\Lambda \rightarrow \infty$, беря два первые члена в разложении логарифма выражения (7.67) по степеням величины $u/\sqrt{\Lambda}$ получим

$$\Psi(\Lambda + u\sqrt{\Lambda}, \Lambda) = \frac{u^2}{2} + O\left(\frac{1}{\sqrt{\Lambda}}\right). \quad (7.77)$$

Задаваясь некоторой близкой к единице величиной P нижней оценки вероятности $P^*_{R}(T)$ [см. соотношение (7.72)], получим с учетом (7.77) необходимое значение

$$u = u_p = \sqrt{2 \ln [T/(1-P)]}. \quad (7.78)$$

Таким образом, полагая $u = u_p$ и $K = \Lambda + u_p\sqrt{\Lambda}$ при больших значениях Λ и K , можно гарантировать непереполнение памяти с вероятностью $P^*_{R}(T) > P$. Но при этом $K \approx \Lambda$ и с учетом (7.70) приходим к основному ограничению, накладываемому на величины $p_i^{(0)}$ и $E_0 v_i$

$$\sum_{i=1}^N p_0^{(i)} E_0 v_i = K. \quad (7.79)$$

Пусть среди N элементов имеется лишь один сигнальный элемент и \mathcal{P}_i — априорная вероятность того, что это i -й элемент. Тогда с учетом соотношения (7.57') среднее время $E\tau$ обнаружения сигнального элемента (средняя задержка) имеет вид

$$E\tau \approx \sum_{i=1}^N \mathcal{P}_i \left(\frac{1}{p_1^{(i)}} + E_1 v_i \right), \quad (7.80)$$

где величина $p_1^{(i)}$ определяется соотношением (7.56), а $E_1 v_i$ — среднее время обнаружения сигнального характера i -го сигнального элемента.

Оптимальная задача выбора порогов X_i ставится так. Найти такие оптимальные значения $X_i = X_i^{(0)}$ ($i = \overline{1, N}$), для которых величина $E\tau$, определяемая соотношением (7.80), обращается в минимум при условии (7.79). Так как величины \mathcal{P}_i , $E_0 v_i$ и $E_1 v_i$ не зависят от порогов X_i , а от них зависят лишь величины $p_0^{(i)}$ и $p_1^{(i)}$ согласно соотношениям

(7.56), то достаточно найти минимум функции

$$f = f(X_1, \dots, X_N) = \sum_{i=1}^N \mathcal{P}_i / p_1^{(i)}(X_i) \quad (7.81)$$

при условии

$$F = F(X_1, \dots, X_N) = \sum_{i=1}^N E_0 v_i p_0^{(i)}(X_i) = K. \quad (7.82)$$

Используя метод множителей Лагранжа, будем искать безусловный экстремум функции $Q = f + \lambda(F - K)$.

С учетом соотношений (7.56), (7.81) и (7.82) получим N уравнений для экстремальных значений

$$\frac{\partial Q}{\partial X_i} = \frac{\mathcal{P}_i}{[p_1^{(i)}(X_i)]^2} f_1^{(i)}(X_i) - \lambda f_0^{(i)}(X_i) E_0 v_i = 0 \quad (i = \overline{1, N})$$

или

$$\frac{f_0^{(i)}(X_i) [p_1^{(i)}(X_i)]^2}{f_1^{(i)}(X_i)} = \frac{\mathcal{P}_i}{\lambda E_0 v_i} \quad (i = \overline{1, N}), \quad (7.83)$$

где неопределенный множитель определяется с помощью соотношения (7.82).

Асимптотическое решение уравнений (7.83) можно найти для случая больших значений $X_i \rightarrow \infty$. Этот случай представляет основной практический интерес, так как обычно $K \ll N$ и здесь [см. соотношение (7.79)] необходимы малые значения $p_0^{(i)}(X_i)$ (большие значения X_i).

Кроме того, рассмотрим широкий класс плотностей вероятностей $\{f_a(x)\}$ вида

$$f_a(x) = \exp[-\omega_a(x)],$$

где функция $\omega_a(x) \rightarrow \infty$, причем

$$\frac{\omega'_a(x)}{\omega_a(x)} \xrightarrow{x \rightarrow \infty} 0.$$

Например, функция $\omega_a(x) = c(a)x^\alpha$, где $c(a)$ — функция параметра a , а α — константа, обладает указанными свойствами.

Тогда интегрированием по частям легко доказать для плотностей вероятностей $f_a(x)$ класса $\{f_a(x)\}$ асимп-

отическое соотношение

$$p(x) = \int_x^\infty f_a(\alpha) d\alpha \rightarrow f_a'(x). \quad (7.84)$$

Если плотности вероятностей, фигурирующие в уравнении (7.83), из класса $\{f_a(x)\}$, то асимптотически при $X_i \rightarrow \infty$ уравнение (7.83) примет вид

$$p_0^{(i)}(X_i) p_1^{(i)}(X_i) = \frac{\mathcal{P}_i}{\lambda E_0 v_i} \quad (i = \overline{1, N}). \quad (7.85)$$

Обыкновенно при $X_i \rightarrow \infty$ $p_0^{(i)}(X_i) < p_1^{(i)}(X_i)$ (распределение „сигнала+шум“ смещено вправо по отношению к распределению одного „шума“).

Поэтому из (7.85) имеем

$$[p_0^{(i)}(X_i)]^2 \leq \frac{\mathcal{P}_i}{\lambda E_0 v_i} \leq [p_1^{(i)}(X_i)]^2 \quad (i = \overline{1, N}).$$

или

$$p_0^{(i)}(X_i) \leq \sqrt{\frac{\mathcal{P}_i}{\lambda E_0 v_i}} \leq p_1^{(i)}(X_i) \quad (i = \overline{1, N}). \quad (7.86)$$

Будем рассматривать случай близких гипотез, когда $a_0^{(i)} - a_1^{(i)} \rightarrow 0$. Этот случай имеет основной практический интерес. Тогда $p_0^{(i)}(X_i) \approx p_1^{(i)}(X_i)$ и соотношение (7.86) примет вид:

$$p_0^{(i)}(X_i) \approx \sqrt{\frac{\mathcal{P}_i}{\lambda E_0 v_i}}. \quad (7.87)$$

Подставив выражение (7.87) в (7.79), получим

$$\lambda = \left(\frac{1}{K} \sum_{i=1}^N \sqrt{\mathcal{P}_i E_0 v_i} \right)^2. \quad (7.88)$$

Подставив выражение (7.88) в (7.87), получим экстремальное значение

$$p_0^{(i)} = p_0^{(i)}(X_i^{(0)}) = \frac{K \sqrt{\mathcal{P}_i E_0 v_i}}{\sum_{i=1}^N \sqrt{\mathcal{P}_i E_0 v_i}} \quad (i = \overline{1, N}), \quad (7.89)$$

откуда можно получить экстремальное значение $X_i^{(0)}$.

Подставив выражение (7.89) в (7.81), получим

$$f_0 = \sum_{i=1}^N \mathcal{P}_i / p_1^{(i)} \lesssim \frac{1}{K} \left(\sum_{i=1}^N \sqrt{\mathcal{P}_i E_0 v_i} \right)^2, \quad (7.90)$$

где экстремальная величина f_0 , оцениваемая сверху правой частью соотношения (7.90) из-за $p_0^{(i)} \lesssim p_1^{(i)}$ оказывается минимумом.

Таким образом, используя (7.80) и (7.90), получим для минимальной задержки обнаружения сигнального элемента

$$\min E\tau \lesssim \frac{1}{K} \left(\sum_{i=1}^N \sqrt{\mathcal{P}_i E_0 v_i} \right)^2 + \overline{E_1 v} \leq \frac{N}{K} \overline{E_0 v} + \overline{E_1 v}, \quad (7.91)$$

где

$$\overline{E_0 v} = \frac{1}{N} \sum_{i=1}^N E_0 v_i \quad \text{и} \quad \overline{E_1 v} = \sum_{i=1}^N \mathcal{P}_i E_1 v_i,$$

и последняя оценка следует из неравенства Гельдера (см. приложение 1). В частном однородном случае одинаково распределенных случайных функций $\xi_i(t)$ и случайных величин v_i соотношения (7.79), (7.89) и (7.91) упрощаются

$$\left. \begin{aligned} N p_0(X^{(0)}) E_0 v &= K \\ p_0(X^{(0)}) &= K / N E_0 v \end{aligned} \right\} \quad (7.92)$$

и

$$\min E\tau = \frac{N}{K} E_0 v + E_1 v. \quad (7.93)$$

Соотношения (7.92) и (7.93) выведены в стационарном случае ($t \rightarrow \infty$). Аналогичные соотношения в нестационарном случае ($t = \text{const}$) позволили проанализировать загрузку памяти системы A при различении шумовых и сигнальных элементов [37, 38]. Соотношения (7.92) и (7.93) были использованы для решения конкретных радиолокационных задач в работе [39].

7.5.3. Оптимальная иерархическая структура самоконтроля системы A с учетом ограниченности объема памяти. В гл. 2 была рассмотрена иерархическая структура поиска неисправностей (самоконтроля системы A) без учета возможностей использования памяти системы. Однако большие объемы памяти и быстродействие современных электронных систем могут быть с успехом использованы в целях их самоконтроля. Очевидно, что аппаратура, выделяемая под самоконтроль, должна

быть менее сложной и соответственно более надежной, чем сама система. Далее будем считать эту аппаратуру абсолютно надежной. Независимо и параллельно с функциональной иерархической структурой системы введем надежностную иерархическую структуру системы. Основным ее принципом является выделение таких максимально крупных подсистем (модулей), разладка каждой из которых приводит к разладке системы в целом. Вводится естественное допущение о невозможности (из-за маловероятности) одновременной разладки более одного первичного элемента и независимость таких разладок в различные моменты времени.

Рассматриваются два режима функционирования системы и соответственно два определения отказа системы в каждом режиме. В режиме работы отказ системы определяется как неуспевание отыскать разлаживающийся первичный элемент до вызова этого элемента «на работу». В режиме готовности (активной консервации) отказ системы определяется, как неуспевание отыскать разлаживающийся первичный элемент до полного выхода его из строя. Здесь первопричиной разладки элементов системы считается разладка составляющих их первичных элементов, а не нарушение структуры соединения последних на более высоких уровнях (соединения зачисляются в первичные элементы).

Непрерывный контроль за всеми первичными элементами системы практически невозможен и не нужен, так как последние выходят из строя сравнительно редко. Поэтому естественно сочетать непрерывный контроль за n_1 модулями (элементами высшей степени иерархии), используя часть памяти системы объема K , с эпизодическим поисковым внутримодульным контролем разлаживающихся первичных элементов с помощью мобильной поисковой аппаратуры. (рис. 7.5).

В соответствии со сказанным, общий алгоритм самоконтроля \bar{A}^{**} (внутреннее поведение системы A) разбивается на три подалгоритма $\bar{A}^{**} = (S_1, S_2, S_3)$ (рис. 7.6).

Подалгоритм S_1 состоит в непрерывном контроле за модулями и отборе части «подозрительных» из них за каждый просмотр (за один такт работы системы) для жесткой коммутации со свободными ячейками памяти системы, выделенными под самоконтроль. Это предполагает наличие индикаторов $\xi_i(t)$ ($i=1, n_1$) качества работы модулей.

Подалгоритм S_2 состоит в накоплении в последующих тактах в ячейках памяти показаний $\xi_i(t)$ от закомутированных с ними модулей для вынесения окончательного статистического решения об их исправности или приближении разладки. После вынесения решения ячей-

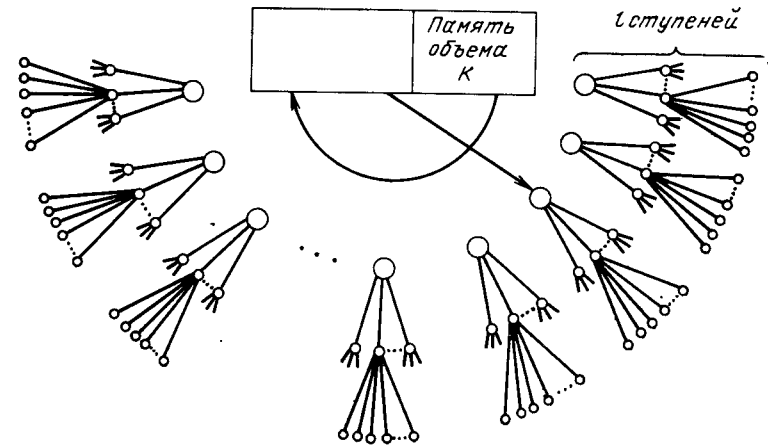


Рис. 7.5. l -ступенная иерархическая структура поиска неисправностей с использованием памяти объема K .

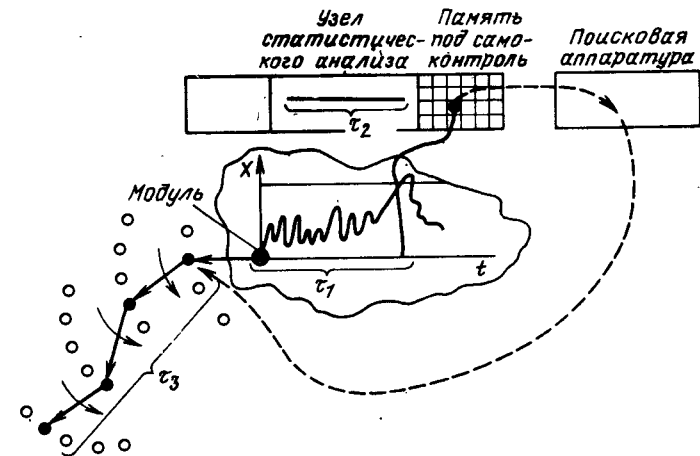


Рис. 7.6. Расчленение общего алгоритма самоконтроля на три подалгоритма.

ка памяти очищается и может быть закомутирована с другими модулями.

В случае решения о надвигающейся разладке модуля подается сигнал мобильной поисковой аппаратуре, которая согласно подалгоритму S_3 проводит поиск в нем разлаживающегося первичного элемента. Дальнейшая судьба последнего не уточняется. В зависимости от конкретных условий он может быть заменен исправным элементом или исправлен и время, идущее на это, не учитывается.

Алгоритм \bar{A}^{**} в целом основное время тратит на устранение «ложных тревог» о «подозрительности» нормально функционирующих модулей. Оптимизация алгоритма \bar{A}^{**} проводится максимизацией вероятности Q успеть осуществить поиск разлаживающегося первичного элемента до его вызова (в режиме работы) или до его полной разладки (в режиме готовности). Далее показано, что при редких вызовах (или медленной разладке) указанная оптимизация эквивалентна минимизации среднего времени до обнаружения разлаживающегося первичного элемента.

В соответствии с расчленением алгоритма $\bar{A}^{**} = (S_1, S_2, S_3)$ на три подалгоритма случайное время τ обнаружения надвигающейся разладки первичного элемента распадается на сумму

$$\tau = \tau(\bar{A}^{**}) = \tau_1(S_1) + \tau_2(S_2) + \tau_3(S_3) \quad (7.94)$$

случайных времен, идущих на осуществление соответствующих подалгоритмов (можно показать, что, вообще говоря, первые два слагаемых зависимы, а третье не зависит от них).

Из соотношения (7.94) следуют соотношения для производящих функций и средних [см. (П.1.42)]

$$g_\tau(u) = g_{\tau_1 + \tau_2}(u) g_{\tau_3}(u) \quad (7.95)$$

и

$$E\tau = E\tau_1 + E\tau_2 + E\tau_3. \quad (7.96)$$

Так как $g'_\tau(1) = E\tau$, то из (7.95) имеем для $u \approx 1$ и $g''_\tau(1) < \infty$

$$\begin{aligned} g_\tau(u) &= 1 - (1-u)E\tau + O((1-u)^2) = \\ &= \exp[-(1-u)E\tau + O((1-u)^2)]. \end{aligned} \quad (7.97)$$

Напомним (см. гл. 1), что имеет место p -напряженный поиск неисправных первичных элементов, если в режиме работы любой первичный элемент независимо от других в каждый дискретный момент может быть вызван на работу с вероятностью p .

В случае режима готовности вероятность p можно интерпретировать как вероятность для любого разлаживающегося первичного элемента выйти из строя в любой дискретный момент.

Тогда согласно лемме П.3 приложения 3 и соотношению (7.97) вероятность Q успеть обнаружить разлаживающийся первичный элемент до вызова (до выхода из строя) имеет вид:

$$Q = g_\tau(q) = 1 - pE\tau + O(p^2), \quad (7.98)$$

где $q = 1 - p$ и $pE\tau < 1$.

Для вычисления надежности системы задается случайная величина κ , характеризующая интенсивность вызовов (в режиме работы) или интенсивность выхода из строя (в режиме готовности) ее первичных элементов. Случайную величину κ определим как число вызовов за время T . Для случая независимых вызовов с вероятностью p имеем $E\kappa = pT$ и из соотношения (2.69) будем иметь при $T \gg E\tau \approx (1-Q)p$ согласно (7.74) выражение для надежности

$$P^{**}_R(T) \approx \exp(-T/f_{R^{**}}) \approx \exp(-p^2 E\tau T), \quad (7.99)$$

где

$$f_{R^{**}}^{-1} \approx p(1-Q) \approx p^2 E\tau. \quad (7.100)$$

Из соотношений (7.76) и (7.100) следует выражение для коэффициента внутренней эффективности f_R в форме

$$f_R \approx \frac{1}{p(1-Q) + f_{R^{**}}^{-1}} \approx \frac{1}{p^2 E\tau + f_{R^{**}}^{-1}}, \quad (7.101)$$

где параметр $f_{R^{**}}$, зависящий от объема памяти K , определен соотношением (7.73).

Для обращения в максимум величины f_R необходимо найти алгоритм $\bar{A}^{**} = (S_1, S_2, S_3)$, обращающий в максимум величину Q или, что то же, в минимум величину $E\tau$ (считается, что параметры p и $f_{R^{**}}$ зафиксированы).

7.5.3. Оптимизация алгоритма самоконтроля. Примеры. Будем вести максимизацию величины f_R , обращая в минимум величину $E\tau = E\tau_1 + E\tau_2 + E\tau_3$. Заметим, что минимизация первых двух слагаемых вариацией порогов была осуществлена в п. 7.5.2. Известные оптимальные процедуры выбора между гипотезами [32] и теории поисков (см. приложение 3) обеспечивают минимальность второго и третьего слагаемых соответственно. Ниже дальнейшая минимизация проводится вариацией иерархической структуры самоконтроля. Рассмотрим однородный случай l -ступенной иерархической структуры с постоянными числами n_r подэлементов $r+1$ -й ступени в элементе r -й ступени. Тогда общее фиксированное число элементов n системы имеет вид:

$$n = \prod_{r=1}^l n_r. \quad (7.102)$$

Рассмотрим простейший случай безошибочного поиска равновероятно неисправных элементов r -й ступени, требующих среднее время t_r ($r = \overline{2, l}$) для обнаружения неисправности. Тогда среднее время поиска $E\tau_r$ на r -й ступени имеет вид [см. приложение 3 (П.3.22)]

$$\begin{aligned} E\tau_r &= \frac{1}{n_r} \left[\sum_{i=1}^{n_r-2} i + 2(n_r - 1) \right] t_r = \\ &= \left(\frac{n_r + 1}{2} - \frac{1}{n_r} \right) t_r \approx \frac{n_r}{2} t_r, \end{aligned} \quad (7.103)$$

причем имеет место точное равенство при $n_r = 2$ и приближенное равенство при $n_r \gg 2$.

Из (7.91), (7.96) и (7.103) при допущении $K \ll n_1$ (откуда $\overline{E_1 v} \ll \frac{n_1}{K} \overline{E_0 v}$) имеем

$$E\tau \approx \frac{n_1}{K} \overline{E_0 v} + \sum_{r=2}^l \frac{t_r}{2} n_r. \quad (7.104)$$

Найдем условный минимум $E\tau$ по переменным $\{n_r\}$ ($r = \overline{1, l}$), связанным условиями (7.102). Введем новые переменные:

$$x_1 = \frac{\overline{E_0 v}}{K} n_1, \quad x_r = \frac{t_r}{2} n_r \quad (r = \overline{2, l}). \quad (7.105)$$

В новых переменных необходимо обратить в минимум величину

$$E\tau = \sum_{r=1}^l x_r \quad (7.106)$$

при условии (7.102), эквивалентном условию

$$\prod_{r=1}^l x_r = n \frac{\overline{E_0 v}}{K} \prod_{r=2}^l \frac{t_r}{2}. \quad (7.107)$$

Как известно (в этом легко убедиться методом множителей Лагранжа), $\min E\tau$ достигается при постоянных значениях

$$x_r = \sqrt[l]{n \frac{\overline{E_0 v}}{K} \prod_{r=2}^l \frac{t_r}{2}} = \frac{g_l}{2} n^{1/l}, \quad (7.108)$$

где

$$g_l = \sqrt[l]{\frac{2\overline{E_0 v}}{K} \prod_{r=2}^l t_r}. \quad (7.109)$$

Из соотношений (7.106) и (7.108) получим минимум

$$(E\tau)_* = \min_{\{n_r\}} E\tau = \frac{g_l}{2} l n^{1/l}, \quad (7.110)$$

который согласно соотношениям (7.105) и (7.108) достигается при

$$n_1 = \frac{K g_l}{2 \overline{E_0 v}} n^{1/l}, \quad n_r = \frac{g_l}{t_r} n^{1/l} \quad (r = \overline{2, l}). \quad (7.111)$$

Рассмотрим случай больших значений l и небольших отклонений величин t_r от постоянного значения t ($r = \overline{2, l}$). Тогда величина g_l тоже близка к этому общему для t_r значению t ($g_l \approx t_r \approx t$).

В этом случае минимум величины $(E\tau)_*$ по l [как легко убедиться дифференцированием по ней выражения (7.110)] достигается при целых значениях l , ближайших к величине $l = \ln n$, растущей с ростом n . При этом из (7.110) имеем минимум

$$(E\tau)_{**} = \min_l (E\tau)_* = \frac{tc}{2} \ln n \approx t \ln n, \quad (7.112)$$

которой согласно (7.111) достигается при $l = \lceil \ln n \rceil$,

$$n_1 \approx \frac{Kl}{2E_0v} e \quad \text{и} \quad n_r \approx e \approx 2 \quad \text{или} \quad 3 \quad (r = \overline{2, l}).$$

(7.113)

Приведенное приближенное решение экстремальной задачи можно дополнить точным решением, приводящим для случая $t_r = 1$ к $n_2 = \dots = n_l = 2$. В самом деле, независимо от того, как проходил поиск на первой ступени, на последующих он должен проходить оптимально, а это достигается, как было показано ранее, двоичной структурой.

Поэтому общее среднее время в нашем случае [см. соотношения (7.103) и (7.104)] имеет вид:

$$E\tau = an_1 + \sum_{r=2}^l \left(\frac{n_r + 1}{2} - \frac{1}{n_r} \right) = an_1 + \\ + \left(\frac{2+1}{2} - \frac{1}{2} \right) (l-1) = an_1 + l - 1,$$

где $a = \overline{E_0v}/K$. Величина l зависит от n_1 и может быть найдена из соотношения $n_1 2^{l-1} = n$. Отсюда

$$l - 1 = \left(\ln \frac{n}{n_1} \right) / \ln 2.$$

Оптимальное значение n_1 найдем из условия обращения

$$E\tau = an_1 + \frac{\ln n - \ln n_1}{\ln 2}$$

в минимум. Дифференцируя $E\tau$ по n_1 , найдем

$$n_1 = \frac{1}{a \ln 2}$$

и

$$l = 1 + \frac{\ln n - \ln (1/a \ln 2)}{\ln 2},$$

где необходимо брать ближайшие целые значения правых частей, определяющих n_1 и l . Последние асимптотически при больших значениях n и a совпадают с ранее приближенно полученными значениями n_1 и l .

Соотношения (7.113) определяют оптимальную иерархическую структуру, которую начиная со второй ступени можно принять бинарной (рис. 7.7).

В отличие от этого рассмотренная в § 2.5 невырожденная иерархическая структура поиска неисправностей оказалась четырехарной. Полученное отличие связано с введением в последнюю модель памяти объема K для накопления данных о «подозрительных» модулях первой ступени иерархической структуры и невозможности пропуска индикатора с сигналом неисправности ($\beta = 0$).

Рассмотрим некоторые примеры.

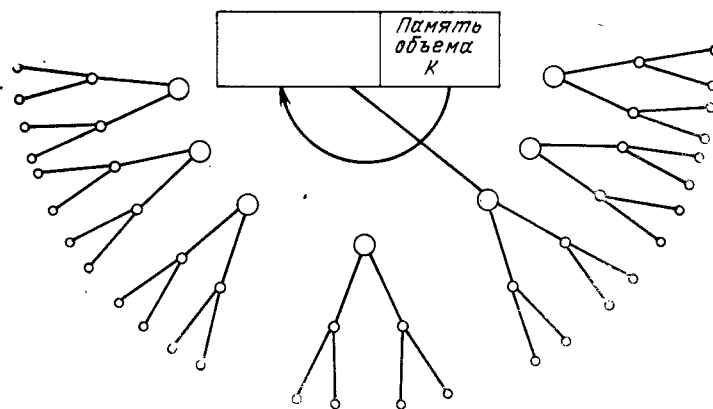


Рис. 7.7. Оптимальная бинарная иерархическая структура поиска неисправностей.

Пример 7.6. Пусть система A имеет $n = 10^{16}$ первичных элементов и отводит $K = 10^6$ ячеек памяти под самоконтроль.

Пусть в среднем требуется 50 испытаний для решения об исправности действительно исправного модуля. На каждое испытание так же как и на проверку на подозрительность, требуется время $t_r = h = 10^{-3}$ сек, $E_0v = 50h$. Вычислить минимальное среднее время задержки $(E\tau)_*$, а также параметры иерархической структуры самоконтроля n_1 и n_r ($r = \overline{2, l}$) при $l = 2, 3, 4$.

Решение. Необходимые вычисления проводятся по формулам (7.110) и (7.111). Их результаты удобно расположить в таблицу:

l	$(E\tau)_*$	n_1	n_r ($r = \overline{2, l}$)
2	10 сек	10^{10}	10^6
3	0,15 сек	10^8	10^4
4	0,02 сек	10^7	10^3

Пример 7.7. Рассмотрим ту же задачу, что и в предыдущем примере, но при других значениях параметров. Пусть $n=10^8$, $K=10^2$,

$$t_r = h, \quad \overline{E_0 v} = 128 h, \quad \text{где } h = 10^{-6} \text{ сек и } l = 2,3.$$

Решение. Результаты вычислений по формулам (7.110) и (7.111) приведены в таблице:

l	$(E_0)_*$	n_1	n_r ($r = \overline{2, l}$)
2	0,016 сек	6250	16 000
3	0,001 сек	250	640

Прокомментируем рассмотренные примеры.

Прежде всего остановимся на исходных данных примера 7.6. В нем они взяты из биологических соображений. Примерное число клеток человека $n=10^{16}$. Число нейронов у него порядка $10^{10} \gg K=10^6$. Величина $h=10^{-5}$ сек времени элементарного акта выбрана исходя из скорости $v=150$ м/сек передачи нервного импульса на расстояние порядка 1 мм.

Рассмотрение результатов расчетов параметров примера 7.6 убедительно показывает целесообразность трех- или четырехступенной иерархической структуры самоконтроля ($l=3,4$), так как двухступенная структура имеет слишком медленную для организма реакцию (10 сек) и требует чрезмерного измельчения первой ступени ($n_1=10^{10}$).

Перейдем к примеру 7.7. В нем исходные данные имеют техническое происхождение и связаны с близкими перспективами развития ЭВМ. Взятые в нем количество первичных элементов $n=10^8$ не слишком превосходит современное значение этой величины $n \approx 10^6$. Объем памяти $K=10^2$, выделенный под самоконтроль, составляет малую долю объемов памяти современных ЭВМ (сейчас они имеют порядок 10^3); параметр $h=10^{-6}$ сек взят как обратная величина достигнутого быстродействия ЭВМ $V=10^6$ опер/сек.

Результаты расчета параметров примера 7.7 убеждают в большей целесообразности использования трехступенной иерархической структуры по сравнению с двухступенной как по скорости реакции (на порядок большей), так и по «соразмерности» структуры различных уровней иерархии.

Из полученных оптимальных соотношений (7.112) и (7.113) можно вывести ограничения на предельные размеры самоконтролирующихся систем A с учетом реальных физических скоростей проходящих в них процессов.

Пусть скорость передачи сигнала о неисправности равна ω_c [см/сек], а скорость перемещения при его поиске равна ω_n [см/сек]. Зададим равномерную пространственную плотность первичных элементов ρ штук/см³ и модулей r штук/см³. Пусть «линейный» размер системы A равен R см, время разладки первичного элемента равно T_0 сек и среднее время задержки на обследование одного модуля равно t сек.

Соотношение (7.112) дает среднее время поиска tl , связанное с задержкой на обследование неисправных модулей, где $l = \lceil \ln n \rceil$ — их число по ходу поиска, а n — число первичных элементов. В нашем случае примерно линейного расположения модулей по пути поиска общий путь следования сигнала о неисправности и поиска равен $S = l/\sqrt[3]{r}$.

Время его прохождения сигналом и поиском равны $l/\omega_c \sqrt[3]{r}$ и $l/\omega_n \sqrt[3]{r}$ соответственно. Таким образом, общее среднее время T от начала разладки первичного элемента до его обнаружения не должно превосходить величину T_0 :

$$T = l/\omega_c \sqrt[3]{r} + l/\omega_n \sqrt[3]{r} + tl = \left(\frac{2}{\sqrt[3]{r}} + tw \right) \frac{l}{\omega} < T_0, \quad (7.114)$$

$$\text{где } \omega = \frac{2}{\omega_c^{-1} + \omega_n^{-1}}. \quad (7.115)$$

Так как в нашем случае число первичных элементов $n=R^3\rho$ и $l = \lceil \ln n \rceil$, то общий вид ограничений для обнаружения разлаживающегося первичного элемента имеет вид

$$\left(\frac{2}{\sqrt[3]{r}} + tw \right) \ln R^3\rho < T_0\omega, \quad (7.116)$$

или

$$R < \frac{1}{\sqrt[3]{\rho}} \exp \left[T_0\omega \left/ \left(\frac{2}{\sqrt[3]{r}} + tw \right) \right. \right]. \quad (7.117)$$

Примем, исходя из биологических соображений (см. пример 7.7), что объем системы не превосходит $R^3 = 1 \text{ м}^3 = 10^6 \text{ см}^3$, плотность ее первичных элементов $\rho = 10^9$ штук/см³ на три порядка большую плотности ее модулей $r = 10^6$ штук/см³ и время задержки на обследование одного модуля

$$t \ll 1/\omega \sqrt[3]{r}.$$

Тогда соотношение (7.116) приводит к ограничению

$$0,7 \text{ сек} \leq T_0\omega, \quad (7.118)$$

выполняемому при медленных разладках (большие T_0) сложных организмов.

Соотношения (7.118) в ЭВМ могут выполняться в основном за счет больших скоростей ω . Если исходить из скорости $\omega = v = 15000$ см/сек, равной скорости передачи нервного импульса, то при тех же параметрах, что и выше, ограничение (7.118) имеет вид

$$0,5 \cdot 10^{-4} \text{ сек} \leq T_0. \quad (7.119)$$

§ 7.6. ПРИМЕНЕНИЯ

7.6.1. Технические применения. Пример. Первые предельные законы теории потенциальной эффективности сложных систем были открыты в результате исследова-

ния предельной эффективности технических систем. Это относится к надежности и помехоустойчивости. Последующие предельные законы, связанные с управляемостью, самоорганизацией и живучестью, имеют не техническое происхождение, но без труда допускают технические интерпретации.

В ходе предыдущего изложения делались некоторые ссылки на технические приложения развитой теории. Однако эти ссылки ни в коей мере не претендовали на полноту. Полное освещение этого вопроса потребовало бы слишком много места. Укажем лишь, что большинство технических приложений касается расчетов оптимальных систем по некоторым отдельным X -сечениям.

Основной целью настоящей главы и всей развитой теории является оценка потенциальной эффективности системы в целом. В связи с этим расчет оптимальных X -сечений можно рассматривать как промежуточный для окончательного расчета оптимальной системы в целом.

Проиллюстрируем такой расчет на примере.

Рассмотрим систему $A=(A_R, A_I, A_C)$, состоящую из трех подсистем (X -сечений), выявляющих три ее X -качества: надежность (R -качество), помехоустойчивость (I -качество) и управляемость (C -качество).

Пусть рассматривается ситуация средней ответвленности (см. пример 7.4), когда пороги осуществимости $(P^{(0)}, T^{(0)})=(1-10^{-6}, 10^5 \text{ сек})$. По последним с помощью формулы (7.16) в указанном примере вычислены значения коэффициентов внутренней и внешней эффективности:

$$f_R=10^{11} \text{ сек} \quad \text{и} \quad f_{\bar{R}}=10^{-3,75} \text{ сек}^{-1}. \quad (7.120)$$

Здесь по условию оптимальное время функционирования системы A имеет значение $T_1=T^{(0)}=10^5 \text{ сек}$.

Пусть подсистемы A_I и A_C действуют последовательно во времени и первая в девять раз дольше другой. Тогда времена их действия равны

$$t_I=0,9 T_1=0,9 \cdot 10^5 \text{ сек} \quad \text{и} \quad t_C=0,1 \cdot T_1=10^4 \text{ сек}$$

соответственно. Пусть эффективности подсистем A_I и A_C экспоненциальны. Тогда согласно соотношению (7.22') оптимальные значения частных коэффициентов эффективности должны быть назначены следующим образом:

$$\left. \begin{aligned} f_I &= T_1 f_{\bar{R}} / t_I = f_{\bar{R}} / 0,9 = 10^{-3,75} / 0,9 \text{ сек}^{-1}, \\ f_C &= T_1 f_{\bar{R}} / t_C = f_{\bar{R}} / 0,1 = 10^{-2,75} \text{ сек}^{-1}. \end{aligned} \right\} \quad (7.121)$$

Рассмотрим два варианта подсистемы $A_I=A'_I$ или $A_I=A''_I$. В первом варианте пусть подсистема A'_I должна осуществлять ки-

бернетическую оценку k -мерного параметра $a=(a_1, \dots, a_k)$, первая компонента которого принимает $N_1=10^5$ значений, проходящих статистический отбор с величиной $k_1^1=10^{-4}$, определяющей вероятность β_1 согласно (7.44).

Тогда в соответствии с (7.55) для обеспечения величины $f_I=10^{-3,75}/0,9 \text{ сек}^{-1}$ (см. (7.121)) необходимо быстрое действие подсистемы

$$V = \frac{f_I N_1}{k_1^1} = \frac{10^{-3,75} \cdot 10^5}{0,9 \cdot 10^{-4}} = 1,87 \cdot 10^5 \text{ опер/сек}.$$

Во втором варианте пусть система A''_I должна осуществить оптимальное в смысле Шеннона декодирование сообщения, принятого по БСК, определяемого вероятностью искажений $p=0,07$ при величине параметра скорости передачи $R=0,35$. Оставляя в стороне вопрос об осуществлении такого декодирования, выясним лишь вопрос о необходимой длине t передаваемых слов (эта величина определяет сложность декодирования). Ввиду экспоненциальной зависимости помехоустойчивости от величины t [см. (3.18)], будем иметь в рассматриваемом случае

$$E_p(R) = E_{0,07}(0,35) = 0,08$$

и

$$E_{0,07}(0,35) t = f_I t_I \log_2 e = f_{\bar{R}} T_1 \log_2 e = 10^{1,25} \cdot 1,43 \approx 24.$$

Отсюда как и в примере 3.2, имеем $t=300$.

Рассмотрим теперь подсистему A_C . Пусть эта подсистема ответственна за игру на выживание, которую ведет система A с системой B в конфликтной ситуации, описанной в § 4.4. Пусть требуется оценить число шагов t в этой игре, если параметры, определенные в примере 4.2 θ_1, p и $k(\theta_1, p)$, имеют значения $\theta_1=0,75; p=0,5$ и $k(\theta_1, p)=k(0,75; 0,5)=0,12$.

Имеем в рассматриваемом случае

$$k(0,75; 0,5) t = f_C t_C \log_2 e = f_{\bar{R}} T_1 \log_2 e = 10^{1,25} \cdot 1,43 \approx 24.$$

Отсюда как и в примере 4.2, имеем $t=200$.

Перейдем наконец к расчету параметров подсистемы A_R , ответственной за обеспечение надежности всей системы A . Пусть зафиксируем некоторые значения вероятностей: вызовов элементов системы A «на работу» $p=10^{-3}$, успеть обнаружить разлагающийся элемент $Q=1-0,9 \cdot 10^{-6}$ и среднего значения числа элементов памяти, занятых в стационарном режиме поиска неисправностей $\Lambda=10^3$.

Тогда согласно формуле (7.101) с учетом значения $f_R=10^{11} \text{ сек}$ [см. выражение (7.120)], получим значение параметра

$$f_{R^*} = \frac{1}{f_R^{-1} - p(1-Q)} = 10^{12} \text{ сек},$$

который связан с объемом памяти $K = \Lambda + u \sqrt{\Lambda}$ соотношениями

(7.73) и (7.77) через параметр u . Имеем

$$u = \sqrt{2 \ln f_{R^*}} = 6,8.$$

Таким образом, объем памяти K системы A , необходимый для обеспечения указанной ее надежности, должен равняться $K = 10^3 + 6,8 \sqrt{10^3} \approx 1211$ ячеек.

7.6.2. Биологические применения. По ходу изложения неоднократно приводились биологические примеры, иллюстрирующие общие положения предельных законов потенциальной эффективности сложных систем и следствий из них. Возможность такой интерпретации не случайна, так как многие модели теории, например модель живучести, возникли на основе обобщения биологических моделей.

По-видимому, сейчас биологами осознана возможность наряду с физикой и химией использовать для исследования биологических систем теорию сложных систем (см., например, последнюю часть книги [40]). В конкретной форме эта попытка сделана в монографии [41], где проблемы эволюции трактуются в терминах теории информации.

В § 1.5 была подчеркнута специфика предельных законов сложных систем, состоящая во все более определяющей роли их по мере возрастания сложности объектов исследования.

Биологические системы обогащены по сравнению с неживыми системами высшими иерархическими уровнями: клетками, органами, особями, популяциями, биоценозами и, наконец, всей биосферой.

Они представляют благодатное поле для исследований методами теории сложных систем, особенно на высших иерархических уровнях, начиная с биоценологического. Эту область «макробиологии» называют экологией [42]. В ней легко найти проявления общих законов потенциальной эффективности сложных систем. Важно, что эти проявления как чисто биологические законы уже давно имеют свои качественные биологические формулировки. Укажем на некоторые из них, известные нам.

Определение жизни и ее устойчивости, в частности, на биоценотическом уровне как постоянство структуры при переменных вещественно-энергетических носителях в современной отчетливой форме приведено в работе [43].

Предельный R -закон приводит количественные условия одного из возможных вариантов такой устойчивости в виде расширяющейся системы. RC -закон устанавливает другой вариант такой устойчивости.

Биологические проявления I -закона, правда, в несколько отвлеченной «энтропийной» форме, весьма оживленно обсуждаются во многих биологических приложениях, в частности в уже упоминавшейся теории эволюции [41].

Почти точная словесная формулировка C -закона содержится в так называемом законе конкурентного вытеснения Гаузе [44].

Современный этап развития биологии связан с попыткой прогнозирования поведения и активного регулирования биологических систем. Для этого в настоящее время широко используются методы математического моделирования (аналитический подход) [40]. На пути этих методов, и даже в большей степени, чем при исследовании технических систем, стоит «проклятие» многомерности (см. § 1.2).

Использование рассмотренных в книге простых моделей сложных систем, связанное с понятием оптимальной системы (синтетический подход), как это было показано в § 1.2, избавляет от «проклятия» многомерности. Однако на пути его использования в биологии возникает ряд новых осложнений, на которых следует остановиться подробнее.

В самом деле, все было бы хорошо, если бы «рассогласование» между траекториями биосистемы $A^*(t)$ и оптимальной системы $A_{opt}(t)$ оказалось бы незначительным (см. § 1.2).

Рассмотрим факторы, влияющие на это рассогласование.

Во-первых, на него влияет степень соответствия выбранного целевого функционала v действительному v^* , зависящая от нашего знания цели биосистемы.

Во-вторых, в случае правильно выбранного целевого функционала $v \approx v^*$, рассогласование зависит от близости биосистемы $A^*(t)$ к оптимальной системе $A_{opt}(t)$.

Но основной тезис биологии как раз и состоит в констатации такой близости, возникшей в результате прогрессивной эволюции.

Таким образом, определяющим источником рассогласования может оказаться лишь неправильный выбор

целевого функционала v , связанный с ошибочными представлениями о цели биосистемы.

К счастью, множество $\underline{a} = \{\underline{A}_r, r = \overline{1, m}\}$ возможных правдоподобных тактических целей \underline{A}_r биосистемы A^* не очень велико и они допускают экспериментальную проверку и уточнение.

Можно считать, что между разнящимися целями \underline{A}_r и соответствующими оптимальными системами A_{opt}^r имеет место взаимно однозначное соответствие

$$A_{opt}^r = g(\underline{A}_r), \quad r = \overline{1, m}.$$

Таким образом, мы приходим к ограниченному набору возможных оптимальных систем, траектории которых $A_{opt}^r(t)$ вместе с траекторией биосистемы $A^*(t)$ располагаются в пространстве „возможностей“. Траектория $A_{opt}^0(t)$, имеющая минимальное рассогласование с траекторией биосистемы $A^*(t)$, может рассматриваться, как подходящая модель биосистемы $A^*(t)$. Одновременно цель \underline{A}_0 может рассматриваться как наиболее правдоподобная цель биосистемы.

В своей повседневной практике биологи весьма эффективно по структуре и поведению биосистем определяют их тактические цели, такие, например, как хищничество, паразитизм, мимикрию и др. Стратегическая цель системы — выжить или способствовать выживанию высшего уровня иерархии.

В решении рассмотренной телеологической проблемы (проблемы цели) существенное значение имеет современное понимание целенаправленности, целесообразности и разумности просто как эффективности. В этой книге мы идем еще дальше, отождествляя понятие эффективности системы с понятием ее оптимальности. Таким образом, телеологические трудности, связанные с использованием концепции оптимальности, не являются принципиально непреодолимыми, в чем могут убедить результаты следующего пункта.

7.6.3. Теоретическое обоснование существующей структуры видов животных. Рассмотрим животных с точки зрения трех иерархических уровней. Выделим составляющие их клетки, сами особи и их совокупности, образующие вид. Обозначим через n число клеток

в особи данного вида, а через M число особей в данном виде.

Одной из характеристик сложности особи данного вида можно считать число n . Одним из проявлений устойчивости особей данного вида можно считать пространственность его клеток, которая характеризуется произведением nN . Г. Фёрстером [45] была установлена экспериментальная зависимость

$$\log_{10} nN = 15 + 0,52 \log_{10} n \quad (7.122)$$

для известных видов животных, начиная от одноклеточных и кончая человеком.

Таким образом, с ростом сложности вида животного (с ростом величины n) увеличивается его устойчивость (растет величина nN). Заметим, что с ростом атомного веса (сложности) химических элементов их распространность (устойчивость) понижается. Так, водород, имеющий атомный вес, равный единице, наиболее распространен во Вселенной.

Для получения теоретической зависимости, соответствующей экспериментальной зависимости (7.122), естественно обратиться к модели живучести (система A_{RC}). Следует заметить, что именно животные (в большей мере, чем растения) наделены этим активным качеством устойчивости.

Используем для наших целей соотношения (6.36) и (6.41), определяющие оптимальные значения жизненно важных элементов

$$n_a^1 = \theta F \beta \left(\frac{1 - \sqrt{1 - \omega}}{2} \right)^2, \quad (7.123)$$

где $\omega = E_{кр}/E$.

Число n_a^1 можно интерпретировать числом N особей вида ($n_a^1 = N$). Число E можно интерпретировать как биомассу вида. Будем считать a - и R -элементы локализованными в одной особи в виде ее жизненно важной и защитной части. Возможность такой интерпретации связана с тем, что в случае оптимальной структуры $|A_{RC}|$ числа a - и R -элементов пропорциональны [см. соотношение (6.21) при $\alpha = 1$].

Далее будем считать, что a - и R -элементы, составляющие особь, состоят из n подэлементов — клеток, причем биомасса одной клетки равна величине b .

Тогда для биомассы вида будем иметь

$$E = nNb. \quad (7.124)$$

Рассмотрим биологически интересный случай больших значений $E \gg E_{кр}$. Тогда согласно соотношению (6.42) таким значением E будут соответствовать малые значения

$$\omega = (E_{кр}/E)^\delta = \frac{4(2-\delta)}{\theta F B E^\delta} = \frac{4\gamma}{\theta F \beta} \ll 1. \quad (7.125)$$

Рассмотрим соотношение (7.123) в асимптотическом случае $\omega \rightarrow 0$. Тогда, используя выражение (6.41) для $\beta = BE^\delta/(\delta - 1)$ и $\gamma = (2 - \delta)/(\delta - 1)$, будем иметь

$$N = n_a^1 \approx \frac{\theta F \beta}{4} \frac{\omega^2}{4} = \frac{\gamma^2}{\theta F \beta} = \left(\frac{2-\delta}{\delta-1} \right)^2 / \theta F \frac{BE^\delta}{\delta-1} = \frac{C_1}{E^\delta}, \quad (7.126)$$

где $C_1 = (2 - \delta)^2 / \theta F B (\delta - 1)$.

Подставив выражение (7.124) в (7.126), получим

$$N \approx C_1 / b^\delta (nN)^\delta = C_2 / (nN)^\delta, \quad (7.127)$$

где $C_2 = C_1 / b^\delta$.

Умножая обе части соотношения (7.127) на n и логарифмируя, после несложных преобразований получим

$$\log_{10} nN \approx \frac{1}{1+\delta} \log_{10} C_2 + \frac{1}{1+\delta} \log_{10} n.$$

Беря значения δ , близкие к 1 как и в числовом примере, соответствующем рис. 6.3, получим

$$\log_{10} nN \approx C + 0,5 \log_{10} n, \quad (7.128)$$

где $C \approx 0,5 \log_{10} C_2$ — константа, смысл которой состоит в следующем. Для одноклеточных животных ($n=1$) параметр C интерпретируется как параметр критической «тесноты». При $C=14 \div 15$ имеем примерно одно одноклеточное животное на один кв. метр земной поверхности. Указанное значение $C \approx 15$, будучи подставленным в соотношение (7.128), делает его весьма близким к экспериментальному соотношению (7.122). Таким образом, оно может быть объяснено, исходя из модели живучести сложных систем.

7.6.4. Эволюционное моделирование. В предисловий упоминалось самое далекое по сложности используемых моделей к рассматриваемому направлению теоретической кибернетики — эвристическое программирование. Совсем недавно для решения задач из этой области было предложено так называемое эволюционное моделирование [46]. В то время как первое направление пытается моделировать то что есть, появившееся направление пытается моделировать условия, приводящие к появлению интересующих нас X -качеств сложных систем. При этом возможно возникновение и новых, не существующих и заранее не известных вариантов оптимальных систем. Таким образом, исследуется не то что есть, а то что может быть.

Последнее свойство эволюционного моделирования неожиданно сблизило его с теорией потенциальной эффективности сложных систем [54, 55].

Говоря об эволюционном моделировании, нельзя не упомянуть о цикле пионерских работ [47, 48] по моделированию биологического процесса эволюции. Эти работы помимо их биологического значения в свете идей эволюционного моделирования приобретают общесистемное значение.

ЗАКЛЮЧЕНИЕ

Системы определяются своей структурой и поведением. Они делятся на простые и сложные в зависимости от отсутствия или присутствия в их поведении акта решения. Простые системы успешно изучаются физикой без использования понятия цели.

Сравнительно недавно начато количественное изучение сложных систем с привлечением понятия цели. Целесообразность структуры и целенаправленность поведения системы определяются эффективностью, с которой она достигает своей цели.

Стратегическая цель системы (выжить вопреки разрушающим воздействиям внешней среды) обуславливает отдельные ее тактические цели: сохранение целостности, информированность о среде, внешняя активность и адаптивное переустройство.

Достижение указанных тактических целей обеспечивается формированием специфических качеств сложных систем: надежности, помехоустойчивости, управляемости

и самоорганизации соответственно (гл. 1—5). Целостное исследование системы требует совместного рассмотрения указанных качеств (гл. 6).

Возможны два количественных подхода при исследовании сложных систем. Один из них состоит в построении отдельных моделей, приближающихся по сложности к оригиналам. Здесь адекватность модели объекту достигается за счет понижения общности результатов. Основной преградой на пути этого подхода является так называемое «проклятие многомерности».

Второй подход состоит в построении широких классов простых моделей, описывающих отдельные качества сложных систем. Здесь общность результатов достигается за счет неадекватности модели объекту, носящей оценочный характер.

Теория потенциальной эффективности сложных систем использует второй из указанных подходов. Для возможности количественного определения тактических целей в ней используется понятие (u, v) -обмена между системой и средой (система обменивает некоторое количество u своих ресурсов одного качества на некоторое количество v ресурсов другого качества). Тактическая цель системы состоит в выгодном (u, v) -обмене, т. е. в получении за фиксированную величину u как можно большую величину v . Система, производящая выгодный (u, v) -обмен, называется оптимальной.

Эффективность системы удобно оценивать вероятностью $P(u, v)$ выгодного (u, v) -обмена. Имеет место предельный закон эффективности сложных систем. Именно для асимптотического случая больших значений u и v вероятность $P(u, v)$ близка либо к нулю, либо к единице, в зависимости от «жадности» системы ($v > v_0(u)$) или ее «умеренности» ($v < v_0(u)$) соответственно, где $v_0(u)$ — некоторая фундаментальная величина для рассматриваемого класса моделей систем.

Весьма примечательным обстоятельством является общий характер предельного закона для разнообразных моделей сложных систем, возникающих независимо друг от друга в теориях надежности, информации, игр и других областях. Это связано с проявлением вероятностного закона больших уклонений, присущего асимптотическому поведению всех рассматриваемых моделей.

Вероятностная форма определения эффективности системы при достижении отдельных ее тактических це-

лей позволяет оценить вероятность достижения ею стратегической цели с помощью неравенства Буля (см. гл. 7).

Развитая теория позволяет оценить потенциальную эффективность сложных технических систем в целом и в зависимости от этого назначить рациональные требования к эффективности составляющих их подсистем. Эта же теория может быть использована при исследовании биологических систем. Обнадеживающим результатом в этом направлении служит обоснование эмпирических данных о распространенности клеток животных различных видов, основанное на теории живучести (см. гл. 7).

ПРИЛОЖЕНИЯ

Приложение 1. Производящие и некоторые оценки.

П.1.1. Максимальная лемма. Рассмотрим два произвольных множества $R = \{r\}$ и $U = \{u\}$, из которых первое конечно.

Зададим на элементах r и u указанных множеств две действительные функции $A(r)$ и $B(r, u)$ — неотрицательную и положительную соответственно. Определим функцию

$$g(u) = \sum_{r \in R} A(r) B(r, u), \quad (\text{П.1.1})$$

которую будем называть *производящей* функции $A(r)$.

Пусть $\mathcal{E} \subseteq R$ — произвольное подмножество R . Рассмотрим сумму

$$A(\mathcal{E}) = \sum_{r \in \mathcal{E}} A(r).$$

Верхнюю оценку суммы $A(\mathcal{E})$ дает следующая «максимальная»

Лемма П.1

$$A(\mathcal{E}) \leq \max_{r \in \mathcal{E}} \min_{u \in U} [g(u)/B(r, u)], \quad (\text{П.1.2})$$

где в случае отсутствия \min берется соответствующий \inf .

Доказательство. При произвольном фиксированном $u \in U$ имеем из (П.1.1)

$$g(u) = \sum_{r \in R} A(r) B(r, u) \geq \sum_{r \in \mathcal{E}} A(r) B(r, u) \geq \min_{r \in \mathcal{E}} B(r, u) A(\mathcal{E}),$$

отсюда

$$A(\mathcal{E}) \leq g(u)/\min_{r \in \mathcal{E}} B(r, u) = \max_{r \in \mathcal{E}} [g(u)/B(r, u)].$$

В частности, эти оценки имеют место и для значения $u = u_0 \in U$, при котором дробь под знаком \max в последнем выражении обращается в минимум, что и требовалось доказать.

Лемма П.2. Пусть задано $2a$ действительных неотрицательных чисел $g_i (i = \overline{1, a})$ и $b_i (i = \overline{1, a})$ с условием $\sum_{i=1}^a b_i = 1$ (числа b_i будем называть весами). Пусть произвольное целое неотрицательное чи-

сло n и действительное число v удовлетворяют условию

$$0 < v < \min_{1 \leq i \leq a} g_i^{1/n}.$$

Тогда имеют место оценки

$$\prod_{i=1}^a g_i^{b_i} \leq \sum_{i=1}^a b_i g_i \leq \left\{ v + \left[\sum_{i=1}^a b_i (g_i^{1/n} - v)^n \right]^{1/n} \right\}^n. \quad (\text{П.1.3})$$

Доказательство. При любых действительных числах $x > 0$ и $0 \leq b \leq 1$ функция

$$f(x) = x^b - b x - (1-b) \leq 0. \quad (\text{П.1.4})$$

Это следует из того, что $\max f(x) = f(1) = 0$. Положим теперь $x = g_i/g_2$. Тогда из (П.1.4) получим

$$g_1^b g_2^{1-b} \leq b g_1 + (1-b) g_2. \quad (\text{П.1.5})$$

С помощью математической индукции оценка (П.1.5) легко обобщается на случай $2a$ чисел g_i и b_i , что приводит к нижней оценке (П.1.3).

Положим в оценке (П.1.5)

$$g_1 = x_i^{1/b} \left/ \sum_{j=1}^a x_j^{1/b} \right. \quad \text{и} \quad g_2 = y_i^{1/(1-b)} \left/ \sum_{j=1}^a y_j^{1/(1-b)} \right.$$

и сложим получившиеся a неравенств.

Тогда после несложных преобразований получим известные неравенства Гельдера [49]

$$\sum_{i=1}^a x_i y_i \leq \left(\sum_{i=1}^a x_i^{1/b} \right)^b \left(\sum_{i=1}^a y_i^{1/(1-b)} \right)^{1-b}.$$

Применим к каждому слагаемому правой части тождество

$$\sum_{i=1}^a (x_i + y_i)^{1/b} = \sum_{i=1}^a x_i (x_i + y_i)^{1/b-1} + \sum_{i=1}^a y_i (x_i + y_i)^{1/b-1}$$

неравенство Гельдера с показателями $1/b$ и $1/(1-b)$. Тогда с учетом тождества

$$\left(\frac{1}{b} - 1 \right) \frac{1}{1-b} = \frac{1}{b}$$

получим известное неравенство Минковского [49]

$$\left[\sum_{i=1}^a (x_i + y_i)^{1/b} \right]^b \leq \left(\sum_{i=1}^a x_i^{1/b} \right)^b + \left(\sum_{i=1}^a y_i^{1/b} \right)^b.$$

Подставив в неравенство Минковского значения $b = 1/n$,

$$x_i = b_i^{1/n} v \quad \text{и} \quad y_i = b_i^{1/n} (g_i^{1/n} - v),$$

после несложных преобразований получим верхнюю оценку (П.1.3). Таким образом, лемма П.2 доказана полностью.

П.1.2. Оценки больших уклонений. Используем доказанные леммы для оценки двух сумм:

$$A_l(\mathcal{E}) = \sum_{r=0}^m A_l(r), \quad (l=1, 2), \quad (\text{П.1.6})$$

где r, m и $n(0 \leq r \leq m \leq n)$ — целые числа,

$$\mathcal{E} = \{r : 0 \leq r \leq m\} \subseteq R = \{r : 0 \leq r \leq n\}.$$

Функции $A_l(r)$ ($l=1, 2$) определяются следующими соотношениями для первой суммы ($l=1$):

$$A_1(r) = \sum_{\nu=1}^m p_\nu^{\varepsilon_\nu} q_\nu^{1-\varepsilon_\nu}, \quad (\text{П.1.7})$$

$$\sum_{\nu=1}^n \varepsilon_\nu = r,$$

где $\varepsilon_\nu = 0, 1$, а действительные числа p_ν и q_ν удовлетворяют условиям $0 \leq p_\nu, q_\nu = 1 - p_\nu \leq 1$ ($\nu = \overline{1, n}$); для второй суммы ($l=2$):

$$A_2(r) = \sum_{i=1}^a b_i C_n^r p_i^r q_i^{n-r}, \quad (\text{П.1.8})$$

где b_i, p_i и q_i — действительные числа, удовлетворяющие условиям $b_i \geq 0, 0 \leq p_i, q_i = 1 - p_i \leq 1$ ($i = \overline{1, a}$);

$$\sum_{i=1}^a b_i = 1.$$

Положим

$$B(r, u) = u^r, \quad (\text{П.1.9})$$

где u — произвольное действительное число.

Тогда, используя определение производящей (П.1.1), выражение (П.1.9) и определения функций $A_l(r)$ и $A_2(r)$ соотношениями (П.1.7) и (П.1.8), получим соответствующие производящие:

$$g_1(u) = \prod_{\nu=1}^m (p_\nu u + q_\nu) \quad (\text{П.1.10})$$

и

$$g_2(u) = \sum_{i=1}^a b_i (p_i u + q_i)^n. \quad (\text{П.1.11})$$

Используем теперь оценки (П.1.3) для оценок полученных производящих сверху. Для оценки $g_1(u)$ используем нижнюю оценку (П.1.3), полагая $g = p_\nu u + q_\nu$ и $b_\nu = 1/n$. Тогда после несложных преобразований получим

$$g_1(u) \leq [1 + \bar{p}(u-1)]^n, \quad (\text{П.1.12})$$

где

$$\bar{p} = \frac{1}{n} \sum_{\nu=1}^m p_\nu. \quad (\text{П.1.12})$$

Для оценки $g_2(u)$ используем верхнюю оценку (П.1.3), полагая $g_i = (p_i u + q_i)^n$ и

$$v = \begin{cases} u, & \text{если } u < 1, \\ 1, & \text{если } u > 1. \end{cases} \quad (\text{П.1.13})$$

Тогда после несложных преобразований с учетом соотношений (П.1.13) получим

$$g_2(u) \leq \begin{cases} [1 + (1 - q^{(n)})(u-1)]^n, & \text{если } u < 1, \\ [1 + p^{(n)}(u-1)]^n, & \text{если } u > 1, \end{cases} \quad (\text{П.1.14})$$

где

$$p^{(n)} = \left(\sum_{i=1}^a b_i p_i^n \right)^{1/n}; \quad q^{(n)} = \left(\sum_{i=1}^a b_i q_i^n \right)^{1/n}. \quad (\text{П.1.14}')$$

Таким образом, во всех случаях $A(r) = A_1(r)$ и $A_2(r)$ нами получена оценка

$$g(u) \leq [1 + p(u-1)]^n, \quad (\text{П.1.15})$$

где

$$g(u) = g_1(u), \quad g_2(u) \quad \text{и} \quad p = \bar{p}, \quad p^{(n)}, \quad 1 - q^{(n)}.$$

Используем теперь оценку (П.1.2) леммы П.1 для оценки суммы $A(\mathcal{E}) = A_1(\mathcal{E}), A_2(\mathcal{E})$, определяемой соотношениями (П.1.6). Имеем с учетом оценки (П.1.15) и соотношения (П.1.9), полагая $U = \{u : u < 1\}$:

$$A(\mathcal{E}) \leq \max_{r \leq m} \min_{u < 1} \{u^{-r} [1 + p(u-1)]^n\} \leq \min_{u < 1} \{u^{-m} [1 + p(u-1)]^n\} \quad (\text{П.1.16})$$

(в рассматриваемом случае \max и \min можно поменять местами).

Для отыскания последнего максимума в оценке (П.1.16) приравняем производную фигурной скобки по u к нулю. После несложных преобразований получим уравнение

$$\frac{m}{n} - \frac{pu}{1 + p(u-1)} = 0. \quad (\text{П.1.17})$$

Отсюда экстремальное значение $u = u_*$ имеет вид

$$u_* = \frac{\frac{m}{n}(1-p)}{\left(1 - \frac{m}{n}\right)p} < 1, \quad (\text{П.1.18})$$

при котором согласно соотношению (П.1.17) имеем

$$1 + p(u_* - 1) = \frac{1 - p}{1 - \frac{m}{n}}, \quad (\text{П.1.19})$$

и из соотношения (П.1.18) следует, что

$$\frac{m}{n} < p. \quad (\text{П.1.20})$$

Подставив соотношения (П.1.18) и (П.1.19) в выражение оцениваемой фигурной скобки, получим, как легко показать, ее минимальное значение

$$\min_{u < 1} \{u^{-m} [1 + p(u - 1)^n]\} = \left[\left(\frac{m}{n} \right)^{m/n} \left(\frac{1 - \frac{m}{n}}{1 - p} \right)^{1 - m/n} \right]^{-n}. \quad (\text{П.1.21})$$

Введя удобное обозначение функции $k(x, y)$ двух действительных переменных $0 \leq x, y \leq 1$:

$$k(x, y) = x \ln \frac{x}{y} + (1 - x) \ln \frac{1 - x}{1 - y}, \quad (\text{П.1.21}')$$

объединяя оценку (П.1.16) и соотношение (П.1.21), получим при $\frac{m}{n} < p$:

$$A(\mathcal{E}) \leq \exp(-kn), \quad (\text{П.1.22})$$

где величина $k = k\left(\frac{m}{n}, p\right)$ определяется согласно соотношению (П.1.22).

Аналогично для множеств $\bar{\mathcal{E}} = R \setminus \mathcal{E} = \{r : r > m\}$ и $\bar{U} = \{u : u > 1\}$ получим при $\frac{m}{n} > p$ ту же оценку

$$A(\bar{\mathcal{E}}) \leq 1 - A(\mathcal{E}) \leq \exp(-kn). \quad (\text{П.1.23})$$

Полученные оценки (П.1.22) и (П.1.23) можно записать в следующей компактной форме с учетом выражений (П.1.7) и (П.1.8) и условий $u \leq 1$, эквивалентных условиям $\frac{m}{n} \leq p$:

$$A(\mathcal{E}) = A_1(\mathcal{E}) \begin{cases} \leq \exp\left[-k\left(\frac{m}{n}, p\right)n\right], & \text{при } \frac{m}{n} < p \\ \geq 1 - \exp\left[-k\left(\frac{m}{n}, p\right)n\right], & \text{при } \frac{m}{n} > p, \end{cases} \quad (\text{П.1.24})$$

где функция $k(x, y)$ определена соотношением (П.1.21'). При этом для $A(\mathcal{E}) = A_1(\mathcal{E})$,

$$p = \begin{cases} \bar{p} & \\ 1 - q^{(n)} & \text{при } \frac{m}{n} < p \\ p^{(n)} & \text{при } \frac{m}{n} > p \end{cases} \quad \text{для } A(\mathcal{E}) = A_2(\mathcal{E}), \quad (\text{П.1.25})$$

где величины \bar{p} , $p^{(n)}$ и $q^{(n)}$ определяются соотношениями (П.1.12') и (П.1.14') соответственно. Так как для $n > 1$ $p^{(1)} < p^{(n)}$, $q^{(1)} < q^{(n)}$ [49, стр. 41] и $p^{(1)} + q^{(1)} = 1$ (откуда $1 - q^{(n)} \leq p^{(n)}$), то оценка (П.1.24) не определена для значений m/n в интервале $(1 - q^{(n)} \leq m/n \leq p^{(n)})$ в случае $A(\mathcal{E}) = A_2(\mathcal{E})$.

В частном случае постоянных значений $p_v = p_i = p$, когда с ними совпадают значения $\bar{p} = p^{(n)} = 1 - q^{(n)} = p$, а также совпадают значения $n = a$,

$$A_1(r) = A_2(r) = A(r) = C_n^r p^r q^{n-r}$$

и

$$A_1(\mathcal{E}) = A_2(\mathcal{E}) = A(\mathcal{E}) = \sum_{r=0}^m C_n^r p^r q^{n-r},$$

из оценок (П.1.24) получим известные оценки Чернова [14] для биномиальных сумм

$$A(\mathcal{E}) = \sum_{r=0}^m C_n^r p^r q^{n-r} \begin{cases} \leq \exp\left[-k\left(\frac{m}{n}, p\right)n\right] & \text{при } \frac{m}{n} < p, \\ \geq 1 - \exp\left[-k\left(\frac{m}{n}, p\right)n\right] & \text{при } \frac{m}{n} > p, \end{cases} \quad (\text{П.1.26})$$

где $q = 1 - p$ и функция $k(x, y)$ определена соотношением (П.1.21').

Вероятностная интерпретация полученных соотношений очевидна. Пусть целочисленная случайная величина μ принимает целочисленные значения $r = \overline{0, n}$ с вероятностями $\mathcal{P}(\mu = r) = A(r)$. Тогда ее функция распределения имеет вид $\mathcal{P}(\mu \leq m) = A(\mathcal{E})$, а вероятность превышения μ величины m $\mathcal{P}(\mu > m) = A(\bar{\mathcal{E}})$ оценивает "большие" отклонения μ от среднего значения, так как при фиксированном значении $\frac{m}{n} = \theta$ имеем $m = \theta n$ [при нормальных отклонениях $m = O(\sqrt{n})$].

П.1.3. Оценки на основе формулы Стирлинга. Дополним полученные оценки другими, основанными по формуле Стирлинга [50]:

$$\sqrt{2\pi n} (n/e)^n < n! < \sqrt{2\pi n} (n/e)^n \exp(1/12n). \quad (\text{П.1.27})$$

Пусть $n\theta$ ($0 < \theta < 1$) — целое число. Тогда, используя оценку (П.1.27), получим

$$C_n^{n\theta} = C_n^{n(1-\theta)} = \frac{n!}{(n\theta)! [n(1-\theta)]!} > \frac{\sqrt{2\pi n} (n/e)^n \exp\{-[1/12n\theta + 1/12n(1-\theta)]\}}{\sqrt{2\pi n\theta} (n\theta/e)^{n\theta} \sqrt{2\pi n(1-\theta)} [n(1-\theta)/e]^{n(1-\theta)}}. \quad (\text{П.1.28})$$

Но при $n\theta \geq 1$ и $n(1-\theta) \geq 3$ имеем

$$1/12n\theta + 1/12n(1-\theta) \leq 1/9.$$

Поэтому здесь

$$\begin{aligned} \exp \{-[1/12n\theta + 1/12n(1-\theta)]\} &\geq e^{-1/9} = \\ &= 0,895 > \sqrt{\pi}/2 = 0,887\dots \end{aligned} \quad (\text{П.1.29})$$

Используя оценку (П.1.29), получим из соотношения (П.1.28) оценку Шеннона [36] для $n\theta \geq 1$ и $n(1-\theta) \geq 3$:

$$C_n^{n\theta} = C_n^{n(1-\theta)} > \frac{1}{\sqrt{8n\theta(1-\theta)}} \exp \left[\theta \ln \frac{1}{\theta} + (1-\theta) \ln \frac{1}{1-\theta} \right]. \quad (\text{П.1.30})$$

Непосредственный просчет показывает, что неравенство (П.1.30) верно и для значений $n\theta < 1$ и $n(1-\theta) < 3$.

Из оценки (П.1.30) следует, что

$$\begin{aligned} &C_n^m p^m (1-p)^{n-m} > \\ &> \frac{1}{\sqrt{8n \frac{m}{n} \left(1 - \frac{m}{n}\right)}} \exp \left[-k \left(\frac{m}{n}, p \right) n \right], \end{aligned} \quad (\text{П.1.31})$$

где величина $k \left(\frac{m}{n}, p \right)$ определяется соотношением (П.1.21').

Используем теперь соотношение (П.1.31) для оценки биномиальной суммы (П.1.26).

Возьмем в качестве нижней оценки всей суммы последнее ее слагаемое. Тогда в сочетании с оценками (П.1.26) получим двусторонние оценки:

$$\begin{aligned} &\sum_{r=0}^m C_n^r p^r (1-p)^{n-r} = \\ &= A(\mathcal{E}) \begin{cases} \left. \begin{aligned} &\leq \exp(-kn) \\ &\geq \left[8n \frac{m}{n} \left(1 - \frac{m}{n}\right) \right]^{-1/2} \exp(-kn) \end{aligned} \right\} \text{ при } \frac{m}{n} < p \\ \left. \begin{aligned} &\geq 1 - \exp(-kn) \\ &\leq 1 - \left[8n \frac{m}{n} \left(1 - \frac{m}{n}\right) \right]^{-1/2} \exp(-kn) \end{aligned} \right\} \text{ при } \frac{m}{n} > p, \end{cases} \end{aligned} \quad (\text{П.1.32})$$

где коэффициент $k = k \left(\frac{m}{n}, p \right)$ определяется соотношением (П.1.21').

Перейдем теперь к двусторонним оценкам сумм $A(\mathcal{E}) = A_1(\mathcal{E})$, $A_2(\mathcal{E})$ общего вида. Пусть фигурирующие в них вероятности p_v и p_i заключены в пределах

$$0 \leq p = p - \Delta \leq p_v, p_i \leq p + \Delta = \tilde{p} \leq 0,5. \quad (\text{П.1.33})$$

Тогда, как легко показать,

$$p = p - \Delta \leq \underline{p} = \frac{p - \Delta}{1 - 2\Delta} \leq p \leq \frac{p + \Delta}{1 + 2\Delta} = \tilde{p} \leq p + \Delta = \tilde{p}$$

и

$$\underline{p} = p - \Delta \leq \bar{p} = \frac{1}{n} \sum_{v=1}^n p_v, p^{(n)} = \left(\sum_{i=1}^a b_i p_i^n \right)^{1/n} \leq p + \Delta = \tilde{p}, \quad (\text{П.1.34})$$

причем при $n=2$ имеем $p = \bar{p}$.

Имеют место следующие легко проверяемые оценки:

$$\begin{aligned} &(1-2\Delta) p^{\varepsilon_v} (1-p)^{1-\varepsilon_v} \leq p_v^{\varepsilon_v} (1-p_v)^{1-\varepsilon_v} \leq \\ &\leq (1+2\Delta) \tilde{p}^{\varepsilon_v} (1-\tilde{p})^{1-\varepsilon_v} \\ &(1-2\Delta)^n p^m (1-p)^{n-m} \leq p_i^m (1-p_i)^{n-m} \leq \\ &\leq (1-2\Delta)^n \tilde{p}^m (1-\tilde{p})^{n-m}. \end{aligned} \quad (\text{П.1.35})$$

Теперь, используя оценки (П.1.32) и (П.1.35), легко получим двусторонние оценки сумм $A(\mathcal{E}) = A_l(\mathcal{E})$ ($l=1, 2$) (П.1.6), определяемых соотношениями (П.1.7) и (П.1.8)

$$\begin{aligned} &A(\mathcal{E}) = \\ &= A_l(\mathcal{E}) \begin{cases} \left. \begin{aligned} &\leq (1+2\Delta)^{\alpha_l} \exp(-kn) \text{ при } \frac{m}{n} < \tilde{p}, \\ &\geq (1-2\Delta)^{\alpha_l} \left[8n \frac{m}{n} \left(1 - \frac{m}{n}\right) \right]^{-1/2} \exp(-kn) \text{ при } \frac{m}{n} < p, \\ &\geq 1 - (1+2\Delta)^{\alpha_l} \exp(-kn) \text{ при } \frac{m}{n} > \tilde{p}, \\ &\leq 1 - (1-2\Delta)^{\alpha_l} \left[8n \frac{m}{n} \left(1 - \frac{m}{n}\right) \right]^{-1/2} \exp(-kn) \text{ при } \frac{m}{n} > p, \end{aligned} \right\} \end{cases} \end{aligned} \quad (\text{П.1.36})$$

где

$$\alpha_l = \begin{cases} 1 & \text{при } l=1, \\ n & \text{при } l=2 \end{cases}$$

и

$$\tilde{k} = k \left(\frac{m}{n}, \tilde{p} \right) \leq k = k \left(\frac{m}{n}, p \right) \leq \underline{k} = k \left(\frac{m}{n}, \underline{p} \right),$$

причем функция $k(x, y)$ определена соотношением (П.1.21'). При $\Delta=0$ оценки (П.1.36) переходят в оценки (П.1.32). Преимущество оценок (П.1.36) перед оценками (П.1.24) состоит в том, что они, в отличие от последних, двусторонние. Однако им присущ недоста-

ток оценок (П.1.24) суммы $A_2(\mathcal{E})$ — их неопределенность для значений величины m/n в пределах

$$\underline{p} \leq \frac{m}{n} \leq \bar{p}.$$

Величина этого интервала по сравнению с интервалом неопределенности оценок (П.1.24) $1 - q^{(n)} \leq \frac{m}{n} \leq p^{(n)}$, а также сила сравниваемых оценок вне их интервалов неопределенности не могут быть найдены в общем случае, так как зависят от конкретных значений величин p_i ($i = \overline{1, n}$) и p_i ($i = \overline{1, a}$).

П.1.4. Некоторые свойства производящих. Далее будут использованы некоторые известные свойства производящих, определяемых в общем случае соотношением (П.1.1). Здесь будет рассмотрен его частный случай при $B(r, u) = u^r$, где u действительное число, а r — целое неотрицательное число. Часто удобно связывать производящую не с функцией $A(r)$, а с дискретной случайной величиной ρ , принимающей значения r с вероятностями $\mathcal{P}(\rho=r) = A(r)$, которые будем называть распределением случайной величины ρ . Расширим старое определение производящей, считая множество $R = \{r\}$ счетным $r=0, 1, 2, \dots$

В соответствии с этим определим производящую случайной величины в виде ряда

$$g_\rho(u) = \sum_{r=0}^{\infty} A(r) u^r, \quad (\text{П.1.37})$$

сходящегося хотя бы для одного значения $u = u_0$. Величину u будем называть формальным аргументом производящей функции.

Рассмотрим сумму σ_n фиксированного числа n независимых случайных величин

$$\sigma_n = \sum_{i=1}^n \rho_i \quad (\text{П.1.38})$$

с различными распределениями

$$\mathcal{P}(\rho_i = r_i) = A_i(r_i) \quad (i = \overline{1, n}). \quad (\text{П.1.39})$$

Легко видеть, что распределение σ_n $\mathcal{P}(\sigma_n = s) = B_n(s)$ имеет вид

$$B_n(s) = \sum_{\sum_{i=1}^n r_i = s} \prod_{i=1}^n A_i(r_i). \quad (\text{П.1.40})$$

Введем обозначения для соответствующих производящих случайных величин ρ_i ($i = \overline{1, n}$) и σ_n :

$$g_{\rho_i}(u) = \sum_{r_i=0}^{\infty} A_i(r_i) u^{r_i} \quad (i = \overline{1, n}), \quad g_{\sigma_n}(u) = \sum_{s=0}^{\infty} B_n(s) u^s. \quad (\text{П.1.41})$$

Соотношение (П.1.40) и обозначения (П.1.41) позволяют легко проверить справедливость тождества

$$g_{\sigma_n}(u) = \prod_{i=1}^n g_{\rho_i}(u), \quad (\text{П.1.42})$$

которое известно как свойство *мультипликативности* производящих [50].

Рассмотрим сумму σ_v случайного числа v независимых случайных величин ρ_i :

$$\sigma_v = \sum_{i=1}^v \rho_i, \quad (\text{П.1.43})$$

независимых также и от случайной величины v .

Пусть задано распределение

$$\mathcal{P}(v = n) = p_n \quad (n = 0, 1, 2, \dots) \quad (\text{П.1.44})$$

случайной величины v .

Тогда на основании формулы полной вероятности для распределения $\mathcal{P}(\sigma_v = s) = B(s)$ случайной величины σ_v получим выражение

$$B(s) = \sum_{n=0}^{\infty} p_n B_n(s). \quad (\text{П.1.45})$$

Обозначим производящие случайных величин v и σ_v :

$$g_v(u) = \sum_{n=0}^{\infty} p_n u^n \quad \text{и} \quad g_{\sigma_v}(u) = \sum_{s=0}^{\infty} B(s) u^s. \quad (\text{П.1.46})$$

Умножая обе части соотношения (П.1.45) на u^s и суммируя по s с учетом обозначений (П.1.41) и (П.1.46), получим

$$g_{\sigma_v}(u) = \sum_{n=0}^{\infty} p_n g_{\sigma_n}(u). \quad (\text{П.1.47})$$

Подставив в соотношение (П.1.47) выражение (П.1.42) для $g_{\sigma_n}(u)$, получим окончательное выражение

$$g_{\sigma_v}(u) = \sum_{n=0}^{\infty} p_n \prod_{i=1}^n g_{\rho_i}(u) \quad (\text{П.1.48})$$

для производящей функции случайной величины σ_v , которое будем называть свойством *итерации* производящих (см. далее).

В частном случае одинаково распределенных случайных величин ρ_i , когда

$$\mathcal{P}(\rho_i = r_i) = A(r_i) \quad (i = \overline{1, n}) \quad (\text{П.1.49})$$

все их производящие тождественны, имеем

$$g_{\rho_i}(u) = g_\rho(u) = \sum_{r=0}^{\infty} A(r) u^r. \quad (\text{П.1.50})$$

Подставив выражение (П.1.50) в (П.1.48) с учетом (П.1.46), получим

$$g_v(u) = g_v[g_p(u)]. \quad (\text{П.1.51})$$

Соотношение (П.1.51) известно как свойство итерации производящих [50] [можно считать, что это свойство заложено и в общем соотношении (П.1.48)].

В качестве примера распределения случайной величины v рассмотрим геометрическое распределение

$$\mathcal{P}(v=n) = q^{n-1}p \quad (n=1, 2, \dots), \quad (\text{П.1.52})$$

где $0 \leq p, q = 1 - p \leq 1$. Здесь производящая

$$g_v(u) = \sum_{n=1}^{\infty} q^{n-1} p u^n = \frac{pu}{1-qu}. \quad (\text{П.1.53})$$

Заметим, что с помощью производящей $g_p(u)$ (П.1.37) случайной величины p можно находить факториальные моменты

$$E_p(p-1) \dots (p-s+1) = E_p[s!] = \sum_{r=s}^{\infty} r(r-1) \dots (r-s+1) A(r) \quad (\text{П.1.54})$$

порядка s по формуле

$$E_p[s!] = \frac{d^s}{du^s} g_p(u) |_{u=1}. \quad (\text{П.1.55})$$

В частности, среднее и дисперсия случайной величины имеют вид

$$\begin{aligned} E_p &= g'_p(1) \text{ и } D_p = E_p(p-1) + E_p - (E_p)^2 = \\ &= g''_p(1) + g'_p(1) - [g'_p(1)]^2 \end{aligned} \quad (\text{П.1.56})$$

Приложение 2. Неравенства Буля.

Рассмотрим произвольное множество $R = \{r\}$ элементов r и два его произвольных подмножества $\mathcal{E}_1, \mathcal{E}_2 \subseteq R$.

Легко доказать следующее соотношение:

$$R \setminus (R \setminus \mathcal{E}_1) \cup (R \setminus \mathcal{E}_2) = \overline{\mathcal{E}_1} \cup \overline{\mathcal{E}_2} = \mathcal{E}_1 \cap \mathcal{E}_2 \subseteq \mathcal{E}_i \quad (i=1, 2). \quad (\text{П.2.1})$$

Его справедливость следует из того, что допущение о принадлежности элемента к одной из его частей влечет за собой его принадлежность к другой, и наоборот.

Далее с помощью математической индукции легко доказывается обобщение соотношения (П.2.1) на произвольное число n подмножеств

$$\mathcal{E}_1, \dots, \mathcal{E}_i, \dots, \mathcal{E}_n \subseteq R.$$

Оно имеет вид

$$R \setminus \bigcup_{i=1}^n (R \setminus \mathcal{E}_i) = \bigcap_{i=1}^n \overline{\mathcal{E}_i} = \bigcap_{i=1}^n \mathcal{E}_i \subseteq \mathcal{E}_i \quad (i = \overline{1, n}). \quad (\text{П.2.2})$$

Если на элементах r множества $R = \{r\}$ определена вероятность $\mathcal{P}(r)$, то из теоретико-множественного соотношения (П.2.2) следуют вероятностные оценки

$$1 - \sum_{i=1}^n [1 - \mathcal{P}(\mathcal{E}_i)] \leq \mathcal{P}\left(\bigcap_{i=1}^n \mathcal{E}_i\right) \leq \min_{1 \leq i \leq n} \mathcal{P}(\mathcal{E}_i), \quad (\text{П.2.3})$$

которые известны как неравенства Буля.

Говорят, что множества \mathcal{E}_i ($i = \overline{1, n}$) независимы в вероятностном смысле, если

$$\mathcal{P}\left(\bigcap_{i=1}^n \mathcal{E}_i\right) = \prod_{i=1}^n \mathcal{P}(\mathcal{E}_i). \quad (\text{П.2.4})$$

Множества \mathcal{E}_i ($i = \overline{1, n}$), для которых соотношения (П.2.4) не выполняются, называются зависимыми.

Верхняя оценка (П.2.3) для независимых множеств очевидна, так как $0 \leq \mathcal{P}(\mathcal{E}_i) \leq 1$.

Перейдем к выводу нижней оценки. Имеем из соотношения (П.2.4)

$$\mathcal{P}\left(\bigcap_{i=1}^n \mathcal{E}_i\right) = \prod_{i=1}^n \{1 - [1 - \mathcal{P}(\mathcal{E}_i)]\} \geq 1 - \sum_{i=1}^n [1 - \mathcal{P}(\mathcal{E}_i)], \quad (\text{П.2.5})$$

причем оценка (П.2.5) тем лучше, чем вероятности $\mathcal{P}(\mathcal{E}_i)$ ближе к единице.

Сравнение оценок Буля (П.2.3) и (П.2.5) показывает, что нижняя оценка вероятности пересечения зависимых множеств \mathcal{E}_i с вероятностями $\mathcal{P}(\mathcal{E}_i)$, близкими к единице, ведет себя так же, как и для независимых множеств.

Для приложений теории вероятностей основное значение имеют множества с вероятностями, близкими к единице (или к нулю — дополнительных множеств).

Поэтому неравенство Буля имеет важнейшее значение для приложений, позволяя оценивать вероятность пересечения зависимых множеств, не вдаваясь в сложную природу их зависимости.

Приложение 3. Теория дискретного поиска*.)

П.3.1. Постановка задачи и общие соотношения. До сих пор имелось лишь два общих результата теории дискретного поиска, полученные методами динамического программирования (см. далее). Развита в этом приложении элементарными методами теория [7] приводит к общим соотношениям, из которых помимо упомянутых следуют и новые результаты, использовавшиеся в предыдущем изложении.

*) Здесь используется термин «дискретный поиск» для того, чтобы отличить развиваемую теорию от теории непрерывного поиска Купмана [51], связанную с оптимальным распределением поисковых усилий.

Теория дискретного поиска, как это будет видно далее, занимает промежуточное положение между теорией обнаружения сигналов на фоне шумов и теоретико-игровой моделью преследования (в последней объект может прятаться (убегать) активно, здесь — пассивно). «Выходные» параметры теории обнаружения сигналов на фоне шумов (теории выбора между гипотезами), такие, как вероятность пропуска сигнала β и среднее время его обнаружения $E\tau$, здесь являются исходными параметрами. Следует заметить, что оптимизация поисковых процедур в этом приложении проводится для достаточно широкого класса. Некоторое расширение этого класса в работе [52] не приводит к поисковым процедурам, существенно отличающимся от полученных для рассматриваемого класса (см. далее).

Рассмотрим дискретные моменты времени $t=1, 2, \dots$, в которые происходит поиск одного объекта, находящегося в одном из n мест. Обследование i -го места с целью выявления наличия или отсутствия в нем объекта в общем случае требует случайное время τ_i также дискретно в рассматриваемой нами схеме. При этом возможен пропуск места с объектом с вероятностью β .

Будем считать, что в любой момент времени независимо от других моментов поисковая процедура может быть прервана с вероятностью $p > 0$ (не прерывается с вероятностью $q=1-p$). Такого рода поиск будем называть *p-напряженным поиском с β -пропуском* или просто *p-напряженным поиском*.

Если поиск не прерывается до своего успешного окончания (объект находится), то будем говорить, что он успел закончиться. В противном случае будем говорить, что он не успел закончиться (он прерван до отыскания объекта).

В частности 0-напряженный поиск всегда успевает закончиться, а 1-напряженный поиск не успевает закончиться не начинаясь.

Приведенные определения охватывают случай поиска одного разлаживающегося элемента из n элементов системы. Прерывание поисковой процедуры здесь можно интерпретировать вызовом разлаживающегося элемента (в режиме работы) или полным выходом разлаживающегося элемента из строя (в режиме готовности).

Это же определение охватывает случай поиска объекта, который может прервать поиск своими активными действиями.

Во всех этих случаях цель поиска состоит в обнаружении объекта до прерывания поиска, которое мы будем называть вызовом объекта.

Связь вероятностей успешного окончания p -напряженного и 0-напряженного поиска содержит следующая

Лемма П.3. Вероятность $p_s(q)$ того, что p -напряженный поиск успеет закончиться в случайный момент времени $\sigma(q)=s$, имеет вид:

$$p_s(q) = p_s(1) q^s / g_{\sigma(1)}(q). \quad (\text{П.3.1})$$

Производящая и среднее значение случайной величины $\sigma(q)$ имеют вид:

$$g_{\sigma(q)}(u) = \sum_{s=1}^{\infty} p_s(q) u^s = \frac{g_{\sigma(1)}(qu)}{g_{\sigma(1)}(q)} \quad (\text{П.3.2})$$

и

$$E\sigma(q) = q'_{\sigma(q)}(1) = q \frac{g'_{\sigma(1)}(q)}{g_{\sigma(1)}(q)}. \quad (\text{П.3.3})$$

При этом вероятность $Q=Q(q)$ того, что p -напряженный поиск вообще успеет закончиться (в любой момент времени), имеет вид:

$$Q(q) = g_{\sigma(1)}(q). \quad (\text{П.3.4})$$

Доказательство. Введем случайную величину x , равную времени до прерывания поиска (вызова). Ясно, что x имеет геометрическое распределение

$$\mathcal{P}(x=k) = q^{k-1} p.$$

Отсюда

$$\mathcal{P}(x \geq k) = q^k.$$

Далее имеем тождества

$$\mathcal{P}[\sigma(1)=s, x \geq s] \equiv \mathcal{P}[\sigma(1)=s, x \geq \sigma(1)]$$

или

$$\mathcal{P}[\sigma(1)=s] \mathcal{P}(x \geq s) \equiv \mathcal{P}[\sigma(1)=s/x \geq \sigma(1)] \mathcal{P}(x \geq \sigma(1)).$$

Но $\mathcal{P}(x \geq s) = q^s$ и по определению

$$\mathcal{P}[\sigma(1)=s/x \geq \sigma(1)] = \mathcal{P}[\sigma(q)=s] = p_s(q).$$

Поэтому из двух предыдущих соотношений имеем

$$p_s(q) = \mathcal{P}[\sigma(q)=s] = \frac{\mathcal{P}[\sigma(1)=s] \mathcal{P}(x \geq s)}{\mathcal{P}[x \geq \sigma(1)]} = \frac{p_s(1) q^s}{\mathcal{P}[x \geq \sigma(1)]}. \quad (\text{П.3.5})$$

Но

$$\mathcal{P}[x \geq \sigma(1)] = \sum_{s=1}^{\infty} \mathcal{P}[\sigma(1)=s] \mathcal{P}(x \geq s) = \sum_{s=1}^{\infty} \mathcal{P}[\sigma(1)=s] q^s = g_{\sigma(1)}(q).$$

Отсюда следуют соотношения (П.3.1) — (П.3.3) леммы.

П.3.2. Оптимизация поисковых процедур. Доказанная лемма касалась общего распределения $p_s(q)$ окончания поиска или другой произвольной процедуры. Рассмотрим теперь поисковую процедуру S , связанную с поиском одного объекта, расположенного в одном из n мест.

Пусть имеет место p -напряженный поиск с β -пропуском.

При этом с каждым местом i связано случайное время τ_i его обследования, с вероятностью β приводящее к пропуску объекта и с вероятностью $1-\alpha=1$ устанавливающее отсутствие объекта в i -м месте. Рассмотрение общего случая $1-\alpha < 1$ приводит к чрезвычайно аналитическим трудностям. В задачах обнаружения сигналов на фоне шумов допущение $\alpha \approx 0$ часто бывает оправданным.

Будем полагать^{*}, что в случае пропуска объекта при опреде-

^{*} В работе [52] предварительно не делается такого допущения, но это следует в асимптотическом случае $t \rightarrow \infty$.

ленном порядке осмотра n мест (ни на одном из них объект не найден) процедура поиска возобновляется в том же порядке (циклический поиск).

В связи с этим случайное время $\tau(1)$ успешного окончания 0-напряженного поиска с β -пропуском имеет следующее основное представление:

$$\tau(1) = \sum_{j=1}^{\mu-1} \sum_{i=1}^n \tau_i^j + \sum_{i=1}^{\kappa} \tau_i^{\mu}, \quad (\text{П.3.6})$$

где τ_i^j — независимые и одинаково с τ_i распределенные случайные величины с производящими функциями $g_{\tau_i}(u)$ соответственно, μ и κ — независимые от них и между собой случайные величины с производящими функциями:

$$g_{\mu}(u) = \frac{(1-\beta)u}{1-\beta u}$$

(производящая функция геометрического распределения) и

$$g_{\kappa}(u) = \sum_{i=1}^n p_i u^i \quad (\text{П.3.7})$$

соответственно, где p_i — априорная вероятность того, что объект находится в i -м месте.

Заметим, что, как легко показать,

$$g_{\mu-1}(u) = \frac{1}{u} g_{\mu}(u) = \frac{1-\beta}{1-\beta u}. \quad (\text{П.3.8})$$

Из (П.3.6), (П.3.7) и (П.3.8), используя свойства мультипликативности (П.1.42) и итерации (П.1.48), имеем для производящей функции $g_{\tau(1)}(u)$ выражение

$$g_{\tau(1)}(u) = g_{\mu-1} \left[\prod_{i=1}^n g_{\tau_i}(u) \right] \sum_{i=1}^n p_i \prod_{l=1}^i g_{\tau_l}(u). \quad (\text{П.3.9})$$

В частности, для одинаково распределенных случайных величин τ_i имеем

$$g_{\tau(1)}(u) = g_{\mu-1} [g_{\tau}^n(u)] g_{\kappa} [g_{\tau}(u)]. \quad (\text{П.3.10})$$

Из соотношения (П.3.4) леммы П.3 и (П.3.9) следует выражение для вероятности $Q(q)$ успешного p -напряженного поиска с β -пропуском

$$Q(q) = g_{\tau(1)}(q) = g_{\mu-1} \left[\prod_{i=1}^n g_{\tau_i}(q) \right] \sum_{i=1}^n p_i \prod_{l=1}^i g_{\tau_l}(q) \quad (\text{П.3.11})$$

Заметим, что от стратегии S поиска (порядка обхода мест) зависит лишь второй множитель соотношения (П.3.11).

Используя это обстоятельство, взяв разность двух значений $Q(q)$, разнящаяся транспозицией мест обхода, может быть легко доказана

Теорема П.1. Оптимальная [максимизирующая $\bar{Q}(q)$] стратегия $S'(q)$ обхода мест в p -напряженном поиске с β -пропуском должна производиться так, чтобы

$$\frac{p_1 g_{\tau_1}(q)}{1-g_{\tau_1}(q)} \geq \frac{p_2 g_{\tau_2}(q)}{1-g_{\tau_2}(q)} \geq \dots \geq \frac{p_n g_{\tau_n}(q)}{1-g_{\tau_n}(q)}, \quad (\text{П.3.12})$$

при этом

$$Q(q) = \bar{Q}(q) = \max_{S(q)} g_{\tau(1)}(q) \quad \text{и} \quad E\tau(q) = q \frac{g'_{\tau(1)}(q)}{g_{\tau(1)}(q)}. \quad (\text{П.3.13})$$

Доказательство. В самом деле, сравним две стратегии обхода $S(q)$ и $S'(q)$.

Пусть места обхода пронумерованы в порядке их обхода согласно стратегии $S(q)$ и вторая стратегия $S'(q)$ разнится от первой лишь тем, что вместо перехода от осмотра i -го места к $i+1$ -му в ней сначала осматривается $i+1$ -е место и потом i -е (разнится в одной транспозиции).

Тогда согласно (П.3.11) имеем для $S(q)$

$$Q(q) = g_{\tau(1)}(q) = g_{\mu-1} \left[\prod_{i=1}^n g_{\tau_i(1)}(q) \right] \left[\sum_{u \neq i, i+1} p_u \prod_{l=1}^u g_{\tau_l}(q) + p_i \prod_{l=1}^i g_{\tau_l}(q) + p_{i+1} \prod_{l=1}^{i+1} g_{\tau_l}(q) \right] \quad (\text{П.3.14})$$

и для $S'(q)$

$$Q'(q) = g_{\tau'(1)}(q) = g_{\mu-1} \left[\prod_{i=1}^n g_{\tau_i}(q) \right] \left[\sum_{u \neq i, i+1} p_u \prod_{l=1}^u g_{\tau_l}(q) + p_i \prod_{l=1}^{i-1} g_{\tau_l}(q) g_{\tau_{i+1}}(q) + p_{i+1} \prod_{l=1}^i g_{\tau_l}(q) g_{\tau_i}(q) g_{\tau_{i+1}}(q) \right]. \quad (\text{П.3.15})$$

Отсюда, вычитая (П.3.15) из (П.3.14), получим

$$Q(q) - Q'(q) = g_{\mu-1} \left(\prod_{i=1}^n g_{\tau_i}(q) \right) \prod_{l=1}^{i-1} g_{\tau_l}(q) [p_i g_{\tau_i}(q) + p_{i+1} g_{\tau_i}(q) g_{\tau_{i+1}}(q) - p_{i+1} g_{\tau_{i+1}}(q) - p_i g_{\tau_i}(q) g_{\tau_{i+1}}(q)]. \quad (\text{П.3.16})$$

Но неотрицательное значение $Q(q) - Q'(q)$ совпадает с неотрицательным значением квадратной скобки соотношения (П.3.16), т. е.

$$p_i g_{\tau_i}(q) [1 - g_{\tau_{i+1}}(q)] - p_{i+1} g_{\tau_{i+1}}(q) [1 - g_{\tau_i}(q)] \geq 0.$$

Отсюда для максималности $Q(q)$ достаточно требовать выполнения неравенства

$$\frac{p_i g_{\tau_i}(q)}{1 - g_{\tau_i}(q)} \geq \frac{p_{i+1} g_{\tau_{i+1}}(q)}{1 - g_{\tau_{i+1}}(q)}. \quad (\text{П.3.17})$$

Таким образом, оптимальный обход $S(q)$ должен для любых i удовлетворять соотношениям (П.3.17), в этом случае соответствующее $Q(q)$ будет не меньше $Q'(q)$, соответствующего любому другому из $n!$ других обходов $S'(q)$, который может быть представлен в виде последовательности транспозиции. Этим теорема полностью доказана.

Следствие П.1. Для случая 0-напряженного поиска с β -пропуском оптимальная в смысле вероятности $Q(q) \rightarrow Q(1) = 1$ стратегия $S(q)$ обхода переходит в оптимальную по среднему времени поиска $E\tau(q) \rightarrow E\tau(1)$ [минимизируется $E\tau(1)$] стратегию $S(1)$ обхода, которая должна проводиться так, чтобы

$$\frac{p_1}{E\tau_1} \geq \frac{p_2}{E\tau_2} \geq \dots \geq \frac{p_n}{E\tau_n}, \quad (\text{П.3.18})$$

при этом

$$E\tau(1) = \min_{S(1)} E\tau(1) = \frac{\beta}{1-\beta} \sum_{i=1}^n E\tau_i + \sum_{i=1}^n p_i \sum_{l=1}^i E\tau_l. \quad (\text{П.3.19})$$

Из соображений непрерывности при $q \rightarrow 1$ критерии оптимальности, максимизирующие $Q(q)$ и минимизирующие $E\tau(q)$, приводят к одной и той же стратегии обхода $S(1)$.

Доказательство. В самом деле, при $u \rightarrow 1$ имеем

$$\begin{aligned} g_{\xi}(u) &= g_{\xi}(1) - (1-u)g'_{\xi}(1) + O((1-u)^2) = \\ &= 1 - (1-u)E\xi + O((1-u)^2) = \exp[-(1-u)E\xi] + O((1-u)^2). \end{aligned} \quad (\text{П.3.20})$$

Отсюда при $q \rightarrow 1$ имеем

$$Q(q) = g_{\tau(1)}(q) = 1 - pE\tau(1) + O(p^2) = \exp[-pE\tau(1)] + O(p^2) \quad (\text{П.3.20'}$$

и ясно, что с ростом $Q(q)$ уменьшается $E\tau(1)$ и этим доказываем последнее утверждение следствия.

Далее, при $q \rightarrow 1$ ($p \rightarrow 0$) имеем

$$\frac{1 - g_{\tau}(1-p)}{p} \rightarrow g'_{\tau}(1) = E\tau.$$

Поэтому, помножив все члены цепочки неравенств (П.3.12) на p и переходя к пределу при $p \rightarrow 0$, получим неравенство (П.3.18). Соотношение (П.3.19) можно получить из соотношения (П.3.13) или непосредственно из соотношения (П.3.6). В свою очередь, соотношение (П.3.18) можно получить непосредственно из соотношения (П.3.19) тем же методом, что и более общее соотношение (П.3.12).

Следствие П.2 (концевой эффект). В случае 0-напряженного поиска с 0-пропуском поиск следует ограничить $n-1$ -местом, в ре-

зультате чего соотношение (П.3.18) переходит в соотношение

$$\frac{p_1}{E\tau_1} \geq \frac{p_2}{E\tau_2} \geq \dots \geq \frac{p_{n-2}}{E\tau_{n-2}} \geq \max\left(\frac{p_{n-1}}{E\tau_{n-1}}, \frac{p_n}{E\tau_n}\right), \quad (\text{П.3.21})$$

причем два последних места выбираются так, чтобы $E\tau_{n-1} \leq E\tau_n$ и при $n > 2$ соотношение (П.3.19) переходит в

$$\min_{S(1)} E\tau(1) = \sum_{i=1}^{n-2} p_i \sum_{l=1}^i E\tau_l + (p_{n-1} + p_n) \sum_{l=1}^{n-1} E\tau_l. \quad (\text{П.3.22})$$

Доказательство. Очевидно, что если пропуск объекта невозможен, то достаточно просмотреть лишь $(n-1)$ -но место, так как если объект при этом не обнаружен, то с вероятностью 1 он находится на n -м месте. Предпоследнее для осмотра место, очевидно, нужно выбрать для минимизации $E\tau(1)$ таким, чтобы $E\tau_{n-1} \leq E\tau_n$. При этом соотношение (П.3.22) очевидно.

Для случая геометрически распределенного времени τ_i при $\beta = 0$ соотношения (П.3.21) и (П.3.22) содержатся в [6].

Более общая постановка задачи дискретного поиска содержится, как уже отмечалось, в работе [52]. В ней учитывается перераспределение апостериорных вероятностей в ходе поиска. Следует заметить, что стратегия $S(q)$ здесь после некоторого нестационарного начала делается циклической (совпадающей с указанной выше). За счет нестационарного начала среднее время поиска может быть несколько снижено по сравнению с рассмотренным случаем. Однако сам алгоритм поиска не может быть задан как здесь в явной форме. Кроме того, при $\beta \rightarrow 1$, когда среднее время поиска асимптотически стремится к бесконечности, оба оптимальных алгоритма поиска асимптотически эквивалентны.

Некоторые из приведенных в этом приложении задач решены в более общей постановке в прекрасной монографии [53].

ЛИТЕРАТУРА

1. Котельников В. А. Теория потенциальной помехоустойчивости. Госэнергоиздат, 1956.
2. Глушков В. М. Введение в кибернетику. Киев, 1964.
3. Черчмен У., Акоф Р., Арнов Л. Введение в исследование операций. Изд-во «Наука», 1968.
4. Рейтман У. Познание и мышление. Моделирование на уровне информационных процессов. Изд-во «Мир», 1968.
5. Эшби У. Р. Введение в кибернетику. Изд-во иностранной литературы, 1959.
6. Беллман Р. Динамическое программирование. Изд-во иностранной литературы, 1960.
7. Флейшман Б. С. Статистические пределы эффективности сложных систем. В сб. «Прикладные задачи технической кибернетики». Изд-во «Советское радио», 1966.
8. Флейшман Б. С. Статистический последовательный анализ и автоматический самоконтроль. Ученые записки по статистике, т. 8. Изд-во «Наука», 1964.
9. Флейшман Б. С. Об оптимальной структуре сложных решающих систем. «Известия АН СССР», «Техническая кибернетика», № 3, 1963.
10. Рудерман С. Ю. Вопросы надежности и поиска неисправностей в системах с учетом вероятностного режима использования элементов. «Известия АН СССР», «Техническая кибернетика», № 6, 1963.
11. Рудерман С. Ю. Надежность систем при случайных режимах использования их элементов. В сб. «Анализ и синтез систем автоматического управления». Изд-во «Наука», 1968.
12. Шеннон К. Математическая теория связи. Работы по теории информации и кибернетике. Изд-во иностранной литературы, 1963.
13. Питерсон У. Коды, исправляющие ошибки. Изд-во «Мир», 1965.
14. Фано Р. Передача информации. Статистическая теория связи. Изд-во «Мир», 1965.
15. Флейшман Б. С. Конструктивные методы оптимального кодирования для каналов с шумами. Изд-во АН СССР, 1963.
16. Бриллюэн. Наука и теория информации. Изд-во физмат. литературы, 1960.
17. Мак Кинси Дж. Введение в теорию игр. Изд-во физмат. литературы, 1960.
18. Флейшман Б. С., Крапивин В. Ф. Регулярный метод решения игр с кусочно-постоянной функцией выигрыша. «Известия АН СССР», «Техническая кибернетика», № 3, 1965.
19. Лефевр В. А. Конфликтующие структуры. Изд-во «Высшая школа», 1967.
20. Розенблат Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. Изд-во «Мир», 1965.
21. Буш Р. и Мостеллер Ф. Стохастические модели обучаемости. Изд-во физмат. литературы, 1962.
22. Стратанович Р. Л. Существует ли теория синтеза оптимальных, адаптивных, самоорганизующихся и самонастраивающихся систем? «Автоматика и телемеханика», т. 29, № 1, 1968.
23. Данков П. П. Оптимальные процедуры обучения принятию статистических решений: «Радиотехника и электроника», т. 10, № 10, 1965.
24. Нильсон Н. Обучающиеся машины. Изд-во «Мир», 1967.
25. Виноград С., Коуэн Дж. Надежные вычисления при наличии шумов. Изд-во «Наука», 1968.
26. «Системы обработки и передачи информации». Сб. статей под ред. проф. Н. А. Железнова. ЛИАП, Труды, вып. 48, 1966.
27. Крапивин В. Ф. Регулирование объектов, подверженных случайным возмущениям. В сб. «Прикладные задачи технической кибернетики». Изд-во «Советское радио», 1966.
28. Флейшман Б. С. О живучести сложных систем. «Известия АН СССР», «Техническая кибернетика», № 5, 1966.
29. Флейшман Б. С., Крапивин В. Ф. Примечание теоретико-игровых методов к оценке живучести сложных систем. «Известия АН СССР», «Техническая кибернетика», № 6, 1967.
30. Эшби У. Р. Несколько замечаний. В сб. статей «Общая теория систем». Изд-во «Мир», 1966.
31. Пирс У. Построение надежных вычислительных машин. Изд-во «Мир», 1968.
32. Кульбак С. Теория информации и статистика. Изд-во «Наука», 1967.
33. Хазен Э. М. Методы оптимальных статистических решений и задачи оптимального управления. Изд-во «Советское радио», 1968.
34. Флейшман Б. С., Крапивин В. Ф. О процедуре выбора многомерного параметра на цифровой вычислительной машине. «Известия АН СССР», «Техническая кибернетика», № 2, 1965.
35. Флейшман Б. С. Параллельное декодирование для дискретного канала с шумами. «Радиотехника и электроника», т. 10, вып. 5, 1965.
36. Возенкрафт Дж. Рейффен В. Последовательное декодирование. Изд-во иностранной литературы, 1963.
37. Флейшман Б. С. О переходном режиме загрузки памяти кибернетического устройства воспроизведения образов. «Радиотехника и электроника», т. 6, вып. 7, 1961.
38. Зятицкий В. А. Моделирование загрузки памяти методом самоанализа. «Радиотехника и электроника», т. 6, вып. 7, 1961.
39. Скловский Н. Б. Применение вычислительной машины с памятью для обнаружения радиолокационного сигнала методом статистического последовательного анализа. «Радиотехника и электроника», т. 9, вып. 9, 1964.
40. «Теоретическая и математическая биология». Изд-во «Мир», 1968.
41. Шмальгаузен И. И. Кибернетические вопросы биологии. Изд-во «Наука». Сибирское отделение, 1968.
42. Гробстайн К. Стратегия жизни. Изд-во «Мир», 1968.
43. Беклемишев В. Н. Об общих принципах организации жизни. Бюллетень МОИП, сер. биологическая, т. 69, 1964.
44. Гаузе Г. Математическая теория борьбы за существование и ее применение к популяции дрожжевых клеток. Бюллетень МОИП, отд. биол., т. 43 (1), 69, 1934.
45. Ферстер Г. Биологика. В сб. «Проблемы бионики». Изд-во «Мир», 1965.

46. Фогель Л., Оуэнс А., Уолш М. Искусственный интеллект и эволюционное моделирование. Изд-во «Мир», 1969.
47. Кулагина О. С., Ляпунов А. А. К вопросу о моделировании эволюционного процесса. В сб. «Проблемы кибернетики», вып. 16. Изд-во «Наука», 1966.
48. Булгакова Т. И., Кулагина О. С., Ляпунов А. А. К вопросу о моделировании эволюционного процесса с учетом отбора I. В сб. «Проблемы кибернетики», вып. 20. Изд-во «Наука», 1968.
49. Харди Г. Г., Литтлвуд Д. Е., Полиа Г. Неравенства. Изд-во иностранной литературы, 1948.
50. Феллер В. Введение в теорию вероятностей и ее приложения, Ч. I, 2-е изд. Изд-во «Мир», 1964.
51. Koopman. Theory of the search. Operation Research, v. 5, № 5, 1957.
52. Старверов О. В. Об одной задаче поиска. Теория вероятностей и ее применения, т. 8, вып. 2, 1963.
53. Кузнецов П. И., Пчелинцев Л. А., Гайденок В. С. Контроль и поиск неисправностей в сложных системах. Изд-во «Советское радио», 1969.
54. Флейшман Б. С. Теория потенциальной эффективности сложных систем. Итоги и проблемы. «Известия АН СССР», Техническая кибернетика, № 4, 1970.
55. B. S. Fleishman. The State of Theory of Potential Efficiency of Complex Systems. Cybernetica, № 4, 1970.

АЛФАВИТНЫЙ УКАЗАТЕЛЬ

- Беллман Р. 90, 91, 101
фон Бергаланфи 3
Бинарная игра 98
Бинарный символ 57, 74
— симметричный канал (БСК) 57
— стирающий канал (БСтК) 73
Биномальное распределение 12, 74
Биологическая система (биосистема) 194
Большая система 6
Большие уклонения 12, 200
Быстродействие 165
Булль 151, 157, 201
Варшавов 64
Величина игры 86
— — верхняя 85
— — нижняя 85
Вероятность больших уклонений 11
Вероятность осуществления цели 148
Винер Н. 3
Входное слово 57
- Гаузе Г. 195
Гельдер 181, 203
Гильберт 64
Данков П. П. 109
Декодирование 58
— в целом 61
— паразитное 170
— посимвольное 58
— мажоритарное 58
— оптимальное 78
— хэмминговское 61
Дихотомия 110, 115
Живучесть 120
Жизнеспособность 24, 165
Загрузка памяти 173
Закон конкурентного вытеснения 195
Игра 83
— безобидная 87
— бинарная 98
Игра дискретная (матричная) 84
— — симметричная 93
— — многошаговая 83, 88

- пульсумная 84
— преследования 99
— рефлексивная 105
— тринарная 92
Игры непрерывные 84, 95
Избыточность
Индивидуальная ситуация 9
Канал
— бинарный симметричный (БСК) 57
— — стирающий (БСтК) 73
— гауссовский
Качества системы смешанные 118
—, R — качество (надежность) 16
—, I — качество (помехоустойчивость) 118
—, C — качество (управляемость) 118
—, L — качество (самоорганизация) 11, 107
—, RC — качество (живучесть) 120
Кибернетическая оценка параметра 193
Классификация 116
Ковер 117
Код 57
Код идеальный 64
— оптимальный 72
— случайный 72
— плотноупакованный 73
Кодирование 58
— случайное 66
Конкурентное вытеснение 195
Конфликтная ситуация 99
Концевой эффект 218
Константа Бреммермана 147
— фундаментальная 200
Корректирующая способность кода 62
Котельников В. А. 3, 56, 105
Коэффициент живучести 123
— жизни
— неустойчивости 139
— смерти системы 22, 26
— устойчивости
— эффективности 192
— — внешней 192
— — внутренней 192
Крапивин В. Ф. 92, 119
Купман 213
Левин Б. Р. 108
Локальная функция выигрыша 88
Макробиология 194
Максимальная лемма 202
Мельников Ю. Н. 119
Минимакс 90
Минимаксные уравнения 89, 149
Множество жизни системы 19
— смерти системы 20
Модель 8, 120
Моделирование на ЭВМ 148
— эволюционное 199
Надежность (вероятность) 16—18
— (R — качество) 13, 16
— системы в целом 23, 53
фон Нейман 3, 87, 101
Неравенство Буля 14, 212, 213
— Гельдера 203
— Минковского 203
— Чебышева 72
— Чернова 207
Нижняя и верхняя величина игры 85
Обмен веществ 17
— (u, v) — обмен 10
Обучение различению сигналов 56, 108
Объем памяти системы 171, 185, 190, 194, 147
Опасность отказа 31
Оптимизация поведения 108
— структуры 47, 129
Отказ системы 18
— — физический 34
— — функциональный 35, 38
— элемента 18, 33
Открытость системы 7
Осуществимость цели 149
Оценка больших уклонений
— Варшавова—Гильберта 64
— на основе формулы Стирлинга — 207
— — Шеннона 208
— — эффективности системы 157, 192
Параметр эффективности 153
— — внешней 151
— — внутренней 152
Пирс 163
Плоткин 64
Поведение 5
— оптимальное 108, 124
Подход аналитический 8
— поведенческий 3
— синтетический 8
— структурный 3
Поиск p -напряженный 38, 214
— неисправностей 189
Поиска теория (см. теория дискретного поиска)
Помехоустойчивость (вероятность) 56
— (I — качество) 56
— потенциальная 56
Пороги осуществимости 149
Правило среднего гармонического 157
Предельный закон живучести 139
Предельный закон жизнеспособности 164
— — классификация 116
— — надежности 26
— — помехоустойчивости 66
— — управляемости 101
— — самоорганизации (классификация) 116
— — эффективности сложных систем 200
Проблема «сторожей сторожей» 235
Программирование динамическое 41
— эвристическое 3, 199
Производительность 202, 210
«Проклятие» многомерности 9, 195, 200
Прочность 17
Различие сигналов на фоне шумов 108
Расположение общего типа
Расогласование между траекториями биосистем 195
Расстояние Хэмминга 167, 109, 64, 60
Режим готовности 184
— работы 184

Рефлексивные игры 105
 Решение игры 88
Рудерман С. Ю. 41

Самоконтроль 183
 Самоорганизация (*L* — качество) 107
 Свойство неограниченного расширения 24
 Свойство производящих 210
 — итерации 211
 — мультипликативности 211
 Седловая точка 86
 Сечения системы 105
 — *X* — сечения 13
 — *XY* — сечения 13
 Сигнал 190
 Символ бинарный 57, 74
 Система 5
 — бессмертная 30
 — биологическая 194, 17
 — большая 6
 —, взаимодействие со средой 122
 — живая 19
 — жизнеспособная 21
 — идеальная 23, 24
 — нежизнеспособная 21
 — неоптимальная 58
 — неуязвимая
 — оптимальная 8, 37, 149, 200
 — простая 6, 199
 — автоматическая 5
 — реальная 5
 — биологическая 17, 194
 — техническая 7, 9
 — симметричная 21
 — сложная 5, 6, 199
 — преобразующаяся 6
 — предвидящая 6
 — решающая 5, 57
 — самоорганизующаяся 5
 — смертная 29
 — уязвимая 139
 Ситуация дифференциальная 9
 — конфликтная 9
 Скорость передачи 79
 Смешанные *XY* — качества 118
 — стратегии 87
 Сообщения 57
 Среда 9, 121
 Средняя вероятность жизни элемента 22
 Среднее время поиска неисправности 47
 Стратегия оптимальная 85, 87
 — смешанная 87
 — чистая 84

Структура иерархическая 36, 47
 — оптимальная 48, 58
 — подоптимальная

Теорема о минимаксе 87
 Теория дискретного поиска 213
 — игра 104
 — информации 78
 — надежности 3
 — осуществимости 147
 — потенциальной помехоустойчивости 3
 — эффективности 3
 Техническая система (техсистема) 7, 9

Управляемость (*C* — качество) 82
 104
 Устойчивость 6, 16, 56
Ферстер Г. 197
Флейшман Б. С. 12
 Формальный аргумент 6
 Фундаментальная константа 104, 200
 Функция выигрыша 84
 — локальная 88
 — тотальная 88
 — тринарная 92
Хэмминг 61

Цель 57
 — стратегическая 199
 — тактическая 199
 Целевой функционал 8

Чебышев П. Л. 72, 73
Чернов 207
Шеннон К. 3, 61, 67, 78, 105, 107, 119, 170, 193, 208
 Шум 56
 — гауссовский 56

Эволюция прогрессивная 195
 Эволюционное моделирование 199
 Эвристическое программирование 3, 199
 Экология 194
 Экспоненциальный закон 150, 33, 31
 Экспоненциальные трудности 78, 166
 Элемент 120
 — *a* — элемент 121
 — *c* — элемент 121
 — *R* — элемент 121
 Энтропия 78
Эфрон 117
 Эффект концевой 218
 Эффективность 11, 56, 110
Эшби У. Р. 165, 147

ОГЛАВЛЕНИЕ

Предисловие	3
1. Введение	5
§ 1.1. Системы	5
§ 1.2. Два подхода к построению моделей	7
§ 1.3. Оптимальность в терминах (<i>u</i> , <i>v</i>)-обмена	9
§ 1.4. Проблема части и целого	13
§ 1.5. Законы сложных систем	14
2. Надежность (<i>R</i>-качество)	16
§ 2.1. Вводные замечания	16
§ 2.2. Идеальная система	18
§ 2.3. Предельный закон надежности	26
§ 2.4. Иерархическая структура поиска неисправностей	31
§ 2.5. Оптимизация иерархической структуры поиска неисправностей	38
§ 2.6. Надежность системы в целом. Примеры	53
3. Помехоустойчивость (<i>I</i>-качество)	56
§ 3.1. Вводные замечания	56
§ 3.2. Неоптимальные системы	58
§ 3.3. Предельный закон помехоустойчивости	66
§ 3.4. Теория информации. Примеры	78
4. Управляемость (<i>C</i>-качество)	82
§ 4.1. Вводные замечания	82
§ 4.2. Одношаговые и многошаговые игры	83
§ 4.3. Простейшие симметричные игры	92
§ 4.4. Предельный закон управляемости	101
§ 4.5. Теория игр. Примеры	104
5. Самоорганизация (<i>L</i>-качество)	107
§ 5.1. Вводные замечания	107
§ 5.2. Обучение различению сигналов на фоне шумов	108
§ 5.3. Предельное число дихотомий	110
§ 5.4. Предельный закон классификации. Примеры	116
6. Смешанные <i>RI</i>-, <i>RC</i>- и <i>IC</i>-качества	118
§ 6.1. Вводные замечания	118
§ 6.2. Понятие живучести. Модель системы	120
§ 6.3. Оптимизация поведения среды	124
§ 6.4. Оптимизация структуры системы	129
§ 6.5. Предельный закон живучести	139
7. Теория осуществимости	147
§ 7.1. Вводные замечания	147
§ 7.2. Понятие осуществимости. Основные соотношения в экспоненциальном случае	148
§ 7.3. Оценка эффективности системы по эффективности ее <i>X</i> -сечений. Правило среднего гармонического	157
§ 7.4. Учет ограниченности быстродействия системы	165
§ 7.5. Учет ограниченности объема памяти системы	171
§ 7.6. Применения	191
Заключение	199
Приложения	202
Литература	220
Алфавитный указатель	223