Глава 2. Два подхода к моделированию и прогноз

Приступая к созданию модели, очень важно определиться *с претензиями на ее прогностические возможности*. Сделанный выбор определяет вид требуемого математического аппарата. Если претензии на точный однозначный прогноз будущих состояний по начальному отсутствуют, традиционно используют *вероятностный подход*. При этом все или часть величин, описывающих объект моделирования, *объявляются случайными*, т.е. принципиально непредсказуемыми, стохастическими[1]. Для такого «приговора» могут быть различные основания (см. п. 2.2), но если он «подписан», используется аппарат теории вероятностей и математической статистики. При этом, характеризуя связь между некоторым условием S и событием A, говорят лишь об определенной вероятности P наступления события A при наступлении события S – об условной вероятности P(A|S).

Альтернативный вероятностному *динамический* подход опирается на концепцию детерминизма. *Детерминизм* – учение о закономерности и причинной обусловленности всех явлений природы и общества. При таком подходе считается, что при каждом наступлении события S (причины) обязательно наступает событие A (строго определенное следствие). Наиболее ярким «проповедником» детерминизма считается известный французский астроном, математик и физик Пьер Симон де Лаплас (1749-1827). В своих научных воззрениях он проявил твердость, удивительную при известной его непоследовательности в житейских привязанностях[2] [118, с.711]. Это Лаплас заявил Наполеону, что в своей теории о происхождении Солнечной системы он не нуждается в «гипотезе о существовании Бога». Он видел в небесной механике образец окончательной формы научного познания и пытался объяснить весь мир, в том числе физиологические, психические и социальные явления, с точки зрения механистического детерминизма.

Математическую реализацию детерминистического подхода обеспечил аппарат бесконечно малых, появившийся в XVII веке благодаря усилиям Ньютона и Лейбница. В арсенал исследователей вошел мощный

---

[1] Слово «стохастический» происходит от греческого «умеющий угадывать, проницательный», но в настоящее время используется в несколько ином смысле – для обозначения неопределенности, случайности.

[2] Несколько слов о колоритной личности Лапласа. Последовательность его материалистического мировоззрения резко контрастирует с политической неустойчивостью: при всяком политическом перевороте он переходил на строну победивших. Сначала был республиканцем, после прихода к власти Наполеона – министром внутренних дел; затем был назначен членом и вице-председателем сената, при Наполеоне получил титул графа империи, а в 1814 году подал свой голос за низложение Наполеона; после реставрации Бурбонов получил пэрство и титул маркиза.

инструмент описания поведения систем во времени – обыкновенные дифференциальные уравнения (ОДУ). Теорема о существовании и единственности их решения при заданных начальных условиях обеспечило дифференциальным уравнениям статус эталона для математического описания явлений с позиций детерминизма («данному настоящему соответствует одно будущее!»). В настоящее время кроме ОДУ для построения детерминистических моделей широко используют и другие виды математического аппарата (см. главу 3), например, разностные уравнения, дискретные отображения, интегро-дифференциальные уравнения. Все эти модели независимо от конкретного содержания, даже далекого от механики (динамики), часто называют *динамическими*. Вообще, термин «динамический» в настоящее время часто используется в смысле «детерминированный», а не «силовой» или «подвижный, динамичный». В этом смысле оно будет использоваться далее и нами.

2.1. Основные понятия и особенности динамического моделирования

2.1.1. Определение динамической системы

Итак, в основе детерминированного описания лежит представление о том, что все будущее поведение объекта однозначно определяется его *состоянием* ($\mathbf{x}$) в начальный момент времени. Правило, которым однозначно определяется эволюция объекта из этого состояния, называют *оператором эволюции*[3]. *Состоянием* (вектором состояния) называют совокупность D величин: $\mathbf{x} = (x_1, x_2, ..., x_D)$, где D – размерность описания. Величины x_k называются *динамическими переменными*. Мы ввели понятие состояния на конечномерном примере (D – конечное число), но возможны и *бесконечномерные* модели, например, когда состояние представляет собой пространственное распределение какой-либо величины, т.е. является произвольной гладкой функцией пространственной координаты.[4]

Оператор эволюции Φ_t позволяет по начальному состоянию $\mathbf{x}(t_0)$ определить состояние объекта в любой последующий момент времени

[3] «Оператор – то же, что отображение: закон, по которому элементу x некоторого заданного множества X сопоставляется однозначно определенный элемент y другого заданного множества Y. Термин часто употребляется в функциональном анализе и линейной алгебре, в особенности для отображений векторных пространств… Пример: оператор дифференцирования сопоставляет каждой дифференцируемой функции ее производную» [118].

[4] В этом случае оно тоже называется вектором состояния. Термин «вектор» понимается в обобщенном смысле как элемент некоторого *пространства* [110], см. ниже.

$t_0 + t$: $\mathbf{x}(t_0 + t) = \Phi_t(\mathbf{x}(t_0))$. Математически он может быть задан уравнениями, отображениями, матрицами, графами и любыми другими способами (см. главу 3), обеспечивающими выполнение условия *однозначности* прогноза.

Ключевым для детерминистического подхода является понятие *динамической системы* (ДС), которым пользовался еще в начале 20-го века А. Пуанкаре, но содержание которого до сих пор до конца не устоялось. Анализ литературы показывает, что в термин ДС зачастую вкладывается различный смысл, поэтому полезно обсудить его подробнее. Слово «система»[5] используется обычно в традиционном смысле – «совокупность элементов, находящихся в отношениях и связях друг с другом, которая образует определенную целостность» [168, с.610]. Разночтения больше связаны с пониманием смысла слова «динамический» и тем, о каких элементах и системах идет речь – о реальных объектах, о математических конструкциях, или о том и другом одновременно. Некоторыми авторами это понятие даже выводится за рамки детерминизма и сочетается со случайностью. Для иллюстрации приведем фрагменты из книг известных специалистов по этому вопросу.

«Понятие ДС возникло как обобщение понятия механической системы, движение которой описывается дифференциальными уравнениями Ньютона. В своем историческом развитии понятие ДС, как и всякое другое понятие, постепенно изменялось, наполняясь новым, более глубоким содержанием…. В настоящее время понятие ДС является весьма широким. Оно охватывает системы любой природы: физической, химической, биологической, экономической и др., причем *не только детерминированные системы, но и стохастические*[6]. Описание ДС также допускает большое разнообразие: оно может осуществляться или при помощи дифференциальных уравнений, или такими средствами, как функции алгебры логики, графы, марковские цепи и т.д.» [48, с. 8].

«Под ДС мы понимаем *любую систему*, какой бы ни была ее природа, которая может принимать различные математические формы: обыкновенных дифференциальных уравнений (автономных и неавтономных), дифференциальных уравнений в частных производных, отображений прямой или плоскости» [43].

В параграфе «Что такое динамическая система?» книги [115] отмечается: «Вообще говоря, в разных книгах можно найти различные толкования термина ДС, например такие:

---

[5] От греческого «συστημα» – целое, составленное из частей; соединение [168, с.610].

[6] Здесь и далее в цитируемых определениях ДС курсивом мы обозначили, на наш взгляд, узловые моменты.

– это синоним термина «система обыкновенных дифференциальных уравнений» (ОДУ) $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x})$;

– это синоним термина «автономная система обыкновенных дифференциальных уравнений» $d\mathbf{x}/dt = \mathbf{g}(\mathbf{x})$;

– это математическая модель некоторой механической системы.

Мы[7] будем придерживаться точки зрения …, согласно которой понятие ДС является *обобщением понятия автономной системы дифференциальных уравнений* и включает в себя два основных компонента: фазовое пространство **P** (метрическое пространство или многообразие) и однопараметрическую непрерывную или дискретную группу (полугруппу) $\varphi^t(\mathbf{x})$ или $\varphi(\mathbf{x},t)$ его преобразований. Параметр группы … t – это время».

Другое, не менее формализованное, определение: «ДС – это четверка объектов (X, B, μ, A), где X – топологическое пространство или многообразие, т.е. абстрактный образ пространства состояний, B – некоторые интересные подмножества в X, например, замкнутые орбиты или неподвижные точки. Они образуют алгебру в том смысле, что содержат не только отдельные элементы, но и их объединение и пересечения. Они необходимы, чтобы ввести меру, поскольку само X может и не быть измеримым. μ – это мера, например, объем или частота посещения траекторией какой то области. Желательно, чтобы она была эргодической, единственной и инвариантной относительно группы преобразований Φ_t, которая и определяет эволюцию. Иногда добавляют и типичную (в смысле меры μ) начальную точку. Например, для оператора $\Phi_t, t \in Z$: $x_{t+1} = \Phi_1(x_t) \equiv x_t(1 + x_t)$ точка $x_0 = 0$ не является типичной – она не приводит к эволюции» [112].

«Под ДС условились понимать систему любой природы (физическую, химическую, биологическую, социальную, экономическую и т.д.), *состояние которой изменяется* (дискретно или непрерывно) во времени» [70,с.6].

«Абстрагируясь от конкретной физической природы объекта, о нем говорят как о ДС, если *можно указать такой набор величин*, называемых динамическими переменными и *характеризующих состояние системы*, что *их значения в последующий момент времени получаются из исходного набора по определенному правилу*. Это правило задает, как говорят, оператор эволюции системы» [104, с. 7].

«ДС можно представлять как объект любой природы, состояние которого изменяется во времени в соответствии с некоторым динамическим законом, то есть как результат действия

[7] Авторы – Г.Г. Малинецкий, А.Б. Потапов.

детерминированного оператора эволюции. Таким образом, понятие ДС является следствием определенной идеализации, при которой пренебрегают влиянием случайных возмущений, неизбежно присутствующих в любой реальной системе …. Каждой ДС соответствует некоторая математическая модель …» [6, с. 1-2].

«ДС – система, поведение которой задается некоторым набором правил (алгоритмом). Динамическая система *представляет собой лишь модель какой-либо реальной системы*. Любая реальная система подвержена флуктуациям и потому не может быть динамической» [107].

Целям нашей книги наиболее близко последнее определение, которое не вносит трудностей в классификацию возможных ситуаций. Так, многие реальные явления или объекты могут быть достаточно успешно рассмотрены и как случайные, и с помощью «детерминистического» математического аппарата. Иллюстрацией ситуации, когда при моделировании одного и того же объекта при определенных условиях хороши динамические представления, а при других – их использование бессмысленно, является всем известное подбрасывание монеты (см. п. 2.6). Противоречий не возникает, если связывать название ДС с детерминистическими моделями, а применительно к реальным системам воспринимать его как научный жаргон.

Итак, мы будем называть далее «динамической системой» *математическую эволюционную модель*, *для задания которой необходимо:*

1) *определить состояние объекта* $\mathbf{x}$;

2) *задать оператор эволюции* Φ_t, *который позволял бы по начальному состоянию однозначно определять будущие* $\mathbf{x}(t_0 + t) = \Phi_t(\mathbf{x}(t_0))$.

В отношении реальных систем будем понимать термин ДС как сокращенный вариант выражения «система, описание которой возможно и целесообразно с помощью динамических моделей».

2.1.2. Нестрогий пример. Переменные и параметры

Рассмотрим с позиций ДС обычную домашнюю кошку (рис.2.1). Если цель моделирования – описание эволюции здоровья животного, в качестве динамических переменных можно предложить ее массу $M = x_3$, рост $H = x_2$ и густоту шерсти – число «шерстинок» на единицу площади $N = x_1$. Совокупность $\mathbf{x} = (x_1, x_2, x_3)$ – вектор состояния; при этом размерность описания $D = 3$. Мы не можем здесь предложить выражение для оператора эволюции, но очевидно, что развитие кошки будет зависеть, например, от ее питания (a_1 – масса пищи в суточном рационе в кг/день) и условий существования (a_2 – длительности прогулок на свежем воздухе в час/день). В отличие от динамических переменных, величины a_1, a_2 и

подобные им, которые могут поддерживаться постоянными или целенаправленно регулироваться, называют параметрами. Параметрами динамических систем являются коэффициенты в уравнениях эволюции. Совокупность всех параметров системы также называют вектором параметров. Например, в системе, оператор эволюции которой задается уравнением нелинейного осциллятора с кубической нелинейностью:

$$d^2x/dt^2 + 2\delta\, dx/dt + \omega_0^2(bx^3 - x) = 0 \qquad (2.1)$$

параметрами являются коэффициенты δ, ω_0, b, т.е. вектор параметров $\mathbf{a} = (\delta, \omega_0, b)$ трехмерен. Сама система двухмерна D = 2, т.к. для решения уравнения (2.1) необходимо задать начальные значения x и dx/dt.

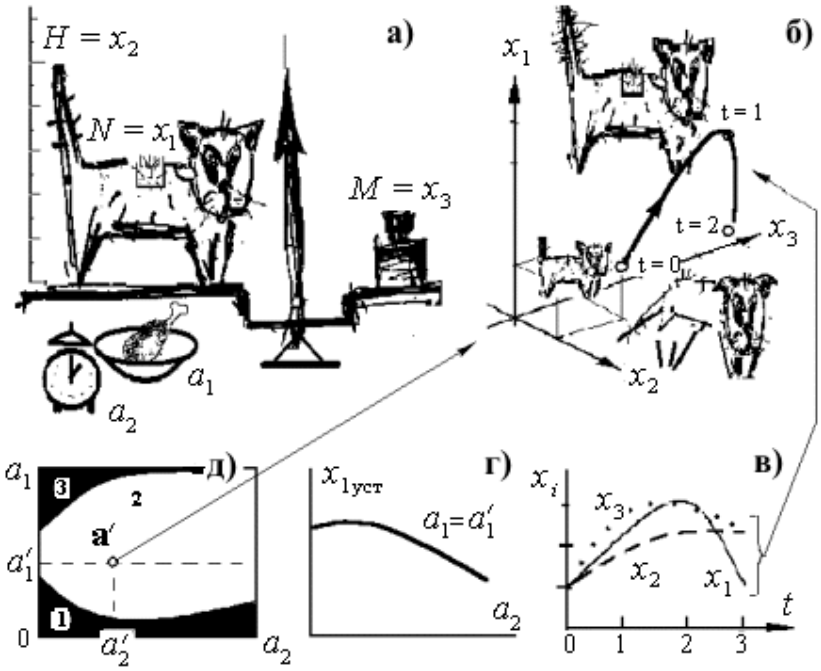

Рис.2.1. Пример с описанием эволюции состояния здоровья кошки: а) переменные и параметры; б) фазовое пространство с выделенной фазовой траекторией при фиксированном значении параметров $\mathbf{a}' = (a_1', a_2')$; в) зависимости проекций вектора состояния на различные оси фазового пространства от времени (временные реализации $x_i(t)$), соответствующие фазовой траектории на рис. б; г) комбинированное пространство параметров и состояний – зависимость установившегося за достаточно длинное время значения x_1 от значения параметра a_2 при фиксированном значении $a_1 = a_1'$; д) пространство параметров, область 2 соответствует нормальным режимам существования кошки, закрашены области параметров, при которых длительное существование невозможно из-за бескормицы (1) или обжорства (3). Точке $\mathbf{a}'$ в пространстве параметров соответствует определенная структура фазового пространства

Разделение характеризующих величин на переменные и параметры диктуется условиями задачи моделирования. Если бы целью моделирования упомянутой кошки было описание ее механического

перемещения в пространстве (а не здоровья, как ранее) разумно было бы выбрать другие переменные. Например, переменными могут быть координаты ее центра масс ($x_1 = x$, $x_2 = y$, $x_3 = z$) и углы поворота продольной оси относительно осей координат ($x_4 = \alpha$, $x_5 = \beta$, $x_6 = \gamma$), а массу животного *M*, его «пушистость» *N* и геометрические размеры *H*, которые не изменятся за время прыжка, но существенно влияют на полет – в качестве параметров ($a_1 = M, a_2 = N, a_3 = H$). Обратите внимание, что величины, при одной цели моделирования служившие динамическими переменными, при другой – стали играть роль параметров.

2.1.3. Фазовое пространство. Консервативные и диссипативные системы. Аттракторы, мультистабильность, бассейны притяжения

Существенным достоинством динамического моделирования является возможность наглядного представления информации, особенно в случаях малой размерности системы *D* и малого числа параметров. Для этого используют формальные *пространства*[8]: *пространство состояний (фазовое пространство*), *пространство параметров* и различные их *гибридные варианты*. В этих случаях на осях координат формального пространства откладываются, соответственно, или значения динамических переменных, или значения параметров, или по одним – параметры, а по другим – переменные.

Состоянию $\mathbf{x}(t)$ в некоторый момент *t* в фазовом пространстве соответствует точка с координатами $x_1(t), x_2(t), x_3(t)$ – *изображающая точка* (она изображает мгновенное состояние). В процессе эволюции изображающая точка с течением времени, смещается вдоль некоторой линии – *фазовой траектории*. Совокупность характерных фазовых

---

[8] «Пространство – логически мыслимая форма (или структура), служащая средой, в которой осуществляются другие формы или конструкции. Например, в элементарной геометрии плоскость или пространство служат средой, в которой строятся разнообразные фигуры. …. В современной математике П. определяют множество каких-либо объектов, которые называют точками…. Отношения между точками определяют «геометрию». Примерами пространств могут служить: 1) метрические пространства…, 2) «пространства событий»…, 3) фазовые пространства. Фазовое пространство физической системы – совокупность всех ее состояний, которые рассматриваются, как точки этого пространства…» [118]. Пространство может быть топологическим (если в нем определено неколичественное понятие «близости»), метрическим (близость определяется «метрикой») и т.д. Выбор определяется уровнем комфорта, который мы желаем иметь при моделировании. Так, при использовании дифференциальных уравнений в качестве модели нам необходимо «гладкое многообразие» (см. п.10.1.1.2). Для определения предельного поведения траекторий ДС необходимо, чтобы пространство было «полным», т.е. каждая предельная точка сходящейся последовательности принадлежала бы этому же пространству.

траекторий называют *фазовым портретом* системы. При определенном навыке по фазовому портрету можно многое сказать о возможных движениях системы. Так, в примере с кошкой из предыдущего параграфа фазовое пространство трехмерно ($D = 3$) (рис.2.1,б), в нем приведена единственная траектория, соответствующая конкретному выбору начального состояния (при $t = 0$). Она позволяет сказать, что животное вначале хорошо развивалось, достигнув отличной формы к моменту $t = 1$, а затем к моменту $t = 2$ похудело и полиняло. Заметим, что в самой фазовой траектории (кривой в фазовом пространстве) информация о моменте времени, когда точка оказывается в том или ином ее месте, отсутствует.

Обычно на фазовом портрете отражают наиболее характерные траектории и объекты. Для иллюстрации рассмотрим элементы фазового портрета системы (2.1), которая моделирует, в частности, колебания в потенциальной яме с двумя минимумами при наличии трения, подобно показанному на рис.2.2,а шарику, катающемуся в двухъямном профиле. Линии на фазовой плоскости $(x, dx/dt)$ (рис.2.2,б) – фазовые траектории, выходящие из точек 1 и 2. Они не могут пересекаться, т.к. последнее означало бы нарушение правила динамического описания – у одного настоящего одно будущее! Ситуации, похожие на пересечения, возможны в *особых* точках – *точках равновесия*, в которых состояние динамической системы остается неизменным сколь угодно долго. Таких точек на портрете три: O, A_1, A_2.

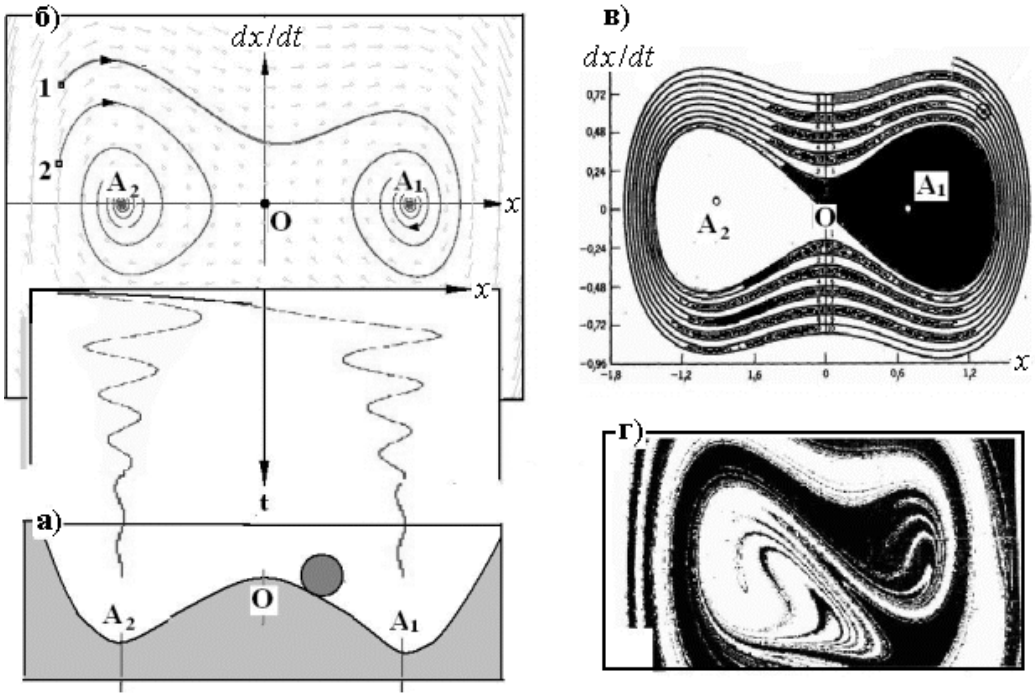

Рис.2.2. Шарик в двойной ямке (а); его фазовый портрет и временные реализации колебаний (б); бассейны притяжения аттракторов (точек равновесия A_1, A_2),

сосуществующих в фазовом пространстве (в); бассейны притяжения неавтономной системы при значениях параметров, когда в фазовом пространстве сосуществуют два хаотических аттрактора (г)

Первая соответствует положению покоящегося шарика на вершине горки (рис.2.2,а), а вторые – в левой и правой ямках. Большинство других точек фазового пространства соответствуют состояниям, в которых изображающие точки в следующий момент не остаются. Их движению в фазовом пространстве соответствует тот или иной вид фазовой траектории и зависимости динамических переменных от времени $x_k(t)$, рис.2.2,б. Обратите внимание, что в типичной фазовой траектории можно выделить начальный участок (*переходный процесс*) и более поздний этап движений, которые отличаются большей степенью повторяемости – *установившиеся движения*. Установившимся движения, которые менее разнообразны, чем переходные процессы, в фазовом пространстве диссипативных систем соответствуют объекты, названные *аттракторами* – от английского «attract» – притягивать, привлекать. В рассматриваемом примере это – состояния устойчивого равновесия – точки равновесия $\mathrm{A}_1, \mathrm{A}_2$. Они действительно как бы притягивают к себе траектории из определенных областей фазового пространства. При старте из разных точек (1 и 2 на рис.2.2,б) фазовые траектории могут попасть на различные аттракторы. Множества точек в фазовом пространстве, из которых система попадает на аттрактор, называется *бассейном притяжения* данного аттрактора.[9] Если аттрактор в фазовом пространстве единственный, его бассейном является все фазовое пространство. При наличии нескольких аттракторов говорят, что имеет место *мультистабильность*. При этом бассейны делят фазовое пространство между собой, как, например, указано штриховкой на рис.2.2,в,г.

Аттракторы могут существовать в пространстве состояний только *диссипативных динамических систем.* Так называют системы,

---

[9] Строгое определение аттрактора – предмет многих дискуссий в научной литературе. До сих пор такого общепризнанного определения нет. Приведем одно из популярных, которое дается в несколько шагов [115, с.76-77]. «… множество A называется … *инвариантным*, если $\Phi_t \mathrm{A} = \mathrm{A}$. *Окрестностью множества* A называется *открытое* множество U, включающее *замыкание* множества A, т.е. A вместе со всеми своими предельными точками, в том числе граничными. … Замкнутое инвариантное множество A называется *притягивающим множеством*, если для него существует окрестность U, такая что для всех $\mathbf{x} \in \mathrm{U}$, $\Phi_t(\mathbf{x}) \to \mathrm{A}$ при $t \to \infty$. Наибольшее U, удовлетворяющее этому определению, называется *областью притяжения* A. … Аттрактором A называется притягивающее множество, содержащее всюду плотную траекторию». Можно переформулировать это определение грубо: аттрактор – наименьшее множество, к которому стремятся почти все траектории ДС из некоторой области ненулевого объема.

обладающие свойством сжатия фазового объема, которое иллюстрирует рис.2.3. На нем представлено множество начальных точек, занимающее объем $V(0)$, стартуя из которого, система через время t оказывается в точках, образующих множество объема $V(t)$. Система называется диссипативной, если с течением времени фазовый объем уменьшается $V(t) < V(0)$. В одномерном случае мерой фазового объема V является длина, в двухмерном – площадь, в многомерном ($D > 3$) – гиперобъем. В конце концов, изображающие точки из стартового объема попадают на аттракторы, объем которых равен 0. Представленное определение диссипативной системы является более широким, чем определение диссипативной системы, используемое в физике, где это – система с трением, в которой механическая (более упорядоченная) энергия переходит в энергию хаотического движения молекул. В *консервативных* системах (в физике это – системы без трения) начальный фазовый объем сохраняется, лишь изменяя свою форму (рис.2.3,б), следовательно, «притягиватели» отсутствуют.

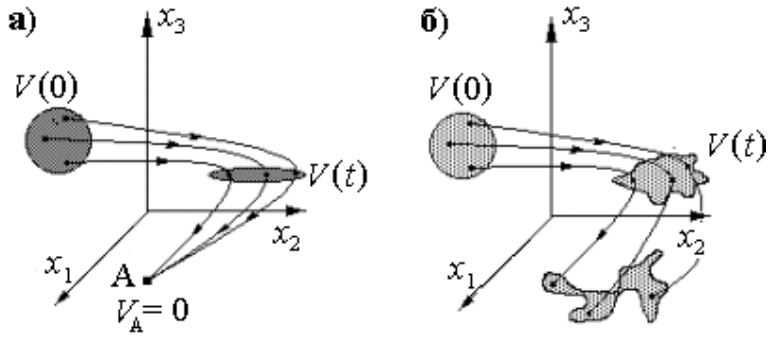

Рис.2.3. Иллюстрация некоторых вариантов деформации фазового объема: а) диссипативная система, б) консервативная система. Линии – фазовые траектории

Перечислим некоторые возможные виды аттракторов в фазовом пространстве и укажем характер установившихся движений, которым они соответствуют (рис.2.4). Кроме состояний равновесия, представляемых точками (рис.2.4,а), аттрактор может представлять собой:

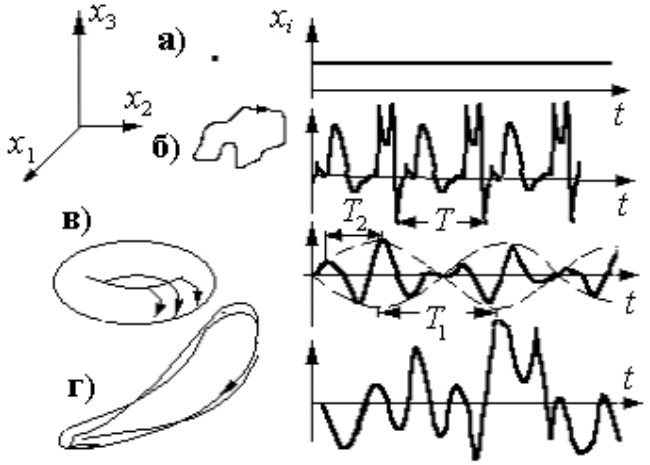

Рис.2.4. Примеры характерных множеств в фазовом пространстве систем с непрерывным временем и временных реализаций движений, которым они соответствуют

- *цикл* – замкнутую кривую в фазовом пространстве, образ движения, повторяющегося с периодом T (рис.2.4,б);
- *тор* – «бесконечно тонкую нитку, наматывающуюся на бублик», образ квазипериодических движений (с двумя характерными периодами T_1 и T_2, находящимися в иррациональном соотношении), рис.2.4,в. Тор может иметь три и более измерений – представлять сложное движение с тремя, четырьмя и т.д. некратными частотами синусоидальных компонент;
- фрактально устроенное множество, сосредоточенное в ограниченной области фазового пространства, образ хаотических колебаний – *странный аттрактор* (рис.2.4,г).[10]

Реализуемые в ДС варианты установившихся движений и соответствующие им аттракторы ограничены размерностью ДС. Так в фазовом пространстве систем с непрерывным временем (с операторами, представленными дифференциальными уравнениями) при $D = 1$ могут существовать лишь состояния равновесия, при $D = 2$ – точки равновесия и циклы, при $D \geq 3$ – все перечисленные ранее предельные множества. Эти сведения могут помочь определиться с выбором размерности модели на практике. Обнаружение, например, хаотических колебаний свидетельствует о том, что для моделирования объекта с помощью нелинейных дифференциальных уравнений первого порядка их потребуется не менее трех.

Несколько другая ситуация с аттракторами складывается в системах с дискретным временем. Вид аттрактора в их фазовом пространстве можно

[10] «Странные» означает «отличные от тех, которые были известны ранее». Обзор видов хаотических аттракторов можно найти в [5].

представить, разрезав левые картинки на рис.2.4 плоскостью (сечение Пуанкаре). Точка и однооборотный цикл дадут в разрезе одну точку. Более сложные циклы – несколько точек. Траектория на торе точками прокола секущей плоскости «нарисует» замкнутую кривую, которая в дискретной системе представляет в фазовом пространстве квазипериодическое движение. Хаотический аттрактор будет представлять собой сложно структурированный (часто, самоподобный) набор точек. Хаотическое движение может происходить в фазовом пространстве даже *одномерных* необратимых отображений.

2.1.4. Характеристики аттракторов

2.1.4.1. Геометрические характеристики. Кроме визуально фиксируемых различий портреты в фазовом пространстве обладают набором количественных мер. Наиболее популярными среди них являются размерности. Целочисленную *топологическую размерность* D_T можно определить по индуктивному принципу [144]: для точки устанавливается размерность $D_T = 0$; множество, которое может быть разделено на непересекающиеся части подмножеством размерности D_T, имеет размерность $D_T + 1$. В соответствии с этими правилами гладкая линия имеет топологическую размерность $D_T = 1$, поверхность – $D_T = 2$, объем – $D_T = 3$. Соответственно, точка равновесия, цикл и тор имеют топологические размерности 0, 1 и 2, подробнее см. [115, с. 208-209]. Структура странных аттракторов качественно отличается от перечисленных множеств. Они фрактальны (самоподобны), что требует введения более сложных мер – *фрактальных размерностей*. Наиболее простая из них – *емкость* – оценивает только геометрию аттрактора. Вводятся и *обобщенные размерности*, учитывающие посещаемость его подмножеств изображающей точкой. Далее мы ограничились лишь короткой информацией о фрактальных мерах и для иллюстрации разместили на сайте [347] учебно-демонстрационную программу, разработанную Т.В. Диканевым. Для более подробного знакомства с фрактальными мерами и методиками их вычисления рекомендуем лекции 11-13 из книги [104], которая доступна на сайте [340], и указанные в них ссылки.

Для определения емкости предельное множество в D-мерном фазовом пространстве покрывается D-мерными кубами (т.е. отрезками, квадратами, трехмерными кубами и т.д.), со стороной ε. Пусть необходимое для

покрытия[11] *минимальное* число (непустых) кубов есть $N(\varepsilon)$. Емкостью множества называется предел

$$D_F = -\lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln \varepsilon}, \tag{2.2}$$

если он существует. Вместо кубов можно использовать *D*-мерные шары или множества другой формы [115, с. 210; 104, с. 170-171]. На рис.2.5 приведены соответствующие иллюстрации и рассмотрен классический пример фрактального множества Кантора, получающегося из единичного отрезка последовательным удалением средней его трети. В соответствии с (2.2) $D_F = -\lim_{\varepsilon \to 0} \frac{\ln 2^N}{\ln (1/3)^N} = \frac{\ln 2}{\ln 3} \approx 0.63$. Большинство известных фрактальных множеств имеет дробную размерность, и их можно «уложить» в пространство, размерность которого равна фрактальной размерности, дополненной до целой величины. Так, Канторово множество уже не конечный набор точек, но еще и не линия.

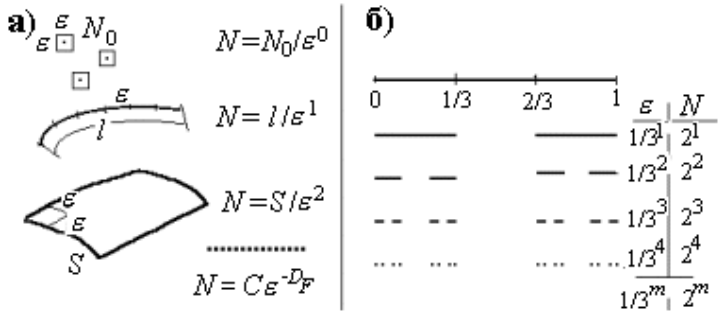

Рис.2.5. Иллюстрация определения емкости (а) и множество Кантора (б)

Более тонкая характеристика – *размерность Хаусдорфа* – обобщает емкость на случай покрытия элементами произвольной формы и размера. Обе величины часто совпадают, но не всегда [104, с. 173; 115, с. 209-211]. Точная численная оценка хаусдорфовой размерности, как правило, невозможна [112].

Обобщенные размерности Реньи D_q «учитывают» частоту посещения изображающей точкой той или иной области аттрактора [115, с. 211-214; 104, с. 176-190]. Пусть аттрактор разбит[12] на *N* непустых кубов (ячеек) размера *ε*. Обозначим долю времени, проводимого изображающей точкой в ячейке номер *i*, через p_i. Это нормированная плотность точек в ячейке –

[11] Покрытием множества A называется семейство подмножеств $\{A_i\}$, таких, что их объединение содержит A.

[12] Разбиение – это покрытие непересекающимися подмножествами $\{A_i\}$.

оценка вероятности ее посещения.[13] Тогда[14]

$$D_q = \frac{1}{q-1}\lim_{\varepsilon\to 0}\frac{\ln\sum_{i=1}^{N(\varepsilon)} p_i^q}{\ln\varepsilon}. \qquad (2.3)$$

Выделяют специальные виды обобщенной размерности: при $q=0$ это – описанная выше емкость, при $q=1$ – информационная размерность (в смысле предела при $q\to 1$), при $q=2$ – корреляционная размерность. Последняя характеризует асимптотическое поведение *пар* точек на аттракторе. Действительно, величины p_i^2 можно интерпретировать как вероятность найти две изображающие точки внутри *i*-го куба размера ε. Как раз эту величину легче всего оценить. Прямое использование определения (2.3) приводит к вычислительным трудностям при $D>3$. Поэтому развиты различные численные методы оценки размерностей по фрагменту дискретизованной во времени фазовой траектории $\mathbf{x}(t_1),\mathbf{x}(t_2),...,\mathbf{x}(t_N)$. Одним из самых известных является алгоритм Грассбергера – Прокаччиа для оценки корреляционной размерности [234]. Он опирается на расчет так называемого корреляционного интеграла $C(\varepsilon)=\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\Theta\left(\varepsilon-\left\|\mathbf{x}(t_i)-\mathbf{x}(t_j)\right\|\right)$, где Θ – функция Хевисайда ($\Theta(s)=0, s\le 0$, $\Theta(s)=1, s>0$), $\|\cdot\|$ – норма вектора (например, евклидова, но можно использовать и любую другую). Нетрудно увидеть, что это – оценка вероятности того, что две точки, произвольно выбранные на аттракторе в соответствии с его вероятностной мерой, окажутся на расстоянии, меньшем ε. При $\varepsilon\to 0$ имеет место $C(\varepsilon)\approx A\varepsilon^{D_2}$, как следует из (2.4).

---

[13] Это строго применимо для аттракторов, обладающих эргодической мерой. [112].

[14] Дополнительный математический комментарий [112]. Предположим, что p_i в каждом (непустом) элементе разбиения подчиняется показательному закону: $p_i \propto \varepsilon^{\alpha}$. Если мы имеем дело с точками на отрезке, то $\alpha=1$ соответствует равномерному распределению точек. Однако, для областей, где точек мало, может возникнуть ситуация, когда $\alpha<1$. В этом случае отношение $p_i/\varepsilon\to\infty$ при $\varepsilon\to 0$, потому такое распределение называют *сингулярным* (для квадрата, сингулярные распределения будут иметь области с показателем $\alpha<2$). Чтобы оценить поведение плотности в различных участках носителя, подсчитывают *функцию разбиения* $\sum_i p_i^q$, где параметр $q\in(-\infty,\infty)$ позволяет «подстраивать» оценку под участки с разной плотностью вероятностей. Если функция разбиения зависит от ε по степенному закону, то вводят определение (2.3) и говорят о *мультифрактальном распределении*. Если D_q принимает различные значения в зависимости от *q*, то аттрактор называют мультифракталом [104, с. 182].

Тогда корреляционную размерность D_2 можно оценить как наклон графика $\ln C(\ln\varepsilon)$ при малом ε. На практике число точек траектории N ограничено, поэтому размер ячейки ε нельзя сделать сколь угодно малым. Причем для надежной оценки размерности требуется тем большее число точек, чем больше размерность. Существуют различные мнения о необходимом числе точек, полученные при различных предположениях [88, 220, 221], см. подробности также в [104, 115].

Для целочисленной оценки размерности наблюдаемого движения сверху используются и другие идеи. Один из наиболее популярных – *метод ложных близких соседей* [258] – основан на проверке того свойства, что фазовая траектория, восстановленная в пространстве достаточной размерности не должна иметь самопересечений. Он применяется для восстановления фазовой траектории по временной реализации единственной переменной (см. п.10.1.2.2).

Другой известный метод – *метод главных компонент* [199] – заключается в выделении в фазовом пространстве направлений, вдоль которых, в основном, развивается движение изображающей точки, на основе анализа корреляций между компонентами вектора состояния (см. п.10.1.2.3).

2.1.4.2. Динамические характеристики. Наиболее популярными из динамических характеристик являются так называемые *ляпуновские* показатели, которые характеризуют скорость разбегания или сближения изначально очень близких фазовых траекторий. Малое отклонение изображающей точки от некоторой траектории на аттракторе, т.е. малое возмущение ε_0, до тех пор, пока оно не достигло значительных величин, эволюционирует приближенно по экспоненциальному закону вида $\varepsilon(\Delta t)=\varepsilon_0 e^{\lambda\Delta t}$ (рис.2.6,а). В результате D-мерная сфера множества начальных возмущений через некоторое время трансформируется в эллипсоид. Не давая системе увеличивать возмущения (эволюционировать по штриховому участку траектории на рис.2.6,а) ограничением отрезка времени наблюдения τ, можно вычислить показатели экспоненты через отношения длин полуосей эллипса возмущений к начальному радиусу: $\lambda_i=(1/\tau)\ln(\varepsilon_i/\varepsilon_0)$. Усредненные по всему аттрактору значения этих коэффициентов называют *показателями Ляпунова*, мы их обозначим $\Lambda_1,\Lambda_2,...,\Lambda_D$. Они характеризуют устойчивость движения на аттракторе в линейном приближении. Упорядоченный по убыванию набор значений Λ_i называют спектром Ляпуновских показателей, а последовательная запись их знаков (+, –, или 0) – *сигнатурой спектра*. Если все показатели отрицательны, т.е. сигнатура имеет вид $\langle -,-,...,-\rangle$, то аттрактором

является точка равновесия. Сигнатура предельного цикла – $\langle 0, -, ..., - \rangle$, для двумерного тора – $\langle 0, 0, -, ..., - \rangle$. В спектре ляпуновских показателей хаотического аттрактора обязательно присутствует хотя бы один положительный показатель, например, $\langle +, 0, -, ..., - \rangle$, определяющий типичное «разбегание» изначально близких траекторий.

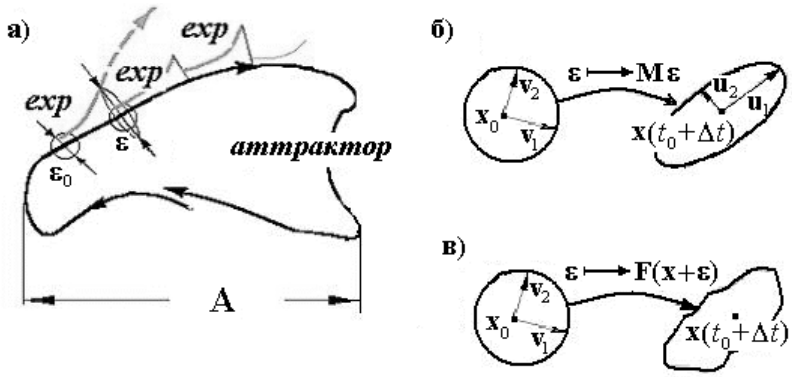

Рис.2.6. Иллюстрация идеи вычисления ляпуновских показателей (а). Эволюция окружности с центром $\mathbf{x}_0$ под действием линейной (б) и нелинейной системы (в)

Опишем теперь некоторые математические детали более подробно. Для этого ограничимся сначала линейной системой обыкновенных дифференциальных уравнений с переменными коэффициентами:

$$d\boldsymbol{\varepsilon}(t)/dt = \mathbf{A}(t)\boldsymbol{\varepsilon}(t), \tag{2.4}$$

где $\boldsymbol{\varepsilon} \in R^D$ и $\mathbf{A}$ – матрица порядка *D*. Обозначим $\boldsymbol{\varepsilon}(t_0) = \boldsymbol{\varepsilon}_0$. Тогда решение уравнения (2.4) в момент времени $t_0 + \Delta t$ можно найти по формуле

$$\boldsymbol{\varepsilon}(t_0 + \Delta t) = \mathbf{M}(t_0, \Delta t) \cdot \boldsymbol{\varepsilon}_0, \tag{2.5}$$

где $\mathbf{M}(t_0, \Delta t)$ – матрица порядка *D*, зависящая от начального момента времени и интервала Δt, которая имеет вид:

$$\mathbf{M}(t_0, \Delta t) = \exp\left(\int_{t_0}^{t_0+\Delta t} \mathbf{A}(t')dt' \right), \tag{2.6}$$

где экспонента с матрицей в показателе понимается в смысле формального разложения в степенной ряд. Для иллюстрации положим $D = 1$ и $\mathbf{A}(t) = a = const$, тогда $d\varepsilon(t)/dt = a\varepsilon(t)$ и решение (2.5) и (2.6) принимает знакомый вид $\varepsilon(t_0 + \Delta t) = \varepsilon_0 e^{at}$. Таким образом, в случае постоянных коэффициентов возмущение изменяется по экспоненциальному закону. Если коэффициенты не постоянны, то ситуация несколько меняется. Например, для $\mathbf{A}(t) = a + b\cos t$ получим $\varepsilon(t_0 + \Delta t) = \varepsilon_0 e^{a\Delta t} e^{-b\sin\Delta t}$.

Чтобы охарактеризовать нарастание (или убывание) $\boldsymbol{\varepsilon}$ в многомерном случае, нужно рассмотреть эволюцию сферы начальных условий с центром в нуле и радиусом $\|\boldsymbol{\varepsilon}_0\|$. Поскольку система линейна, сфера преобразуется в эллипсоид с тем же центром. Величина и ориентация полуосей эллипсоида зависят от матрицы **M**, а значит, и от величины интервала Δt. В зависимости от ориентации начального вектора $\boldsymbol{\varepsilon}_0$ модуль $\boldsymbol{\varepsilon}$ меняется по-разному. Для описания этой ситуации используем так называемое *сингулярное разложение* матрицы **M**. Это разложение вида $\mathbf{M} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathrm{T}}$, где **U** и **V** – ортогональные матрицы, которые удобно записать в виде наборов векторов $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_D]$, $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_D]$. Если **M** невырождена, то векторы $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_D$ (их называют левыми сингулярными векторами матрицы **M**) имеют единичную длину и взаимно ортогональны, т.е. образуют ортонормированный базис пространства R^D. То же самое относится к векторам $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_D$ (правым сингулярным векторам). Матрица $\boldsymbol{\Sigma}$ диагональна; числа $\sigma_1, \ldots, \sigma_D$ на ее диагонали обычно располагают в порядке убывания и называют *сингулярными числами* матрицы **M**. Действие матрицы **M** на вектор $\boldsymbol{\varepsilon}_0$, имеющий направление одного из правых сингулярных векторов $\mathbf{v}_i$, изменяет его длину в σ_i раз и переводит его в вектор, имеющий направление левого сингулярного вектора с номером *i*: $\boldsymbol{\varepsilon}(t_0 + \Delta t) = \sigma_i \|\boldsymbol{\varepsilon}_0\| \mathbf{u}_i$ (рис.2.6,б). Таким образом, если хотя бы одно из сингулярных чисел **M** по модулю больше 1, то для некоторых направлений начальное возмущение возрастает (на рис.2.6,б одно сингулярное число больше 1, а другое – меньше). Быстрее всего оно возрастает для направления $\mathbf{v}_1$. Величины, показывающие, во сколько раз меняется величина возмущения, называют *локальными* ляпуновскими показателями:

$$\lambda_i(t_0, \Delta t) = \frac{1}{\Delta t} \ln \sigma_i . \tag{2.7}$$

Они описывают *средний за конечное время* экспоненциальный рост возмущений. Согласно определению (2.7) для соответствующих направлений начального возмущения имеет место строгое равенство $\|\boldsymbol{\varepsilon}(t_0 + \Delta t)\| = \|\boldsymbol{\varepsilon}_0\| \cdot e^{\lambda_i(t_0, \Delta t)\Delta t}$. А.М. Ляпунов показал, что при стремлении Δt к бесконечности и при определенных условиях, наложенных на матрицу **A**,[15] существуют конечные пределы

[15] Существует число *L*, такое что $\frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} |A_{ij}(t')| dt' \le L$ для любых A_{ij} и Δt [104, с. 140].

$$\Lambda_i = \overline{\lim_{\Delta t \to \infty}} \frac{1}{\Delta t} \ln \frac{\|\boldsymbol{\varepsilon}(t_0 + \Delta t)\|}{\|\boldsymbol{\varepsilon}_0\|}, \; i = 1,2,\ldots,D, \tag{2.8}$$

где величины Λ_i и есть *ляпуновские показатели*. Они показывают *эффективную* скорость нарастания (убывания) возмущений. Какое именно из D значений наблюдается для данного начального возмущения $\boldsymbol{\varepsilon}_0$, зависит от его направления. Для *почти любого* направления $\boldsymbol{\varepsilon}_0$ возмущение изменяется со скоростью, определяемой старшим показателем Λ_1. Если Λ_1 положителен, то он и определяет эффективную скорость нарастания возмущений. Следовательно, он может быть связан с пределом предсказуемости системы (2.4) в ситуации, когда само уравнение известно точно, а начальное условие известно с некоторой погрешностью. Аналогичный анализ можно провести для линейных разностных уравнений.

Линеаризованная динамика и ляпуновские показатели. Для нелинейных систем анализ на устойчивость проводят с помощью исследования линеаризованной системы уравнений. Пусть задана нелинейная система:

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}). \tag{2.9}$$

Рассмотрим одну из ее траекторий $\mathbf{x}(t)$ с начальным условием $\mathbf{x}(t_0) = \mathbf{x}_0$ и исследуем поведение траектории, стартующей очень близко от нее в точке $\mathbf{x}(t_0) = \mathbf{x}_0 + \boldsymbol{\varepsilon}_0$. Будем называть траекторию с $\mathbf{x}(t_0) = \mathbf{x}_0$ опорной. Эволюция очень малых возмущений, которые остаются малыми на всем интервале рассмотрения, определяется линеаризованной в окрестности опорной траектории системой:

$$\frac{d\boldsymbol{\varepsilon}}{dt} = \frac{\partial \mathbf{f}(\mathbf{x}(\mathbf{x}_0, t))}{\partial \mathbf{x}_0} \boldsymbol{\varepsilon}_0. \tag{2.10}$$

Это уравнение совпадает с (2.4), если положить $\mathbf{A}(t) = \dfrac{\partial \mathbf{f}(\mathbf{x}(\mathbf{x}_0, t))}{\partial \mathbf{x}_0}$. Для него можно записать решение в виде (2.5), где матрица **M** отображает бесконечно малую сферу начальных условий с центром $\mathbf{x}_0$ в эллипсоид с центром $\mathbf{x}(t_0 + \Delta t)$, см. рис.2.6,б. Строго говоря, если возмущение не бесконечно мало, а конечно, образом сферы будет не эллипсоид, а другое множество. Линеаризованная динамика лишь приближенно будет описывать эволюцию этих конечных возмущений, см. рис.2.6,в. Для любой опорной траектории существует набор ляпуновских показателей, характеризующих линеаризованную динамику в ее окрестности.

В 1968 г. Оселедец показал, что для любого типичного $\mathbf{x}_0$ на рассматриваемом аттракторе набор ляпуновских показателей один и тот же. Это утверждение составляет содержание так называемой мультипликативной эргодической теоремы, см., например, [115, с. 224-227; 157].

Итак, ляпуновские показатели характеризуют эволюцию малых (настолько малых, чтобы для их описания была применима линеаризованная система (2.10)) возмущений не только на индивидуальной опорной траектории, но и на всем аттракторе ДС. Старший ляпуновский показатель оценивает эффективную скорость роста бесконечно малых возмущений.

2.1.5. Пространство параметров. Бифуркации. Комбинированные пространства, бифуркационные диаграммы

При изменении параметров аттракторы в фазовом пространстве эволюционируют (деформируются, меняют свои размеры), и при некоторых значениях параметров теряют устойчивость. В результате происходит качественное изменение возможных движений в системе, изменение фазового портрета (например, изменение числа аттракторов в фазовом пространстве) – ситуация, получившая название *бифуркации* [16] [42, 114]. Следует подчеркнуть, что плавная деформация аттрактора и сопровождающие ее изменения формы колебаний не считаются качественным изменением. Используя геометрические представление в *пространстве параметров*, когда на осях координат откладываются значения параметров, можно наглядно представить всю совокупность видов установившихся движений и переходов между ними, возможные в рассматриваемой динамической системе. Это делается с помощью некоторых методических приемов.

Основной прием состоит в выделении множеств точек, разделяющих области с качественно различным поведением – *бифуркационных* множеств. В наглядном примере с кошкой пространством параметров является плоскость $\mathbf{a}_1, \mathbf{a}_2$ (рис.2.1,д), а линии границ областей с различной штриховкой являются бифуркационными: область 2 соответствует здоровью, полноценному существованию, а в областях 1 и 3 скоро существование заканчивается трагически из-за бескормицы или обжорства. Значительно сложнее устроены пространства параметров, представленные, например, на рис.3.6, 3.7, 3.11, 3.19. Бифуркационные множества – здесь это линии –делят всю плоскость параметров на области существования различных аттракторов в фазовом пространстве. Обратите

[16] Первоначально слово бифуркация использовалось для обозначения раздвоения пути эволюции, но в настоящее время так стали называть любое качественное изменение.

внимание на использованный при построении рис.3.6, 3.11, 3.19 способ сделать наглядной ситуацию мультистабильности (наличие нескольких вариантов движений, аттракторов в фазовом пространстве) на плоскости двух параметров. Область значений параметров, при которых существует каждый из возможных колебательных режимов, изображается на отдельном листе, так что наличие многолистности при некоторых значениях параметров означает существование мультистабильности. Например, на рис.3.6 в области пересечения листов А и В, имеет место бистабильность – существование двух аттракторов. Третий лист в области бистабильности принадлежит неустойчивому циклу. Мультистабильность на рис.3.19 также имеет место при значениях параметров, взятых в области пересечения листов, на которых существуют различные моды исследуемой системы.

В заключение отметим некоторые возможности, возникающие при использовании *комбинированных пространств*. Например, по оси абсцисс можно отложить значение параметра, а по оси ординат – значения динамической переменной в установившемся режиме. Построенная таким образом бифуркационная диаграмма для квадратичного отображения (рис.3.8,д) широко используется для демонстрации универсальных закономерностей подобия (скейлинга) при переходе к хаосу через каскад удвоений периода. В отображении диссипативного нелинейного осциллятора такая диаграмма хорошо иллюстрирует явления резонанса, гистерезиса, бистабильности, последовательность бифуркаций (рис.3.10, внизу).

Наконец, следует обратить внимание на существенные возможности представления информации в фазовых и параметрических пространствах с помощью цвета. Так часто представляют бассейны притяжения различных аттракторов, области существования и эволюции колебательных режимов (рис.2.2 и 3.11).

Сказанное в данном параграфе является лишь кратким введением в реализацию динамического подхода. Тем, кто хочет получить глубокое представление по проблеме, следует обратиться к классическим трудам по качественной теории дифференциальных уравнений, отображений, теории колебаний, нелинейной динамике, см., например, [4, 13, 14, 24, 48, 175].

2.2. Основания для объявления процессов случайными

Использование вероятностного подхода обычно связано с признанием некоторой величины «случайной». Но что такое «случайная величина» и чем она отличается от «неслучайной»? В настоящее время существует несколько точек зрения на случайность, позволяющих ввести количественные меры. Они в основном согласуются друг с другом, но не всегда, и могут приводить даже к противоположным результатам при

оценке случайности или не случайности на практике. Рассмотрим проблему в соответствии со схемой, использованной в очень полезных статьях [96, 97].

2.2.1. Теоретико-множественный подход

Теоретико-множественный подход, лежащий в основе современной теории вероятностей [60,145], связывает понятие случайности с возможностью задать для данной величины закон распределения вероятностей. Оценка отсутствия или присутствия закономерности характеризуется возможным разбросом значений величины:

- детерминированным величинам соответствует плотность распределения вероятностей в виде δ-функции;
- непредсказуемость значения случайной величины определяется «шириной», «размазанностью» распределения.

2.2.1.1. Случайные события и вероятность. При описании многих явлений исследователь часто сталкивается с невозможностью однозначно предсказать ход событий, даже если все условия, которые можно контролировать, поддерживаются «одинаковыми»[17]. Для исследования таких явлений ввели понятия случайного события и вероятности, которыми оперирует теория вероятностей. В теории эти понятия являются неопределимыми, задаются аксиоматически только некоторые их свойства, а связь их с практикой и наглядная интерпретация – задача тех, кто использует этот аппарат. Здесь мы напомним эти основные понятия, но не строго, а на интуитивном уровне.

Событием называют исход *испытания*. Рассмотрим классический пример – подбрасывание монеты (см. также п. 2.6). Пусть она подбрасывается только один раз. Тогда испытанием будет этот единственный бросок. В результате испытания могут реализоваться только два события – выпадение орла (событие A) или решки (событие B).[18] A и B – события, взаимно исключающие друг друга. Событие, которое состоит в том, что реализуется либо A, либо B, называют *объединением* событий A и B и обозначают $A \cup B$. В данном случае оно обязательно произойдет в результате любого испытания. Такое событие называют *достоверным* и говорят, что его вероятность равна единице: $P\{A \cup B\} = 1$. Поскольку объединение A и B – достоверное событие, то говорят, что A и B образуют полную группу событий. Из соображений симметрии следует, что если вещество монеты распределено равномерно и она имеет правильную

[17] Мы поставили слово «одинаковыми» в кавычки, чтобы подчеркнуть его условность, модельность. Реально можно говорить об одинаковости лишь в той степени, которая допускается условиями наблюдения или измерения.

[18] Теоретически мыслимым случаем, когда монета встанет «на ребро», пренебрегаем.

форму, то события A и B равновероятны (шансы на выпадение орла и решки равны). Равновероятные взаимно исключающие друг друга события, образующие полную группу, называют *элементарными*. Вероятность объединения взаимно исключающих друг друга событий равна сумме их вероятностей. В рассматриваемом примере события A и B образуют полную группу, поэтому $\mathrm{P}\{\mathrm{A} \cup \mathrm{B}\} = \mathrm{P}\{\mathrm{A}\} + \mathrm{P}\{\mathrm{B}\} = 1$. Отсюда и из условия равновероятности получаем, что индивидуальные вероятности $\mathrm{P}\{\mathrm{A}\} = \mathrm{P}\{\mathrm{B}\} = 1/2$.

Не всегда удается выделить элементарные события. В некоторых случаях помогают геометрические соображения (геометрическое определение вероятности). Пусть испытание состоит в том, что в область A плоскости случайно бросается точка. В область A точка попадет достоверно, и все участки A «равноправны». Точка может попасть в область $\mathrm{B} \subset \mathrm{A}$, а может и не попасть. Вероятность того, что точка попадет в область B определяется через отношение площадей B и A: $\mu(\mathrm{B})/\mu(\mathrm{A})$, где μ означает (лебегову) *меру* множества. В данном примере – это площадь, но формулу можно использовать без изменений для «бросания точки» в пространство любой размерности. Это определение можно связать с выделением элементарных событий, если назвать ими попадания точки в малые квадраты, покрывающие A, и устремить длину сторон квадратов к нулю.[19]

Наиболее привычным для физиков является статистическое определение вероятности. Если в результате N независимых испытаний событие A реализовалось M раз, то отношение M/N называют частотой появления события A. Если при увеличении числа испытаний до бесконечности частота M/N стремится к некоторому пределу, то этот предел и называют вероятностью события A. Это наиболее наглядный (физический) смысл понятия вероятности. Заметим также, что свойство стабилизации частот появления событий называют статистической

[19] Следует заметить, что для геометрических вероятностей важно четко определить, что понимается под «случайной» точкой, прямой или плоскостью. Например, оценим вероятность того, что «случайная» хорда превышает длину стороны правильного треугольника, вписанного в единичную окружность. Хорду можно выбрать «случайно» разными способами. Первый вариант: совместим начало хорды с одной из вершин треугольника, оставив второй конец свободным. Доля благоприятных исходов, когда длина хорды превысит длину стороны, составляет 1/3. Другой вариант: будем выбирать случайно точку – середину «случайной» хорды в круге. Хорда длиннее стороны треугольника, когда ее середина лежит внутри круга вписанного в треугольник. Радиус этого круга равен половине радиуса описанной окружности и, следовательно, доля благоприятных исходов, которая оценивается как отношение площадей двух кругов, равна 1/4. Мы получили два разных ответа, в зависимости от выбора понятия «случайной» хорды.

устойчивостью. Аппарат теории вероятностей пригоден для описания явлений, удовлетворяющих условию статистической устойчивости.

2.2.1.2. Случайные величины и их характеристики. Случайной величиной называется числовая функция ξ, определенная на множестве случайных событий. В случае с монетой значения случайной величины можно определить как $\xi = 1$ (выпал орел) и $\xi = 0$ (выпала решка). Вероятность того, что $\xi = 1$, есть вероятность выпадения орла.

Для полной характеристики случайной величины следует задать вероятности, с которыми она принимает свои значения. Например, можно использовать так называемую *функцию распределения* $F_\xi(x) \equiv \mathrm{P}\{\xi \le x\}$. Если ξ непрерывнозначна и функция распределения дифференцируема, то можно ввести функцию *плотности распределения вероятностей* $p_\xi(x) \equiv dF_\xi(x)/dx$: вероятность попадания ξ в бесконечно малый отрезок $[x, x+dx]$ равна $p_\xi(x)dx$. Далее для определенности будем говорить о случайных величинах, для которых можно ввести плотность вероятности.

Перечислим некоторые часто встречающиеся законы распределения. Нормальный (гауссовский) закон:

$$p_\xi(x) = \left(1\Big/\sqrt{2\pi\sigma^2}\right)\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{2.11}$$

где a и σ^2 – параметры. Это одно из наиболее часто используемых в теории вероятностей распределений. Причина его популярности в том, что оно имеет многие полезные теоретические свойства и допускает получение ряда аналитических результатов. Кроме того, на практике величины, которые являются результатом действия многочисленных факторов, часто распределены примерно по нормальному закону. Это имеет теоретическое обоснование – центральную предельную теорему,[20] которая утверждает, что сумма большого числа одинаково распределенных независимых случайных величин распределена асимптотически нормально, т.е. закон ее распределения стремится к нормальному при увеличении числа слагаемых.

Экспоненциальный закон (распределение Лапласа):

$$p_\xi(x) = \begin{cases} (1/a)\exp(-x/a), & x \ge 0, \\ 0, & x < 0. \end{cases} \tag{2.12}$$

Равномерное распределение на отрезке $[a,b]$:

[20] Многие авторы пишут с иронией о частом использовании нормального закона для описания данных и ссылках на центральную предельную теорему: «… инженеры думают, что применимость ЦПТ на практике – это строго доказанное утверждение, а математики полагают, что это экспериментальный факт…», см., например, [289].

$$p_\xi(x) = \begin{cases} 1/(b-a), a \le x \le b, \\ 0, \qquad x < a, x > b. \end{cases} \tag{2.13}$$

Случайную величину ξ часто характеризуют моментами ее распределения. Начальным моментом порядка n называется величина

$$M\left[\xi^n\right] \equiv \int_{-\infty}^{\infty} x^n p(x)dx\,. \tag{2.14}$$

Через M будем обозначать математическое ожидание величины, стоящей в квадратных скобках. Момент первого порядка есть математическое ожидание самой величины ξ. Его физический смысл – среднее значение по бесконечно большому множеству независимых испытаний. Центральные моменты определяются как моменты отклонения ξ от среднего:

$$M\left[(\xi - M[\xi])^n\right] \equiv \int_{-\infty}^{\infty} (x - M[\xi])^n p(x)dx\,. \tag{2.15}$$

Центральный момент второго порядка называется *дисперсией*. Это наиболее часто используемая мера разброса значений, обозначим ее σ_ξ^2. Величину σ_ξ называют среднеквадратичным (стандартным) отклонением ξ. Центральный момент третьего порядка называют асимметрией, четвертого – эксцессом. Для нормального закона (2.11) асимметрия равна 0, а эксцесс равен $3\sigma_\xi^4$. Если существуют все начальные моменты ξ (при $n = 1,2,...$) и они известны, то по их значениям однозначно восстанавливается функция распределения. Параметры закона распределения связаны с его моментами. Например, для нормального закона (2.11) $M[\xi] = a$ и $\sigma_\xi^2 = \sigma^2$; для экспоненциального закона (2.12) $M[\xi] = a$ и $M\left[\xi^2\right] = a^2$, для равномерного закона (2.13) $M[\xi] = (a+b)/2$ и $\sigma_\xi^2 = (b-a)^2/12$.

Для двух случайных величин ξ_1 и ξ_2 рассматриваются их совместные характеристики. Эти две величины можно рассматривать как компоненты двумерного *случайного вектора* $\boldsymbol{\xi}$. Вводится совместная плотность распределения вероятностей $p_{\boldsymbol{\xi}}(x_1, x_2)$: вероятность того, что значения величин попадут *одновременно* (в результате одного испытания) в малые отрезки $[x_1, x_1 + dx_1]$ и $[x_2, x_2 + dx_2]$ равна $p_{\boldsymbol{\xi}}(x_1, x_2)dx_1dx_2$. Вводится условная плотность распределения одной величины при условии, что другая принимает фиксированное значение x^*, например, $p_{\xi_1}\left(x_1 \middle| x_2 = x^*\right)$.

Величины ξ_1 и ξ_2 называют (статистически) *независимыми*, если $p_\xi(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(x_2)$. В последнем случае условное распределение каждой из величин ξ_1 и ξ_2 совпадает с ее безусловным распределением.

Случайная величина, зависящая от времени (например, последовательность значений ξ), называется *случайным процессом*, см. главу 4.

2.2.1.3. Понятие статистической оценки. На практике мы не располагаем, как правило, известным законом распределения, а должны *оценить* математическое ожидание наблюдаемой величины или параметры ее распределения по результатам испытаний. Это задача математической статистики [145, 87, 51, 81] – обратная задаче теории вероятностей, где определяют свойства случайной величины по заданному закону распределения. Набор значений, которые приняла случайная величина ξ в результате N испытаний обозначим $\{x_1, ..., x_N\}$, он называется *выборкой*.[21]

Величина, значение которой получено в результате обработки данных $\{x_1, ..., x_N\}$, называется функцией выборки. *Оценкой* некоторого параметра распределения величины ξ называют функцию выборки, значения которой в некотором смысле близки к истинному значению этого параметра.[22] Будем обозначать оценки «крышечкой» сверху: $\hat{a}$ и т.п.

Пусть выборка $\{x_1, ..., x_N\}$ получена в результате *независимых* испытаний. Пусть математическое ожидание ξ равно a. Значение a неизвестно, известно только, что оно принадлежит некоторому множеству A. Нужно получить оценку $\hat{a}$, которая была бы как можно ближе к a (для любого истинного значения a из A). Любая оценка – это случайная величина, поскольку она является

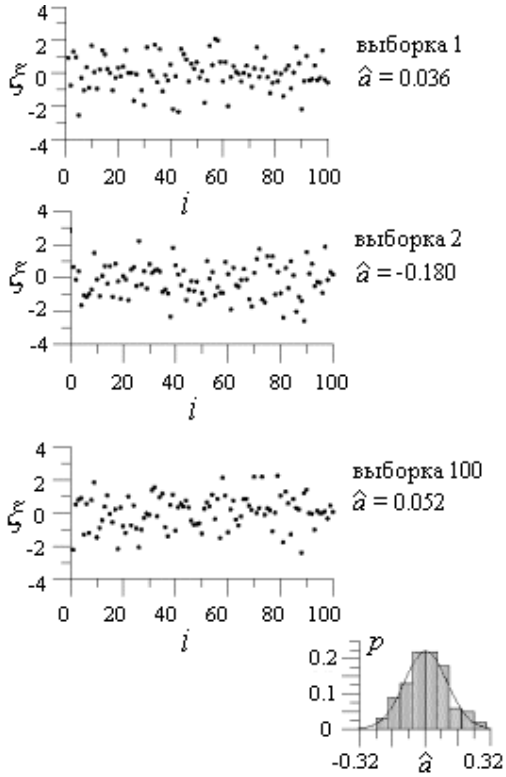

Рис.2.7. Выборки объема 100 из нормального закона с нулевым средним и единичной дисперсией. Значения оценки среднего для различных выборок. Оценка распределения вероятностей этой оценки, полученная по 100 выборкам (теоретически – это нормальный закон с нулевым средним и

[21] Выборка является N-мерным случайным векто

[22] Теоретически любую измеримую функци значения оценки не близки к истинному значени

функцией случайных величин $\hat{a} = f(x_1,...,x_N)$. По одной выборке получим одно значение $\hat{a}$, а по другой – другое, т.е. $\hat{a}$ характеризуется своей плотностью распределения $p_f(\hat{a})$ (см. рис.2.7), которая определяется законом распределения $p_\xi(x)$ и выбранным способом расчета $\hat{a}$, т.е. функцией *f*. Разным *f* соответствуют оценки с разными законами распределения и вероятностными свойствами.

2.2.1.4. Смещение и дисперсия оценок. Наиболее важным свойством оценки является близость ее значений $\hat{a}$ к истинному значению оцениваемой величины *a*. Эту близость можно характеризовать по-разному. Наиболее удобным и широко применяемым способом определения *погрешности оценки* является средний квадрат разности между $\hat{a}$ и *a*:

$$M\left[(\hat{a}-a)^2\right] \equiv \int_{-\infty}^{\infty} (\hat{a}-a)^2 p_f(\hat{a}) d\hat{a}. \qquad (2.16)$$

Нетрудно показать, что погрешность равна сумме двух слагаемых:

$$M[(\hat{a}-a)^2] = \left(M[\hat{a}]-a\right)^2 + \sigma_{\hat{a}}^2. \qquad (2.17)$$

Разность $M[\hat{a}]-a$ называется *смещением* или систематической ошибкой. Таким образом, квадрат погрешности оценки равен сумме квадрата смещения и дисперсии оценки. «Хорошей» оценкой будет та, для которой малы и смещение, и дисперсия. Но при фиксированном объеме выборки требования минимизации дисперсии и смещения, как правило, противоречат друг другу, и приходится искать компромисс между ними. Конкретная форма этого компромисса определяется спецификой задачи, например, можно минимизировать величину (2.17).

Оценка, смещение которой равно нулю, т.е. $M[\hat{a}] = a$ для любого $a \in A$, называется *несмещенной*. При усреднении значений такой оценки по различным выборкам получится величина, более близкая к истинному значению *a*, т.к. случайные ошибки взаимно компенсируются.

Несмещенных оценок величины *a* (с различными функциями *f*) можно «придумать» много. Они различаются по дисперсии. Можно показать, что несмещенная оценка с минимальной дисперсией единственна, т.е. если минимально возможное значение дисперсии есть $\sigma_{\min}^2$, то только для одной оценки оно достигается в точности. Несмещенная оценка с минимальной дисперсией $\sigma_{\min}^2$, является привлекательным вариантом, хотя минимальный квадрат погрешности (2.17) может достигаться для другой оценки – несколько смещенной, но зато с меньшей дисперсией.

Несмещенной оценкой математического ожидания по выборке из независимых значений является так называемое *эмпирическое*

(*выборочное*) среднее, будем обозначать его угловыми скобками и нижним индексом N: $\langle x \rangle_N$. Это просто арифметическое среднее:

$$\langle \xi \rangle_N = f(x_1,...,x_N) = \frac{1}{N}\sum_{i=1}^{N} x_i \,. \tag{2.18}$$

При нормальном законе – это оценка с минимальной дисперсией. Для симметричного распределения ξ, но с большим эксцессом и прочими отклонениями от нормальности, меньшую дисперсию имеет оценка, рассчитываемая как *выборочная медиана*.[23] Она более устойчива и по отношению к вариациям закона распределения ξ. Свойство устойчивости по отношению к тем или иным возмущениям часто называют *робастностью*. Для расчета выборочной медианы упорядочивают значения в выборке по возрастанию: $x_{i_1} < x_{i_2} < ... < x_{i_N}$. Тогда при нечетном N выборочная медиана есть $x_{i_{(N+1)/2}}$, а при четном $\left(x_{i_{N/2}} + x_{i_{N/2+1}}\right)/2$.

Выборочное среднее (2.18) – оценка, несмещенная при любом распределении ξ, а выборочная медиана смещена для несимметричных законов распределения.

Оценкой начального момента $M[\xi^n]$ может служить эмпирический (выборочный) момент порядка n:

$$\langle \xi^n \rangle_N = \frac{1}{N}\sum_{i=1}^{N} x_i^n \,. \tag{2.19}$$

При оценивании центральных моментов ситуация несколько отличается, т.к. неизвестно значение $M[\xi]$, которое входит в определение. Так, оценку дисперсии можно получить как *эмпирическую* (*выборочную*) *дисперсию*:

$$\hat{\sigma}_\xi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle \xi \rangle_N\right)^2 . \tag{2.20}$$

Но она несколько смещена из-за замены $M[\xi]$ на эмпирическое среднее. Можно показать, что несмещенной оценкой будет

$$\hat{\sigma}_\xi^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \langle \xi \rangle_N\right)^2 . \tag{2.21}$$

2.2.1.5. Состоятельность оценок. Важно проследить, как меняются свойства оценки при увеличении объема выборки N. В общем случае при

[23] Медианой распределения называется такое число b, которое делит ось x на две равновероятных области: $\mathrm{P}\{\xi < b\} = \mathrm{P}\{\xi > b\} = 1/2$. Для симметричного распределения медиана совпадает с математическим ожиданием.

этом меняется закон распределения оценки, следовательно, могут меняться ее смещение и дисперсия. Как правило, значения оценки, более близкие к истинному, можно получить при достаточно больших *N*. Если при $N \to \infty$ смещение оценки $M[\hat{a}] - a$ стремится к нулю (для любого *a* из *A*), то она называется *асимптотически несмещенной*. Если $\hat{a}$ сходится к *a* по вероятности (т.е. вероятность того, что значение оценки отличается от истинного на величину большую ε, стремится к нулю для любого сколь угодно малого ε: $\forall \varepsilon > 0 \;\; \mathrm{P}\{|\hat{a} - a| > \varepsilon\} \underset{N\to\infty}{\to} 0$), то говорят, что оценка является *состоятельной*. Состоятельность является очень важным свойством оценок, гарантирующим их высокое качество при больших объемах выборки. Эмпирические моменты (2.19) являются состоятельными оценками теоретических начальных моментов [145, 93].

2.2.1.6. Метод статистических моментов. Рассмотрим теперь задачу оценивания параметров распределения, когда известна его функциональная форма $p_\xi(x, \mathbf{c})$, где $\mathbf{c} = (c_1, ..., c_P)$ – вектор параметров, принимающий значения на множестве $A \subset R^P$. Одним из возможных подходов является *метод статистических моментов*. Он состоит в следующем. Первые *P* теоретических начальных моментов выражают как функции параметров, см. примеры для нормального, экспоненциального и равномерного распределений, где первые два момента выражались через параметры с помощью простых функций. Получают систему

$$\begin{aligned} M[\xi] &= g_1(c_1, ..., c_P), \\ M[\xi^2] &= g_2(c_1, ..., c_P), \\ &..., \\ M[\xi^P] &= g_P(c_1, ..., c_P). \end{aligned} \tag{2.22}$$

Подставляя в (2.22) эмпирические моменты вместо теоретических, получают систему уравнений для параметров

$$\begin{aligned} \langle \xi \rangle_N &= g_1(c_1, ..., c_P), \\ \langle \xi^2 \rangle_N &= g_2(c_1, ..., c_P), \\ &..., \\ \langle \xi^P \rangle_N &= g_P(c_1, ..., c_P), \end{aligned} \tag{2.23}$$

решение которой $\hat{c}_1, ..., \hat{c}_P$ и называют оценками, полученными методом статистических моментов. Такие оценки могут иметь не самые лучшие свойства при малых объемах выборки, но являются асимптотически

несмещенными и состоятельными [93], так что при больших объемах выборки вполне могут использоваться.

2.2.1.7. Метод максимального правдоподобия. В большинстве случаев оценки с наилучшими свойствами обеспечивает метод максимального правдоподобия (ММП). Он состоит в следующем. Выборка есть случайный вектор размерности N: $\mathbf{x} = (x_1,...,x_N)$, который характеризуется некоторой плотностью распределения, зависящей от значения вектора параметров $\mathbf{c}$. Обозначим эту условную плотность $p_N(\mathbf{x}|\mathbf{c})$. Подберем такие значения параметров $\mathbf{c} = \hat{\mathbf{c}}$, при которых плотность вероятности появления *имеющейся* выборки максимальна, т.е. когда ее появление есть наиболее правдоподобное событие. Эти значения $\hat{\mathbf{c}}$ и называются оценками максимального правдоподобия (МП-оценками).

Функция $L(\mathbf{c}) = p_N(\mathbf{x}|\mathbf{c})$, где $\mathbf{x}$ – фиксированный вектор (имеющаяся выборка) называется *функцией правдоподобия*. Заметим, что ее нельзя трактовать как плотность вероятностей для значений параметров $\mathbf{c}$, т.к. эти параметры по постановке задачи являются фиксированными числами, а не случайными величинами. Поэтому и был введен специальный термин – функция правдоподобия. МП-оценки доставляют ей максимальное значение: $L(\hat{\mathbf{c}}) = \max\limits_{\mathbf{c}\in A} L(\mathbf{c})$. Необходимое условие максимума имеет вид:

$$\partial L(\mathbf{c}) / \partial c_j = 0,\ j = 1,..., P. \tag{2.24}$$

Часто удобнее работать с логарифмом функции правдоподобия. Он имеет максимум в той же точке, что и $L(\mathbf{c})$, поэтому МП-оценки ищут из уравнений

$$\partial \ln L(\mathbf{c}) / \partial c_j = 0,\ j = 1,..., P, \tag{2.25}$$

которые и называют *уравнениями правдоподобия*.

В рассматриваемом случае выборки, состоящей из независимых значений, функция правдоподобия равна произведению плотностей вероятности каждого значения x_i, а логарифмическая функция правдоподобия равна сумме логарифмов:

$$\ln L(\mathbf{c}) = \sum_{i=1}^{N} \ln p(x_i|\mathbf{c}). \tag{2.26}$$

В рассматриваемом случае МП-оценки асимптотически не смещены, состоятельны, и имеют минимальную дисперсию.

Для нормального распределения величины ξ логарифмическая функция правдоподобия имеет вид:

$$\ln L(a,\sigma^2) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(X_i - a\right)^2 . \qquad (2.27)$$

Нетрудно убедится, что МП-оценки параметров a и σ^2 совпадают с эмпирическим средним (2.18) и эмпирической дисперсией (2.20). МП-оценка дисперсии смещена, но при увеличении N она стремится к несмещенной оценке (2.21), т.е. является асимптотически несмещенной. Можно показать, что МП-оценка $\hat{a}$ тоже распределена в этом случае по нормальному закону с математическим ожиданием a и дисперсией σ^2/N (рис.2.7). Отсюда видно, что при большом числе испытаний значение эмпирического среднего приближается к истинному значению a, поскольку дисперсия распределения уменьшается с ростом N. В частности, вероятность того, что $\left|\hat{a} - a\right| > 1.96\sigma/\sqrt{N}$, равна 0.95.

Интервал $[\hat{a} - 1.96\sigma/\sqrt{N}, \hat{a} + 1.96\sigma/\sqrt{N}]$ называется 95%-ным доверительным интервалом для величины a. Чем больше N, тем уже этот интервал. Для его оценки можно заменить истинное значение σ на оценку $\hat{\sigma}$. Оценка $\hat{a}$ называется *точечной* (это одно число – одна точка). Если же указывается интервал наиболее правдоподобных значений параметра, то оценка называется *интервальной*. Интервальные оценки очень желательны, т.к. по значению точечной оценки нельзя сказать, насколько оно может отклоняться от истинного.

2.2.1.8. Когда несостоятелен ММП? Обсудим одну из ситуаций, когда ММП может дать асимптотически смещенные оценки. Например, это задача исследования зависимости между двумя переменными, когда значения *обеих* известны с ошибками. Она изучается в разделе статистики, называемом конфлюэнтным анализом [1, 93]. Пусть имеется некоторая случайная величина Z, а также величины X и Y, связанные с Z как [287]:

$$\begin{aligned} X &= Z + \xi, \\ Y &= Z + \eta, \end{aligned} \qquad (2.28)$$

где ξ и η – независимые друг от друга и от X, Y нормально распределенные случайные величины с нулевым математическим ожиданием и одинаковой дисперсией σ^2. Можно сказать, что X и Y – это величины, полученные при измерении Z двумя независимыми способами. Есть выборка значений величин X и Y, полученная в результате независимых испытаний $\{x_i, y_i\}_{i=1}^{N}$. Нужно оценить дисперсию ошибок измерений σ^2.

Наиболее простой способ получения оценки – заметить, что величина $X - Y = \xi - \eta$ является нормально распределенной случайной величиной с нулевым математическим ожиданием и дисперсией $2\sigma^2$ (дисперсия суммы

двух независимых величин равна сумме дисперсий). Тогда по выборке $\{x_i - y_i\}_{i=1}^N$ нетрудно получить состоятельную оценку дисперсии $X - Y$:

$$\hat{\sigma}^2_{X-Y} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 , \tag{2.29}$$

следовательно, σ^2 можно оценить как

$$\hat{\sigma}^2 = \frac{1}{2N}\sum_{i=1}^{N}(x_i - y_i)^2 . \tag{2.30}$$

При непосредственном применении ММП (без перехода к новой переменной) приходится записать функцию правдоподобия, включив в число оцениваемых величин ненаблюдаемые значения величины Z:

$$L(x_1, y_1, ..., x_N, y_N | z_1, ..., z_N, \sigma) = \frac{1}{\left(2\pi\sigma^2\right)^N} \exp\left(-\sum_{i=1}^{N} \frac{(x_i - z_i)^2 + (y_i - z_i)^2}{2\sigma^2} \right).$$

Решая уравнение правдоподобия, получим оценки

$$\hat{z}_i = (x_i + y_i)/2 , i = 1, ..., N , \tag{2.31}$$

$$\hat{\sigma}^2_{МП} = \frac{1}{4N}\sum_{i=1}^{N}(x_i - y_i)^2 . \tag{2.32}$$

Таким образом, МП-оценка дисперсии вдвое меньше несмещенной оценки (2.30) при любом N, т.е. она смещена и несостоятельна. В чем же коренное отличие этой задачи? Оно состоит в том, что число оцениваемых величин (равное $N + 1$) с ростом объема выборки растет! А в предыдущем случае оценивалось фиксированное число параметров. Вообще говоря, чем меньше величин оценивается, тем лучше свойства получаемых оценок.

2.2.1.9. Байесовское оценивание. Очень широкая область теории статистического оценивания относится к случаю, когда истинные значения параметров $\mathbf{c}_0$ также представляют собой случайные величины, т.е. они могут меняться от выборки к выборке в соответствии с плотностью распределения $p(\mathbf{c}_0)$, которую называют *априорной*. Если априорная плотность известна, то разумно учесть эту информацию при оценивании. Соответствующие подходы называют *байесовскими*.[24]

В наиболее распространенном варианте стремятся восстановить закон распределения параметров $\mathbf{c}_0$ при том условии, что реализовалась выборка $x_1, ..., x_N$. Это так называемая *апостериорная* плотность вероятности

[24] По имени английского священника Томаса Байеса (1702-1761), который предложил идею в работе, опубликованной после его смерти.

$p(\mathbf{c}_0|x_1,...,x_N)$. Получить его можно, если известна плотность распределения выборки $p(x_1,...,x_N|\mathbf{c}_0)$ при данном $\mathbf{c}_0$, т.е. известен закон распределения ξ. Тогда апостериорную плотность находят по *правилу Байеса*:[25]

$$p(\mathbf{c}_0|x_1,...,x_N) = \frac{p(\mathbf{c}_0)p(x_1,...,x_N|\mathbf{c}_0)}{\int p(\mathbf{c}_0)p(\eta_1,...,\eta_N|\mathbf{c}_0)d\mathbf{c}_0}. \qquad (2.33)$$

Заметим, что знаменатель не зависит от искомых параметров (по ним проведено интегрирование).

Если получен апостериорный закон распределения, то конкретную точечную оценку $\hat{\mathbf{c}}$ можно найти различными способами, например, как математическое ожидание $\hat{\mathbf{c}} = \int \mathbf{c}_0 p(\mathbf{c}_0|x_1,...,x_N)d\mathbf{c}_0$ или как точку его максимума (моду). При отсутствии сведений об априорной плотности, ее заменяют константой $p(\mathbf{c}_0)$, что соответствует равномерному распределению на очень широком отрезке. Тогда с точностью до множителя, не зависящего от $\mathbf{c}_0$, апостериорное распределение совпадает с функцией правдоподобия. Если при этом байесовскую оценку рассчитывать как моду апостериорного распределения, то получим ММП.

Как правило, при оценивании на практике выдвигают гипотезу о том, какому закону распределения починяется наблюдаемая величина, являются ли независимыми испытания и т.д. Приняв эти предположения, применяют соответствующие методы. Справедливость предположений проверяется статистическими средствами после получения оценки (п. 7.3).

2.2.2. Признаки случайности, традиционные для физиков

Все перечисленные ниже признаки базируются в той или иной степени на понимании случайности как отсутствия «повторяемости» в процессе.

а) Нерегулярная (непериодическая) форма временной реализации. Это наиболее примитивный признак случайности. Здесь она непосредственно противопоставляется периодичности: *нет периода – случайность, есть – детерминированность.*

б) Спадающие корреляции. Этим признаком является убывание автокорреляционной функции $\rho(\tau)$ до 0 с ростом τ. Для стационарного процесса (подробнее см. главу 4) с нулевым средним она определяется как

[25] Фактически, это запись совместной вероятности двух событий А и В в двух эквивалентных вариантах $P\{A \cap B\} = P\{A\}P\{B|A\} = P\{B\}P\{A|B\}$, откуда следует, что $P\{B|A\} = P\{B\}P\{A|B\}/P\{A\}$.

$\rho(\tau) = \langle x(t)x(t+\tau)\rangle / \langle x(t)x(t)\rangle\langle x(t+\tau)x(t+\tau)\rangle$. Угловые скобки означают усреднение по ансамблю, для эргодических процессов оно совпадает с усреднением по времени (п. 4.1.3). Обычно используют нормировку: $\rho(0) = 1$. Этот признак также дает, по существу, количественную меру отклонения процесса от периодического. С его помощью нельзя выявить периодичность, если период $T > T_H$, где T_H – время наблюдения.

в) Сплошной спектр. Согласно этому признаку случайным называют процесс со сплошным, непрерывным спектром; спектр периодического сигнала дискретен. Конечность времени наблюдения T_H накладывает ограничение на спектральное разрешение: $\Delta\omega_{\min} = 2\pi/T_H$. Увеличивая время наблюдения T_H, наблюдатель, в конце концов, установит конечность ширины спектральной линии любого реального процесса и, строго говоря, по всем перечисленными признаками (а-в) будет вынужден считать случайным любой реальный процесс.

г) Неупорядоченность множеств в восстановленном «фазовом пространстве»: отсутствие признаков конечной размерности и т.д. Это уже более тонкие характеристики, которые не связаны с выявлением только непериодичности.

Существуют и некоторые скорее качественные критерии: невоспроизводимость процесса и его неконтролируемость, т.е. невозможность создать условия, при которых процесс протекал бы одинаковым или заранее предписанным образом, соответственно.

2.2.3. Алгоритмический подход

Алгоритмический подход интерпретирует «отсутствие закономерностей», как чрезмерную сложность алгоритма, требуемого для воспроизведения данного процесса в цифровом виде. Впервые, идея связать случайность со сложностью была выдвинута А.Н. Колмогоровым и независимо –Чейтиным и Соломоновым.

Любой процесс можно представить последовательностью нулей и единиц, т.е. записать в двоичной системе: $\{y_i\}$, i = 1, 2, …, N. Колмогоров предложил считать мерой сложности такой последовательности длину минимальной программы l (в битах), которая воспроизводит последовательность $\{y_i\}$. Например, программа для репродуцирования последовательности 101010…10 (сто пар символов «10») будет совсем короткой: «напечатать “10” сто раз». Если же единицы и нули расположены случайно, программа будет состоять в том, чтобы транслировать последовательность посимвольно: она не сжимается. Таким образом, $l \sim N$ для случайных последовательностей и $l << N$ для неслучайных.

К сожалению, в общем случае не существует способа установить минимальную длину программы на практике.[26] На основе идеи алгоритмической сложности развивались новые подходы к понятиям сложности и случайности. В последние годы все более популярен взгляд, связывающий эти понятия с *предсказуемостью* [96, 305, 188].

2.2.4. Случайность как непредсказуемость

Случайность или детерминированность процесса связывается в [96, 97] с возможностью его предсказания с помощью имеющейся модели. Пусть имеется регистрируемый процесс $x(t)$ и модельный процесс $z(t)$. Для простоты примем $\langle x(t)\rangle = \langle z(t)\rangle = 0$. В текущий момент времени $t = t_0$ эти величины принимают значения $x = x_0, z = z_0$. Естественно задавать модельный процесс так, чтобы $z_0 = x_0$, а качество прогноза оценивать разницей $x(t) - z(t) = \Delta(t)$ – ошибкой прогноза, $\Delta(t_0) = 0$. Подход основан на статистическом описании пары x, z.

В типичном случае значения x и z с течением времени «разойдутся», так что абсолютная величина $\Delta(t)$ возрастет. Повторив эксперименты и процедуру сравнения $x(t)$ и $z(t)$, можно набрать ансамбль и найти распределение вероятностей $p(x,z,t,x_0,z_0,t_0)$ и $p(\Delta,t,x_0,z_0,t_0)$. При таком описании модель $z(t)$ включается в статистику наряду с регистрируемым процессом. Мерой степени предсказуемости могут выступать:

1) средний квадрат ошибки $\sigma_\Delta^2(\tau) \equiv \langle \Delta^2(\tau)\rangle = \langle |x(t) - z(t)|^2 \rangle$, где $t = t_0 + \tau$. $\sigma_\Delta^2(0) = 0$, а при $\tau \to \infty$, если процессы $x(t)$ и $z(t)$ станут статистически независимыми $\langle x(t)z(t)\rangle = 0$, то $\sigma_\Delta^2(\tau) = \langle x^2(t)\rangle + \langle z^2(t)\rangle$. Здесь предполагается, что x и z ограничены. Относительную ошибку разумно определить как $E(\tau) = \sigma_\Delta^2(\tau) \Big/ \left(\langle x^2(t)\rangle + \langle z^2(t)\rangle\right)$; тогда $E \to 1$ при $t \to \infty$.

2) взаимная корреляционная функция исходного и модельного процессов $D(\tau) = \langle x(t_0+\tau)z(t_0+\tau)\rangle \Big/ \sqrt{\langle x^2(t_0+\tau)\rangle\langle z^2(t_0+\tau)\rangle}$. $D(0) = 1$ и всегда $|D(\tau)| \le 1$. Используя известные статистические соотношения, можно выразить: $D(\tau) = \dfrac{\langle x^2(t_0+\tau)\rangle + \langle z^2(t_0+\tau)\rangle}{2\sqrt{\langle x^2(t_0+\tau)\rangle\langle z^2(t_0+\tau)\rangle}}\left(1 - E(t)\right)$.

---

[26] По тем же фундаментальным причинам, что и в теореме Геделя о неполноте любой системы аксиом, см., например, работу [305] и ссылки в ней.

Таким образом, качество предсказуемости можно выразить через различные схожие величины. Оценка процесса как случайного или детерминированного определяется возможностью его предсказания с помощью имеющейся модели. Здесь случайно то, что мы по какой-то причине не можем предсказать (из-за свойств системы – характера $x(t)$, или из-за вида $z(t)$, определяемого моделью, или из-за отсутствия модели). Такой подход к случайности популярен в методологии, он развивался в теории различения гипотез для нужд радиолокации.

2.3. Концепция частичной детерминированности

Концепция частичной детерминированности основана на соглашении, что в качестве признака случайности (детерминированности) выбирается непредсказуемость (предсказуемость) наблюдаемого процесса $x(t)$ на основе определенной прогностической модели $z(t)$. При этом случайность и детерминированность не противопоставляются друг другу, а рассматриваются как полюса единого свойства – частичной детерминированности.

В качестве величины, характеризующей степень детерминированности (предсказуемости) удобно использовать взаимную корреляцию $D(\tau)$ наблюдаемого и модельного процессов. Ее характерный график изображен на рис.2.8. Здесь D – область полной детерминированности; DC – область частичной детерминированности, C – область случайного (непредсказуемого) поведения. Наблюдаемый процесс $x(t)$ выступает как детерминированный (предсказуемый) при $D \approx 1$; как непредсказуемый при $|D| << 1$; как частично предсказуемый при $0 < |D| < 1$.

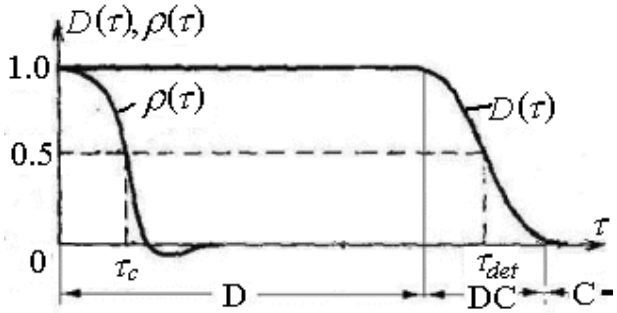

Рис.2.8. Типичное соотношение между степенью детерминированности $D(\tau)$ и автокорреляционной функцией

Время $\tau = \tau_{\text{det}}$, при котором степень предсказуемости спадает до определенной пороговой величины, например, $D(\tau_{\text{det}}) = 1/2$, называют временем детерминированного поведения. Что влияет на эту величину? Для реальной системы оно всегда ограничено из-за следующих причин:

- наблюдаемый процесс $x(t)$ всегда отличается от исследуемого процесса $\eta(t)$ из-за влияния приборов – регистрирующих устройств (измерительный шум $\nu(t)$);

- существуют неучтенные внешние воздействия $\mu(t)$ (так называемый «динамический шум»), случайные и/или неслучайные. Последние, в принципе, можно ввести в качестве поправки в модель;
- модель не адекватно отражает свойства объекта («шумы незнания» $\Delta M(t)$). Это связано с ее структурой или значениями параметров.

Поэтому $\tau_{\det} = f(\nu, \mu, \Delta M)$. Даже, если свести к нулю влияние приборов и погрешности детерминированной компоненты модели, то останутся неустранимые флуктуационные внешние воздействия. Они могут быть связаны хотя бы с бесконечномерностью микроструктуры реальных объектов с шумами различного происхождения, с процессами старения и т.п. и принципиально ограничивают время предсказуемости. Предел $\tau_{\lim} = \lim\limits_{\nu \to 0, \mu \to 0} \tau_{\det} = f(0, \mu, 0)$ называют «горизонтом предсказуемости».

При $\tau >> \tau_{\lim}$ x и z становятся, как правило, статистически независимыми, при этом $D(\tau) \to 0$. Время детерминированного поведения $\tau_{\det}$ может превышать время автокорреляции τ_c процесса $x(t)$, характеризующее скорость спадания автокорреляционной функции: $\rho(\tau) = \langle x(t)x(t+\tau)\rangle / \langle x(t)x(t)\rangle\langle x(t+\tau)x(t+\tau)\rangle$, где угловые скобки означают усреднение по ансамблю. Время τ_c можно оценить как $\tau_c \approx \dfrac{1}{\Delta\omega}$, где $\Delta\omega$ – ширина спектра (например, для белого шума $\Delta\omega \to \infty, \tau_c = 0$, т.е. система сразу «забывает» о прошлом). Время автокорреляции τ_c выступает как наименьшее время детерминированности. Это надо понимать следующим образом: если мы не располагаем динамическим уравнением для модели $z(t)$, то прогноз приходится строить, только опираясь на сам наблюдаемый процесс $x(t)$ в предыдущие моменты времени. Самый простой принцип – «завтра будет то же, что сегодня», т.е. модель $z(t+\tau) = x(t)$. В этом случае $D(\tau) = \rho(\tau)$ и $\tau_{\det} = \tau_c$, т.е. степень детерминированности переходит в степень когерентности. Вообще говоря, может быть $\tau_{\det} > \tau_c$, как это представлено на рис.2.8. Одно и то же явление может демонстрировать вполне детерминированное поведение в рамках одной модели и полностью недетерминированное – в рамках другой.

2.4. Ляпуновские показатели и пределы предсказуемости

2.4.1. Практическая оценка дальности прогноза

Прогноз является одной из очень широко распространенных и наиболее интригующих научных задач. Пределы прогноза многих

процессов, являются, по-видимому, принципиально ограниченными и не очень большими с практической точки зрения. Если исследуемый процесс – хаотический, т.е. близкие траектории экспоненциально разбегаются, то естественно ожидать, что дальность прогноза этого процесса связана со скоростью разбегания близких траекторий. Последняя определяется величиной его старшего ляпуновского показателя Λ_1 (п. 2.1.4). В качестве оценки дальности прогноза для некоторой динамической модели разумно взять величину интервала, на котором малое возмущение (которое определяется различными источниками возмущений в системе и ошибками модели) нарастает до характерных масштабов изменения наблюдаемой величины. Дальность прогноза может быть грубо оценена по формуле:

$$\tau_{\text{pred}} = \frac{1}{2\Lambda_1} \ln \frac{\sigma_x^2}{\sigma_\nu^2 + \sigma_\mu^2 + \sigma_{\Delta M}^2}, \tag{2.34}$$

где σ_μ^2 – дисперсия динамического шума, σ_ν^2 – дисперсия измерительного шума, $\sigma_{\Delta M}^2$ – погрешности модели (дисперсия «шумов незнания»), σ_x^2 – дисперсия наблюдаемой величины [96], а старший ляпуновский показатель Λ_1 положителен. Получить эту формулу можно из следующих соображений. Пусть нам точно известны уравнения исходной системы, но начальные условия мы знаем только с погрешностью ε (измерительный шум). Тогда, взяв в качестве начальных условий модели эти «неправильные значения», получим, что в среднем ошибка прогноза растет как $\varepsilon \cdot e^{\Lambda_1 t}$. Если дальность прогноза определить как время, за которое ошибка прогноза достигает σ_x, то получим $\tau_{pred} = \frac{1}{\Lambda_1} \ln \frac{\sigma_x}{\varepsilon}$. Величину $\tau_\Lambda = \frac{1}{\Lambda_1}$ называют ляпуновским временем. Пусть теперь в системе имеется не только шум измерений с дисперсией $\sigma_\nu^2 = \varepsilon^2$, но и внешние случайные воздействия, и погрешности модели. Если считать все эти факторы примерно независимыми, то дисперсия итогового возмущения равна сумме дисперсий компонент. Заменив ε в последней формуле для τ_{pred} на корень квадратный из суммы квадратов дисперсий компонент, получим выражение (2.34).

Если шумы и погрешности модели не велики по сравнению с уровнем сигнала, то (2.34) может значительно превышать время корреляции процесса, которое в ряде случаев может быть грубо оценено как $\tau_c \sim 1/\Lambda_1$. Так, если уровень сигнала больше уровня шумов по среднеквадратичному отклонению в 1000 раз, то дальность прогноза по (2.34) превосходит время корреляции примерно в 7 раз.

Однако формула (2.35) в качестве оценки дальности прогноза применима не всегда. Дело в том, что любое конечное возмущение в хаотическом режиме через определенный промежуток времени нарастает до размеров, при которых линеаризованная система (2.10) уже не применима. Следовательно, дальнейшая эволюция возмущения уже, строго говоря, не связана с ляпуновским показателем. Так что, если нас интересует прогноз с точностью не очень высокой, а только практически приемлемой, то ляпуновский показатель в общем случае не имеет отношения к ситуации и не может накладывать ограничения на дальность прогноза. Но если предположить, что и на больших масштабах по порядку величины ляпуновский показатель верно характеризует скорость роста возмущений (что во многих случаях имеет место), то с его помощью можно охарактеризовать дальность прогноза даже если нас интересуют конечные масштабы.

Однако при строгом рассмотрении оказывается, что даже в пределе бесконечно малых возмущений величина ляпуновского времени не всегда связана со временем предсказуемости системы. Рассмотрим подробнее этот интересный факт.

2.4.2. Предсказуемость и ляпуновский показатель: случай малых возмущений

В принципе временем предсказуемости можно по определению назвать величину (2.34). Но могут быть и другие подходы. Один из них, весьма простой и логичный, состоит в следующем [318]. Рассмотрим, как ведет себя возмущение при заданном начальном условии $\mathbf{x}_0$. Согласно определению локальных ляпуновских показателей имеем в худшем (максимальный рост возмущения) случае $\|\boldsymbol{\varepsilon}(t_0+\Delta t\| = \|\boldsymbol{\varepsilon}_0\| e^{\lambda_1(\mathbf{x}_0,\Delta t)\cdot\Delta t}$. Определим время предсказуемости как время увеличения начального малого возмущения в q раз. Смысл этого утверждения тот же самый, что и в определении его как времени, за которое начальное возмущение нарастает до некоторого порогового значения. Обозначим это время $\tau_q(\mathbf{x}_0) = \dfrac{\ln q}{\lambda_1(\mathbf{x}_0,\Delta t)}$; оно зависит от начального условия $\mathbf{x}_0$. Чтобы получить общую характеристику предсказуемости, усредним $\tau_q(\mathbf{x}_0)$ по инвариантной мере $p(\mathbf{x}_0)$, т.е. плотности распределения вероятностей на аттракторе. Получим

$$\tau_q \equiv \int p(\mathbf{x}_0)\tau_q(\mathbf{x}_0)d\mathbf{x}_0 . \qquad (2.35)$$

Это определение существенно отличается от (2.34). Если бы мы определяли время увеличения ошибки в q раз через старший ляпуновский показатель, то получили бы, что оно равно:

$$\tau_{q,\Lambda} \equiv \frac{\ln q}{\Lambda_1} = \frac{\ln q}{\int p(\mathbf{x}_0)\lambda_1(\mathbf{x}_0)d\mathbf{x}_0} = \frac{1}{\int p(\mathbf{x}_0)\dfrac{1}{\tau_q(\mathbf{x}_0)}d\mathbf{x}_0}. \qquad (2.36)$$

Здесь для ляпуновского показателя, т.е. величины в знаменателе, использовано выражение, полученное усреднением с естественной мерой,[27] что для эргодической системы эквивалентно усреднению по времени.

Теперь ситуация аналогична следующей. Пусть имеются значения случайной величины $x_1, x_2, ..., x_N$, и нужно оценить ее математическое ожидание $M[x]$. Наиболее простой путь – рассчитать эмпирическое среднее, которое является «хорошей» оценкой: $\langle x \rangle = (x_1 + ... + x_N)/N$. Это – аналог формулы (2.35) для среднего времени предсказуемости. Но можно составить и много других формул для оценки. Например, если подсчитаны обратные значения $1/x_1, 1/x_2, ..., 1/x_N$, то можно оценить величину $1/M[x]$ как их эмпирическое среднее, и взять обратную величину. Итоговая оценка $\langle x' \rangle = 1/[N(1/x_1 + 1/x_2 + ... + 1/x_N)]$ является аналогом (2.34). Но среднее обратных величин – это в общем случае смещенная оценка $1/M[x]$! Поэтому и $\langle x' \rangle$ – «плохая» оценка для $M[x]$. Величины $\langle x \rangle$ и $\langle x' \rangle$ совпадают только при $x_1 = x_2 = ... = x_N$. Для нашей задачи это означает, что ляпуновское время совпадает с τ_q (с точностью до множителя $\ln q$), только если локальный ляпуновский показатель не зависит от $\mathbf{x}_0$, т.е. траектории разбегаются с одной и той же скоростью в любой области фазового пространства. Это – условие применимости формулы (2.34) даже в линейном случае.

Итак, время предсказуемости можно и даже более логично определить не через ляпуновский показатель. Мы покажем ниже, что ляпуновский показатель не связан со временем предсказуемости τ_q, т.е. система с большим ляпуновским показателем (более хаотичная) может иметь большее τ_q, чем менее хаотическая система. Кроме того, системы с одинаковыми значениями ляпуновского показателя могут иметь самые различные τ_q. Покажем это *аналитически* на простом примере, следуя

[27] Грубо, это плотность вероятности p посещения различных областей аттрактора изображающей точкой, см., например, [104].

[318]. Будем говорить для определенности о времени удвоения возмущения τ_2.

Рассмотрим сначала пример, когда ляпуновское время и τ_2 совпадают (с точностью до множителя $\ln 2$). Возьмем двумерное нелинейное отображение, которое является одной из базовых моделей нелинейной динамики, – отображение пекаря:

$$x_{n+1} = \begin{cases} \dfrac{1}{\alpha} x_n, & 0 \le x_n < \alpha, \\ \beta(x_n - \alpha), & \alpha \le x_n < 1, \end{cases}$$
$$y_{n+1} = \begin{cases} \alpha y_n, & 0 \le x_n < \alpha, \\ \alpha + \dfrac{1}{\beta} y_n, & \alpha \le x_n < 1, \end{cases} \qquad (2.37)$$

где $\alpha = 1/\beta = 1/2$. Это отображение сохраняет площадь (консервативное) и отображает множество $[0,1) \times [0,1)$ на себя. Инвариантная мера удовлетворяет соотношению $p(x, y) = 1$. Оно названо отображением пекаря, потому что его действия над единичным квадратом напоминают действия пекаря с куском теста. Сначала тесто сжимается вдоль оси *y* в два раза и в два раза растягивается вдоль оси *x*; затем разрезается пополам и правый кусок кладется сверху на левый (параллельным переносом). Все это умещается в одной итерации отображения (см. рис.2.9). Две близкие точки, отличающиеся только значениями координаты *x*, переходят в две точки, расстояние между которыми в два раза больше, для любого начального условия на плоскости. Аналогично, расстояние вдоль оси *y* в два раза уменьшается за одну итерацию. Так что для любой точки квадрата старшему локальному ляпуновскому показателю соответствует направление оси *x*. Он не зависит от интервала Δt и просто равен старшему ляпуновскому показателю: это система с равномерной скоростью разбегания близких траекторий. Поскольку в данном случае $\Lambda_1 = \ln 2$, то ляпуновское время равно $\tau_\Lambda = 1/\ln 2$.

Время $\tau_2(\mathbf{x}_0)$ для каждого начального условия равно 1, т.е. возмущение удваивается за одну итерацию. Соответственно, и среднее $\tau_2 = 1$.

Рис.2.9. Одна итерация отображения пекаря, квадрат раскрашен черным цветом, чтобы показать, куда попадают точки из разных областей

Рассмотрим теперь модификацию системы (2.37), названную отображением подмастерья пекаря:

$$x_{n+1} = \begin{cases} \dfrac{1}{\alpha} x_n, & 0 \le x_n < \alpha, \\ \beta(x_n - \alpha) \bmod 1, & \alpha \le x_n < 1, \end{cases}$$
$$y_{n+1} = \begin{cases} \alpha y_n, & 0 \le x_n < \alpha, \\ \alpha + \dfrac{1}{\beta}\left(\left[\beta(x_n - \alpha)\right] + y_n\right), & \alpha \le x_n < 1, \end{cases} \qquad (2.38)$$

где квадратные скобки означают целую часть числа, $\alpha = \left(2^N - 1\right)/2^N$, $\beta = 2^{2^N}$. Действие этого отображения состоит в том, что больший кусок теста $[0,1/\alpha)\times[0,1)$ сжимается очень слабо в α раз вдоль оси y и растягивается во столько же раз вдоль оси x, переходя в кусок $[0,1)\times[0,1/\alpha)$. Правая же узкая полоска $[1/\alpha,1)\times[0,1)$ сжимается очень сильно в β раз. При этом получается узкая лента длиной в целое число единиц, эта лента разрезается на ленточки единичной длины, которые укладываются сверху «штабелем» на большой кусок $[0,1)\times[0,1/\alpha)$. Левые ленточки располагаются ниже, чем правые (рис.2.10).

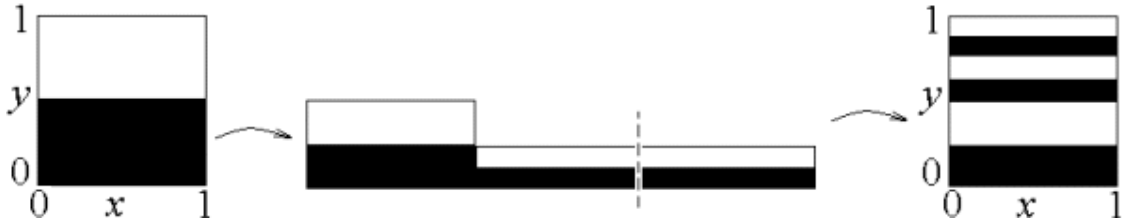

Рис.2.10. Одна итерация отображения подмастерья пекаря для N = 1 (самый «мастеровитый» подмастерье)

Эта система также сохраняет площадь и имеет инвариантную меру $p(x,y) = 1$. Ее направления, отвечающие большему и меньшему локальным показателям, также совпадают с направлениями координатных осей. Можно показать, что ее старший ляпуновский показатель равен $\Lambda_1 = \alpha \ln 1/\alpha + (1-\alpha)\ln\beta$, и при данных значениях параметров $\Lambda_1 = 1 - \alpha \ln\alpha > 1$. Таким образом, система (2.38) является более хаотической, чем (2.37), в смысле большего ляпуновского показателя.

Ее локальные ляпуновские показатели сильно зависят от начального условия. Для области с малыми значениями x они очень малы (время предсказуемости очень велико), зато для области с большими x они очень велики (но эта область мала). Пример действия обоих отображений на множество точек внутри квадрата представлен на рис.2.11. Слева показан

результат 4 итераций отображения пекаря (2.37). Картинка полностью исказилась (т.е. предсказуемость плохая). Справа показан результат 4 итераций отображения (2.38) с $N = 4$. Значительная часть картины хорошо сохранилась, лишь слегка деформировалась: в этой области предсказуемость хорошая.

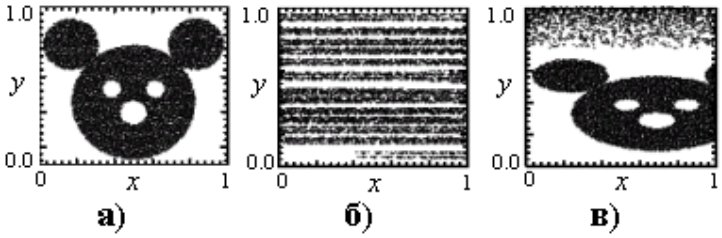

Рис.2.11. Иллюстрация динамики отображений (2.37) и (2.38): а) исходное множество точек; б) его образ под действием 4 итераций отображения пекаря (2.37); в) под действием 4 итераций отображения подмастерья пекаря (2.38) с $N = 4$

Особенно интересно в этом примере следующее. Не только локальные времена предсказуемости $\tau_2(\mathbf{x}_0)$ в некоторых областях больше для отображения (2.38), чем для (2.37). Среднее время τ_2 для (2.38) также больше, хотя это отображение имеет больший ляпуновский показатель! Величину τ_2 можно найти аналитически: $\tau_2 = \dfrac{1-\alpha^j}{1-\alpha}$, где $j = \left[-\dfrac{\ln 2}{\ln \alpha}\right]$ и квадратные скобки означают наименьшее целое число, большее или равное дроби в скобках. Можно показать, что при $N \to \infty$ время предсказуемости $\tau_2 \approx 2^{N-1} \to \infty$, а $\Lambda_1 \to \ln 2$. Результаты аналитических выкладок для некоторых N сведены в табл.2.1: для систем столь же хаотичных, как и (2.37) (и даже с чуть большим ляпуновским показателем) время предсказуемости может быть сколь угодно велико!

Таблица 2.1.

Характеристики отображений (2.38) при разных N: максимальное локальное время предсказуемости, среднее время предсказуемости, ляпуновский показатель

N	$\tau_{2,\max}$	τ_2	Λ_1	N	$\tau_{2,\max}$	τ_2	Λ_1
1	1	1.00	$1.5 \cdot \ln 2$	5	22	16.09	$1.04 \cdot \ln 2$
2	3	2.31	$1.31 \cdot \ln 2$	6	45	32.49	$1.02 \cdot \ln 2$
3	6	4.41	$1.17 \cdot \ln 2$	7	89	64.32	$1.01 \cdot \ln 2$
4	11	8.13	$1.09 \cdot \ln 2$				

Это означает, что ляпуновский показатель еще не исчерпывает вопроса о предсказуемости системы. Хотя, конечно, определенную информацию он несет, а если скорость разбегания траекторий однородна по пространству, то он уместен для характеристики предсказуемости.

2.5. Масштабы рассмотрения: как они определяют оценку свойств процесса (сложная динамика или случайность)

На практике данные измеряются с конечной точностью, т.е. сколь угодно малые масштабы рассмотрения недоступны. При этом ответить на вопрос, является ли наблюдаемое нерегулярное поведение детерминированным хаотическим или стохастическим (случайным) не просто. Строго говоря, ответить можно, только если данные сгенерированы нами на компьютере, и потому известно, какому закону они подчиняются. В реальности же следует говорить о том, какое из двух модельных представлений более адекватно. В работе [210] предложен конструктивный подход, согласно которому ответ на вопрос о детерминированности/случайности зависит от выбора масштаба рассмотрения.

Для количественной характеристики эволюции возмущений масштаба ε в ДС (2.9) предложено использовать ляпуновский показатель для конечных отклонений $\lambda(\varepsilon)$. Он показывает, как быстро разбегаются траектории, изначально разделенные величиной ε. Причем конечные отклонения уже не описываются в общем случае линеаризованным уравнением (2.10). Для его расчета сначала необходимо ввести норму (длину) векторов состояния. В отличие от случая бесконечно малых возмущений, здесь численное значение $\lambda(\varepsilon)$ зависит от выбора нормы. Будем для определенности говорить о евклидовой норме и обозначим норму начального возмущения $\|\boldsymbol{\varepsilon}(0)\| = \varepsilon_0$. Будем следить, за какое время величина возмущения достигнет некоторых пороговых значений $\varepsilon_1, \varepsilon_2, ..., \varepsilon_P$. Например, зададим пороги как $\varepsilon_n = 2\varepsilon_{n-1}$, $n = 1, ..., P-1$, и будем говорить о времени удвоения возмущения для разных масштабов $\tau_2(\varepsilon_n)$ (все эквивалентно для рассмотрения времен нарастания возмущения в q раз). Проведем N экспериментов, «запуская» из различных начальных условий соседние траектории, разделенные расстоянием ε_0. Для каждой отдельной траектории получим свое время удвоения $\tau_2^{(j)}(\varepsilon_n)$, j = 1, …, N. Среднее время удвоения определим как $\tau_2(\varepsilon_n) = (1/N)\sum_{j=1}^{N}\tau_2^{(j)}(\varepsilon_n)$. Ляпуновский показатель для конечных отклонений равен $\lambda(\varepsilon_n) = \ln 2/\tau_2(\varepsilon_n)$.

Если процесс детерминированный хаотический и скорость разбегания фазовых траекторий одинакова по всему пространству, то $\lim_{\varepsilon\to 0}\lambda(\varepsilon) = \Lambda_1$

(см. п. 2.4.2).[28] Важно заметить, что для детерминированного процесса $\lambda(\varepsilon)$ при малых ε практически перестает зависеть от масштаба: $\lambda(\varepsilon) = const$. Для стохастических процессов величина $\lambda(\varepsilon) \to \infty$ при $\varepsilon \to 0$. Закон роста $\lambda(\varepsilon)$ с уменьшением ε может быть различным, например, для броуновского движения (винеровский процесс, см. главу 4) имеет место $\lambda(\varepsilon) \propto \varepsilon^{-2}$.

Авторы [210] предлагают следующий подход к разделению сигналов на детерминированные хаотические и шумовые (случайные). Если при анализе реального процесса в некотором интервале масштабов наблюдается $\lambda(\varepsilon) = const$, то в этом интервале масштабов процесс разумно описать как детерминированный. Если в каком-то интервале масштабов $\lambda(\varepsilon)$ растет с уменьшением ε, то в данном интервале процесс целесообразно рассматривать как шум. Проиллюстрируем подход на простом примере.

Рассмотрим детерминированное отображение, которое демонстрирует «случайное блуждание» (диффузию) на больших масштабах:

$$x_{n+1} = [x_n] + F(x_n - [x_n]), \tag{2.39}$$

где квадратные скобки означают целую часть числа, $F(y) = \begin{cases} (2+\delta)y, & 0 \le y < 0.5, \\ (2+\delta)y - (1+\delta), & 0.5 \le y < 1.0. \end{cases}$ График функции F показан на рис.2.12 для $\delta = 0.4$. Ляпуновский показатель равен $\Lambda_1 = \ln|F'| = \ln|2+\delta|$. Если рассматривать масштабы $\varepsilon > 1$, то процесс ведет себя как винеровский. Например, $\varepsilon = 1$ означает слежение только за целой частью x. Изменение целой части на ± 1 определяется детерминированной хаотической динамикой дробной части x (а она игнорируется при рассмотрении на большом масштабе) и поэтому выглядит как случайное блуждание. На рис.2.13 видно, что в интервале масштабов $\varepsilon < 1$ имеет место $\lambda(\varepsilon) \approx 0.9$ и процесс классифицируется как детерминированный. А при $\varepsilon > 1$ имеем $\lambda(\varepsilon) \propto \varepsilon^{-2}$ и процесс

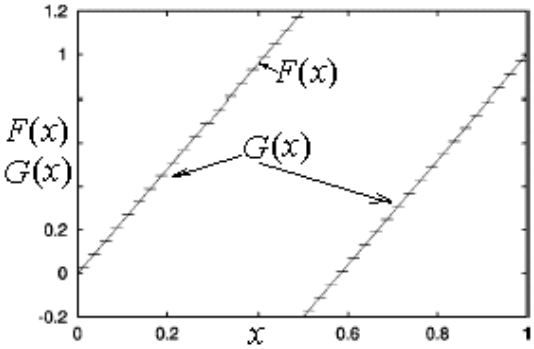

Рис.2.12. Функция $F(x)$ из (2.39). Горизонтальные линии – ее аппроксимация $G(x)$ из (2.40), состояшая из 40 кусков с наклоном 0

[28] Заметим, что для процессов с $\Lambda_1 < 0$ ляпуновский показатель для конечных отклонений, определенный через времена удвоения, равен нулю.

классифицируется как случайный.

Модифицируем отображение (2.39), добавив в него шум ξ_n, равномерно распределенный на $[-1,1]$, и заменив F на ее аппроксимацию G (10000 линейных кусков с наклоном 0.9 вместо двух с наклоном 2.4):

$$x_{n+1} = [x_n] + G(x_n - [x_n]) + \sigma\xi_n, \qquad (2.40)$$

величина $\sigma = 10^{-4}$ определяет дисперсию шума. На интервале $\varepsilon > 1$ процессы (2.39) и (2.40) не различаются и выглядят как случайное блуждание (рис.2.13). В интервале $10^{-4} < \varepsilon < 1$ они выглядят, как детерминированные, причем с одинаковым ляпуновским показателем, несмотря на различный наклон линейных участков – 2.4 в (2.39) и 0.9 в (2.40). Это результат усреднения локальной линейной динамики (2.40) на масштабах $\varepsilon > 10^{-4}$. На масштабах $\varepsilon < 10^{-4}$ процессы различаются. Теперь процесс (2.40) опять ведет себя как случайный с точки зрения $\lambda(\varepsilon)$, теперь уже из-за наличия «настоящего» шума ξ. Таким образом, свойства динамики могут быть различны на различных масштабах, что важно учитывать при описании реальных сложных процессов.

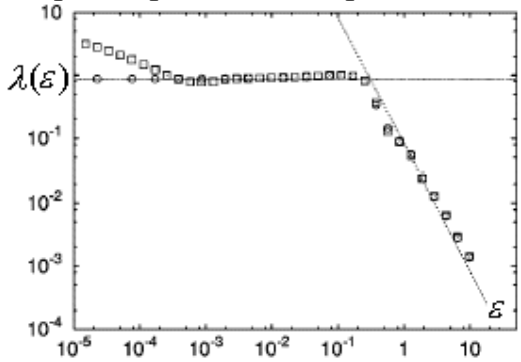

Рис.2.13. Зависимость ляпуновского показателя для конечных отклонений от масштаба. Кружки – для системы (2.39), квадраты – для (2.40) с $G(x)$, состоящей 10000 кусков с наклоном

Наконец, отметим, что те же авторы с опорой на описанный подход предложили остроумные термины для характеристики некоторых нерегулярных процессов: «хаотический шум» и «шумовой хаос». А именно, если процесс на больших масштабах выглядит как детерминированный (хаос), а на маленьких как случайный (шум), то это «шумовой хаос», т.е. макроскопический хаос, вызванный шумом на микроуровне. Аналогично, если процесс случаен на больших масштабах и детерминирован на малых, то это «хаотический шум».

2.6. Пример с монетой

Наверное, всем случалось положить монету на согнутые пальцы (рис.2.14,а), загадать – «орел» или «решка», подбросить и … снять с себя ответственность за решение вопроса. Маленький диск, падающий с вращением, во все времена у многих народов популярен как символ непредвзятости, олицетворение случая. Используем его для иллюстрации

сказанного о детерминированности, случайности, разных подходах к моделированию.

Начнем с содержательной модели. В типичном случае при броске рука придает монете поступательное движение со скоростью v_0, а толчок большим пальцем обеспечивает ей еще и вращение с начальной угловой скоростью ω_0. Далее диск летит, взаимодействуя с Землей и воздухом, и падает на поверхность. Если она твердая, то он несколько раз подпрыгивает и, наконец, успокаивается на одной из своих сторон. Без специальной тренировки вряд ли кому удастся повторить бросок несколько раз так, чтобы монета все время падала, скажем, «орлом» вверх. Это становится невозможным при сильном броске, когда она до приземления успевает сделать много оборотов. Основная причина – большой разброс начальных значений скорости и положения в пространстве. Повторяемости отчасти удается добиться, используя специальное устройство для подбрасывания, например, стальную линейку с приспособлением для воспроизводимого задания деформации[29] (рис.2.14,б). Но и такая установка – не панацея: уверенно предсказать результат удается лишь при слабых бросках, если монета сделала половину, один, от силы два оборота (рис.2.14,в). Чем дальше путь до падения, тем больше неопределенность конечного состояния – частоты появления и «орла» и «решки» сравниваются, хотя условия экспериментов, казалось бы, одинаковы.

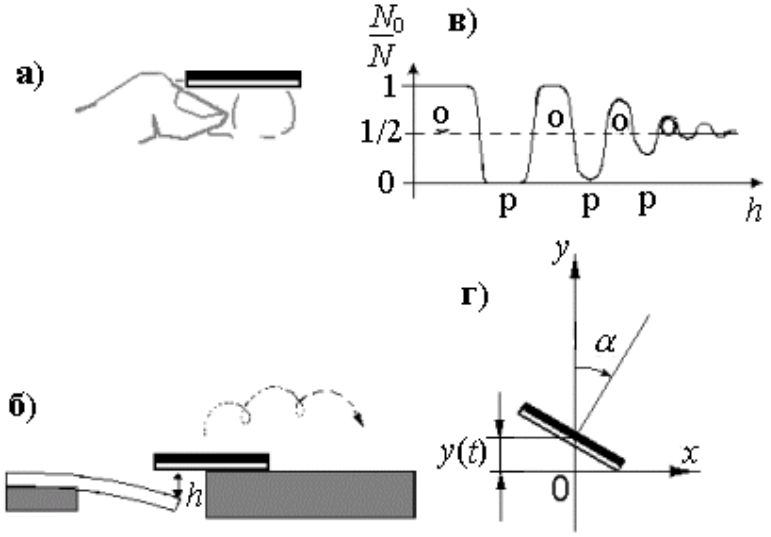

Рис.2.14. Упражнения с монетой: а) стандартная ситуация; б) физическая модель с регулируемой «силой удара» (h – изгиб линейки); в) качественный вид экспериментальной зависимости частоты выпадания «орла» от «силы удара» (в

[29] Упорный студент, предоставивший нам эти экспериментальные данные, подбрасывал монету по 100 раз с помощью линейки. Изгиб ее он регулировал, изменяя число страниц книги, на которую она опиралась.

экспериментах при фиксированном h проводилось N опытов, N_0 – число исходов с «орлом»); г) к содержательной модели. Закрашенная сторона – «орел»

При динамическом моделировании будем характеризовать состояние монеты координатой y и скоростью v ее центра тяжести, а также углом поворота α вокруг оси z, перпендикулярной осям x и y (рис.2.14,г), и угловой скоростью ω. Выделим три качественных этапа эволюции системы, введя на каждом из них свои приближения.

Старт. Начальные условия: монета начинает двигаться, находясь «орлом» вверх, с линейной скоростью v_0, направленной вертикально, вращение происходит по часовой стрелке с угловой скоростью ω_0 (рис.2.14,г). Если при старте $2y < d\sin\alpha$ (край монеты после начала движения давит на опору – «вращение опережает взлет»), будем считать, что она остается на «орле». При $2v_0/\omega_0 > d$ монета улетает, не задевая плоскость $y = 0$.

Полет. Пренебрежем взаимодействием монеты с воздухом. Пусть она взаимодействует только с Землей. Тогда угловая скорость вращения остается неизменной и равной ω_0, а центр тяжести движется с постоянным ускорением g.

Финиш. В момент t_f происходит касание стола, $2y(t_f) = d\sin\alpha(t_f)$, вращение мгновенно прекращается и монета «ложится» на ту или иную сторону в зависимости от значения угла поворота: при $0 < \left(\alpha(t_f)\bmod 2\pi\right) < \pi/2$ и $3\pi/2 < \left(\alpha(t_f)\bmod 2\pi\right) < 2\pi$, выпадает «орел», а при $\pi/2 < \left(\alpha(t_f)\bmod 2\pi\right) < 3\pi/2$ – «решка».

Задание единого оператора эволюции системы на всех этапах движения – задача слишком сложная, поэтому мы ограничимся его рассмотрением лишь на этапе полета, а на первом и последнем этапах ограничимся качественными соображениями. Так, очевидно, что в фазовом пространстве системы имеется множество аттракторов – точек равновесия с координатами $y = v = \omega = 0$, $\alpha = n\pi$, $n = 0,1,2,...$, соответствующих конечному положению монеты, лежащей на одной из сторон (на рис.2.15,а для «орла» точка выделена черными кружками, а для «решки» – белыми). Аттракторам соответствует разное число оборотов до момента падения. В заштрихованных областях фазового пространства, соответствующих финальной стадии и движению с малой начальной скоростью v_0, в соответствии с принятой содержательной моделью имеет место сильная диссипация и изображающая точка попадает на один из аттракторов. Границы бассейнов притяжения аттракторов можно определить из модели движения на стадии полета. Получим ее аналогично [257] асимптотически из системы дифференциальных уравнений Ньютона $\mathbf{F} = m \cdot \mathbf{a}$ и $\mathbf{M} = I \cdot \boldsymbol{\beta}$,

где **F** и **M** – равнодействующая действующих на монету сил и их вращающих моментов, соответственно, **a** и **β** – линейное и угловое ускорения, m и I – масса и момент инерции монеты. В рассматриваемом случае модель имеет вид

$$dy/dt = v, dv/dt = -g, d\alpha/dt = \omega_0, d\omega/dt = 0. \qquad (2.41).$$

Решение (2.41) с заданными начальными условиями дает траекторию

$$y(t) = v_0 t - gt^2/2, v(t) = v_0 - gt, \alpha(t) = \omega_0 t. \qquad (2.42)$$

Отсюда получается зависимость $\alpha(t_f) = f(v_0, \omega_0)$ и выражения для границ бассейнов притяжения аттракторов на плоскости начальных условий ω_0, v_0 (в сечении фазового пространства плоскостью $\alpha = 0$, $y = 0$, рис.2.15,б):

$$\alpha(t_f) = 2\omega_0 v_0 / g = \pi/2 + \pi n. \qquad (2.43)$$

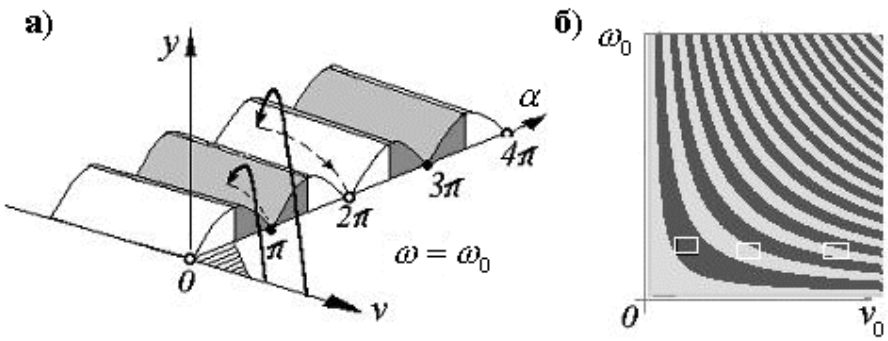

Рис.2.15. а) Трехмерная проекция фазового пространства динамической системы (2.41) при $\omega = \omega_0$. Волнообразной поверхностью ограничена область финальной стадии, в которой имеет место сильная диссипация. Заштрихована область, старт из которой не описывается использованной моделью «полета». Линии со стрелками – примеры фазовых траекторий. б) Сечение фазового пространства плоскостью ($y = 0$, $\alpha = 0$). Заштрихованы бассейны притяжения аттракторов, соответствующих конечному положению «решка». Белые прямоугольники иллюстрируют точность задания начальных условий (уровень шумов), их площадь $\Delta v \times \Delta \omega$

При точном задании начальных условий, что обязательно при динамическом подходе, монета оказывается в определенном конечном состоянии. Согласно этому подходу можно прогнозировать конечное положение монеты, что иллюстрирует рис.2.16,а, где частота выпадания «орла» в зависимости от v_0 принимает только значения 0 и 1. Это соответствует действительности лишь при малых v_0 (рис.2.16,б). Но если бросок достаточно силен, и монета делает несколько оборотов, то такой подход только вводит в заблуждение. Эксперименты показывают, что, даже предприняв меры по увеличению точности задания начальных условий, удается гарантированно обеспечить выпадение орла или решки только при малом числе оборотов монеты. Значительно более

правдоподобная модель получается, если отказаться от динамического описания и ввести в рассмотрение случайные величины. Предположим, что $v_0 = V_0 + \xi$, где V_0 – детерминированная составляющая, ξ – случайная величина, распределенная, например, равномерно в некотором интервале шириной Δv с центром V_0. Такая стохастическая модель демонстрирует зависимость от V_0, качественно совпадающую с экспериментом. При большом числе вращений частота повторения стремится к 0.5, вертикальные и горизонтальные участки графика сглаживаются. При равномерном законе распределения ξ наблюдаемые закономерности удобно объяснить, выделив на плоскости (рис.2.15,б) площадку размером $\Delta v \times \Delta\omega$ с центром V_0. Если площадка целиком укладывается в бассейн притяжения того или иного аттрактора, то всегда происходит событие, которому этот аттрактор соответствует (частота выпадения одной из сторон монеты равна 1). Если площадка включает в себя бассейны притяжения обоих аттракторов («орла» и «решки»), частота выпадения определяется долей их площади. В общем случае частота выпадения «орла» определяется интегралом $P\{O\} = \iint\limits_{O} p(v_0, \omega_0) dv_0 d\omega_0$ по всей области, занимаемой «бассейном притяжения орла», где p – плотность распределения вероятностей для начальных условий.

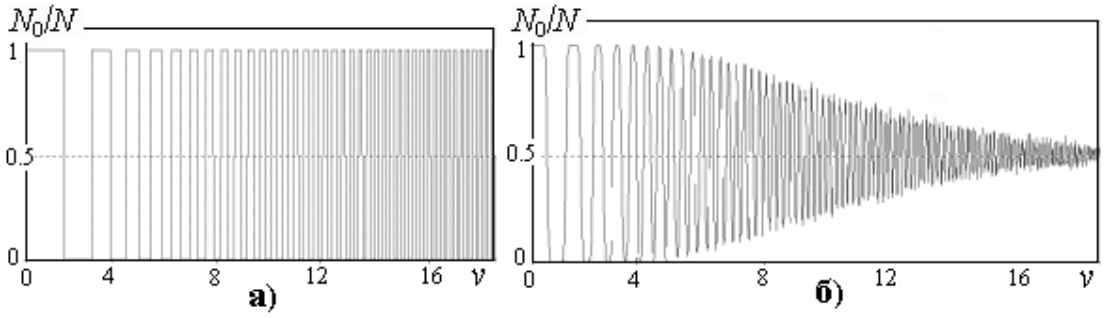

Рис.2.16. Частота выпадания орла в зависимости от начальной скорости v_0 при фиксированном значении ω_0: а) точное задание начальных условий; б) есть погрешность в задании начальных условий

Кроме рассмотренной асимптотической модели и стохастической модели на ее основе, можно предложить и чисто эмпирическую вероятностную модель. Например, можно аппроксимировать экспериментальную зависимость вероятности выпадения «орла» от начальной скорости (или силы подбрасывания) типа рис.2.14,в или рис.2.16,б формулой

$$N_0/N = \begin{cases} z(v), & 0 < z < 1, \\ 1, & z > 1, \\ 0, & z < 0, \end{cases} \tag{2.44}$$

$$z(v) = 0.5 + ae^{-bv}\cos(cv).$$

Мы проиллюстрировали возможность описания одного и того же реального объекта различными моделями: динамическими и вероятностными. Каждая из них имеет право на существование и полезна для достижения определенных целей. Это лишний раз показывает условность ярлыков «динамическая система» или «случайная величина» в приложении к реальным ситуациям. Даже «международный» символ случайности – монету, подброшенную с вращением, – целесообразно рассматривать в общем случае с позиций концепции частичной детерминированности.