Chapter 2
Two Approaches to Modelling and Forecast

Before creation of a model, one should specify one's *intentions* in respect of its *predictive ability*. Such a choice determines which mathematical tools are appropriate. If one does not pretend to a precise and unique forecast of future states, then a *probabilistic approach* is traditionally used. Then, some quantities describing an object under investigation are *declared random*, i.e. fundamentally unpredictable, stochastic.[1] Such a "verdict" may be based on different reasoning (Sect. 2.2) but if it is accepted, one uses a body of the theory of probability and mathematical statistics. At that, to characterise dependence between a condition S and an event A, one speaks only of a probability P of A if S has occurred, i.e. of a conditional probability $P(A|S)$.

A *dynamical approach*, which is an alternative to the probabilistic one, relies on the conception of determinism. *Determinism* is a doctrine about regularity and causation of all phenomena in nature and society. Therein, one assumes that each occurrence of an event S (a cause) inevitably leads to an occurrence of an event A (a consequence). Famous French astronomer, mathematician and physicist Pierre Simon de Laplace (1749–1827) was reputed as the brightest proponent of determinism. In respect of his scientific views, he showed solidity which seemed surprising in view of his inconsistency in everyday attachments[2] (Mathematical dictionary, 1988, p. 117). It was Laplace who told Napoleon that he did not need "a hypothesis about the existence of God" in his theory of the Solar system origin. He saw an etalon of a complete system of scientific knowledge in celestial mechanics and tried to explain the entire world including physiological, psychological, and social phenomena, from the viewpoint of mechanistic determinism.

---

[1] "Stochastic" originates from a Greek word which means "capable of guessing, acute". However, it is currently used in a somewhat different sense to denote uncertainty, randomness.

[2] Several words about the picturesque personality of Laplace. Consistency of his materialistic world view stands in a sharp contrast to his political instability; he took a victor's side at each political upheaval. Initially, he was a republican. After Napoleon came to power, he became a Minister of the Interior and, then, was appointed as a member and vice-president of Senate. In the time of Napoleon, he got the title of count of the empire. He voted for dethronement of Napoleon in 1814. After restoration of Bourbons, he got peerage and a title of marquis.

B.P. Bezruchko, D.A. Smirnov, *Extracting Knowledge From Time Series*, Springer Series in Synergetics, DOI 10.1007/978-3-642-12601-7_2,

Mathematical realisation of the dynamical (deterministic) approach was provided by the apparatus of infinitesimals which appeared in the seventeenth century due to the efforts of Newton and Leibniz. An arsenal of researchers got a powerful tool for the description of temporal evolution: ordinary differential equations (ODEs). A theorem about unique existence of their solution at fixed initial conditions made differential equations an etalon for deterministic mathematical description of various phenomena: "a unique future corresponds to a given present!". Currently, apart from ODEs one widely uses other mathematical tools for construction of deterministic models (Chap. 3) including difference equations, discrete maps and integro-differential equations. All those models regardless of their concrete meaning, which may be far from mechanics (dynamics), are often called *dynamical models*. In general, the term "dynamical" is currently often used to denote "deterministic" rather than "force" or "mobile".

2.1 Basic Concepts and Peculiarities of Dynamical Modelling

2.1.1 Definition of Dynamical System

The basis of deterministic description is an idea that the entire future behaviour of an object is *uniquely* determined by its state at an initial time instant. A rule determining an evolution from an initial state is called *evolution operator*.[3] *State* or *state vector* is a collection of D quantities $\mathbf{x} = (x_1, x_2, \ldots, x_D)$, where D is called dimension. The quantities x_k are called *dynamical variables*. A state may be both finite dimensional (D is a finite number) and infinite dimensional. The latter is the case, e.g., when a state is a spatial distribution of some quantity, i.e. a smooth function of a spatial coordinate.[4]

Evolution operator Φ_t determines a state at any future time instant $t_0 + t$ based on an initial state $\mathbf{x}(t_0)$: $\mathbf{x}(t_0 + t) = \Phi_t(\mathbf{x}(t_0))$. Mathematically, it can be specified with equations, maps, matrices, graphs and any other means (Chap. 3) under the only condition of a *unique* forecast.

The concept of a *dynamical system* (DS) is a key one in the deterministic approach. It was used already by Poincare at the beginning of the twentieth century but its meaning is still not completely established. The term DS is often understood in different ways. Therefore, it is useful to discuss it in more detail. The word

[3] In general, operator is the same as mapping, i.e. a law which relates some element x of a certain given set X to a uniquely determined element y of another given set Y. The term "operator" is often used in functional analysis and linear algebra, especially for mappings in vector spaces. For instance, operator of differentiation relates each differentiable function to its derivative (Mathematical dictionary, 1988).

[4] In this case the state is also called "state vector". The term "vector" is understood in a general sense as an element of some *space* (Lusternik and Sobolev, 1965).

"system"[5] is typically used in a traditional sense as "a collection of elements being in some relations to each other and forming a certain integrity" (Philosophic dictionary, 1983, p. 610). Alternative interpretations relate mainly to the understanding of the word "dynamical" and to what elements and systems are implied (real-world objects, mathematical constructions or both). Some authors even take the term "dynamical" out of the deterministic framework and combine it with randomness. For the sake of illustration, we cite below several selected definitions formulated by known specialists in the field (see also Alligood et al., 2000; Guckenheimer and Holmes, 1983; Katok and Hasselblat, 1995; Loskutov and Mikhailov, 2007):

> The concept of a DS appeared as a generalisation of the concept of a mechanical system whose motion is described with Newton's differential equations. In its historical development, the concept of a DS similarly to any other concept gradually changed getting new and deeper contents... Nowadays, the concept of a DS is quite broad. It covers systems of any nature (physical, chemical, biological, economical, etc) *both deterministic and stochastic*.[6] Description of a DS is very diverse. It can be done with differential equations, functions from algebra of logic, graphs, Markov chains, etc. (Butenin et al., 1987, p. 8).
>
> When speaking of a DS, we imply *a system* of any nature which can take different mathematical forms including ordinary differential equations (autonomous and non-autonomous), partial differential equations, maps on a straight line or a plane (Berger et al., 1984).

In the section "What is a dynamical system?" of the monograph Malinetsky and Potapov (2000), the authors note: "In general, in different books one can find different interpretations of the term DS, e.g. like the following ones:

- a synonym of the term "a set of ODEs $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, t)$";
- a synonym of the term "a set of autonomous ODEs $d\mathbf{x}/dt = \mathbf{g}(\mathbf{x})$";
- a mathematical model of some mechanical system.

We[7] will adhere to the viewpoint according to which the concept of a DS is *a generalisation of the concept of a set of autonomous differential equations* and includes two main components: phase space $\mathbf{P}$ (metric space or manifold) and continuous or discrete one-parametric group (semigroup) $\varphi^t(\mathbf{x})$ or $\varphi(\mathbf{x}, t)$ of its transforms. A parameter t of the group is time."

Another formalised definition is as follows: "A DS is a quadruple (X, B, μ, Φ), where X is a topological space or a manifold, i.e. an abstract image of a state space, B are some interesting subsets in X, e.g. closed orbits or fixed points. They form an algebra in the sense that they include not only separate elements but also their unions and intersections. They are necessary to introduce a measure, since X itself can be immeasurable. μ is a measure, e.g. a volume of some domain or a frequency of an orbit visitations to it. μ is desired to be ergodic, unique, and invariant under the group of transforms Φ_t which defines an evolution. Sometimes, one adds also a

[5] From the Greek word "συστημα", i.e. "a whole consisting of parts" (Philosophic dictionary, 1983, p. 610).

[6] Highlighting with italic is ours in all the cited definitions.

[7] The authors G.G. Malinetsky and A.B. Potapov.

typical (in the sense of the measure μ) initial point. For example, the point $x_0 = 0$ is not typical for the operator $\Phi_t, t \in Z : x_{t+1} = \Phi_1(x_t) \equiv x_t(1+x_t)$, since it does not lead to an evolution" (Makarenko, 2002).

> One settled to understand a DS as a system of any nature (physical, chemical, biological, social, economical, etc.) *whose state changes* discretely or continuously in time (Danilov, 2001, p. 6).
>
> By abstracting from a concrete physical nature of an object, one speaks of it as of DS if it is *possible to specify such a set of quantities* called dynamical variables and *characterising a system state* whose *values at subsequent time instant are obtained from an initial set according to some rule*. This rule is said to determine an evolution operator for the system (Kuznetsov, 2001, p. 7).
>
> A DS can be thought of as an object of any nature whose state changes in time according to some dynamical law, i.e. as a result of a *deterministic* evolution operator action. Thus, the concept of DS is a consequence of a certain idealisation when one neglects influences of random perturbations inevitably present in any real-world system... Each DS corresponds to some mathematical model... (Anishchenko et al., 2002, pp. 1–2).
>
> A DS is a system whose behaviour is specified by a certain set of rules (an algorithm). A DS *represents only a model of some real-world system*. Any real-world system is prone to fluctuations and, therefore, cannot be dynamical (Landa, 1996).

The last definition is the closest one to the considerations in our book. It does not lead to difficulties in classification of possible situations. Thus, many real-world phenomena and objects can be successfully considered both with the aid of "probabilistic" (random) and "deterministic" mathematical tools. To illustrate that dynamical ideas can be fruitful under certain conditions and meaningless under different ones in modelling of the same object, we refer to the well-known "coin flips" (Sect. 2.6). There are no contradictions if the name of DS is related only to deterministic models and perceived as a kind of scientific jargon in application to real-world systems.

Further, we call DS *a mathematical evolution model for which one specifies* (i) a state $\mathbf{x}$ and (ii) an evolution operator Φ_t allowing a unique prediction of future states based on an initial one: $\mathbf{x}(t_0 + t) = \Phi_t(\mathbf{x}(t_0))$. In relation to real-world systems, we understand the term DS as a brief version of a statement "a system whose description with a dynamical model is possible and reasonable".

2.1.2 Non-rigorous Example: Variables and Parameters

Let us consider different dynamical systems which could describe an object which is familiar to many people – a usual cat (Fig. 2.1). The choice of quantities playing a role of dynamical variables or parameters of a model is determined by the purpose of modelling. If the purpose is to describe an evolution of the state of the cat's health, one can use its mass $M = x_3$, height $H = x_2$ and hair density $N = x_1$ (number of strands per a unit area) as dynamical variables. The collection $\mathbf{x} = (x_1, x_2, x_3)$ is a state vector of a dimension $D = 3$. Of course, one can imagine a number of other variables, such as blood haemoglobin concentration (x_4) and pulse rate (x_5). It would increase a model dimension D and make an investigation of the model

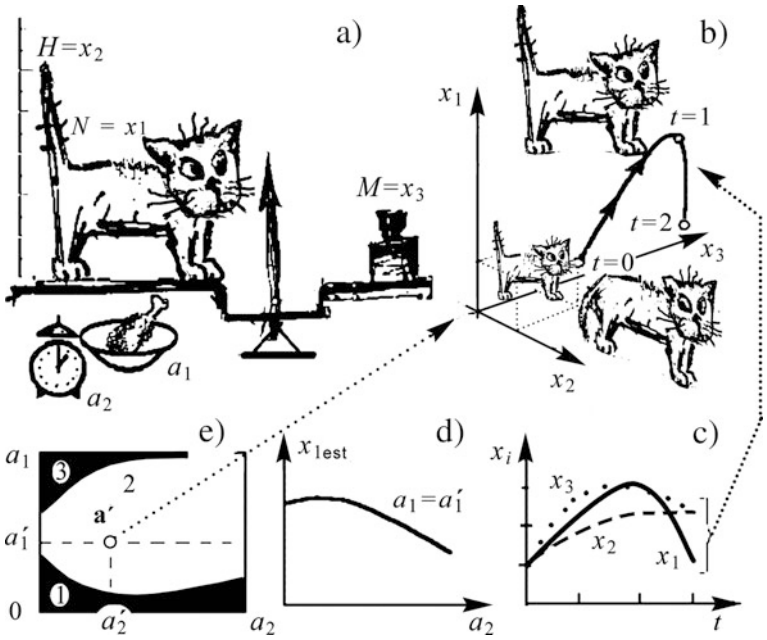

Fig. 2.1 Description of evolution of the cat's health: (**a**) variables and parameters; (**b**) the phase space and a phase orbit at fixed values of parameters $\mathbf{a}' = (a_1', a_2')$; (**c**) time realisations $x_i(t)$, i.e. projections of the phase orbit onto the phase space axes; (**d**) a combined space of parameters and states presenting an established value of x_1 versus a_2 at fixed $a_1 = a_1'$; (**e**) a parameter space, the area 2 corresponds to a normal life of a cat and areas where its prolonged existence is impossible due to either hunger (the area 1) or gluttony (the area 3) are painted over. A point $\mathbf{a}'$ in the parameter space corresponds to a definite structure of the entire phase space

more complicated. For the sake of illustration, it is sufficient for us to use the three dynamical variables and consider dynamics of the object in the three-dimensional phase space (Fig. 2.1b). Each point of the phase space corresponds to a vector $\mathbf{x} = (x_1, x_2, x_3)$ reflecting an object state. Thus, the cat is too young and feeble at the point $t = 0$ (Fig. 2.1b), it is young and strong at the point $t = 1$, and it is already beaten by the life at $t = 2$.

Obviously, a current health state of the cat and its variations depend on the quantities which we can keep constant or change as we want. Such quantities are called *parameters*. For instance, these can be nutrition (the mass of food in a daily ration a_1, kg/day) and life conditions (the duration of walks in fresh air a_2, h/day). The number of model parameters as well as the number of dynamical variables is determined by the problem at hand and by the properties of an original. Thus, the health of a cat depends not only on the mass of food but also on the calorie content of food (a_3), amount of vitamins (a_4), concentration of harmful substances in the air (a_5), etc. For simplicity, we confine ourselves to two parameters and consider behaviour of the object in a two-dimensional parameter space, i.e. on a parameter plane (a_1, a_2), see Fig. 2.1e. Each point of the parameter plane corresponds to a

certain kind of the temporal evolution of the object, i.e. to a certain kind of a phase orbit passing through an initial point in the phase space. Regions in the parameter space which correspond to different qualitative behaviour are separated with bifurcational sets of points. Bifurcational sets on the parameter plane (Fig. 2.1e) are boundary curves between white and black areas (bifurcational curves).

Just to illustrate the terms introduced above, without pretensions of strict description of such a complex biological object, one can consider the following set of first-order ordinary differential equations as a dynamical system modelling the health state of the cat:

$$\begin{aligned} \mathrm{d}x_1/\mathrm{d}t &= f_1(x_1, x_2, x_3, a_1, a_2), \\ \mathrm{d}x_2/\mathrm{d}t &= f_2(x_1, x_2, x_3, a_1, a_2), \\ \mathrm{d}x_3/\mathrm{d}t &= f_3(x_1, x_2, x_3, a_1, a_2). \end{aligned}$$

Relying upon everyday-life experience and imagination, one could suggest different forms for the functions f_k, e.g. algebraic polynomials whose coefficients are expressed via a_1 and a_2. It is a very common situation when model parameters enter evolution equations just as polynomial coefficients. According to the theorem of existence and uniqueness of a solution, the set of ordinary differential equations at fixed values of parameters and initial conditions has a unique solution under some general conditions. It means that the set of ODEs specifies a single phase orbit passing through a given initial point in the phase space.

Division of characterising quantities into dynamical variables and parameters is dictated by a modelling task. If the purpose of the cat modelling were description of its mechanical movements in space (rather than the state of its health as above), it would be reasonable to choose different variables and parameters. Thus, neither the animal mass M nor its "downiness" N and height H (previous dynamical variables) change during a jump of the cat. However, these quantities essentially affect its flight and must be taken into account as parameters $a_1 = M,\ a_2 = N,\ a_3 = H$, along with other quantities which influence mechanical motion (e.g. the shape of the cat's body). As dynamical variables, one can consider coordinates of the centre of the mass of the cat ($x_1 = x,\ x_2 = y,\ x_3 = z$) and angular displacements of its longitudinal axis in relation to coordinate axes ($x_4 = \alpha,\ x_5 = \beta,\ x_6 = \gamma$). Further, one can write down an evolution operator based on Newton's equations for progressive and rotational movements in contrast to the above semi-intuitive invention of model equations for the state of the cat health. Thus, depending on the purpose of modelling, the same physical quantities serve as dynamical variables in one case and play a role of parameters in another one.

Dynamical variables and parameters can be recognised in the evolution equations for a dynamical system. For instance, in the system specified by the classical equation of non-linear oscillator with cubic non-linearity (Duffing oscillator)

$$\mathrm{d}^2x\Big/\mathrm{d}t^2 + 2\delta\,\mathrm{d}x\big/\mathrm{d}t + \omega_0^2(bx^3 - x) = 0, \tag{2.1}$$

one of the dynamical variables is the quantity x and the parameters are δ, ω_0, b, i.e. the *parameter vector* $\mathbf{a} = (\delta, \omega_0, b)$ is three dimensional. The system itself is two dimensional ($D = 2$) since one must specify initial values of x and $\mathrm{d}x/\mathrm{d}t$ to find a particular solution to equation (2.1). The latter becomes clearer if one rewrites equation (2.1) equivalently as a set of two first-order equations for the variables $x_1 = x$ and $x_2 = \mathrm{d}x/\mathrm{d}t$:

$$\mathrm{d}x_1/\mathrm{d}t = x_2;\ \mathrm{d}x_2/\mathrm{d}t = -2\delta\, x_2 - \omega_0^2\left(bx_1^3 - x_1\right).$$

Thus, the derivative $\mathrm{d}x/\mathrm{d}t$ serves as the second dynamical variable of the system (2.1).

2.1.3 Phase Space. Conservative and Dissipative Systems. Attractors, Multistability, Basins of Attraction

A significant merit of dynamical modelling is a possibility of a vivid representation, especially in the case of low dimension D and small number of parameters. For such a representation, one uses formal *spaces*[8]: *state space* (or *phase space*), *parameter space* and their *hybrid versions*. Along the axes of a formal space, one indicates the values of dynamical variables or parameters. In a hybrid version, parameters are shown along certain axes and variables along others.

A state vector $\mathbf{x}(t)$ at some time instant t corresponds to a point in a phase space with coordinates $x_1(t), x_2(t), x_3(t)$ called *a representative point* since it represents an instantaneous state. In evolution process, a representative point moves along a certain curve called *a phase orbit*. A set of characteristic phase orbits is called *phase portrait* of a system. Having some experience, one can extract a lot of information about possible motions of a system from its phase portrait. Thus, a phase space is three dimensional in the above example with a cat. A single orbit corresponding to a concrete choice of an initial state at $t = 0$ is shown in Fig. 2.1b. It evidences that the animal developed well at the beginning and achieved excellent conditions at $t = 1$. Then, it grew thin and cast the coat up to an instant $t = 2$. We note that a

[8] "Space is a logically conceivable form (structure) serving as a medium where other forms or constructions are realised. For instance, a plane or a space serve in elementary geometry as media where various figures are constructed. ... In contemporary mathematics, a space defines a set of objects called points. ... Relations between points define "geometry". As examples of spaces, one can mention: (1) metric spaces ..., (2) "spaces of events" ..., (3) phase spaces. Phase space of a physical system is a set of all its states which are considered as points in that space. ..." (Mathematical dictionary, 1988). A space can be topological (if a certain non-quantitative concept of "closeness" is defined), metric (closeness is determined by "metrics"), etc. The choice of a phase space is determined by what one wants to use in modelling. For example, one needs "a smooth manifold" (Sect. 10.1.1) to use differential equations as a model. To define a limit behaviour of DS orbits, one needs a "complete" space, i.e. each limit point of a convergent sequence should belong to the same space.

phase orbit itself does not carry information about a time instant when a point visits a certain area.

Usually, one shows the most characteristic orbits on a phase portrait. For illustration, let us consider elements of the phase space of a system (2.1), which models oscillations in a potential well with two minima in the presence of friction similarly to a ball rolling in a double-pit profile shown in Fig. 2.2a. The curves on a phase plane $(x, \mathrm{d}x/\mathrm{d}t)$ in Fig. 2.2b are phase orbits starting from points 1 and 2. They cannot intersect since it would violate a dynamical description: a unique present must lead to a unique future! Situations resembling intersections can be found at *singular* points, i.e. *points of equilibrium* where a state of a DS remains constant for arbitrarily long time. There are three of them on the portrait: O, A_1, A_2. The first one corresponds to the location of a resting ball on the top of the hill (Fig. 2.2a), while the others show the left and right pits. Other points of the phase space correspond to states which are left by a representative point at a future time instant. Each of them corresponds to a certain phase orbit and time realisations of dynamical variables $x_k(t)$, Fig. 2.2b. We note that in a typical phase orbit, one can distinguish between a starting interval (*a transient process*) and a later stage with greater degree of repeatability (*an established motion*). Established motions are less diverse than transient processes and correspond to objects called *attractors* in a phase space of a dissipative system. In our example these are states of stable equilibrium: points A_1, A_2. Indeed, they seem to attract orbits from certain areas of the phase space. Starting from different points (1 and 2 in Fig. 2.2b), phase orbits can get to different attractors.

A set of points in a phase space from which a system gets to a certain attractor is called *basin of attraction*.[9] If an attractor in a phase space is unique, then its basin of attraction is the entire phase space. If there are several attractors, one says that *multistability* takes place. Then, their basins divide a phase space between each other, e.g. as shown with shading in Fig. 2.2c, d. Attractor can exist only in a phase space of a *dissipative dynamical system*. This is a system exhibiting phase volume compression illustrated in Fig. 2.3. A set of initial points occupies a volume $V(0)$. Starting from $V(0)$, a system gets to a volume $V(t)$ after some time interval t. A system is called dissipative if a phase volume decreases with time, $V(t) < V(0)$. In a one-dimensional case, a measure of a phase volume V is an interval length, it is a surface area in a two-dimensional case and a hyper-volume in a multidimensional case of $D > 3$. Finally, representative points get from an initial volume to attractors

[9] Strict definition of attractor is a subject of multiple discussions. A universally accepted definition is still lacking. One of the popular ones is given in several steps (Malinetsky and Potapov, 2000, pp. 76–77). "... a set A is called ... *invariant* ... if $\Phi_t \mathrm{A} = \mathrm{A}$. *Neighbourhood of a set* A is an *open* set U containing the *closure* of A, i.e. A together with all its limit points including boundary points. ... A closed invariant set A is called an *attracting set* if there exists its neighbourhood U such that $\Phi_t(\mathbf{x}) \to \mathrm{A}$ for all $\mathbf{x} \in \mathrm{U}$ and $t \to \infty$. A maximal U satisfying this definition is called *basin of attraction* of A. ... An attracting set containing an everywhere dense orbit is called an attractor A." This definition can be roughly reformulated as follows: an attractor is the least set to which almost all orbits of a DS from some area of non-zero volume tend.

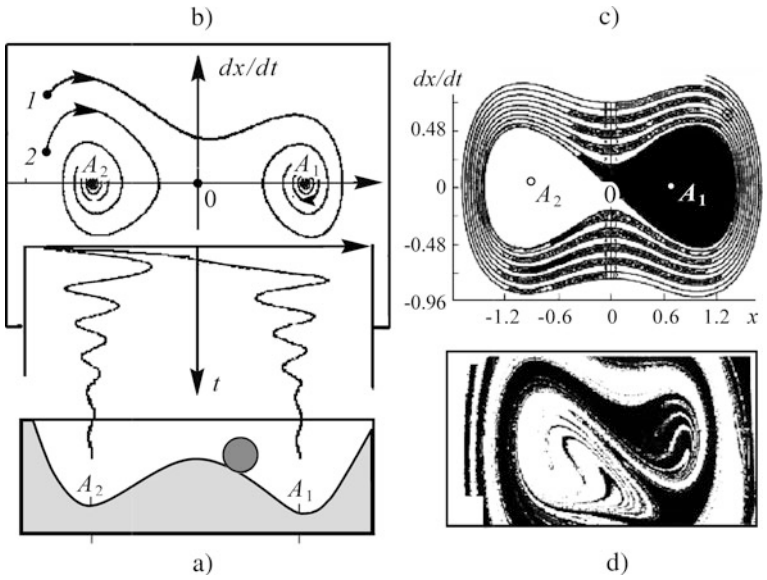

Fig. 2.2 A ball in a double pit: an illustration (**a**); a phase portrait and time realisations (**b**); basins of attraction of the equilibrium points A_1, A_2, i.e. of the two attractors coexisting in the phase space (**c**); basins of attraction in a non-autonomous case at the values of parameters for which two chaotic attractors coexist in the phase space (**d**)

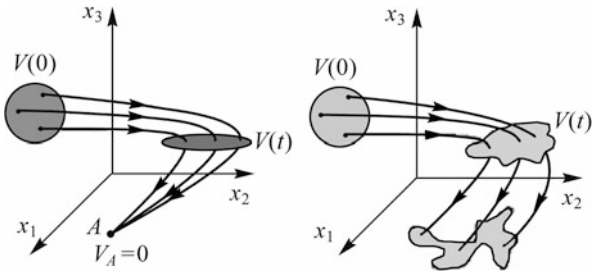

Fig. 2.3 Illustration of some phase volume deformations: (**a**) a dissipative system; (**b**) a conservative system. The curves are phase orbits

whose volume is equal to zero. Such a definition of a dissipative system is broader than that used in physics where a dissipative system is a system with friction in which mechanical energy turns into energy of chaotic molecular motion. In *conservative systems* (friction-free systems in physics) an initial phase volume is preserved and only its form changes, hence attractors are absent.

Some possible kinds of attractors and the character of the corresponding established motions are shown in Fig. 2.4. Apart from equilibrium states represented by points, an attractor can be

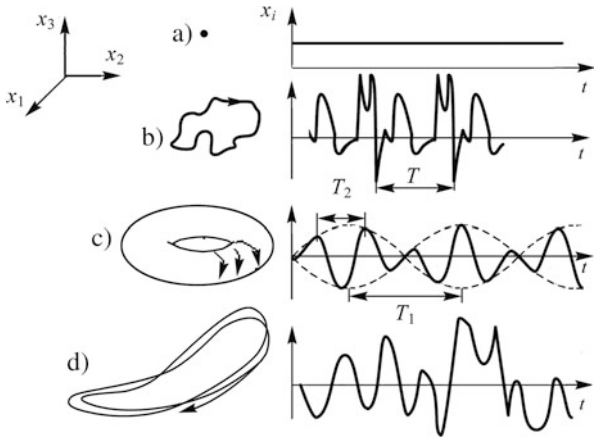

Fig. 2.4 Examples of characteristic sets in a phase space of a continuous-time system and corresponding time realisations

- a *limit cycle*, i.e. a closed curve, an image of a motion repeating itself with some period T (Fig. 2.4b);
- a *torus*, i.e. "an infinitely thin tread winding up on a bagel", an image of a quasi-periodic motion (with two characteristic periods T_1 and T_2 whose ratio is an irrational number) (Fig. 2.4c). A torus can be three- and multidimensional, i.e. represent complex behaviour with three, four, and more incommensurable frequencies of periodic components;
- a fractal set concentrated in a bounded area of a phase space, an image of *chaotic oscillations* called *a strange attractor* (Fig. 2.4d).[10]

Kinds of established motion realised in a DS and corresponding attractors are limited by its dimension. Thus, a phase space of a continuous-time system (e.g. with operators represented by differential equations) can contain only equilibrium points for $D = 1$, equilibrium points and limit cycles for $D = 2$, all the limit sets listed above for $D \geq 3$. Such considerations can help in practice to choose a model dimension. For instance, detection of a chaotic motion indicates that one needs at least three first-order non-linear ordinary differential equations to model an object. A somewhat different situation is found in a discrete-time system. An outlook of an attractor in its phase space can be imagined if one dissects the left pictures in Fig. 2.4 with a plane (a Poincare cross section). A single-turn cycle gives a single point in such a section. More complex cycles give several points. An orbit on a torus "draws" a closed curve in a section representing a quasi-periodic motion in a phase space of a discrete-time system. A chaotic attractor is represented by a set of points structured in a complicated (often self-similar) manner. A chaotic motion can be observed even in a phase space of *one-dimensional* non-invertible maps.

[10] "Strange" means here "different from previously known". An overview of kinds of chaotic attractors is given, e.g. in Anishchenko (1997), Katok and Hasselblat (1995) and Kuznetsov (2001).

2.1.4 Characteristics of Attractors

2.1.4.1 Geometrical Characteristics

Apart from visually detected differences, phase portraits are characterised by a number of quantitative measures. The most popular among them are dimensions. An integer-valued *topological dimension* D_T can be defined via an inductive principle (Poincare, 1982): $D_T = 0$ for a point; $D_T + 1$ is the dimension of a set which can be divided into non-intersecting parts with a subset of dimension D_T. According to those rules, a smooth curve has topological dimension $D_T = 1$, a surface $D_T = 2$, a volume $D_T = 3$. In particular, an equilibrium point, a cycle and a torus have topological dimensions 0, 1 and 2, respectively (see, e.g., Malinetsky, 2000, pp. 208–209). Structure of strange attractors differs qualitatively from the above sets. The former are *fractal* (self-similar) so that one needs more complicated measures called *fractal dimensions*. The simplest among them is *capacity* which characterises only geometry of an attractor. One also introduces *generalised dimensions* to take into account a frequency of a representative point visitations to subsets of an attractor. Below, we present only brief information about fractal measures. An educational computer program providing additional illustrations is located at our website (http://www.nonlinmod.sgu.ru). For more detailed study of fractal measures and techniques of their computation, we recommend the lectures 11–13 in the monograph Kuznetsov (2001) and references therein.

To define capacity, one covers a limit set in a D-dimensional phase space with D-dimensional cubes (i.e. line segments, squares, three-dimensional cubes, etc.) with an edge ε. Let a minimal number of cubes sufficient to provide covering be $N(\varepsilon)$.[11] Capacity of a set is

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

if the limit exists. One can use D-dimensional balls or sets of another shape instead of cubes (Kuznetsov, 2001, pp. 170–171; Malinetsky and Potapov, 2000, p. 210). Corresponding illustrations are given in Fig. 2.5, where we also present a classical example of a fractal *Cantor set* obtained from a unit segment by subsequent removal of middle thirds. In the latter case, one gets the capacity

$$D_F = -\lim_{\varepsilon \to 0} \frac{\ln 2^N}{\ln\left(1/3\right)^N} = \frac{\ln 2}{\ln 3} \approx 0.63$$

according to the definition (2.2). Majority of fractal sets are of non-integer dimension and can be embedded into spaces whose dimension equals the smallest integer exceeding a fractal dimension. Thus, the Cantor set is not already a finite set of points, but it is still not a line.

[11] Covering of a set A is a family of its subsets $\{A_i\}$ such that their union contains A.

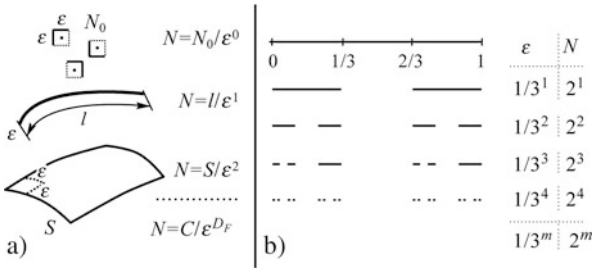

Fig. 2.5 Illustrations to (**a**) definition of capacity; (**b**) the Cantor set

A more subtle characteristic is *Hausdorff dimension*, which generalises capacity to the case of covering with elements of an arbitrary shape and size. Both quantities often coincide, but not always (Kuznetsov, 2001, p. 173; Malinetsky and Potapov, 2000, p. 209). As a rule, accurate numerical estimation of the Hausdorff dimension is impossible (Makarenko, 2002).

Generalised dimensions of Renyi D_q take into account a frequency of a representative point visitation to different attractor areas (Kuznetsov, 2001, pp. 176–190; Malinetsky and Potapov, 2000, pp. 211–214). Let an attractor be partitioned[12] into N non-empty cubes (cells) of size ε. Let us denote a portion of time spent by a representative point at a cell number i as p_i. It is a normalised density of points in a cell, i.e. an estimate of the probability of a visitation to a cell.[13] Then, one defines[14]

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

One distinguishes special kinds of generalised dimension: capacity at $q = 0$; *information dimension* at $q = 1$ (in the sense of limit for $q \to 1$); *correlation dimension* at $q = 2$. The latter characterises an asymptotic behaviour of *pairs* of points on an attractor. Indeed, a quantity p_i^2 can be interpreted as a probability to find two representative points within an ith cube of size ε. It is this quantity that can be easily estimated. Direct usage of the formula (2.3) leads to computational

[12] A partition is a covering with non-overlapping subsets $\{A_i\}$.

[13] It is strictly applicable to attractors supplied with an ergodic measure (Makarenko, 2002).

[14] A mathematical comment (Makarenko, 2002). Let us assume that p_i in each non-empty element of a partition follows an exponential law: $p_i \propto \varepsilon^\alpha$. If we deal with points on a line segment, then $\alpha = 1$ corresponds to a uniform distribution of points. However, $\alpha < 1$ may appear for rarely populated areas. Then, the ratio $p_i/\varepsilon \to \infty$ for $\varepsilon \to 0$. Therefore, such a distribution is called *singular*. For a square, areas with an exponent $\alpha < 2$ support singular distributions. One calculates *a partition function* $\sum_i p_i^q$, where a parameter q allows "to adjust" an estimator to locations with different probability density. If a partition function depends on ε via a power law, one introduces the definition (2.3) and speaks of *multifractal distribution*. If D_q differs for different q, an attractor is called multifractal (Kuznetsov, 2001, p. 182).

difficulties at $D > 3$. Therefore, a number of numerical techniques for the estimation of dimensions from a fragment of a phase orbit sampled discretely in time $(\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_N))$ have been developed. One of the most popular ones is the *algorithm of Grassberger and Procaccia* for the correlation dimension estimation (Grassberger and Procaccia, 1983). It relies on the calculation of the so-called correlation integral

$$C(\varepsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta\left(\varepsilon - \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|\right),$$

where Θ is the Heavyside function ($\Theta(s) = 0,\ s \le 0;\ \Theta(s) = 1,\ s > 0$) and $\|\cdot\|$ is a norm of a vector (Euclidean or any other). One can easily see that it is an estimate of the probability that two points, arbitrarily chosen on an attractor according to its probability measure, are separated by a distance less than ε. As it follows from equation (2.3), $C(\varepsilon) \approx A\varepsilon^{D_2}$ holds true for $\varepsilon \to 0$. Correlation dimension can be estimated as a slope on the plot $\ln C(\ln \varepsilon)$ at small ε. In practice, the number of orbit points N is limited. Therefore, the size ε of a cell cannot be selected arbitrarily small. Furthermore, the greater the dimension, the greater the number of points required for its reliable estimation. There are different recommendations in respect of the necessary number of points obtained under different assumptions (Eckmann and Ruelle, 1985, 1992; Kipchatov, 1995).

To get integer-valued estimates of dimension of an observed motion, one uses several ideas. One of the most popular is the *false nearest neighbour technique* (Kennel et al., 1992). According to it, one checks the property that a phase orbit restored in a space of a sufficient dimension should not exhibit self-intersections. The technique is applied to reconstruct a phase orbit from a time realisation of a single variable (Sect. 10.1.2).

Another widely known method is the *principal component analysis* (Broomhead and King, 1986), where one distinguishes directions in a phase space along which the motion of a representative point develops more intensively. It is done via the analysis of correlations between state vector components (Sect. 10.1.2).

2.1.4.2 Dynamical Characteristics

The most widely used are *Lyapunov exponents* which characterise a speed of divergence or convergence of initially nearby phase orbits. A weak deviation of a representative point from an orbit on an attractor, i.e. a weak perturbation ε_0, evolves approximately according to an exponential law $\varepsilon(\Delta t) = \varepsilon_0 \mathrm{e}^{\lambda \Delta t}$ until it gets large (Fig. 2.6a). As a result, a D-dimensional sphere of initial perturbations transforms into an ellipsoid after some time interval. If one prevents a system from a significant rise of perturbations (from an evolution along the grey arrow in Fig. 2.6a) by limiting an observation time interval τ, it is possible to estimate the exponents via the ratios of an ellipsoid semi-axis length to an initial radius: $\lambda_i = (1/\tau) \ln(\varepsilon_i/\varepsilon_0)$. These values averaged over an entire attractor are called Lyapunov exponents. Let

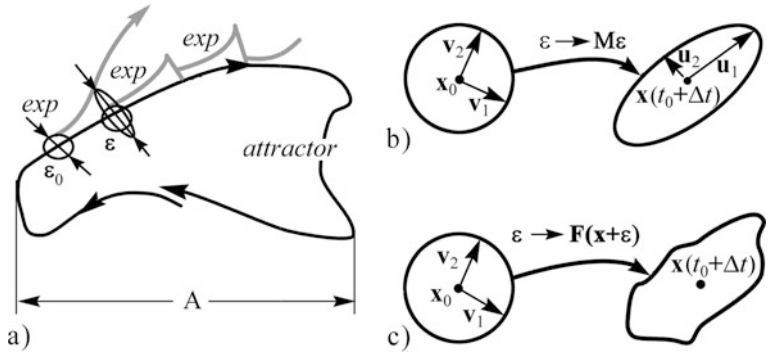

Fig. 2.6 Illustrations of Lyapunov exponents: (**a**) idea of calculation; evolution of a circle with a centre $\mathbf{x}_0$ (**b**) for a linear system; (**c**) for a non-linear system

us denote them $\Lambda_1, \Lambda_2, \ldots, \Lambda_D$. They characterise stability of the motion on an attractor in a linear approximation. The set of values Λ_i in descending order is called *spectrum of Lyapunov exponents*, while sequence of their signs (+, − or 0) is called the *spectrum signature*. If all the exponents are negative, i.e. the signature is $\langle -, -, \ldots, - \rangle$, then an attractor is an equilibrium point. The signature of a limit cycle is $\langle 0, -, \ldots, - \rangle$ and that of a two-dimensional torus is $\langle 0, 0, -, \ldots, - \rangle$. Spectrum of Lyapunov exponents for a chaotic attractor contains at least one positive exponent, e.g. $\langle +, 0, -, \ldots, - \rangle$, which determines the speed of divergence of initially close orbits.

Let us now describe some mathematical details. We start with a set of linear ordinary differential equations with variable coefficients:

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

where $\boldsymbol{\varepsilon} \in R^D$ and $\mathbf{A}$ is a matrix of an order D. Let us denote $\boldsymbol{\varepsilon}(t_0) = \boldsymbol{\varepsilon}_0$. Then, a solution to equation (2.4) at a time instant $t_0 + \Delta t$ is

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

where $\mathbf{M}(t_0, \Delta t)$ is a matrix of order D which depends on the initial time instant and the interval Δt and takes the form

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

where the matrix exponent is understood in the sense of formal expansion in a power series. For example, if $D = 1$ and $\mathbf{A}(t) = a = \text{const}$, then $\mathrm{d}\varepsilon(t)/\mathrm{d}t = a\varepsilon(t)$ and the solution to equation (2.5) takes a familiar form $\varepsilon(t_0 + \Delta t) = \varepsilon_0 \mathrm{e}^{a\Delta t}$. Thus, in the case of constant coefficients, a perturbation evolves according to an exponential

law. If coefficients are not constant, then a situation changes to some extent. For instance, one gets $\varepsilon(t_0 + \Delta t) = \varepsilon_0 e^{a\Delta t} e^{b \sin \Delta t}$ for $\mathbf{A}(t) = a + b \cos t$.

To characterise increase (or decrease) in $\boldsymbol{\varepsilon}$ in a multidimensional case, one should consider an evolution of a sphere of initial conditions with a centre at the origin and a radius $\|\boldsymbol{\varepsilon}_0\|$. Since the system is linear, a sphere transforms into an ellipsoid. Lengths and orientations of semi-axes of the ellipsoid depend on the matrix $\mathbf{M}$ and, hence, on the value of Δt. An absolute value of $\boldsymbol{\varepsilon}$ changes in a different manner depending on the orientation of the initial vector $\boldsymbol{\varepsilon}_0$. To describe it, one can use the so-called *singular value decomposition* of the matrix $\mathbf{M}$. This is a decomposition of the form $\mathbf{M} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathrm{T}}$, where $\mathbf{U}$ and $\mathbf{V}$ are mutually orthogonal matrices which can be conveniently written in the form of vectors sets $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_D]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, .., \mathbf{v}_D]$. If the matrix $\mathbf{M}$ is non-singular, then vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_D$ (called left singular vectors of the matrix $\mathbf{M}$) are mutually orthogonal and of unit length, i.e. they form an orthonormal basis in the space R^D. The same considerations apply to vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_D$ (right singular vectors). The matrix Σ is diagonal. Its diagonal elements $\sigma_1, \ldots, \sigma_D$ are listed in descending order. They are called *singular values* of the matrix $\mathbf{M}$. Action of the matrix $\mathbf{M}$ on the vector $\boldsymbol{\varepsilon}_0$ parallel to one of the right singular vectors $\mathbf{v}_i$ multiplies its length by σ_i and transforms it to a vector parallel to an ith left singular vector: $\boldsymbol{\varepsilon}(t_0 + \Delta t) = \sigma_i \|\boldsymbol{\varepsilon}_0\| \mathbf{u}_i$ (Fig. 2.6b). Thus, if at least one singular value of $\mathbf{M}$ exceeds 1 in absolute value, then an initial perturbation rises for some directions (one singular value is greater than 1 and another one is less than 1 in Fig. 2.6b). It rises in the fastest way for the direction of $\mathbf{v}_1$. The quantities showing how a perturbation changes are called *local Lyapunov exponents*:

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

They describe an exponential growth in perturbations *averaged over a finite time interval*. According to the definition (2.7), a strict equality $\|\boldsymbol{\varepsilon}(t_0 + \Delta t)\| = \|\boldsymbol{\varepsilon}_0\| \cdot e^{\lambda_i(t_0, \Delta t)\Delta t}$ holds true for respective directions of an initial perturbation. A.M. Lyapunov proved that under certain conditions[15] imposed on a matrix $\mathbf{A}$, there exist finite limits:

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

where the quantities Λ_i are exactly the *Lyapunov exponents*. They show an *efficient* speed of increase (decrease) in perturbations. Which of D exponents is realised for a given $\boldsymbol{\varepsilon}_0$ depends on the direction of the latter. A perturbation changes at a speed determined by the largest Lyapunov exponent Λ_1 for *almost any* direction

[15] There exists a number L such that $\frac{1}{\Delta t} \int_{t_0}^{t_0 + \Delta t} |A_{ij}(t')| \mathrm{d}t' \leq L$ for all i, j and Δt (Kuznetsov, 2001, p. 140).

of $\boldsymbol{\varepsilon}_0$. If Λ_1 is positive, then a typical perturbation rises. Hence, Λ_1 can be related to a predictability horizon of a system (2.4) in a situation where the equation itself is known precisely but initial conditions are specified at a certain error. A similar analysis can be done for linear difference equations.

Linearised dynamics and Lyapunov exponents. The analysis of stability for non-linear systems is performed via investigation of the linearised equations. Let us consider a non-linear system

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

Let $\mathbf{x}(t)$ be one of its orbits with an initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$. Let us call the orbit with $\mathbf{x}(t_0) = \mathbf{x}_0$ a reference orbit and consider an orbit starting at a very close initial condition $\mathbf{x}(t_0) = \mathbf{x}_0 + \boldsymbol{\varepsilon}_0$. Evolution of very small perturbations remaining small over an entire time interval considered is described with a set of equations linearised in a vicinity of the reference orbit:

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

This equation coincides with equation (2.4) if one assigns

$$\mathbf{A}(t) = \frac{\partial \mathbf{f}(\mathbf{x}(\mathbf{x}_0, t))}{\partial \mathbf{x}}.$$

One can write down its solution in the form of equation (2.5), where the matrix $\mathbf{M}$ maps an infinitesimal sphere of initial conditions with a centre $\mathbf{x}_0$ to an ellipsoid with a centre $\mathbf{x}(t_0 + \Delta t)$ (Fig. 2.6b). Strictly speaking, if a perturbation is not infinitesimal but finite, an image of a sphere will not be an ellipsoid, but another set. Linearised dynamics only approximately describes an evolution of finite perturbations (Fig. 2.6c). For any reference orbit, there exists a set of Lyapunov exponents characterising linearised dynamics in its vicinity.

In 1968, Oseledets showed that a set of Lyapunov exponents is the same for any generic point $\mathbf{x}_0$ on an attractor. This statement is an essence of multiplicative ergodic theorem (see, e.g., Kornfel'd et al., 1982; Malinetsky and Potapov, 2000; pp. 224–227; Sinai, 1995). Thus, Lyapunov exponents characterise evolution of infinitesimal perturbations not only for a given reference orbit but also for an entire attractor of a DS. The *largest Lyapunov exponent* assesses an efficient speed of growth of infinitesimal perturbations (see also Sect. 2.4).

2.1.5 Parameter Space, Bifurcations, Combined Spaces, Bifurcation Diagrams

Attractors in a phase space evolve (change their shape, size, etc.) under parameter variations and loose stability at certain parameter values. As a result, one observes

qualitative changes in a system motion, changes in its phase portrait, e.g. changes in the number of attractors in a phase space. Such a situation has got a name of *bifurcation*[16] (Belykh, 1997; Guckenheimer and Holmes, 1983; Malinetsky, 2000). We stress that a smooth deformation of an attractor and corresponding variations in an oscillation waveform are not regarded as a qualitative change.

To represent vividly the entire picture of possible kinds of established motions and transitions between them, one can use a geometrical representation in a *parameter space* where the values of parameters are shown along the axes. Some special methods are applied for that. The main idea is to single out sets of points separating areas with qualitatively different behaviour, i.e. *bifurcation sets*. In a pictorial example with a cat, a parameter space is the plane $\mathbf{a}_1$, $\mathbf{a}_2$ (Fig. 2.1e), while the boundaries between areas with different shading are bifurcation curves: the area 2 corresponds to a healthy life, while in the areas 1 and 3 the existence tragically and quickly stops due to hunger or gluttony. Parameter spaces shown below (e.g. Figs. 3.6, 3.7, 3.11 and 3.19a) are structured in a much more complicated manner. Bifurcation sets (curves) divide an entire parameter plane into areas where different attractors in a phase space exist. A way to represent vividly a situation of multistability (coexistence of several kinds of motion, several attractors in a phase space) on a parameter plane (e.g. in Figs. 3.6, 3.11 and 3.19a) is to show an area where a certain attractor exists as a separate sheet. Then, overlapping of many sheets at some parameters values is equivalent to multistability. For instance, bistability (coexistence of two attractors) takes place in the domain of intersection of sheets A and B in Fig. 3.6. The third sheet in that domain relates to an unstable cycle. Similarly, multistability in Fig. 3.19a takes place in the domain of intersection of sheets representing different modes of a system under investigation.

It is relevant to note some opportunities provided by the use of *combined spaces*. For instance, one can show a parameter value as an abscissa and a dynamical variable value in an established regime as an ordinate. A *bifurcation diagram* obtained in such a way for a quadratic map (Fig. 3.8e) is widely used to demonstrate universal laws of similarity (scaling) in transition to chaos via the period-doubling cascade. For a map describing a dissipative non-linear oscillator, such a diagram illustrates phenomena of resonance, hysteresis, bistability and bifurcation cascade (Fig. 3.10c–e). Moreover, one can present information in the phase and parameter spaces with colours. Basins of different attractors or areas of existence and evolution of different oscillatory regimes are often shown in such a way (Figs. 2.2 and 3.11).

We have given a very brief introduction to realisation of the dynamical approach. For readers who want to get deep knowledge in the field, we refer to classical works on qualitative theory of differential equations, theory of oscillations and non-linear dynamics (e.g. Andronov et al., 1959, 1967; Arnold, 1971, 1978; Bautin and Leontovich, 1990; Butenin et al., 1987; Guckenheimer and Holmes, 1983; Katok and Hasselblat, 1995; Shil'nikov et al., 2001).

[16] Initially, the word "bifurcation" meant division of an evolution pathway into two branches. However, currently any qualitative change is called call "bifurcation".

2.2 Foundations to Claim a Process "Random"

The use of the probabilistic approach is typically related to recognition of some quantity as "random". However, what is "random quantity" and what is its difference from a "non-random" one? Currently, there are several points of view on randomness which allow introduction of quantitative measures. For the most part, they agree with each other, but not always. Sometimes they can even lead to opposite results in the assessment of randomness or non-randomness of some quantity in practice. Here, we consider the problem according to a scheme suggested in Kravtsov (1989, 1997).

2.2.1 Set-Theoretic Approach

Set-theoretic approach underlying contemporary theory of probability (Gnedenko, 1950; Hoel, 1971; Pugachev, 1979, 1984; von Mises, 1964) associates the concept of randomness with possibility to specify a probability distribution law for a given quantity. Absence or presence of regularity is assessed via possible scattering of the values of a quantity: (i) probability distribution density in the form of Dirac δ function corresponds to a deterministic quantity; (ii) a non-zero "width", "smeared character" of distribution corresponds to unpredictable, random quantity.

2.2.1.1 Random Events and Probability

In description of many phenomena, researchers face impossibility to predict a course of events uniquely, even if all controllable conditions are held "the same".[17] To investigate such phenomena, the concepts of random event and probability were introduced in the theory of probability. These concepts are indefinable in theory, only some of their properties are defined via axioms. Their vivid interpretation and connection to practice are the tasks for the users. Below, we remind these basic concepts on an intuitive level, rather than rigorously.

An *event* is an outcome of a *trial*. Let us consider a classical example of the "coin flip" (see also Sect. 2.6). Let a coin be flipped only once. Then, the single flip is a trial. As a result, two events are possible: "a head" (an event A) and "a tail" (an event B). A and B are mutually exclusive events. An event in which either A or B occurs is called a *union of events* A and B and designated as $\mathrm{A} \cup \mathrm{B}$. In our case, it inevitably occurs as a result of any trial. Such an event is called *sure* and its probability is said to be equal to unity: $P\{\mathrm{A} \cup \mathrm{B}\} = 1$. Since a union of A and B is a sure event, one says that A and B constitute *a complete group* of events. It follows from an idea of symmetry that the events A and B are equiprobable, i.e.

[17] We mark the words "the same" with inverted commas to stress their conventional nature. To be realistic, one can speak of sameness only to a degree permitted by the conditions of observations or measurements.

the chances to observe a head or a tail are equal for a usual coin with a uniform density of metal. Equiprobable mutually exclusive events constituting a complete group are called *elementary events*. Probability of a union of mutually exclusive events is equal to the sum of their probabilities. In our case, A and B constitute a complete group, therefore, one can write down: $P\{\mathrm{A} \cup \mathrm{B}\} = P\{\mathrm{A}\} + P\{\mathrm{B}\} = 1$. From here and the condition of equiprobability, one gets the individual probabilities as $P\{\mathrm{A}\} = P\{\mathrm{B}\} = 1/2$.

Elementary events may not be always singled out so easily. Sometimes, geometrical considerations can help (a geometrical definition of probability). Let a trial consist of a random throwing of a point onto an area A of a plane. A point falls into A for sure and all subsets of A are "equal in rights". A point may either fall into a region $\mathrm{B} \subset \mathrm{A}$ or not. Probability of an event that a point falls into a subset B is defined via the ratio of the areas $\mu(\mathrm{B})/\mu(\mathrm{A})$, where μ stands for the *Lebesgue measure*. The latter is a surface area in our example, but the same formula can be used for a space of any dimension. Such a definition can be interpreted in terms of elementary events if they are introduced as falls of a point into small squares covering A (for a size of squares tending to zero).[18]

The most habitual to physicists is a statistical definition of probability. If an event A is realised M times in a sequence of N independent trials, then the ratio M/N is called a frequency of occurrence of the event A. If a frequency M/N tends to some limit for a number of trials tending to infinity, then such a limit is called a probability of the event A. This is the most vivid (physical) sense of the concept of probability. The property of an event frequency stabilisation is called *statistical stability*. The entire machinery of the theory of probability is appropriate for the phenomena satisfying the condition of statistical stability.

2.2.1.2 Random Quantities and Their Characteristics

Random quantity is any numerical function ξ of a random event. In the case of coin flips the values of a random quantity can be defined as $\xi = 1$ (a head) and $\xi = 0$ (a tail). The probability of $\xi = 1$ is the probability of a head.

For a complete characterisation of a random quantity, one needs to specify probabilities of its possible values. For instance, one uses a *distribution function* $F_\xi(x) \equiv P\{\xi \leq x\}$. If ξ is continuous valued and its distribution function is

[18] We note that it is important to define clearly what is a "random" point, line or plane for the definition of geometrical probabilities. For instance, let us assess the probability of an event that a "random" chord exceeds in length an edge of an equilateral triangle inscribed into a unit circumference. A chord can be chosen "randomly" in different ways. The first way: let us superpose a vertex of a chord with one of the triangle vertices leaving the other chord vertex free. Then, a portion of favourable outcomes when the length of a chord exceeds the length of an edge is 1/3. The second way: let us select randomly a point in a circle which is the middle of a "random" chord. A chord is longer than a triangle edge if its middle belongs to a circle inscribed into the triangle. Radius of that circle equals half the radius of the circumscribed circle and, hence, a portion of favourable outcomes assessed as the ratio of the areas of the two circles equals 1/4. We get two different answers for two different notions of a random chord.

differentiable, then one defines a *probability density function* $p_\xi(x) \equiv \mathrm{d}F_\xi(x)/\mathrm{d}x$. Then, a probability for ξ to take a value from an infinitesimal segment $[x, x + \mathrm{d}x]$ equals $p_\xi(x)\mathrm{d}x$. For the sake of definiteness, we speak of random quantities supplied with probability density functions.

Several often used distributions are the following:

(i) The *normal (Gaussian) law*

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

where a and σ^2 are parameters. This is one of the most often used distributions in the theory of probabilities. The reason is that it possesses many useful theoretical properties and allows obtaining a number of analytical results. Besides, in practice the quantities resulting from influence of multiple factors are often distributed approximately according to the Gaussian law. It finds theoretical justifications: the *central limit theorem* states that a sum of independent identically distributed random quantities is asymptotically normal, i.e. its distribution law tends to the Gaussian one for an increasing number of items.[19]

(ii) The exponential law (*Laplace distribution*):

$$p_\xi(x) = \begin{cases} \left(1\big/a\right)\exp(-x/a), & x \geq 0, \\ 0, & x < 0; \end{cases} \tag{2.12}$$

(iii) The *uniform distribution* on a segment $[a, b]$

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

A random quantity ξ is often characterised by statistical moments of its distribution. An *ordinary moment* of an order n is the quantity

$$E\left[\xi^n\right] \equiv \int\limits_{-\infty}^{\infty} x^n p(x)\mathrm{d}x. \tag{2.14}$$

Here and further, E stands for the mathematical expectation of the quantity in square brackets. The first-order moment is just the expectation of ξ. Its physical meaning is an average over infinitely many independent trials. *Central moments* are defined as ordinary moments for deviations of ξ from its expectation:

[19] Some authors mention ironically the frequent use of the normal law in data analysis and the references to the central limit theorem: engineers think that practical applicability of the central limit theorem is a strictly proven statement, while mathematicians believe that it is an experimental fact (see, e.g., Press et al., 1988).

$$E\left[(\xi - E\left[\xi\right])^n\right] \equiv \int_{-\infty}^{\infty} (x - E\left[\xi\right])^n p(x)\mathrm{d}x. \tag{2.15}$$

The second-order central moment is called *variance*. This is the most often used measure of scattering. Let us denote it as σ_ξ^2. Then, σ_ξ is called a root-mean-squared (standard) deviation of ξ. The third-order central moment is called skewness (a measure of a distribution asymmetry). The fourth-order central moment is called kurtosis. Skewness is equal to 0 and kurtosis is $3\sigma_\xi^4$ for the normal law (2.11). If all ordinary moments of ξ (for $n = 1, 2, \ldots$) exist, then one can uniquely restore the distribution function from their values. Parameters of a distribution law are related to its moments. For instance, $E[\xi] = a$ and $\sigma_\xi^2 = \sigma^2$ for the normal law (2.11); $E[\xi] = a$ and $\sigma_\xi^2 = a^2$ for the exponential law (2.12); $E[\xi] = (a + b)/2$ and $\sigma_\xi^2 = (b - a)^2/12$ for the uniform law (2.13).

For two random quantities ξ_1 and ξ_2, one considers joint characteristics. The two quantities can be regarded components of a two-dimensional *random vector* ξ. A joint probability density function $p_\xi(x_1, x_2)$ is then defined: a probability that the values of ξ_1 and ξ_2 fall *simultaneously* (in the same trial) into infinitesimal segments $[x_1, x_1 + \mathrm{d}x_1]$ and $[x_2, x_2 + \mathrm{d}x_2]$ equals $p_\xi(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2$. One also introduces a conditional probability density for one quantity under the condition that the other one takes a certain value, e.g. $p_{\xi_1}(x_1 | x_2 = x^*)$. The quantities ξ_1 and ξ_2 are called *statistically independent* if $p_\xi(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)$. In the latter case, the conditional distributions of ξ_1 and ξ_2 coincide with the respective unconditional distributions.

A random quantity depending on time (e.g. one deals with a sequence of values of a quantity ξ) is called a *random process*, see Chap. 4.

2.2.1.3 The Concept of Statistical Estimator

As a rule, in practice one does not know a distribution law and must *estimate* the expectation of an observed quantity or parameters of its distribution from results of several trials. This is a problem of mathematical statistics (Hoel, 1971; Ibragimov and Has'minskii, 1979; Kendall and Stuart, 1979; Pugachev, 1979, 1984; Vapnik, 1979, 1995; von Mises, 1964) which is inverse to problems of the theory of probability where one determines properties of a random quantity, given its distribution law. Let us denote a set of values of a random quantity ξ in N trials as $\{x_1, \ldots, x_N\}$. It is called a *sample*.[20]

A quantity whose value is obtained via processing the data $\{x_1, \ldots, x_N\}$ is called a sample function. An *estimator* of some distribution parameter is a sample function, whose values are in some sense close to the true value of that parameter.[21] We denote estimators with a "hat" like $\hat{a}$.

---

[20] A sample is an N-dimensional random vector with its own distribution law.

[21] Theoretically speaking, any measurable sample function is called estimator. If estimator values are not close to a true parameter value, such an estimator is just a "bad" one.

Let a sample $\{x_1, \ldots, x_N\}$ represent *independent* trials. Let the expectation of ξ be equal to a. Let the value of a be unknown but known to belong to a set A. It is necessary to get an estimator $\hat{a}$, which is as close to a as possible for any true value of a from A. Any estimator is a random quantity since it is a function of random quantities: $\hat{a} = f(x_1, \ldots, x_N)$. One gets a certain value of $\hat{a}$ from a certain sample and another value from another sample, i.e. $\hat{a}$ is characterised by its own probability density function $p_f(\hat{a})$ (Fig. 2.7), which is determined by the distribution law $p_\xi(x)$ and the way how $\hat{a}$ is computed (i.e. by the function f). Different functions f correspond to estimators with different distributions and, hence, with different probabilistic properties.

2.2.1.4 Estimator Bias and Variance

The most important property of an estimator $\hat{a}$ is closeness of its values to a true value of an estimated quantity a. Closeness can be characterised in different ways. The most convenient and widely used one is to define an *estimator error* as the

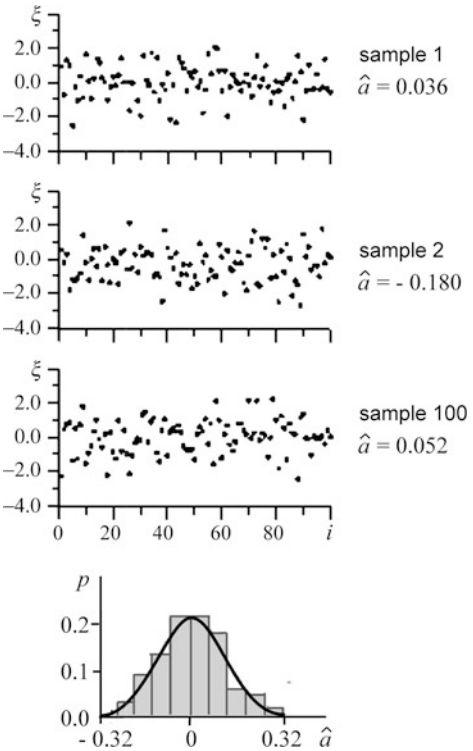

Fig. 2.7 Samples consisting of 100 values taken from a Gaussian law with zero mean and unit variance. The values of an estimator of the expectation for different samples are shown and its probability distribution density (theoretically, it is Gaussian with zero mean and the variance of 0.1) obtained from 100 samples

mean-squared difference between $\hat{a}$ and a:

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

It can be readily shown that the error is equal to the sum of two items:

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

An estimator whose bias is equal to zero, i.e. $E[\hat{a}] = a$ for any $a \in A$, is called *unbiased*. If the values of such an estimator are averaged over different samples, one gets a quantity closer to the true value of a since the random errors in $\hat{a}$ compensate each other. One could derive many unbiased estimators of a quantity a, i.e. different functions f. They would differ in their variance. One can show that an unbiased least-variance estimator is unique, i.e. if the least possible value of the variance is $\sigma_{\min}^2$, then it is exactly achieved only for a single estimator. An unbiased least-variance estimator is an attractive tool, though the least value of the squared error (2.17) may be achieved for another estimator, which is somewhat biased but exhibits significantly smaller variance.

An unbiased estimator of the expectation from a sample of independent values is the *sample mean*. We denote it by angular brackets and a subscript N: $\langle\xi\rangle_N$. This is just an arithmetic mean

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

This is a least-variance estimator of the expectation in the case of the normally distributed quantity ξ. If the distribution of ξ is symmetric and exhibits large kurtosis and/or other deviations form normality, a *sample median*[22] has typically a smaller variance as an estimator of its expectation. As well, the sample median is more stable to variations in the distribution law of ξ. Stability with respect to some perturbations of the distribution law is often called *robustness*. To compute the sample median, one may write down the values in a sample in ascending order: $x_{i_1} < x_{i_2} < \ldots < x_{i_N}$. Then, a sample median is $x_{i_{(N+1)/2}}$ for an uneven N and $\left(x_{i_{N/2}} + x_{i_{N/2+1}}\right)/2$ for an even N. The sample mean (2.18) is an estimator which is unbiased for any distribution of ξ, while a sample median can be biased for asymmetric distribution laws.

The sample moment of an order n can serve as an estimator of the respective ordinary moment $E[\xi^n]$:

[22] A median of a distribution is such a number b which divides the x-axis into two equiprobable areas: $P\{\xi < b\} = P\{\xi > b\} = 1/2$. A median coincides with the expectation for a symmetric distribution.

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

A situation with central moments is somewhat different since the value of $E[\xi]$ entering their definition is unknown. Yet, an estimator of variance can be obtained as the *sample variance*

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

It is biased due to the replacement of $E[\xi]$ with a sample mean. Its bias is of the order of $1/N$. One can show that an unbiased estimator is

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

2.2.1.5 Estimator Consistency

How do estimator properties change under increase in the sample size N? In general, an estimator distribution law varies with N. Hence, its bias and variance may also change. As a rule, one gets estimator values closer to a true value a at bigger N. If the bias $E[\hat{a}] - a$ tends to zero at $N \to \infty$ for any a from A, then the estimator $\hat{a}$ is called *asymptotically unbiased*. If the estimator $\hat{a}$ converges to a in probability (i.e. the probability that the estimator value differs from a true one more than by ε tends zero for arbitrarily small ε: $\forall \varepsilon > 0$ $\mathrm{P}\{|\hat{a} - a| > \varepsilon\} \underset{N \to \infty}{\to} 0$), it is called *consistent*. Consistency is a very important property of an estimator assuring its high goodness for large samples. The sample moments (2.19) are consistent estimators of the ordinary moments (Korn and Korn, 1961; Pugachev, 1979).

2.2.1.6 Method of Statistical Moments

Let us consider the problem of parameter estimation when a functional form of the distribution $p_\xi(x, \mathbf{c})$ is known and $\mathbf{c} = (c_1, \ldots, c_P)$ is a parameter vector taking values from a set $A \subset R^P$. One of the possible approaches is the *method of statistical moments* which is following. The first P theoretical ordinary moments are expressed as functions of parameters. Examples for normal, exponential and uniform distributions are given above, where the first two moments are expressed as simple functions of parameters. Thereby, one obtains a system

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

By substituting the sample moments instead of the theoretical ones into equation (2.22), one gets a set of equations for the parameters

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

whose solution gives estimators $\hat{c}_1, \ldots, \hat{c}_P$. Such moments-based estimators may not possess the best properties for small samples. However, they are asymptotically unbiased and consistent (Korn and Korn, 1961) so that they can be readily used for large samples.

2.2.1.7 Maximum Likelihood Method

Typically, the maximum likelihood method provides estimators with the best properties. According to it, a sample is considered as a random vector $\mathbf{x} = (x_1, \ldots, x_N)$ of dimension N which is characterised by some probability density function depending on a parameter vector $\mathbf{c}$. Let us denote such a conditional probability density as $p_N(\mathbf{x}|\mathbf{c})$. One looks for the parameter values $\mathbf{c} = \hat{\mathbf{c}}$ maximising $p_N(\mathbf{x}|\mathbf{c})$ for an *observed* sample, i.e. an occurrence of the sample $\mathbf{x} = (x_1, \ldots, x_N)$ is the most probable event for the values $\mathbf{c} = \hat{\mathbf{c}}$. They are called *maximum likelihood estimators* (ML estimators).

The function $L(\mathbf{c}) = p_N(\mathbf{x}|\mathbf{c})$ where $\mathbf{x}$ is a fixed vector (an observed sample), is called *likelihood function* or just *likelihood*. It should not be interpreted as a probability density function for parameters $\mathbf{c}$ since the parameters are fixed numbers (not random quantities) according to the problem setting. Therefore, a special term "likelihood" is introduced. ML estimators give a maximal value to the likelihood: $L(\hat{\mathbf{c}}) = \max\limits_{\mathbf{c} \in A} L(\mathbf{c})$. Necessary conditions of the maximum read as

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

It is often more convenient to deal with the likelihood logarithm. It gets maximal at the same point as $L(\mathbf{c})$, therefore, ML estimators are found from equation

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

which is called *likelihood equation*.

For a sample consisting of independent values, the likelihood function equals the product of probability density functions at each value of x_i and the logarithmic likelihood equals the sum of logarithms:

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

In such a case, ML estimators are consistent and asymptotically unbiased. Asymptotically, they are the least-variance estimators.

For the normal distribution of ξ, the logarithmic likelihood reads as

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

One can readily see that the ML estimators of the parameters a and σ^2 coincide with the sample mean (2.18) and the sample variance (2.20). Thus, the ML estimator of σ^2 is biased. However, it tends to the unbiased estimator (2.21) with increasing N, i.e. it is asymptotically unbiased. One can show that the ML estimator $\hat{a}$ is distributed here according to the normal law with the expectation a and the variance σ^2/N (see Fig. 2.7 for a concrete illustration). It follows from these observations that the value of the sample mean gets closer to the true value of a for a large number of trials, since the estimator variance decreases with N. In particular, $|\hat{a} - a| < 1.96\sigma/\sqrt{N}$ holds true with a probability of 0.95.

The interval $[\hat{a} - 1.96\sigma/\sqrt{N}, \hat{a} + 1.96\sigma/\sqrt{N}]$ is called 95% confidence interval for the quantity a. The greater the N, the narrower this interval. To estimate it from observations, one can replace the true value of σ with its estimator $\hat{\sigma}$. An estimator $\hat{a}$ is called a *point* estimator since it gives only a single number (a single point). If an interval of the most probable values of an estimated parameter is indicated, then one speaks of an *interval* estimator. Interval estimators are quite desirable, since from a single value of a point estimator one cannot judge to what extent it can differ from the true value.

2.2.1.8 When the ML Technique is Inconsistent

Sometimes, the ML technique can give asymptotically biased estimators. This is encountered, for instance, in the investigation of a dependence between two variables when the values of *both* variables are known with errors. It is studied by the so-called confluent analysis (Aivazian, 1968; Korn and Korn, 1961). As an example, let us consider the following problem. There is a random quantity Z and quantities X and Y related to Z via the following equations (Pisarenko and Sornette, 2004):

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

where ξ and η are independent of each other and of Z, and are normally distributed random quantities with zero expectation and the same variance σ^2. One can say that X and Y represent measurements of Z in two independent ways. There is a sample of X and Y values obtained from independent trials: $\{x_i, y_i\}_{i=1}^N$. It is necessary to estimate a measurement error variance σ^2.

The simplest way to derive an estimator is to note that a quantity $X-Y = \xi-\eta$ is normally distributed with zero expectation and the variance $2\sigma^2$, since the variance of the sum of two independent quantities is equal to the sum of their variances. Then, one can easily get a consistent estimator of $X - Y$ variance from a sample $\{x_i - y_i\}_{i=1}^N$ as follows:

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

Hence, the value of σ^2 is estimated via the following equation:

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

At the same time, a direct application of the ML technique (without introduction of the above auxiliary variable) gives the likelihood function

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

which contains unobserved values of Z. By solving the likelihood equations, one then gets estimators:

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

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

Thus, the ML estimator of the variance is twice as small as the unbiased estimators (2.30) at any N, i.e. the former is biased and inconsistent. What is a principal difference of this problem? It is as follows: The number of estimated quantities (equal to $N + 1$ under the ML method) rises with the sample size! In the previous cases, we have considered estimation of a fixed number of parameters.

In general, the less the number of quantities estimated, the better the properties of their estimators.

2.2.1.9 Bayesian Estimation

A very broad branch of the theory of statistical estimation is related to the case when true values of parameters $\mathbf{c}$ are also random quantities, i.e. they can vary between different samples according to a probability density function $p(\mathbf{c})$ which is called *prior*. If a prior density is known, then it is reasonable to take it into account in estimation. The corresponding approaches are called *Bayesian*.[23]

In the most widespread version, one tries to find a distribution law for the parameters $\mathbf{c}$ under the condition that a sample $x_1, \ldots, x_N$ has been realised. This is a so-called *posterior* probability density function $p(\mathbf{c}|x_1, \ldots, x_N)$. It can be derived if a probability density function $p(x_1, \ldots, x_N|\mathbf{c})$ at a given $\mathbf{c}$ is known. Then, one finds posterior density via the *Bayesian rule*:[24]

$$p(\mathbf{c}\,|x_1, \ldots, x_N) = \frac{p(\mathbf{c})\,p(x_1, \ldots, x_N\,|\mathbf{c})}{\int p(\mathbf{c})\,p(x_1, \ldots, x_N\,|\mathbf{c})\mathrm{d}\mathbf{c}}. \tag{2.33}$$

We note that the denominator does not depend on the estimated parameters, since integration over them is performed.

If a posterior distribution law is found, then one can get a concrete point estimator $\hat{\mathbf{c}}$ in different ways, e.g., as the expectation $\hat{\mathbf{c}} = \int \mathbf{c}\,p(\mathbf{c}|x_1, \ldots, x_N)d\mathbf{c}$ or as its point of maximum (a mode). In the absence of knowledge about a prior density, it is replaced with a constant $p(\mathbf{c})$ that corresponds to a distribution which is uniform over a very broad segment. Then, to a multiplier independent of $\mathbf{c}$, a posterior distribution coincides with the likelihood function. Further, if a Bayesian estimator is defined as a posterior distribution mode, one comes exactly to the ML technique.

As a rule, in practice one sets up a hypothesis: which distribution law an observed quantity follows, whether trials are independent or not, etc. Accepting such assumptions, one applies corresponding techniques. Validity of the assumptions is checked with statistical tools after getting an estimate (Sect. 7.3).

2.2.2 Signs of Randomness Traditional for Physicists

All the signs listed below rely to some extent on the understanding of randomness as a lack of "repeatability" in a process.

(a) *Irregular* (non-periodic) outlook of a time realisation. This is the most primitive sign of randomness. Here, it is directly opposed to periodicity: *absence of strict period means randomness*, *its presence means determinancy*.

---

[23] From the name of an English priest Thomas Bayes (1702–1761), who suggested the idea in a work published after his death.

[24] In fact, this is a joint probability of two events A and B written down in two versions: $P\{A \cap B\} = P\{A\}P\{B|A\} = P\{B\}P\{A|B\}$. Hence, one deduces $P\{B|A\} = P\{B\}P\{A|B\}/P\{A\}$.

(b) *Decaying correlations*. This is a decrease of an *autocorrelation function* $\rho(\tau)$ (ACF, Sect. 4.1.2) to zero with increasing τ. For a zero-mean stationary process (Sect. 4.1.3), the ACF reads as $\rho(\tau) = \langle x(t)x(t+\tau)\rangle \Big/ \sqrt{\langle x^2(t)\rangle\langle x^2(t+\tau)\rangle}$. Angular brackets denote averaging over an ensemble which coincides with temporal averaging for an ergodic process (Sect. 4.1.3). This sign gives, in essence, a quantitative measure of an observed process deviation from a periodic one. One cannot reveal periodicity with this approach if a period $T > T_o$, where T_o is an observation time.

(c) *Continuous spectrum*. According to this sign, a process with a continuous power spectrum (Sect. 6.4.2) is called random, while a spectrum of a periodic process is discrete. In practice, finiteness of an observation time T_o limits a spectral resolution: $\Delta\omega_{\min} = 2\pi/T_o$. By increasing an observation time T_o, one would finally establish finiteness of spectral lines for any real-world process and, strictly speaking, would have to regard any real-world process random according to any of the signs (a–c).

(d) *Irregularity* of sets of data points in a restored "phase space" (Sect. 10.1): absence of any signs for a finite dimension and so forth. These are more delicate characteristics which are not related just to the detection of non-periodicity.

There are also more qualitative criteria: irreproducibility of a process or its uncontrollability, i.e. impossibility to make conditions under which a process would occur in the same way or in the way prescribed in advance, respectively.

2.2.3 Algorithmic Approach

An *algorithmic approach* interprets "a lack of regularity" as an excessive complexity of an algorithm required to reproduce a given process in a digital form. The idea to relate randomness to complexity was put forward for the first time by A.N. Kolmogorov and independently by Chaitin and Solomonoff.

Any process can be represented as a sequence of 0s and 1s, i.e. written down in a binary system: $\{y_i\}$, $i = 1, 2, \ldots, N$. Kolmogorov suggested to regard a length l (in bits) of the shortest program capable of reproducing the sequence $\{y_i\}$ as a measure of its complexity. For instance, a program reproducing a sequence 1010...10 (a hundred of pairs "10") is very short: *print "10" a hundred times*. If 0s and 1s are located randomly, a program consists of symbol-wise transmission of a sequence which appears uncompressible. Thus, $l \sim N$ for random sequences and $l << N$ for non-random ones.

Unfortunately, there is no generally applicable way to find the minimal length of a program in practice.[25] New approaches to the concepts of complexity and randomness based on the idea of algorithmic complexity have been developed. A view

[25] For the same fundamental reasons that are mentioned in Gödel's theorem stating incompleteness of any system of axioms as discussed, e.g., in Shalizi (2003) and references therein.

relating those concepts to *predictability* is getting more and more popular during the last years (Badii and Politi, 1997; Kravtsov, 1989; Shalizi, 2003).

2.2.4 Randomness as Unpredictability

Randomness or determinancy of a process is related in Kravtsov (1989, 1997) to possibility of its prediction with the aid of an existing model. The author considers a registered process $x(t)$ and a model process $z(t)$. For the sake of simplicity, it is assumed $\langle x(t)\rangle = \langle z(t)\rangle = 0$. At a current time instant $t = t_0$ the quantities take the values $x = x_0, z = z_0$. It is natural to specify a model process so that $z_0 = x_0$ and assess a forecast quality via its error $x(t) - z(t) = \Delta(t)$, $\Delta(t_0) = 0$. The entire approach is based on the statistical description of the pair x, z.

The values of x and z typically diverge as time passes so that the absolute value of $\Delta(t)$ rises. By repeating experiments and comparison of $x(t)$ to $z(t)$, one can form an ensemble of realisations and estimate probability distributions $p(x, z, t, x_0, z_0, t_0)$ and $p(\Delta, t, x_0, z_0, t_0)$. In such a description, the model process $z(t)$ is included into statistical considerations along with the registered process. Measures of predictability can be the following:

(i) The mean-squared error $\sigma^2_\Delta(\tau) \equiv \langle\Delta^2(\tau)\rangle = \langle|x(t) - z(t)|^2\rangle$, where $t = t_0 + \tau$, $\sigma^2_\Delta(0) = 0$. If the quantities $x(t)$ and $z(t)$ become statistically independent at $\tau \to \infty$, then $\langle x(t)z(t)\rangle = 0$ and $\sigma^2_\Delta(\tau) = \langle x^2(t)\rangle + \langle z^2(t)\rangle$. One assumes that x and z are bounded. Then, a relative error can be reasonably defined as $E(\tau) = \sigma^2_\Delta(\tau)/\left(\left\langle x^2(t)\right\rangle + \left\langle z^2(t)\right\rangle\right)$ so that $E \to 1$ for $t \to \infty$.

(ii) The cross-correlation function between an original and a model processes $D(\tau) = \langle x(t_0 + \tau)z(t_0 + \tau)\rangle / \sqrt{\langle x^2(t_0 + \tau)\rangle\langle z^2(t_0 + \tau)\rangle}$. One has $D(0) = 1$ and $|D(\tau)| \le 1$ for any τ. From well-known statistical relationships, one can derive

$$D(\tau) = \frac{\left\langle x^2(t_0 + \tau)\right\rangle + \left\langle z^2(t_0 + \tau)\right\rangle}{2\sqrt{\left\langle x^2(t_0 + \tau)\right\rangle\left\langle z^2(t_0 + \tau)\right\rangle}}\,(1 - E(t))\,.$$

Thus, the degree of predictability can be expressed via different similar quantities. Qualification of a process as random or deterministic is determined by the possibility of its prediction with an available model. Here, random is something that we cannot predict for some reasons: due to the properties of $x(t)$, or due to the kind of a model process $z(t)$, or due to the absence of a model. Such an approach to randomness was developed within a hypothesis distinction theory for the needs of radio-location.

2.3 Conception of Partial Determinancy

Conception of partial determinancy is based on the convention that one chooses unpredictability (predictability) of an observed process $x(t)$ with a certain predictive model $z(t)$ as a sign of randomness (determinancy) of $x(t)$. Randomness and determinancy are not opposed to each other but considered as poles of a single property called partial determinancy.

It is convenient to use cross-correlation between an observed and a model process $D(\tau)$ as a quantity characterising the degree of determinancy (predictability). Its typical plot is shown in Fig. 2.8, where D is an area of full determinancy; DC is an area of partial determinancy; C is an area of random (unpredictable) behaviour. An observed process $x(t)$ appears deterministic (predictable) if $D \approx 1$; unpredictable if $|D| << 1$; partially predictable if $0 < |D| < 1$.

A time interval $\tau = \tau_{\text{det}}$ over which the degree of predictability falls down to a certain threshold value, e.g. $D(\tau_{\text{det}}) = 1/2$, is called an interval of deterministic behaviour. What affects this quantity? For a real-world system, it is always finite for the following reasons:

- An observed process always differs from an investigated process due to the influence of registering devices, a *measurement noise* $v(t)$.
- There are random and/or non-random unaccounted external influences $\mu(t)$, the so-called "dynamical noise".
- A model does not adequately reflect properties of an object. This is a "noise of ignorance" $\Delta M(t)$ depending on a model structure and parameters values.

Therefore, $\tau_{\text{det}} = f(v, \mu, \Delta M)$. Even if one manages to reduce strongly an effect of devices and an error in a deterministic component of a model, unavoidable external fluctuations remain. They can be related to infinite dimension of real-world object microstructure, to noises of different origin, to senescence processes, and so forth and principally limit predictability interval. The limit $\tau_{\text{lim}} = \lim_{v \to 0, \Delta M \to 0} \tau_{\text{det}} = f(v, \mu, \Delta M)$ is called "a predictability horizon".

As a rule, x and z become statistically independent for $\tau >> \tau_{\text{lim}}$ so that $D(\tau) \to 0$. An interval of deterministic behaviour τ_{det} can exceed an autocorrelation

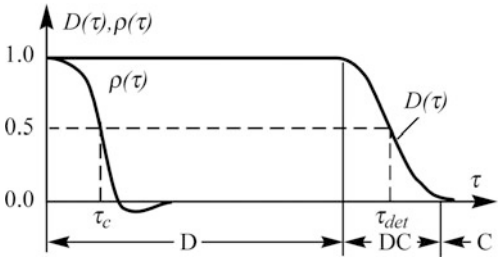

Fig. 2.8 Typical relationship between the degree of determinancy $D(\tau)$ and the autocorrelation function

time τ_c of $x(t)$ characterising the speed of its autocorrelation function decay. The latter can be estimated as $\tau_c \approx 1/\Delta\omega$, where $\Delta\omega$ is the width of the spectrum line. For instance, one gets $\Delta\omega \to \infty$, $\tau_c = 0$ for a white noise, i.e. for process which "forgets" its past at once (Sect. 4.2). An autocorrelation time τ_c can be considered as the least interval of determinancy due to the following consideration. If one has no dynamical equations for a model $z(t)$, then a forecast can be based on previous values of $x(t)$. The simplest principle is "tomorrow is the same as today", i.e. a model $z(t+\tau) = x(t)$. In such a case, one gets $D(\tau) = \rho(\tau)$ and $\tau_{\text{det}} = \tau_c$. In general, it can be that $\tau_{\text{det}} > \tau_c$ (Fig. 2.8). The same phenomenon can be close to a deterministic one from a viewpoint of one model and fully non-deterministic from a viewpoint of another model.

2.4 Lyapunov Exponents and Limits of Predictability

2.4.1 Practical Prediction Time Estimator

Forecast is a widespread and most intriguing scientific problem. A predictability time for many processes is seemingly limited in principle and even not large from a practical viewpoint. If a process under investigation is chaotic, i.e. close orbits diverge exponentially, it is natural to expect its predictability time to be related to the speed of close orbit divergence. The latter is determined by the value of the *largest Lyapunov exponent* Λ_1 (Sect. 2.1.4). For a dynamical model, it is reasonable to take an interval over which a small perturbation (determined both by model errors and different noise sources in a system) rises up to a characteristic scale of an observed oscillations as an estimator of predictability time. A *predictability time* can be roughly estimated via the following formula (Kravtsov, 1989):

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

where σ_μ^2 is the dynamical noise variance, σ_ν^2 is the measurement noise variance, $\sigma_{\Delta M}^2$ is the model error (an "ignorance noise" variance), σ_x^2 is the observable quantity variance, and the largest Lyapunov exponent Λ_1 is positive. The formula can be derived from the following qualitative considerations. Let equations of an original system be known exactly and initial conditions only to an error ε (measurement noise). Then, if those "incorrect values" are taken as initial conditions for a model, one gets a prediction error rising in time as $\varepsilon \cdot e^{\Lambda_1 t}$ on average. If a predictability time is defined as a time interval over which a prediction error reaches the value of σ_x, one gets

$$\tau_{\text{pred}} = \frac{1}{\Lambda_1} \ln \frac{\sigma_x}{\varepsilon}.$$

The quantity $\tau_\Lambda = 1/\Lambda_1$ is called *Lyapunov time*. Now, let us consider external random influences and model errors along with the measurement noise. If all those factors are regarded as approximately independent, an overall perturbation variance is equal to the sum of variances of the components. Replacing ε in the last formula for τ_{pred} by a square root of an overall perturbation variance, one gets the expression (2.34).

If noises and model errors are small as compared with the signal level, a time interval (2.34) can significantly exceed the autocorrelation time of a process which can be roughly estimated as $\tau_c \sim 1/\Lambda_1$ in many cases. Thus, if the signal level is 1000 times as big as the noise level in terms of root-mean-squared deviations, then a predictability time (2.34) is approximately seven times as big as the autocorrelation time.

The formula (2.34) is not always applicable to estimate a predictability time. The point is that after a certain time interval, any finite perturbation in a chaotic regime reaches a scale where the linearised system (2.10) is no longer appropriate. Further evolution is, strictly speaking, not connected with Lyapunov exponents. Thus, if one is interested in a forecast with a practically acceptable accuracy rather than with a very high accuracy, the Lyapunov exponent is not relevant and cannot impose restrictions on a predictability time. Yet, if the Lyapunov exponent characterises a speed of the perturbation rise at large scales correctly (which is often the case), then one can use it to assess a predictability time even for finite perturbations and errors.

However, under stricter considerations it appears that even in the limit of infinitesimal perturbations the Lyapunov time is not always related to a predictability time. Let us consider this interesting fact in more detail.

2.4.2 *Predictability and Lyapunov Exponent: The Case of Infinitesimal Perturbations*

The quantity (2.34) can be called a *predictability time* by definition. However, other approaches are also possible. One of the reasonable ideas consists of the following (Smith, 1997). Let us consider how a perturbation of a given initial condition $\mathbf{x}_0$ evolves. According to the definition of the local Lyapunov exponents (2.7), one gets $\|\varepsilon(t_0 + \Delta t\| = \|\varepsilon_0\| e^{\lambda_1(\mathbf{x}_0, \Delta t)\cdot\Delta t}$ in the worst case, i.e. as largest increase in a perturbation. Let us define a predictability time via time intervals over which an initial small perturbation gets q times greater:

$$\tau_q(\mathbf{x}_0) = \frac{\ln q}{\lambda_1(\mathbf{x}_0, \Delta t)}.$$

Such a time interval depends on $\mathbf{x}_0$. To get an overall characteristic of predictability, one can average $\tau_q(\mathbf{x}_0)$ over an invariant measure $p(\mathbf{x}_0)$, i.e. over probability distribution on an attractor:

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

This definition of a predictability time differs essentially from equation (2.34). Thus, if a time interval over which an error gets q times greater were defined via the largest Lyapunov exponent, then one would get

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

Here, the Lyapunov exponent (the quantity in the denominator) is expressed as an average over a natural measure[26] which is equivalent to temporal averaging for an ergodic system.

Hence, the situation is analogous to the following one. There are values of a random quantity $x_1, x_2, \ldots, x_N$ and one needs to estimate its expectation $E[x]$. The simplest way is to calculate a sample mean which is a "good" estimator: $\langle x\rangle = (x_1+\ldots+x_N)/N$. This is an analogue to the formula (2.35) for a mean predictability time. However, one can imagine many other formulas for an estimator. For instance, one may calculate inverse values $1/x_1, 1/x_2, \ldots, 1/x_N$, estimate a quantity $1/E[x]$ as their sample mean and take its inverse. The resulting estimator $\langle x'\rangle = N/(1/x_1 + 1/x_2 + \ldots + 1/x_N)$ is an analogue to equation (2.34). However, a mean value of the inverse quantities is generally a biased estimator of $1/E[x]$. Therefore, $\langle x'\rangle$ is also a "bad" estimator of $E[x]$. The quantities $\langle x\rangle$ and $\langle x'\rangle$ coincide only when $x_1 = x_2 = \ldots = x_N$. In our case it means that the Lyapunov time coincides with τ_q (up to a multiplier $\ln q$) only if the local Lyapunov exponent does not depend on $\mathbf{x}_0$, i.e. orbits diverge at the same speed at any phase space area. This is a condition of applicability of the formula (2.34) even in the linear case.

Thus, a predictability time can be defined without appealing to the Lyapunov exponent which seems even more reasonable. As shown below, the Lyapunov exponent may not relate to a predictability time τ_q, i.e. a system with a greater value of the Lyapunov exponent (a more chaotic system) can have a greater value of τ_q (to be more predictable) compared to a less chaotic system. Besides, systems with the same values of the Lyapunov exponent can have very different predictability times τ_q. Let us discuss an analytic example from Smith (1997). For the sake of definiteness, we speak of the doubling time τ_2.

An example where the Lyapunov time and τ_2 coincide (up to a multiplier $\ln 2$) is a two-dimensional non-linear map which is one of basic models in non-linear dynamics – a baker's map

[26] Roughly, this is a probability density p of the visitations of a representative point to different areas of an attractor (see, e.g., Kuznetsov, 2001).

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

with $\alpha = 1/\beta = 1/2$. This map is area preserving (conservative). It maps the square $[0, 1) \times [0, 1)$ on itself. An invariant measure satisfies a condition $p(x, y) = 1$ so that fractal dimension of any kind described in Sect. 2.1.4 is equal to 2. It is called a baker's map since its action on a unit square reminds operations of a baker with a piece of pastry. Firstly, pastry is compressed twice along the y-axis and stretched twice along the x-axis. Secondly, it is cut in half and the right piece is located over the left one via a parallel shift. A single iteration of the map involves all those manipulations (Fig. 2.9). For almost any initial condition on the plane, two nearby points differing only in their x-coordinate are mapped to two points separated by a distance twice as big as the initial one. Similarly, a distance along the y-axis becomes twice as small in a single iteration. Thus, for any point within the square, the direction of the x-axis corresponds to the largest local Lyapunov exponent. The latter does not depend on the interval Δt and equals just to the largest Lyapunov exponent. This is a system with a uniform speed of nearby orbit divergence. Since $\Lambda_1 = \ln 2$, the Lyapunov time is equal to $\tau_\Lambda = 1/\ln 2$. The time $\tau_2(\mathbf{x}_0)$ equals 1 for any initial condition, i.e. a perturbation is doubled in a single iteration. Accordingly, an average doubling time is $\tau_2 = 1$.

Let us now consider a modification of the system (2.37) called a baker's apprentice map:

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

where square brackets denote the greatest integer not exceeding the number in the brackets, $\alpha = (2^N - 1)/2^N$ and $\beta = 2^{2^N}$. The action of this map is as follows. A greater piece of pastry $[0, \alpha) \times [0, 1)$ is compressed very weakly along the y-axis and stretched along the x-axis turning into the piece $[0, 1) \times [0, \alpha)$. The right narrow band is compressed very strongly β times along the y-axis. Thereby, one gets a

Fig. 2.9 A single iteration of a baker's map. The square is coloured with *black* and *white* to show where the points from different areas are mapped to

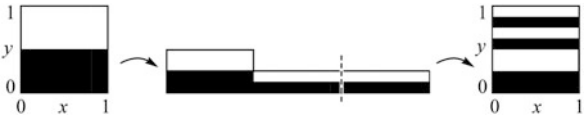

Fig. 2.10 A single iteration of the baker's apprentice map for $N = 1$ (the most "skilful" apprentice)

narrow belt of an integer length which is divided into belts of unit length. The latter ones are located over the big piece $[0, 1) \times [0, \alpha)$ as a pile, the left belts being below the right ones (Fig. 2.10).

This system also preserves an area and has an invariant measure $p(x, y) = 1$ (hence, fractal dimension of any kind equals 2 as for the baker's map). Its directions corresponding to a greater and a smaller local Lyapunov exponents also coincide with the directions of the coordinate axes. It can be shown that its largest Lyapunov exponent is equal to $\Lambda_1 = \alpha \ln 1/\alpha + (1-\alpha) \ln \beta$ and at the above particular values, one gets $\Lambda_1 = \mathbf{ln\,2} - \alpha \ln \alpha > \mathbf{ln\,2}$. Thus, system (2.38) is more chaotic than system (2.37) in the sense of the greater Lyapunov exponent. Its local Lyapunov exponents depend strongly on initial conditions. They are very small for the area of smaller x (a predictability time is big), while they are very big for the area of bigger x (this area is very narrow). An example of operations of both maps with a set of points is shown in Fig. 2.11. A result of four iterates of the baker's map (2.37) is shown in the middle panel. The picture is completely distorted, i.e. predictability is bad. A result of the four iterates of the map (2.38) with $N = 4$ is shown in the right panel. A significant part of the picture is well preserved being just weakly deformed: predictability in this area is good.

The most interesting in this example is the following circumstance. Not only local predictability times $\tau_2(\mathbf{x}_0)$ in some areas are greater for the map (2.38) than for the map (2.37), an average time τ_2 for the map (2.38) is also greater though it has a greater Lyapunov exponent! The value of τ_2 can be found analytically as

$$\tau_2 = \frac{1-\alpha^j}{1-\alpha},$$

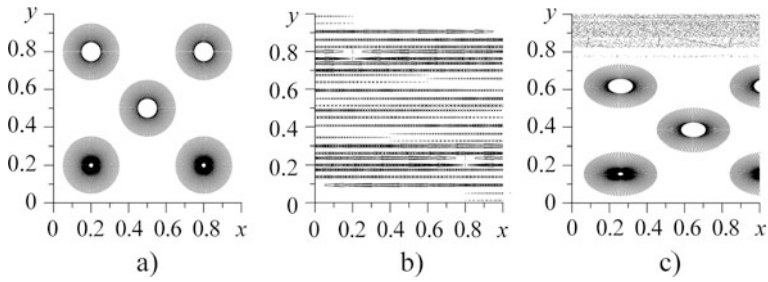

Fig. 2.11 Illustration of dynamics of the maps (2.37) and (2.38) analogous to that presented in Smith (1997): (**a**) an initial set of points; (**b**) an image under the fourth iterate of the baker's map (2.37); (**c**) an image under the fourth iterate of the baker's apprentice map (2.38) with $N = 4$

Table 2.1 Characteristics of the map (2.38) for different N: a maximal local predictability time, a mean predictability time, the largest Lyapunov exponent (Smith, 1997)

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	$1.09 \cdot \ln 2$				

where $j = \left[-\frac{\ln 2}{\ln \alpha}\right]^*$ and $[\cdot]^*$ denote the smallest integer greater than or equal to the number in the brackets. It can be shown that a predictability time $\tau_2 \approx 2^{N-1} \to \infty$ and $\Lambda_1 \to \ln 2$ for $N \to \infty$. The results of analytic manipulations for some N are brought together in Table 2.1: the predictability time can be arbitrarily high for systems as chaotic as (2.37) and even with a bit greater Lyapunov exponent!

Thus, Lyapunov exponents do not exhaust a question about predictability. Still, they carry certain information and become relevant characteristics of predictability if the speed of the divergence of orbits is uniform over a phase space.

2.5 Scale of Consideration Influences Classification of a Process (Complex Deterministic Dynamics Versus Randomness)

In practice, data are measured at finite accuracy, i.e. arbitrarily small scales of consideration are unavailable. At that, it is often difficult to decide whether an observed irregular behaviour is deterministically chaotic or stochastic (random). Strictly speaking, one can answer such a question only if data are generated with a computer and, therefore, it is known what law they obey. For a real-world process, one should ask which of the two representations are more adequate. A constructive approach is suggested in Cencini et al. (2000), where the answer depends on the *consideration scale*.

To characterise quantitatively an evolution of a perturbation with a size ε in a DS (2.9), it is suggested to use a *finite-size* Lyapunov exponent (FSLE) denoted as $\lambda(\varepsilon)$. It indicates how quickly orbits initially separated by a distance ε diverge. In general, finite perturbations may no longer be described with the linearised equation (2.10). To compute a FSLE, one needs first to introduce a norm (length) of state vectors. In contrast to the case of infinitesimal perturbations, a numerical value of $\lambda(\varepsilon)$ depends on the norm used. For the sake of definiteness, let us speak of the Euclidean norm and denote the norm of an initial perturbation as $\|\boldsymbol{\varepsilon}(0)\| = \varepsilon_0$. The value of a perturbation reaches threshold values $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ at certain time instants. For instance, let us specify the thresholds as $\varepsilon_n = 2\varepsilon_{n-1}$, $n = 1, \ldots, P-1$ and speak of a perturbation doubling time for different scales $\tau_2(\varepsilon_n)$. Let us perform N experiments by "launching" neighbouring orbits separated by a distance ε_0 from different initial conditions. We get an individual doubling time of $\tau_2^{(j)}(\varepsilon_n)$, $j = 1, \ldots, N$ for each pair of orbits. A mean doubling time is defined as

$$\tau_2(\varepsilon_n) = \left(1/N\right) \sum_{j=1}^{N} \tau_2^{(j)}(\varepsilon_n)$$

and an FSLE is defined as $\lambda(\varepsilon_n) = \ln 2/\tau_2(\varepsilon_n)$.

If a process is deterministically chaotic and a speed of phase orbit divergence is constant over an entire phase space, then $\lim_{\varepsilon \to 0} \lambda(\varepsilon) = \Lambda_1$ (Sect. 2.4.2).[27] It is important that for a deterministic process, $\lambda(\varepsilon)$ does not depend on ε at small scales: $\lambda(\varepsilon) = \text{const}$. For a stochastic process, $\lambda(\varepsilon) \to \infty$ for $\varepsilon \to 0$. The law of the rise in $\lambda(\varepsilon)$ with decreasing ε may be different, e.g. $\lambda(\varepsilon) \propto \varepsilon^{-2}$ for a Brownian motion (Wiener's process, Sect. 4.2). The authors of Cencini et al. (2000) suggest the following approach to the distinction between deterministically chaotic signals and noisy (random) ones. If one gets for a real-world process that $\lambda(\varepsilon) = \text{const}$ within a certain range of scales, then it is reasonable to describe the process as deterministic in that range of scales. If $\lambda(\varepsilon)$ rises with decreasing ε within a certain range of scales, then the process should be regarded as noisy within that range.

A simple example is a deterministic map exhibiting a "random walk" (*diffusion*) at large scales:

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

where

$$F(y) = \begin{cases} (2+\delta)y, & 0 \le y < 0.5, \\ (2+\delta)y - (1+\delta), & 0.5 \le y < 1.0, \end{cases}$$

and square brackets denote an integer part. The function F is plotted in Fig. 2.12 for $\delta = 0.4$. The Lyapunov exponent equals to $\Lambda_1 = \ln|F'| = \ln|2+\delta|$. The process behaves like Wiener's process (Sect. 4.2) at $\varepsilon > 1$. For instance, $\varepsilon = 1$ means that one traces only an integer part of x. A change of an integer part by ± 1 results from the deterministically chaotic dynamics of a fractional part of x. Since the latter is ignored in consideration at large scales, the former looks like random walk. Figure 2.13 shows that $\lambda(\varepsilon) \approx 0.9$ and the process is classified as deterministic within the range of scales $\varepsilon < 1$. One gets $\lambda(\varepsilon) \propto \varepsilon^{-2}$ for $\varepsilon > 1$ and considers the process as random.

Let us modify the map (2.39) by introducing noise ξ_n uniformly distributed on a segment $[-1, 1]$ and replacing F with its approximation G (10,000 linear pieces with a slope 0.9 instead of the two pieces with a slope 2.4):

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

[27] FSLE defined via doubling times is equal to zero for a process with $\Lambda_1 < 0$.

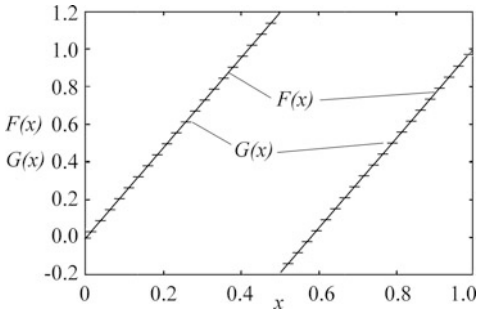

Fig. 2.12 Function $F(x)$ from equation (2.39). Horizontal lines are its approximation $G(x)$ from equation (2.40) consisting of 40 segments with zero slope (Cencini et al., 2000)

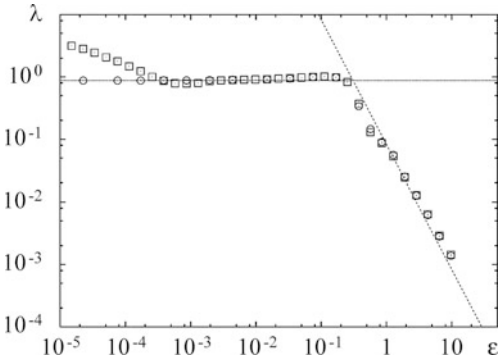

Fig. 2.13 FSLE versus a scale. Circles are shown for the system (2.39), squares for the system (2.40) with $G(x)$ consisting of 10,000 segments with a slope of 0.9 (Cencini et al., 2000)

where the quantity $\sigma = 10^{-4}$ determines the noise standard deviation. The processes (2.39) and (2.40) do not differ for $\varepsilon > 1$ and look like a random walk (Fig. 2.13). They look deterministic with the same Lyapunov exponent in the interval $10^{-4} < \varepsilon < 1$ despite different slopes of their linear pieces: 2.4 in equation (2.39) and 0.9 in equation (2.40). This is the result of averaging of the local linear dynamics of equation (2.40) over the scales $\varepsilon > 10^{-4}$. The processes differ for $\varepsilon < 10^{-4}$ where the process (2.40) behaves again as random from the viewpoint of $\lambda(\varepsilon)$ due to the presence of the noise ξ. Thus, dynamical properties may differ at different scales. It is important to take it into account in describing complex real-world processes.

Based on the described approach, the authors have suggested witty terms to characterise some irregular processes: "*noisy chaos*" and "*chaotic noise*". The first one relates to a process which looks deterministic (chaos) at large scales and random (noise) at small ones, i.e. a macroscopic chaos induced by a micro-level noise. Analogously, the second term describes a process which is random at large scales and deterministic at small ones.

2.6 "Coin Flip" Example

Most likely, everybody used to put a coin on bent fingers (Fig. 2.14a), offer "head" or "tail", flip it and . . .relieve him/herself of responsibility for some decision. A small disk falling with rotation is popular as a symbol of candour, an embodiment of chance for different peoples at different times. We use it below to illustrate the discussion of determinancy, randomness, and different approaches to modelling.

We start with a conceptual model. In a typical case, a hand imparts to a coin both a progressive motion with an initial velocity v_0 and a rotation with an initial angular velocity ω_0. Further, the disk flies interacting with the earth and an air until it falls on a surface. If the latter is solid, then it would jump up several times and finally settle down on one of its sides. Without a special practice, one can hardly repeat a flip several times so as to reproduce the same result, e.g. a head. It gets impossible for a strong flip when a coin has enough time to perform many revolutions before landing. The main cause of irreproducibility is a significant scattering of initial velocities and coordinate. In part, one can reach reproducibility if a special device is used, e.g. a steel ruler with a gadget to adjust a deformation[28] (Fig. 2.14b). However, such a device is not a panacea: one can confidently predict a result only for weak flips when a coin performs half a revolution, a single revolution or at most two revolutions (Fig. 2.14c). The longer is the way before landing, the more is an uncertainty in a

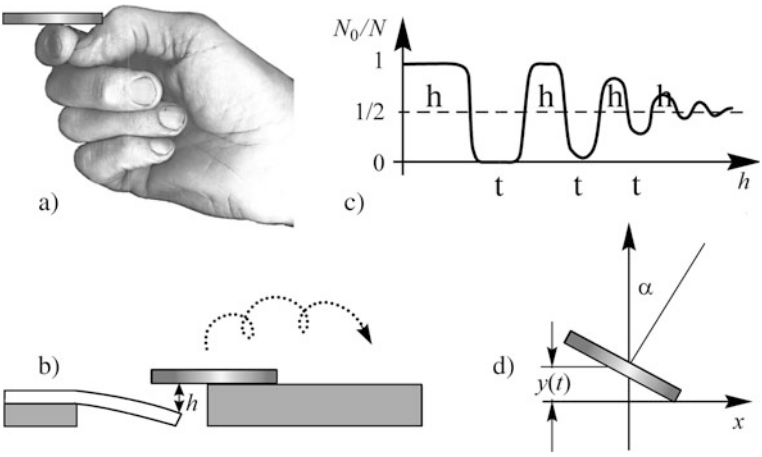

Fig. 2.14 Exercises with a coin: (**a**) a standard situation; (**b**) a physical model with a controllable "strength of a kick" (h is a ruler bend); (**c**) qualitative outlook of an experimental dependency "frequency of a head versus a kick strength" (N experiments were performed at a fixed h, N_0 is the number of resulting heads); (**d**) illustration to a conceptual model

[28] A persistent student providing us with experimental data flipped a coin 100 times per experiment with a ruler. He controlled a bend of the ruler by changing the number of pages in a book serving as a ruler support.

final state. Frequencies of a head and a tail equalise despite conditions of successive experiments seem the same.

For a dynamical modelling, let us characterise a coin state with a coordinate y and a velocity v of its centre of mass along with an angle of rotation α about the z-axis perpendicular to x and y (Fig. 2.14d) and an angular velocity ω. Let us single out three qualitatively different stages in the system evolution and introduce special approximation at each of them.

Start. Initial conditions: a coin starts to move having a head as its upper side with a linear velocity v_0 directed vertically; rotation occurs clockwise with an angular velocity ω_0 (Fig. 2.14d). If $2y < d\sin\alpha$ for the starting conditions (where d is the diameter of the coin), then an edge of a coin touches a support after the start of motion (rotation leaves take-off behind) and we regard an outcome as a head. For $2v_0/\omega_0 > d$, the coin flies away without touching the plane of $y = 0$.

Flight. Let us neglect interaction of a coin with air. Let it interact only with the earth. Then, an angular velocity remains constant and is equal to ω_0, while the centre of mass moves with a constant acceleration g.

Finish. Touching a table happens at a time instant t_f, $2y(t_\mathrm{f}) = d\sin\alpha(t_\mathrm{f})$, and rotation stops immediately. A coin falls on one of its sides depending on the value of a rotation angle. One gets a head for $0 < (\alpha(t_\mathrm{f})\mathrm{mod}2\pi) < \pi/2$ or $3\pi/2 < (\alpha(t_\mathrm{f})\mathrm{mod}2\pi) < 2\pi$ and a tail for $\pi/2 < (\alpha(t_\mathrm{f})\mathrm{mod}2\pi) < 3\pi/2$.

It is too difficult to specify a single evolution operator for all stages of motion. Therefore, we confine ourselves only with the stage of flight and qualitative considerations for the first and the last stages. Thus, it is obvious that there are many attractors in the phase space of the system: equilibrium points with coordinates $y = v = \omega = 0$, $\alpha = n\pi$, $n = 0, 1, 2, \ldots$, corresponding to final states of a coin lying on one of its sides (Fig. 2.15a shows "tail" points by filled circles and "head" points by open ones). Different attractors correspond to different numbers of coin revolutions before landing. According to the conceptual model, strong dissipation takes place in shaded phase space areas, corresponding to the final stage and to a motion with a small initial velocity v_0, and a representative point reaches one of the two attractors. Boundaries of their basins can be determined from a model of flight. Let us derive them in analogy to Keller (1986) asymptotically from a set of Newton's differential equations $\mathbf{F} = m \cdot \mathbf{a}$ and $\mathbf{M} = I \cdot \boldsymbol{\beta}$, where $\mathbf{F}$ and $\mathbf{M}$ are resultants of forces and their moment of rotation, respectively, $\mathbf{a}$ and β are linear and angular accelerations, m and I are the coin mass and moment of inertia. In our case, a model takes the form

$$\mathrm{d}y/\mathrm{d}t = v,\ \mathrm{d}v/\mathrm{d}t = -g,\ \mathrm{d}\alpha/\mathrm{d}t = \omega_0,\ \mathrm{d}\omega/\mathrm{d}t = 0. \tag{2.41}$$

Given initial conditions, a solution to equation (2.41) is an orbit

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

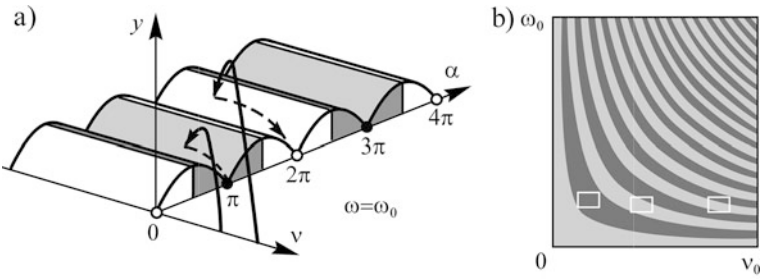

Fig. 2.15 Illustrations to the dynamics of the coin flip model: (**a**) a three-dimensional section of phase space of the dynamical system (2.41) at $\omega = \omega_0$. Wave-shaped surface bounds an area of the final stage where strong dissipation takes place. The considered model of "flight" does not describe the start from the shaded area. Curves with arrows are examples of phase orbits; (**b**) a section of the phase space with a plane ($y = 0,\ \alpha = 0$). Basin of attractors corresponding to the final state "tail" is shaded. *White rectangles* illustrate an accuracy of the initial condition setting (a noise level); their area is $\Delta\nu \times \Delta\omega$

From here, one gets a dependency $\alpha(t_f) = f(\nu_0, \omega_0)$ and expressions for the basin boundaries on the plane ω_0, ν_0 (a section of phase space by a plane $\alpha = 0,\ y = 0$, Fig. 2.15b)

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

Given exact initial conditions, which are obligatory under the dynamical approach, a coin reaches a definite final state. According to this approach, one can predict a final state of a coin, which is illustrated in Fig. 2.16a, where the frequency of "a head" outcome takes only the values of 0 and 1 depending on ν_0. It corresponds to reality only for small ν_0 (Fig. 2.16b). However, if a flip is sufficiently strong so that a coin performs several revolutions, then such an approach only misleads. Experiments show that by even making efforts to improve accuracy of initial condition setting, one can assure "a head" or "a tail" outcome only for small number of a coin revolution. A significantly more plausible model is obtained if one refuses dynamical description and introduces random quantities into consideration. Let us assume that $\nu_0 = V_0 + \xi$, where V_0 is a deterministic component, ξ is a random quantity, e.g. distributed uniformly in some interval of $\Delta\nu$ with a centre at V_0. Such a stochastic model demonstrates dependency on V_0 qualitatively coinciding with an experiment. Frequencies of both outcomes tend to be 0.5 and vertical and horizontal pieces of the plot are smoothed out for a large number of revolutions.

Given a uniform distribution of ξ, it is convenient to explain observed regularities by selecting a rectangular area $\Delta\nu \times \Delta\omega$ with a centre at V_0 (Fig. 2.15b). If the entire area is included into a basin of a certain attractor, then an event corresponding to that attractor occurs for sure, i.e. a frequency of one of the outcomes equals unity. If the area intersects both basins (for "a head" and "a tail"), then a frequency of a certain outcome is determined by a portion of the area occupied by the corresponding basin. In general, a frequency of "a head" is defined by an integral taken over the entire

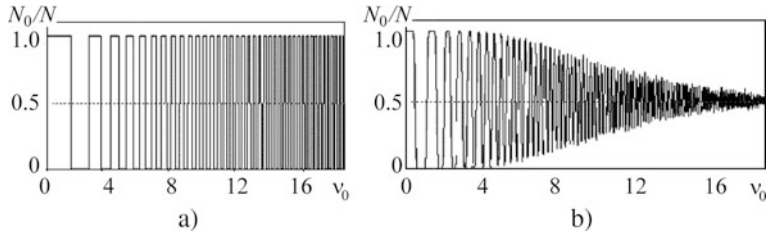

Fig. 2.16 Frequency of "a head" versus an initial velocity v_0 at a fixed value of ω_0: (**a**) exact setting of initial conditions; (**b**) an error in initial conditions

region occupied by its basin of attraction $P\{H\} = \iint\limits_H p(v_0, \omega_0)\mathrm{d}v_0\,\mathrm{d}\omega_0$, where p is the probability density for observing a "head" in respect of the initial conditions.

Apart from the considered asymptotic and stochastic models, one can suggest a purely empirical probabilistic model. For instance, one can approximate an experimental dependency of the frequency of "a head" on the initial velocity (or on the strength of a flip) shown in Figs. 2.14c and 2.16b with a formula

$$N_0/N = \begin{cases} z, & 0 < z(v) < 1, \\ 1, & z(v) > 1, \\ 0, & z(v) < 0, \end{cases} \qquad z(v) = 0.5 + a\,\mathrm{e}^{-bv}\cos(cv)\,. \tag{2.44}$$

Thus, we have illustrated possibility of the description of a single real-world object with different models, both dynamical and stochastic ones. Each of the models can be useful for certain purposes. It proves again a conventional character of the labels "dynamical system" and "random quantity" in application to real-world situations. In general, even an "international" symbol of randomness, a coin flip, should be considered from the viewpoint of a partial determinancy conception.

Finally, we note that apart from the alternative "deterministic models versus stochastic models", there are other, more complex, interactions between the deterministic and stochastic approaches to modelling. In particular, complicated deterministic small-scale behaviour may be appropriately described by stochastic equations and large-scale averages of a random process may exhibit a good deal of deterministic regularity (Sect. 2.5). Traditional statistical approaches, such as methods of statistical moments or Kalman filtering, are successfully used to estimate parameters in deterministically chaotic systems (see Sects. 8.1.2 and 8.2.2 for concrete examples). Concepts of the theory of probability and the theory of random processes are fruitfully used to describe statistical properties of dynamical chaos, (see, e.g. Anishchenko et al., 2005a, b). Therefore, both approaches discussed throughout this chapter are often used together for the description of complex phenomena in nature.

References

Aivazian, S.A.: Statistical Investigation of Dependencies. Moscow, Metallurgiya (in Russian) (1968)
Alligood, K., Sauer, T., and Yorke, J.A.: Chaos: An Introduction to Dynamical Systems. New York: Springer Verlag, New York (2000)
Andronov, A.A., Vitt, A.A., Khaikin, S.E.: Theory of Oscillators. Nauka, Moscow (1959). Translated into English: Pergamon Press, Oxford (1966)
Andronov, A.A., Leontovich, E.A., Gordon, I.I., Mayer, A.G.: Theory of Bifurcations of Dynamic Systems on a Plane. Nauka, Moscow (1967). Translated into English: Wiley, New York (1973)
Anishchenko, V.S.: Attractors of dynamical systems. Izvestiya VUZ. Applied Nonlinear Dynamics (ISSN 0869-6632). **5**(1), 109–127, (in Russian) (1997)
Anishchenko, V.S., Astakhov, V.V., Neiman, A.B., Vadivasova, T.Ye., Schimansky-Geier, L.: Nonlinear Effects in Chaotic and Stochastic Systems. Tutorial and Modern Development. Springer, Berlin (2002)
Anishchenko, V.S., Okrokvertskhov, G.A., Vadivasova, T.E., Strelkova, G.I.: Mixing and spectral correlation properties of chaotic and stochastic systems: numerical and physical experiments. New J. Phys. **7**, 76–106 (2005a)
Anishchenko, V.S., Vadivasova, T.Ye., Okrokvertskhov, G.A., Strelkova, G.I.: Statistical properties of dynamical chaos. Phys. Uspekhi. **48**, 163–179 (2005b)
Arnold, V.I.: Additional Chapters of the Theory of Ordinary Differential Equations. Nauka, Moscow (in Russian) (1978)
Arnold, V.I.: Ordinary Differential Equations. Nauka, Moscow (1971). Translated into English: MIT Press, Cambridge, MA (1978)
Badii, R., Politi, A.: Complexity: Hierarchical Structures and Scaling in Physics. Cambridge University Press, Cambridge (1997)
Bautin, N.N., Leontovich, E.A.: Methods and Techniques for Qualitative Investigation of Dynamical Systems on a Plane. Nauka, Moscow (in Russian) (1990)
Belykh, V.N.: Elementary introduction into qualitative theory and bifurcation theory in dynamical systems. Soros Educ. J. (1), 115–121 (in Russian) (1997)
Berger, P., Pomeau, Y., Vidal, C.: Order Within Chaos. Hermann, Paris (1984)
Broomhead, D.S., King, G.P.: Extracting qualitative dynamics from experimental data. Physica D. **20**, 217–236 (1986)
Butenin, N.V., Neimark, Yu.I., Fufaev, N.A.: Introduction into the Theory of Nonlinear Oscillations. Nauka, Moscow (in Russian) (1987)
Cencini, M., Falcioni, M., Olbrich, E., et al. Chaos or noise: Difficulties of a distinction. Phys. Rev. E. **62**, 427–437 (2000)
Danilov Yu.A.: Lectures in Nonlinear Dynamics. An Elementary Introduction. Postmarket, Moscow (in Russian) (2001)
Eckmann, J.P., Ruelle, D.: Ergodic theory of chaos and strange attractors. Rev. Mod. Phys. **57**, 617–656 (1985)
Gnedenko, B.V.: A Course in Probability Theory. Gostekhizdat, Moscow (1950). Translated into English: Morikita, Tokyo (1972)
Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Physica D. **9**, 189–208 (1983)
Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector-Fields. Springer, Berlin (1983)
Hoel, P.G.: Introduction to Mathematical Statistics, 4th ed. Wiley, New York (1971)
Ibragimov, I.A., Has'minskii, R.Z.: Asymptotic Theory of Estimation. Nauka, Moscow (1979). Translated into English Under the Title Statistical Estimation. Springer, New York (1981)
Katok, A., Hasselblat, B.: Introduction to the modern theory of dynamical systems. Encyclopaedia of Mathematics and its Applications, vol. 54. Cambridge University Press, Cambridge (1995)
Keller, J.B.: The probability of heads. Am. Math. Monthly. **93**(3), 191–197 (1986)

Kendall, M.G., Stuart, A.: The Advanced Theory of Statistics,. vols. 2 and 3. Charles Griffin, London (1979)
Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A. **45**, 3403–3411 (1992)
Kipchatov, A.A.: Estimate of the correlation dimension of attractors, reconstructed from data of finite accuracy and length. Tech. Phys. Lett. **21**(8), 627–629 (1995)
Korn, G., Korn, T.: Handbook of Mathematics for Engineers and Scientists. McGraw-Hill, New York (1961)
Kornfel'd I.P., Sinai Ya.G., Fomin, S.V.: Ergodic Theory. Springer, Berlin (1982)
Kravtsov, Yu.A.: Fundamental and practical limits of predictability. In: Kravtsov, Yu.A. (ed.) Limits of Predictability, pp. 170–200. TsentrCom, Moscow (in Russian) (1997)
Kravtsov, Yu.A.: Randomness, determinancy, predictability. Phys. Uspekhi. **158**(1), 93–115 (in Russian) (1989)
Kuznetsov, S.P.: Complex dynamics of oscillators with delayed feedback (review). Radiophys. Quantum Electr. **25**(12), 1410–1428 (in Russian) (1982)
Kuznetsov, S.P.: Dynamical Chaos. Fizmatlit, Moscow (in Russian) (2001)
Landa, P.S.: Nonlinear Oscillations and Waves in Dynamical Systems. Kluwer Academic Publishers, Dordrecht (1996)
Loskutov, A.Yu., Mikhailov, A.S.: Basics of Complex Systems Theory. Regular and Chaotic Dynamics, Moscow (2007)
Lusternik, L.A., Sobolev, V.I.: Elements of Functional Analysis. Nauka, Moscow (1965). Translated into English: Hindustan Publishing Corpn., Delhi (1974)
Makarenko, N.G.: Fractals, attractors, neural networks and so forth. Procs. IV All-Russian Conf. "Neuroinformatics-2002", Moscow, 2002. M., 2002. Part 2. pp. 121–169 (in Russian)
Malinetsky, G.G.: Chaos. Structures. Numerical Experiment. An Introduction to Nonlinear Dynamics. Editorial URSS, Moscow (in Russian) (2000)
Malinetsky, G.G., Potapov, A.B.: Contemporary Problems of Nonlinear Dynamics. Editorial URSS, Moscow (in Russian) (2000)
Mathematical Encyclopedic Dictionary. 846p. Sov. Encyclopedia, Moscow (in Russian) (1988)
Philosophic Encyclopedic Dictionary. Sov. Encyclopedia, Moscow, 840p. (in Russian) (1983)
Pisarenko, V.F., Sornette, D.: Statistical methods of parameter estimation for deterministically chaotic time series. Phys. Rev. E. **69**, 036122 (2004)
Poincare, H.: Science and Method. Dover, New York (1982)
Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C. Cambridge University Press, Cambridge (1988)
Pugachev, V.S.: Theory of Probabilities and Mathematical Statistics. Nauka, Moscow (in Russian) (1979)
Pugachev, V.S.: Probability Theory and Mathematical Statistics for Engineers. Pergamon Press, Oxford (1984)
Shalizi, C.R.: Methods and techniques of complex systems science: an overview, vol. 3, arXiv:nlin.AO/0307015 (2003). Available at http://www.arxiv.org/abs/nlin.AO/0307015
Shil'nikov L.P., Shil'nikov A.L., Turayev, D.V., Chua, L.: Methods of Qualitative Theory in Nonlinear Dynamics. Parts I and II. World Scientific, Singapore (2001)
Sinai Ya.G.: Contemporary problems of ergodic theory. Fizmatlit, Moscow (in Russian) (1995)
Smith, L.A.: Maintenance of uncertainty. Proc. Int. School of Physics "Enrico Fermi", Course CXXXIII, pp. 177–246. Italian Physical Society, Bologna (1997). Available at http://www.maths.ox.ac.uk/∼lenny
Vapnik, V.N.: Estimation of Dependencies Based on Empirical Data. Nauka, Moscow (1979). Translated into English: Springer, New York (1982)
Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
von Mises, R.: Mathematical Theory of Probability and Statistics. Academic Press, New York (1964)