Chapter 1

The utility of probability

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering,
Faculty of Civil Engineering, National Technical University of Athens, Greece

Summary

Commonly, probability is regarded to be a branch of applied mathematics that provides tools for data analysis. Nonetheless, probability is a more general concept that helps shape a consistent, realistic and powerful view of the world. Historically, the modern science was initiated from deterministic views of the world, in which probability had a marginal role for peculiar unpredictable phenomena. However, in the turn of the nineteenth century, radical developments in physics, and particularly thermodynamics, dynamical systems and quantum physics, as well as in mathematics has given the probability theory a central role in the scientific scene, in the understanding and the modelling of natural phenomena. Furthermore, probability has provided grounds for philosophical concepts such as indeterminism and causality, as well as for extending the typical mathematical logic, offering the mathematical foundation of induction. In typical scientific and technological applications, probability provides the tools to quantify uncertainty, rationalize decisions under uncertainty, and make predictions of future events under uncertainty, in lieu of unsuccessful deterministic predictions. Uncertainty seems to be an intrinsic property of nature, as it can emerge even from pure and simple deterministic dynamics, and cannot been eliminated. This is demonstrated here using a working example with simple deterministic equations and showing that deterministic methods may be good for short-term predictions but for long horizons their predictive capacity is cancelled, whereas the probabilistic methods can perform much better.

1.1 Determinism and indeterminism

The philosophical proposition of *determinism* is widely accepted in science. It is manifested in the idea of a clockwork universe, which comes from the French philosopher and scientist René Descartes (1596-1650) and was perfected by the French mathematician and astronomer Pierre-Simon Laplace (1749-1827). It is vividly expressed in the metaphor of *Laplace's demon*, a hypothetical all-knowing entity that knows the precise location and momentum of every atom in the universe at present, and therefore could use Newton's laws to reveal the entire course of cosmic events, past and future. (Isaac Newton – 1643-1727 – however, rejected cartesian thinking and especially the clockwork idea; he was aware of the fragility the world and believed that God had to keep making adjustments all the time to correct the emerging chaos.) The demon who knows the present perfectly is of course a metaphor; what is more important in this idea is that knowing the present perfectly one can deduce the future and the past using Newton's laws. The metaphor helps us understand that, according to deterministic thinking, the roots of uncertainty about future should be subjective, i.e. rely on the fact that we do not know exactly the present, or we do not have good enough methods and models. It is then a matter of time to eliminate uncertainty, acquiring better data (observations) and building better models.

However, according to *indeterminism*, a philosophical belief contradictory to determinism, uncertainty may be a structural element of nature and thus cannot be eliminated. Indeterminism has its origin in the Greek philosophers Heraclitus (*ca*. 535–475 BC) and Epicurus (341–270 BC). In science, indeterminism largely relies on the notion of *probability*, which according to the Austrian-British philosopher Karl Popper (1902-1994) is the extension (quantification) of the Aristotelian idea of *potentia* (Popper, 1982, p. 133). Practically, the idea is that several outcomes can be produced by a specified cause, while in deterministic thinking only one outcome is possible (but it may be difficult to predict which one). Probability is a quantification of the likelihood of each outcome or of any set of outcomes. In this chapter we use the term probability in a loose manner. In the next chapter we will provide a precise description of the term using the axiomatization introduced by the soviet mathematician Andrey Nikolaevich Kolmogorov (1903-1987).

In everyday problems deterministic thinking may lead to deadlocks, for instance in dealing with the outcome of a dice throw or a roulette spin. The movements of both obey Newton's laws; however, application of these laws did not help anyone to become rich predicting the dice outcomes. In an attempt to rectify such deadlocks, some have been tempted to divide the natural phenomena into two categories, deterministic (e.g. the movement of planets) and random (e.g. the movement of dice). We maintain that this is a fallacy (both planets and dice obey to the same Newton's laws). Another very common fallacy of the same type (in fact, an extension of the former) is the attempt to separate natural processes into deterministic and random components, one superimposed (usually added) to the other. Both fallacies can be avoided by abandoning the premise of determinism.

1.2 Deduction and induction

In mathematical logic, determinism can be paralleled to the premise that all truth can be revealed by *deductive reasoning* or *deduction* (the Aristotelian *apodeixis*). This type of reasoning consists of repeated application of strong syllogisms such as:

If A is true, then B is true
A is true
Therefore, B is true

and

If A is true, then B is true;
B is false
Therefore, A is false

Deduction uses a set of axioms to prove propositions known as theorems, which, given the axioms, are irrefutable, absolutely true statements. It is also irrefutable that deduction is the preferred route to truth; the question is, however, whether or not it has any limits. David Hilbert (1862-1943) expressed his belief that there are no limits in his slogan (from his talk in 1930; also inscribed in his tombstone at Göttingen): "*Wir müssen wissen, wir werden wissen* - We must know, we will know". His idea, more formally known as *completeness*, is that any mathematical statement could be proved or disproved by deduction from axioms.

In everyday life, however, we use weaker syllogisms of the type:

If A is true, then B is true;
B is true
Therefore, A becomes more plausible

and

If A is true, then B is true;
A is false
Therefore, B becomes less plausible

The latter type of syllogism is called *induction* (the Aristotelian *epagoge*). It does not offer a proof that a proposition is true or false and may lead to errors. However, it is very useful in decision making, when deduction is not possible.

An important achievement of probability is that it quantifies (expresses in the form of a number between 0 and 1) the degree of plausibility of a certain proposition or statement. The formal probability framework uses both deduction, for proving theorems, and induction, for inference with incomplete information or data.

1.3 The illusion of certainty and its collapse

Determinism in physics and completeness in mathematics reflect the idea that uncertainty could in principle be eliminated. However, in the turn of the nineteenth century and the first half of the twentieth century this idea proved to be an illusion as it received several blows in four major scientific areas, summarized below.

1.3.1 Statistical physics and maximum entropy

In its initial steps, thermodynamics was based on purely deterministic concepts and particularly on the notion of the caloric fluid, a hypothetical fluid (a weightless gas) that flows from hotter to colder bodies (passes in pores of solids and liquids). The caloric theory was proposed in 1783 by Antoine Lavoisier and persisted in scientific literature until the end of the 19th century. In 1902 the term *statistical thermodynamics* was coined by the American mathematical-engineer, physicist, and chemist J. Willard Gibbs. The statistical theory of thermodynamics is essentially based on the probabilistic description of kinetic properties of atoms and molecules and was very successful in explaining all concepts and phenomena related to heat transfer.

The concept of *entropy* (from the Greek *εντροπία*), which was essential for the formulation of the second law of thermodynamics (by Rudolf Clausius in 1850), was given a statistical interpretation by Ludwig Boltzmann (in 1872). The second law says that the entropy of an isolated system which will tend to increase over time, approaching a maximum value at equilibrium. Boltzmann showed that entropy can be defined in terms of the number of possible microscopic configurations that result in the observed macroscopic description of a thermodynamic system. In 1878, Gibbs extended this notion of entropy introducing the idea of the statistical (or thermodynamic) ensemble, an idealization consisting of a large number (sometimes infinitely many) of mental copies of a system, each of which represents a possible state that the real system might be in. In 1948, Claude E. Shannon generalized the concept of entropy and gave it an abstract probabilistic definition applicable for any random variable, thus essentially showing that entropy is a measure of uncertainty of a system. Kolmogorov and his student Sinai went far beyond and suggested a definition of the metric entropy for dynamical systems (their results were published in 1959). In 1957, the American mathematician and physicist Edwin Thompson Jaynes extended Gibbs' statistical mechanics ideas showing that they can be applied for statistical inference about any type of a system. Specifically, he showed that the *principle of maximum entropy* can be used as a general method to infer the unknown probability distribution of any random variable. For instance, the principle of maximum entropy can easily produce that the probability of the landing of a die in each of its six faces will be 1/6 (any departure from equality of all six probabilities would decrease the uncertainty of the event). It is thus impressive that the principle that predicts that heat spontaneously flows from a hot to a cold body, is the same principle that can give the probability distribution of dice.

Thus, statistical thermodynamics has formed a nice paradigm entirely based on probability as a tool for both explanation and mathematical description of natural behaviours. Furthermore, the second law of thermodynamics essentially shows that nature works in a way that maximizes uncertainty in complex systems. Following nature's behaviour and applying the principle of maximum entropy (maximum uncertainty) to any type of system we can infer useful knowledge about the system's behaviour. This knowledge, however, is no longer expressed in terms of certainty about the sharp states of the system, but rather in terms of probabilities of these states. In large systems however, it turns out that this knowledge can lead to nearly precise descriptions of macroscopical properties, despite the maximum uncertainty at the microscopical level. For instance, we can easily infer that the average of the outcomes of 45 000 dice throws will be between 3.49 and 3.51 with probability 99.99%. From a practical point of view such a statement is almost equivalent to certainty; however, it does not preclude the case that all 45 000 will be sixes (and the average will be also six).

1.3.2 Dynamical systems and chaos

Chaos (from the Greek χάος) is most often referred to as a deterministic notion (deterministic chaos). Yet it offers an excellent insight of uncertainty, even in the case of purely deterministic dynamics. The basic concepts of chaos are due to the French mathematician Jules Henri Poincaré (1854–1912). In 1890, Poincaré's memoir on the three body problem was published in the journal Acta Mathematica as the winning entry in the international prize competition sponsored by Oscar II, King of Sweden and Norway, to mark his 60th birthday. Today this paper is renowned for containing the first mathematical description of chaotic behavior in a dynamical system (Barrow-Green, 1994). It was the first time that the complexity of Newtonian dynamics was demonstrated, even in a system as apparently simple as three gravitational bodies. Poincaré gave the first example of the *sensitive dependence on initial conditions*, a characteristic of chaotic behaviour that is met in unstable dynamical systems.

Ironically, however, the prize winning work of Poincaré was not exactly the published one. In contrast, in his original work Poincaré, had found certain stability results for the three-body problem. After the prize award (1889) and after the prize winning essay had been printed (but not distributed), Poincaré discovered a fatal flaw in his proof that was supposed to show that the universe worked like clockwork. Poincaré then had to spend his monetary prize plus 1000 Crowns to withdraw the printed volumes with the erroneous version of the memoir, as well as several months of work to correct the error. In the final paper he had reinstated the chaos in the movement of the astral bodies and brought down for ever the idea of a clockwork universe.

We can understand the emergence of chaos and chance from purely deterministic dynamics reading his own words (from Henri Poincaré, Science et méthode, 1908; reproduced in Poincaré, 1956, p. 1382):

> *A very small cause, which escapes us, determines a considerable effect which we cannot help seeing, and then we say that the effect is due to chance. If we could know exactly the laws of nature and the situation of the universe at the initial instant, we should be able to predict the situation of this same universe at a subsequent instant. But even when the natural laws should have no further secret for us, we could know the initial situation only* ***approximately****. If that permits us to foresee the succeeding situation* ***with the same degree of approximation****, that is all we require, we say the phenomenon had been predicted, that it is ruled by laws. But it is not always the case; it may happen that slight differences in the initial conditions produce very great differences in the final phenomena; a slight error in the former would make an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon.*

> *... Why have the meteorologists such difficulty in predicting the weather with any certainty? Why do the rains, the tempests themselves seem to us to come by chance, so that many persons find it quite natural to pray for rain or shine, when they would think it ridiculous to pray for an eclipse? We see that great perturbations generally happen in regions where the atmosphere is in unstable equilibrium... Here again we find the same contrast between a very slight cause, inappreciable to the observer, and important effects, which are sometimes tremendous disasters.*

Non-linear chaotic dynamics remained in the backwoods of mathematics and physics until the 1960s, even though some of the leading mathematicians, mostly in Russia/USSR (Lyapunov, Kolmogorov, Andronov), worked on it. Then the use of computers made it possible to experiment with chaos in numerical applications. The American meteorologist Edward Norton Lorenz was an early pioneer of experimenting chaos with computers; also he coined the term *butterfly effect* to encapsulate the notion of sensitive dependence on initial conditions in chaotic systems: a butterfly's wings (a small change in the initial condition of the atmospheric system) might create tiny changes in the atmosphere that ultimately cause a tornado to appear.

Now the mathematical theory of nonlinear complex chaotic dynamic systems is centre stage and mainstream. A prominent characteristic of the notion of chaos is that it is easily understandable, as it may involve simple deterministic dynamics, and allows the experimentation with very simple examples that exhibit chaotic behaviour. Such a simple example we will study in the next section. It is fascinating that a simple nonlinear deterministic system (such as the gravitational movement of three bodies or the hydrological system studied below) can have a complex, erratic evolution. Sadly, however, most of hydrological studies understood this in the opposite direction: they attempted to show, making naïve and mistaken use of tools from dynamical systems, that complexity in hydrological phenomena implies that their dynamics are simple (Koutsoyiannis, 2006).

1.3.3 Quantum physics

While chaotic systems demonstrated that uncertainty can be produced even in a purely deterministic framework, quantum physics has shown that uncertainty is an intrinsic characteristic of nature. In this respect, probability is not only a necessary epistemic addition or luxury for modelling natural phenomena. Rather it is a structural element of nature, an ontological rather than epistemic concept.

Quantum physics has put limits to the knowledge we can obtain from observation of a microscopic system and has shown that exact measurements are impossible. The outcome of even an ideal measurement of a system is not sharp (exact), but instead is characterized by a probability distribution. The Heisenberg uncertainty principle gives a lower bound on the product of the uncertainty measures of position and momentum for a system, implying that it is impossible to have a particle that has an arbitrarily well-defined position and momentum simultaneously. Thus, our familiar deterministic description proves to be impossible for the microscopic world.

A famous example that shows how fundamental the notion of probability is in nature is the double-slit experiment. Light is shined at a thin, solid barrier that has two slits cut into it. A photographic plate is set up behind the barrier to record what passes through slits. When only one slit is open, there is only one possibility for a photon, to pass through the open slit. Indeed the plate shows a single dark line where all the photons are accumulated. However, when both slits are open and only one photon at a time is fired at the barrier, there are two possibilities for the photon, which however are not mutually exclusive because, according to the uncertainty principle, the position of the photon is not sharp. Thus, it seems that the photon passes from both slits simultaneously. This will be recorded in the photographic plate, which shows regions of brightness and darkness (interference fringes). It seems that a single photon materializes the theoretical probability distribution in each case. According to our macroscopic experience the phonon would follow one of the two available options, and at the time it passes though the barrier it would be in one of the two slits with equal probabilities. However, in a quantum physics description the photon is simultaneously in both slits and the two probabilities interfere.

Such phenomena are difficult to describe or explain based on our experience (and language) of the macroscopic world. However, the phenomena of the quantum physics are reflected in the macroscopic world too (e.g. in the double-slit experiment), and thus cannot be irrelevant to our description of macrocosmos. For instance, statistical physics is strongly influenced by quantum physics.

1.3.4 Incompleteness

While the three previous developments eventually deal with physics, this fourth one concerns pure mathematical logic. In 1931 the Austrian mathematician Kurt Gödel proved two important theorems, so-called *incompleteness theorems*, stating inherent limitations of mathematical logic. The theorems are also important in the philosophy of mathematics and in wider areas of philosophy. The first incompleteness theorem practically says that any system with some axioms, containing the natural numbers and basic arithmetic (addition, multiplication) is necessarily incomplete: it contains undecidable statements, i.e. statements that are neither provably true nor provably false. Furthermore, if an undecidable statement is added to the system as an axiom, there will always be other statements that still cannot be proved as true, even with the new axiom. The second theorem says that if the system can prove that it is consistent, then it is inconsistent. That is to say, we can never know that a system is consistent, meaning that it does not contain a contradiction. Note that if the system contains a contradiction, i.e. a case where a proposition and its negation are both provably true, then every proposition becomes true.

Ironically, Gödel had presented his incompleteness results the day before Hilbert pronounced his slogan discussed above (*Wir müssen wissen, wir werden wissen*). Obviously, the slogan received a strong blow by Gödel's results. The conjectured almightiness of deduction was vitiated. In other words, Gödel's results show that uncertainty is not eliminable. Simultaneously, they enhance the role of probability theory, as extended logic, and the necessity of induction (see also Jaynes, 2003, p. 47).

1.3.5 The positive side of uncertainty

Surprisingly, the new role of probability is not well assimilated in the scientific community. The quest of determinism and uncertainty elimination still dominates in science. Another symptom of this type is the exorcism of probability and its replacement with any type of substitutes. One good example for this is provided by the *fuzzy methods* which are regarded much more fashioned than probability. However, no solutions using fuzzy approaches could not have been achieved at least as effectively using probability and statistics (Laviolette, *et al.,* 1995). The Education still promotes deterministic thinking as if all above fundamental changes in science had not happened. Hopes are expressed that these results are flawed and determinism will be reinstated. These results are considered negative and pessimistic by many. We maintain that they are absolutely positive and optimistic. Life would not have any meaning without uncertainty. This is well known by people working in the media, who spend much money to show live (i.e. with uncertain outcome) reportages and sports games; had determinism been more fascinating, they would show recorded versions in the next day, with eliminated uncertainty (e.g. the score of the game would be known). Without uncertainty concepts such as *hope*, *will* (particularly *free will*), *freedom*, *expectation*, *optimism*, *pessimism* etc. would hardly make sense.

1.4 A working example

With this example we will see that, contrary to intuition, pure determinism does not help very much to predict the future, even in very simple systems. The example studies a hydrological system that is fully deterministic and is deliberately made extremely simple. This system is a natural plain with water stored in the soil, which sustains some vegetation. We assume that each year a constant amount of water I = 250 mm enters the soil and that the potential evapotranspiration is also constant, PET = 1000 mm. (Obviously in reality the inflow and potential evapotranspiration – especially the former – vary in an irregular manner but we deliberately assumed constant rates to simplify the example and make it fully deterministic). The actual evapotranspiration is $E \leq$ PET. We assume that a fraction f of the total plain area is covered by vegetation, and that the evapotranspiration rate in this area equals PET and in all other area is zero (assuming no route of soil water to the surface), so that in the entire plain, the average actual evapotranspiration will be

$$E = \mathrm{PET}\, f \tag{1.1}$$

It is easy to see that if $f = I$ / PET = 0.25 then $E = I = 250$ mm, i.e. the input equals the output and the system stays at an equilibrium; the water stored in the soil stays at a constant value. The situation becomes more interesting if at some time $f \neq 0.25$. In this case f may vary in time. It is natural to assume that f will increase if there is plenty of water stored in the soil (the vegetation will tend to expand) and to decrease otherwise. We denote s the water stored in the soil and we assume a certain reference level for which we set $s = 0$, so that $s > 0$ stands for soil water excess and $s < 0$ for soil water deficit.

Our system is described by the two state variables, the soil water s and the vegetation cover f, which can vary in time. If $i = 1, 2, \ldots$ denotes time in years, then the water balance equation for our system is

$$s_i = s_{i-1} + I - \mathrm{PET}\, f_{i-1} \tag{1.2}$$

Since our system is described by two state variables, we need one more equation to fully describe its dynamics (i.e. its evolution in time). Naturally, the second equation should be sought in the dynamics of grow and decay of plants, which however may be too complicated. Here we will approach it in an extremely simplified, conceptual manner. We set a basic desideratum that f should increase when $s > 0$ and decrease otherwise. A second desideratum is the consistency with the fact that $0 \leq f \leq 1$.

Such desiderata are fulfilled by the curves shown in Fig. 1.1. The curves are described by the following equation, which takes an input x and produces an output y, depending on a parameter a that can take any real value, positive or negative:

$$y = g(x, a) := \frac{\max(1 + a, 1)x}{\max(1 - a, 1) + ax} \tag{1.3}$$

By inspection it can be verified that if $0 \leq x \leq 1$, then $0 \leq y \leq 1$, whatever the value of a is. Furthermore, it can be seen that if $a = 0$ then $y = x$, when $a > 0$ then $y > x$, and when $a < 0$ then $y < x$.

Thus, if in equation (1.3) we replace x with f_{i-1}, y with f_i, and a with some increasing function of s_{i-1} such that it takes the value 0 when $s_{i-1} = 0$, then we obtain an equation that is conceptually consistent with our desiderata. For the latter let us set $a \equiv (s_{i-1}/s^*)^3$, where s^* is a standardizing constant assumed to be $s^* = 100$ mm. Hence, the equation that completes the system dynamics becomes

$$f_i = g(f_{i-1}, (s_{i-1}/s^*)^3) \tag{1.4}$$

or

$$f_i = \frac{\max(1 + (s_{i-1}/s^*)^3, 1) f_{i-1}}{\max(1 - (s_{i-1}/s^*)^3, 1) + (s_{i-1}/s^*)^3 f_{i-1}} \tag{1.5}$$

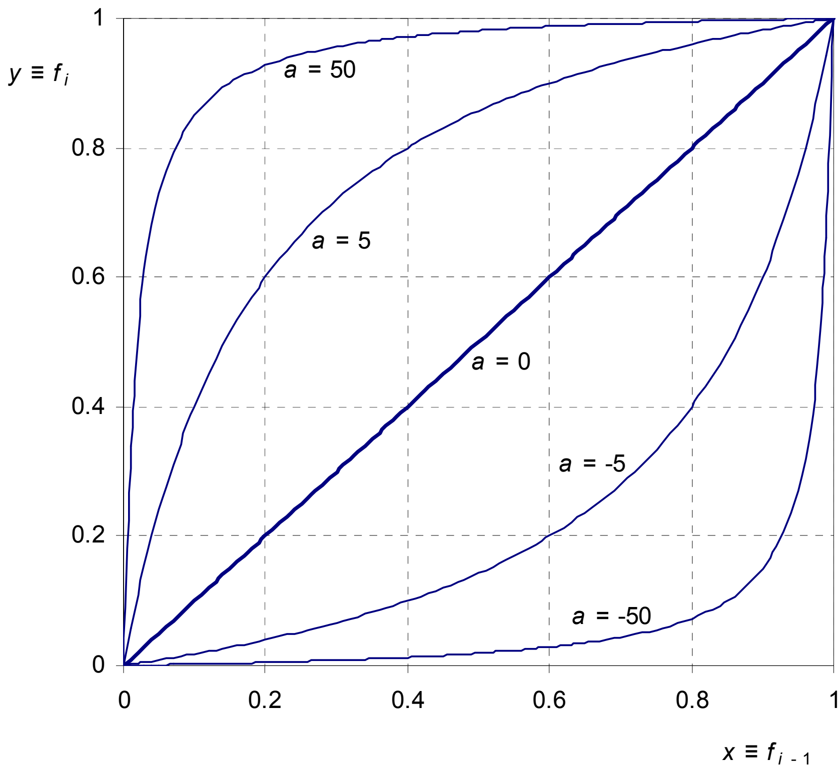

Fig. 1.1 Graphical depiction of equation (1.3) for several values of the parameter a, which is an increasing function of the soil water s.

From the system dynamics (1.2) and (1.5), it can be easily verified that if the initial conditions at time $i = 0$ are $f_0 = 0.25$, $s_0 = 0$, then the system will stay for ever at state $f_i = 0.25$, $s_i = 0$ for any time i. Now let us assume that the initial conditions depart from these conditions of stability. For instance we consider $f_0 = 0.30$, $s_0 = 100$ mm. From the system dynamics (1.2) and (1.5) we can easily find that, at time $i = 1$, $f_1 = 0.462$ (the vegetation cover was increased because of surplus water) and $s_1 = -111.5$ mm (the increased vegetation consumed more water, so that the surplus was exhausted and now there is deficit). Continuing in this manner we can calculate (f_2, s_2), (f_3, s_3) etc. It is a matter of a few minutes to set up a spreadsheet with two columns that evaluate equations (1.2) and (1.5), and calculate the system state (f_i, s_i) at time $i = 1$ to, say, 10 000, given the initial state (f_0, s_0) (homework). Fig. 1.2 depicts the first 100 values of the evolution of system state. It is observed that the system does not tend to the stable state discussed above. Rather, the vegetation cover fluctuates around 0.25 (roughly between 0 and 0.8) and the soil water fluctuates around 0 (roughly between -400 and 400 mm). These fluctuations seem to have a period of roughly 4-5 years but are not perfectly periodic.

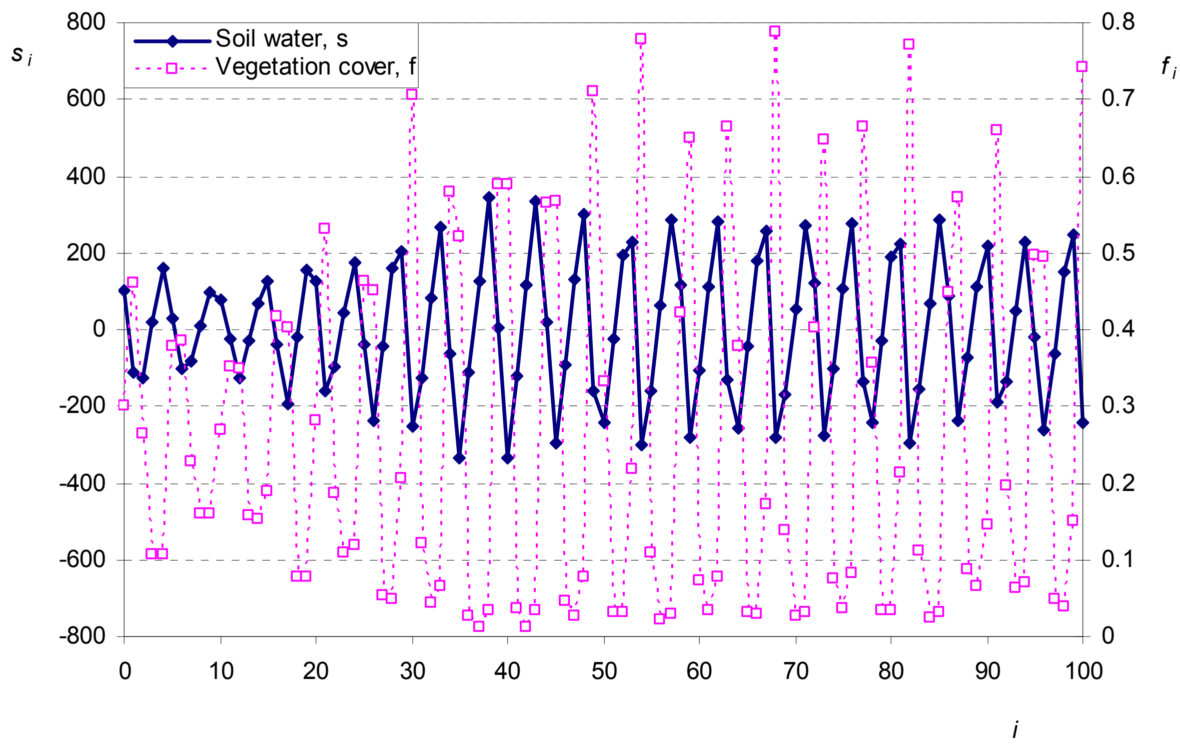

Fig. 1.2 Graphical depiction of the system evolution for time up to 100.

Despite fluctuating behaviour, it appears that we can predict exactly any future state given the initial conditions f_0, s_0. The question we wish to examine here is this: Will predictions represent reality? We can split this question to two: (1) Is our model a perfect representation of reality? (2) Is our knowledge of the initial conditions perfectly accurate? The reply to both questions should be negative. No model can represent nature perfectly; all models are just approximations. Furthermore, our knowledge of initial conditions at the best case comes from measurements and all measurements include some error or uncertainty.

Let us circumvent the first problem, and assume that our model is perfect. Put it in a different way, let us temporarily forget that the mathematical system with dynamics (1.2) and (1.5) aims to represent a natural system, so that we do not care about model errors. What is then the effect of imperfect knowledge of the initial conditions? To demonstrate this, we assume that the initial conditions set above are obtained by rounding off some true values, which introduces some small error. (We suppose that rounding off mimics the measurement error in a natural system). Our true conditions are assumed to be f_0 = 0.2999, s_0 = 100.01 mm and our approximations are f'_0 = 0.30, s'_0 = 100 mm, as above; the errors in f_0 and s_0 are -0.0001 and 0.01 mm, respectively. Repeating our calculations (with our spreadsheet) with the true conditions, we obtain a set of *true* values that are depicted in Fig. 1.3, along with the *approximate* values. By approximate we mean the values that were obtained by the rounded off initial conditions f'_0 and s'_0; these values are precisely those shown in Fig. 1.2.

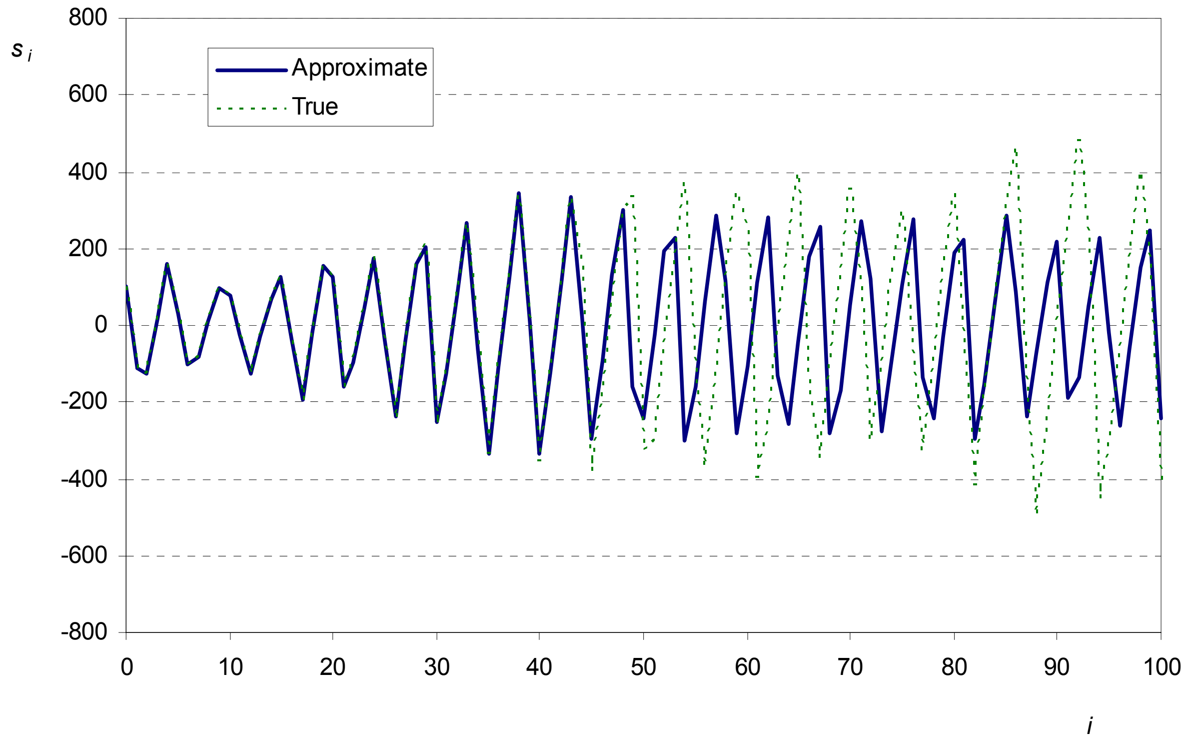

Fig. 1.3 Graphical depiction of the true system evolution and its approximation for time up to 100.

We can observe in Fig. 1.3 that the approximation is almost perfect for times up to 40 but becomes imperfect for larger times. For instance, the true value at time i = 60 is s_{60} = 245.9 mm whereas the approximate value is s'_{60} = –105.4 mm. Thus a small error of 0.01 mm in the initial conditions is magnified to 245.9 – (–105.4) = 491.8 mm in 60 time steps. Here we have used this definition of error:

$$e_i := s_i - s'_i \tag{1.6}$$

This large error clearly suggests that deterministic dynamics, even perfectly known and simple, may be unable to give deterministic future predictions for long lead times.

Nevertheless, in engineering applications it is often necessary to cast predictions for long time horizons. For instance, when we design a major project, we may have a planning horizon of say 100 years and we wish to know the behaviour of the natural system for the next 100 years. However, in most situations we are interested about the events that may occur and particularly about their magnitude while we are not interested about the exact time of occurrence. Such predictions can be obtained in a different manner, which may not need to know the deterministic dynamics of the system. Rather, it is based on the statistical properties of the system *trajectory* as reflected in a time series of the system evolution.

In the simplest case, a statistical prediction is obtained by taking the average of the time series. In our system this average of s is around 0, so that the prediction for any future time is simply s'_i = 0. As strange as it may seem, for large lead times this prediction is better (i.e.

gives a smaller error) than obtained by running the deterministic model. For instance, at time $i = 60$, $e_i = 245.9 - 0 = 245.9 < 491.8$ mm. A graphical depiction of prediction errors of both the deterministic and statistical method (where in the second method $e_i = s_i - 0 = s_i$) is shown in Fig. 1.4

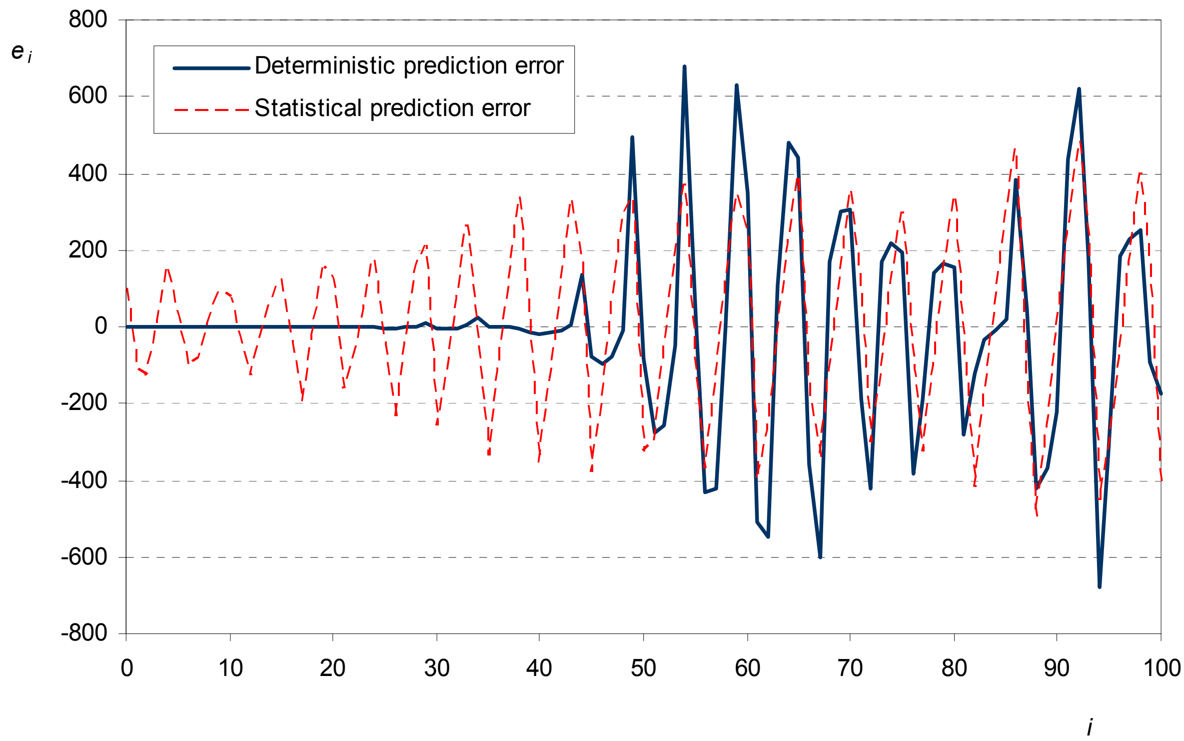

Fig. 1.4 Comparison of prediction errors of the deterministic and statistical methods for time up to 100.

We can observe that the deterministic method yields zero error for time up to 25 and negligible error for time up to 40. Then the error becomes high, fluctuating between about –800 and 800. The error of the statistical prediction fluctuates in a narrower range, between –500 and 500 mm. Statistics gives us a way to give a quantitative global measure of the error and compare the errors quantitatively rather than graphically. Thus for the n-year period $[l, l + n - 1]$ we define the root mean square (RMS) error as

$$e_{\mathrm{RMS}} := \frac{1}{n}\sqrt{\sum_{i=l}^{l+n-1} e_i^2} \tag{1.7}$$

The logic in taking the squares of errors and then summing up is to avoid an artificial cancelling up of negative and positive errors. Thus, for the last 50 years ($l = 51$, $n = 50$) the RMS error (calculated in the spreadsheet) is 342.0 and 277.9 mm, respectively. This verifies that the statistical prediction is better than the deterministic one for times > 50. On the other hand, Fig. 1.4 clearly shows that the deterministic prediction is better than the statistical for times < 50.

This happens in most real world systems, but the time horizon, up to which a deterministic prediction is reliable, varies and depends on the system dynamics. For instance, we know that a weather prediction, obtained by solving the differential equations describing the global atmospheric system dynamics, is very good for the first couple of days but is totally unreliable for more than a week or ten days lead time. After that time, statistical predictions of weather conditions, based on records of previous years for the same time of the year, are more reliable.

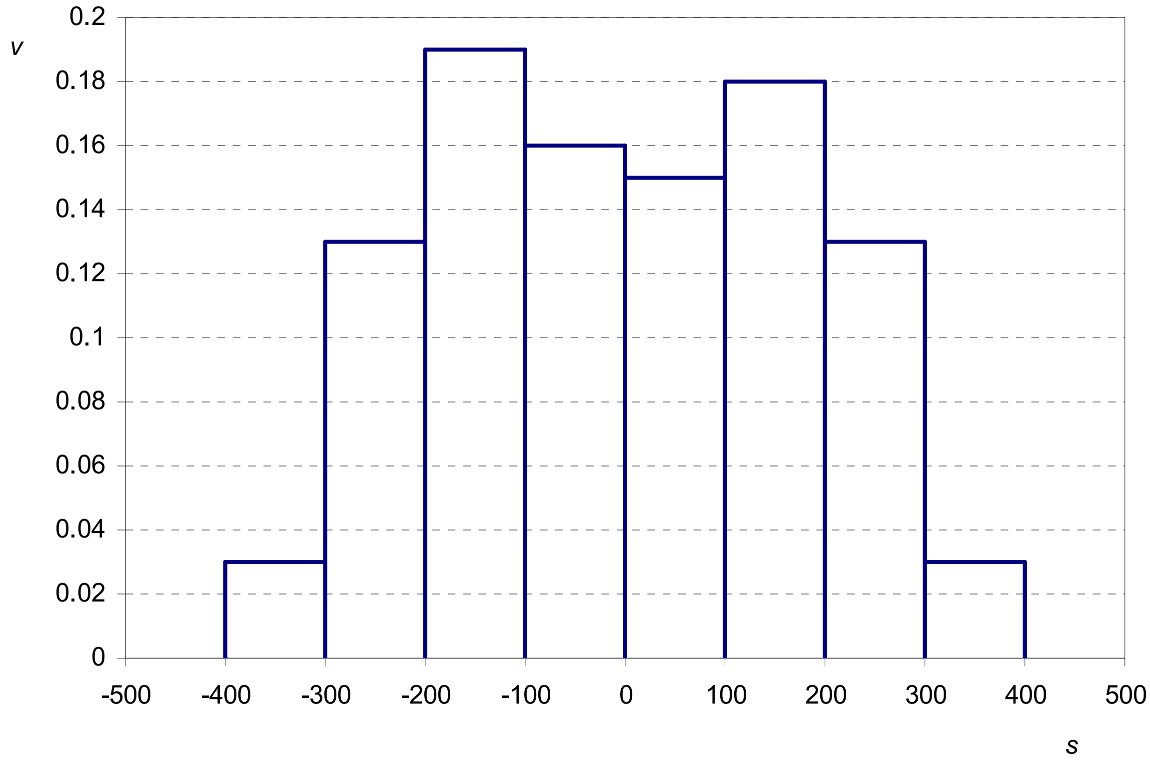

Fig. 1.5 Relative frequency *v* of the intervals of *s,* each with length 100 mm, as determined from the time series shown in Fig. 1.2.

A statistical prediction is generally more powerful than indicated in the example above. Instead of providing a single value (the value 0 in the example) that is a likely future state of the system, it can give ranges of likely values and a likelihood measure for each range. This measure is an empirical estimate of *probability* obtained by analyzing the available time series and using the theory of probability and statistics. That is to say, it is obtained by induction and not by deduction. In our example, analyzing the time series of Fig. 1.2, we can construct the *histogram* shown in Fig. 1.5, which represents empirically estimated probabilities for ranges of values of the soil water *s*. The histogram shows for instance that with probability 16%, *s* will be between –100 mm and 0, or that with probability 3%, *s* will be between –400 and –300 mm. We must be careful, however, about the validity of empirical inferences of this type. For instance, extending this logic we may conclude from Fig. 1.5 that with probability 100% the soil water will be between –400 and 400 mm. This is a mistaken conclusion: we cannot exclude values of soil water smaller than –400 mm or higher than 400 mm. The probabilities of such extreme (very low or very high) events are nonzero. To find them the empirical observations do not suffice and we need some theoretical tools, i.e. deductive reasoning. The tools are provided by the *probability theory* and the related areas of *statistics* and *stochastics*. Particularly, the latter area deals with processes that possess some dependence in time and perhaps cyclical behaviour (as happens in our example), and endeavour to incorporate any known deterministic laws within a unified, probability based, mathematical description.

1.5 Concluding remarks

If we try to make the above example more realistic, we should do several changes. Particularly: (a) the input (soil infiltration) should vary in time (and in space) in a rather irregular (random) manner; and (b) the relationship between soil water and vegetation cover should be revisited in light of some observational data in the specific area. For step (a) we need to build an additional model to simulate the input. This model should utilize infiltration data in the area, if available, or other hydrological data (rainfall, runoff) of the area; in the latter case an additional model that transforms rainfall to infiltration and runoff will be required. In all cases, the building of the model will require tools from probability, statistics, and stochastics. For step (b), which aims at establishing a deterministic relationship, it is wise to admit from the beginning the great difficulty or impossibility to establish the relationship by pure theoretical (deductive) reasoning. Usually a mixed approach is followed: (b1) a plausible (conceptual) mathematical expression is assumed that contains some parameters strongly affecting its shape; and (b2) an available time series of measurements is used to estimate its parameters. Step b2 is clearly based on a statistical/inductive approach and will always give some error; in fact the parameter estimation is done with the target to minimize (but not to eliminate) the error. This error should be modelled itself, again using tools from probability, statistics, and stochastics.

It may seem contradictory, at first glance, that in the establishment of a deterministic relationship we use statistical tools. As strange as it may seem, this happens all the time. The detection of deterministic controls, based on observed field or laboratory data, and the establishment of deterministic relationships, again based on data, is always done using tools from probability, statistics, and stochastics. A variety of such tools, all probability-based, are available: least squares estimation, Bayesian estimation, spectral analysis, time delay embedding (based on the entropy concept) and others. Here it should be added that even purely deterministic problems such as the numerical optimization of a purely deterministic non-convex function and the numerical integration of a multivariate purely deterministic function can be handled more efficiently and effectively by probability-based methods (evolutionary algorithms and Monte Carlo integration, respectively) rather than by deterministic methods.

Obviously, in a realistic setting of our example problem, the system trajectory should look more irregular than demonstrated above and the horizon for a reliable deterministic prediction should decrease significantly, perhaps to zero. In this case, a probabilistic-statistical treatment of the problem should be attempted *from the outset*, not for long horizons only. In this case we need not disregard the deterministic dynamics, if identified. On the contrary, stochastic methods are able to make explicit use of any identified deterministic control, so as to improve predictions as much as possible. That is to say, a stochastic approach *from the outset* does not deny *causality* and deterministic controls; rather it poses them in a more consistent framework admitting that uncertainty is inherent in natural systems. Here we should clarify that causality is conceptually different in a deterministic and a probabilistic approach. In the former case causality (or causation) is a directional relationship between one event (called cause) and another event (called effect), which is the consequence (result) of the first. In a stochastic view of the world, the definition of causality can be generalized in the following way (Suppes, 1970): An event A is the *prima facie* cause of an event B if and only if (i) A occurs earlier than B, (ii) A has a nonzero probability of occuring, and (iii) the conditional probability[*] of B occurring when A occurs is greater than the unconditional probability of B occurring.

It is, however, possible that in a real world problem our attempt to establish a causal relationship between our state variables fails. In a probabilistic framework this is not a tragedy, provided that we have a sufficient series of observations. We can build a model (for instance for the soil water s) without having identified the system dynamics. This is actually done in many cases of hydrological simulations.

All graphs in the above example indicate that the trajectories of the state variables of our system are irregular; simultaneously, they do not look like a purely random phenomenon, such as a series of roulette outcomes. This is very important and should be taken into serious consideration in any modelling attempt using probabilistic tools. In fact, the trajectories of natural systems never look like our more familiar purely random systems.[†] One major difference is the dependence in time, which may be very complex, contrary to the independence of roulette outcomes or to simple type of dependence (e.g. Markovian) encountered in simplistic stochastic models. In one of the next chapters we will examine these properties (revisiting the above example). We note, however, that such properties of natural processes, which seem peculiar in comparison to simple random systems, have been overlooked for years. Even worse, the standard statistical framework that was developed for independent events has been typically used in hydrological and geophysical applications, and this gave rise to erroneous results and conceptions.

Acknowledgment I thank Richard Mackey for his discussions that have influenced and fed this text.

[*] For a formal definition of conditional probability see chapter 2.

[†] We will revisit these differences in chapter 4.

References

Barrow-Green, J., Oscar II's prize competition and the error in Poincaré's memoir on the three body problem, *Archive for History of Exact Sciences*, 48(2), 107-131, 1994.

Jaynes, E.T., *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

Koutsoyiannis, D., On the quest for chaotic attractors in hydrological processes, *Hydrological Sciences Journal*, 51(6), 1065-1091, 2006.

Laviolette, M., J. W. Seaman, Jr., J.D. Barrett and W.H. Woodall, A probabilistic and statistical view of fuzzy methods, *Technometrics*, 37(3), 249-261, 1995.

Poincaré, H. Chance, *The World of Mathematics*, Simon & Schuster, New York, 1956.

Popper, K., *Quantum Theory and the Schism in Physics*, Routledge, 1982.

Suppes, P., *A Probabilistic Theory of Causality*, North-Holland, Amsterdam, 1970.