On Deterministic Chaos and Randomness: a Joint Approach

Massimo Stella

Introduction

The concept of mathematics as a tool to interpret and quantify natural phenomena flourished in the 17th century, when Galileo and Newton revolutionised science, with [16] and [27] respectively. The main legacy of these works was the idea that Nature followed immutable consequential laws, expressible in a procedural mathematical language, a notion later translated in the philosophical paradigm of determinism. Today this Weltanschaung seems to collide with the unpredictable and irregular phenomena of daily experience, such as the weather [24] or asset fluctuations [7] or turbulence [15].

Laplace, strenous defender of determinism, tried to address this issue already in [21], by arguing that the regularity of celestial mechanics phenomena was a property shared with the whole physical reality, even though a human being could not perceive it because of his incomplete or fragmentary experience. Laplace’s position was clearly influenced by the great success of determinism in the field of celestial mechanics at his time, i.e. the predicted discovery of Neptune [35]. From an historical perspective, his concept of “ignorance” represents a prototypical example of epistemological uncertainty: the lack of knowledge of all the interactions and initial conditions of a given system ultimately determines the high variability in its evolution outcomes. Laplace naturally linked this irregularity to the concept of probability [21], even though nowadays the latter would probably be better addressed as an example of chaos [26, 8].

In fact, modern physics calls “chaotic” the very same deterministic yet unpredictable systems that Laplace found to be too difficult to fully understand without recurring to chance and stochasticity. In this way, the modern field of deterministic chaos represents the intersection between the fascinating yet uncertain world of probability and the apparently schematic and consequential “clockwork” universe of determinism.

The main aim of this review is to investigate the link between chaotic dynamical systems (see Section I) and a class of stochastic processes in probability theory, Markov chains (see Section II), also by recurring to the examples of two case studies (see Section III). The closing remarks of the analysis, speculating on the role of epistemological uncertainty in chaos, are contained in the ending part.
I. Chaos Theory

The seminal ideas of chaos theory date back to the end of the 19th century. In 1890 Poincaré’s essay on the three body problem won the international prize competition sponsored by Oscar II, King of Sweden and Norway, patron of mathematics and science [2, 36]. In his memoir [30], originally containing an error later revised, Poincaré showed that the three body problem exhibited a strong sensitivity on initial conditions, in spite of its celestial mechanics nature. In his subsequent work [31], Poincaré developed from scratches many theoretical tools, related to stability, recurrence and qualitative analysis, which ended up being of great relevance in the modern landscape of chaos [8]. Nonetheless, why his discoveries remained neglected for over fifty years is still a debatable issue [8, 36, 2]. On the one hand, Poincaré was a true “universalist”, a visionary mathematical physicist endowed with great intuitions but also limited by little to no interest in focusing on one area only [36]. However, this interpretation critically ignores the socio-cultural context in which Poincaré’s ideas about chaos originated [2, 8]. In the early 20th century, classical mechanics became overshadowed by the rising fields of relativity and quantum mechanics. Additionally, the computational power necessary to efficiently tackle Poincaré’s prescriptions was still unaccessible at that time. Unsurprisingly, a similar fate happened to Lyapunov’s characterisation of local stability analysis for nonlinear systems [25].

Chaos remained a niche topic until 1960s [35, 18, 26]. Interestingly, there is a general claim [8] that many other contributions directly related to Poincaré’s findings, and henceworth to chaos, even much before the sixties, but this is not entirely correct. In fact, the 1920s and 1930s saw the rise of ergodic theory [5, 34], another branch of mathematics analysing the evolution of systems over time, from a measure theoretic perspective. Even if, nowadays, we use results about ergodicity, such as Birkhoff’s theorem¹, in dealing with chaotic behavior, it has to be stressed that these findings originated for different purposes, in the context of statistical mechanics [4].

It was in the 1960s that chaos was brought to the attention of scientific community, thanks to the contributions of Lorenz from meteorology [24], Hénon from astrophysics [19], Sinai from the game of billiards [34] and Chirikov from plasma physics [9]. Additional evidences of the physical relevance of chaos came with the works of Ruelle, Takens and Newhouse on turbulence in fluids on the onset of chaos [15]. Another crucial development was Feigenbaum’s discovery [14] of the period doubling mechanism as a universal source of chaotic behavior. Rényi’s work related information theory and chaos, establishing a common background with statistical mechanics [4]. As already stated, fundamental in providing the suitable (practical) tools for the investigation of chaotic systems were computer aided techniques, such as numerical methods for the computation of Lyapunov exponents (Benettin, Galgani, Giorgilli and Strelcyn), the frac-

¹ Birkhoff’s ergodic theorem [5] outlines the conditions for a property evolving in time $f_t$ of a dynamical system (i.e. energy, entropy, etc.) to have a time average $\langle f_t \rangle$ almost everywhere equal to its space average $\langle f_i \rangle$. In details, $f_i$ can be thought of as the arithmetic average of the values of $f_t$ sampled at the same time but for different initial conditions. Systems satisfying Birkhoff’s theorem are said to be “ergodic”. In the 1955, Fermi, Pasta and Ulam showed that ergodicity is a non-generic property, even for low-dimensional systems [8]. The Fermi-Pasta-Ulam experiment was one of the first examples of simulations developed in synergy with analytical analysis of a given problem.
global communications. The state set was generalised to a countable
and telephonic networks. Erlang successfully applied Markov chains to call loss
state
geneous chains, for which the transition probability from state
the previous outcome (state)
the outcome (state)
incorrect claims
Markov chains were mainly conceived by the author as a reply to the
According to historic evidence (see [33] for references).

\begin{equation}
x_{t+1} = F(x_t)
\end{equation}

exhibits a chaotic behavior if and only if $F : \mathbb{J} \to \mathbb{J}$ is a unidimensional continuous function, having a period three orbit\(^2\). This finding was actually a specific case of the more general Sharkovskiy’s theorem [33]. Through the years, the work in [23] motivated both the motto “[period] three implies chaos” and a fierce debate, with Ruelle being one of the main detractors [32]. His main two objections were the lack of originality in [23] and also the author’s noncanonical definition of chaos. Reading Lü and Yorke’s original paper, the first objection simply does not hold, since the demonstrations in [23] are different\(^3\) from the ordering technique used by Sharkovsky in [33]. Additionally, Ruelle’s definition of chaos was based on the occurrence of exponential divergence sustaining for long-term behavior, an interpretation reminiscent of Poincaré’s results but which hides many problems [4, 32] behind the connotation of the “long-term” behavior (see the side picture). Li and Yorke tried to overcome this difficulty, by defining as “chaos” the presence of an infinite number of periodic orbits “scrambled” altogether [23].

II. Markov Chains

In 1906 Markov extended both the law of large numbers and the central limit theorem [12] to sequences or chains of “weakly” dependent random variables \(\{X_1, X_2, \ldots, X_m\}\), i.e. variables with a nontrivially zero conditional probability

\begin{equation}
P(X_m = s_m | X_1 = s_1, X_2 = s_2, \ldots, X_{m-1} = s_{m-1}).
\end{equation}

According to historic evidence (see [3] for the English references), Markov chains were mainly conceived by the author as a reply to the incorrect claims\(^4\) of Nekrasov, Markov’s main nemesys. The term “chain” originally meant the case with the conditional probability on the outcome (state) \(s_m\) of the random variable \(X_m\) depending only on the previous outcome (state) \(s_{m-1}\). Markov introduced also homogeneous chains, for which the transition probability from state \(i\) to state \(j\), \(p_{ij} = P(X_{t+1} = s_j | X_t = s_i)\), is effectively independent on the discrete transition time \(t\) [4, 8].

Unfortunately, Markov never grasped the full potential of his chains as a mathematical null model for physical systems, also because at his time the theory of stochastic processes [20, 4, 12] was not yet well established. Erlang successfully applied Markov chains to call loss and telephonic networks [13], his results becoming a milestone in global communications. The state set was generalised to a countable

\(^2\) In the discrete dynamical systems jargon an orbit of period \(a\) is a set of \(a\) points \(\{x_1, x_2, \ldots, x_a\}\) such that by applying \(F\) on the initial point \(x_1\) for \(a\) times retrieves \(x_1\). In formulas, a period three orbit is a set of states \(\{x_1, x_2, x_3\}\) and \(F(x_1) = x_2 \neq x_1, F(x_2) = x_3\) and \(F(x_3) = x_1\). This rather simple property can actually have unexpectedly complex consequences, as pinpointed by Li and Yorke [23].

\(^3\) Moreover, Sharkovskii’s results were initially published in Russian and they remained largely neglected until 1977 [8], until their translation in English.

\(^4\) Markov chains constitute a generalisation of earlier works on urn problems, random walks and Brownian motion by Laplace, Bunyakovsky and Einstein respectively [3, 12]. Nonetheless, Markov was unaware of the latter and he developed his chains mainly because of his rivalry towards Pavel Alekseevich Nekrasov, whose work he considered “an abuse of mathematics”.\n
The Lorenz system [24] displays chaotic behavior in three dimensions. For a given set of values, the system states evolve in time in the neighborhood of other attractor points (top), their trajectories shaping a “butterfly”. Additionally, the system displays sensitive dependence from initial configurations. Moving point \(P_1\) to a distance of \(10^{-6}\) to \(P_2\) and letting those two states evolve in time gives rise to two different trajectories (bottom). The problem of this definition of chaos is that attractors are vital to make a fair comparison between the evolved states. Pictures from [8].
infinity by Kolmogorov only in 1936 [12] and this paved the way for many generalised Markovian processes, historically applied to a variety of fields, ranging from bioinformatics to neural networks, from cryptography to chemical networks [17, 28, 29]. In particular, Kolmogorov introduced also the concept of ergodic Markov chains, i.e. chains for which it is possible to go from every state to every state, in a finite number of transition moves. With the advent of symbolic dynamics [18, 20, 12], the original “chain” definition was further modified to include also prohibited state sequences, originating the concept of topological Markov chains [3]. The latter include conditional probabilities, like \( P \), that can be zero also if the entire sequence \( s_1 s_2 \ldots s_{m-1} \) is already forbidden [4]. Additionally, Adler, Konheim and McAndrew linked topological Markov chains to statistical mechanics, by introducing the concept of topological entropy \( h_{top} = \log \eta_0 \) [1], namely equal to the natural logarithm of the maximum eigenvalue \( \eta_0 \) of a given chain transition matrix \( \Pi = \{p_{ij}\} \). This quantity \( h_{top} \) estimates the number of periodic sequences related to the symbolic dynamics of \( \Pi \) [29].

### III. Common Grounds

An ergodic (or irreducible) topological Markov chain can be thought of as a dynamical system updating its state \( s_t \) at time step \( t \) according to its transition matrix \( \Pi = \{p_{ij}\} \) and to its forbidden sequences [4, 8, 18]. Ulam tried to reconcile the symbolic dynamics of topological Markov chains with discrete maps, \( x_{t+1} = F(x_t) \), where \( F : J \to J \), [37]. His procedure introduced a partition \( \{C_i\} \) of \( J \), i.e. a set of disjointed subsets of \( J \) such that \( J = \cup_i C_i \). Furthermore, each cell \( C_i \) was related to a given symbol \( s_i \), from a given alphabet \( \{s_i\} \). In formulas, \( C_i \to s_i \). According to Ulam’s methodology, at each time step \( t \), if \( x_t \) was in \( C_i \), then \( s_t \) had to be the \( t \)-th symbol of the generated symbolic sequence. In formulas, \( x_0 \in C_i, x_1 \in C_j, x_2 \in C_k \ldots \Rightarrow s_is_js_k \ldots \) (see the side picture for an example). The Polish mathematician also provided a prescription to build the transition matrix \( \Pi \) related to such sequences. His suggestion was elegant and intuitive at the same time, inspired by measure theoretic findings of ergodic theory [37, 8]. His prescription, however, was hard to corroborate with numerical results and it remained a conjecture until it was proven by Li in 1976 [22], for piecewise linear expansive maps [6]. Li’s approach was rather convoluted and stronger results were later reobtained [10] for maps of the same type, but with particular partitions, chosen in such a way to preserve the topological Markov property on the induced symbolic dynamics [11].

### Two Not-So-Random Examples

For a unidimensional piecewise linear expansive map \( F : J \to J \), with \( J \) covered by a Markov partition \( \{M_i\} \), Ulam’s approximation [37, 8]...

\[\begin{align*}
\Pi &= \begin{pmatrix}
1-p & p \\
p & q
\end{pmatrix}.
\end{align*}\]

For instance, at each time step, there is a probability \( p \) of transitioning from state 1 to 2. According to Ulam’s procedure, in case of a partitioned map starting from cell \( C_1 \) at time \( t = 0 \), a possible sequence can be 121212... In this case cells were related to states by their index \( (C_1 \to s_1, C_2 \to s_2) \). Notice that the chosen alphabet, \( \{1, 2\} \), labels the Markov chain states.

\[\begin{align*}
1-p & \quad 1 \quad 2 \quad 1-q \\
p & \quad q \\
\end{align*}\]

Graph representation [29] of a Markov chain, with transition matrix:

\[\begin{align*}
\Pi &= \begin{pmatrix}
1-p & p \\
p & q
\end{pmatrix}.
\end{align*}\]

For instance, at each time step, there is a probability \( p \) of transitioning from state 1 to 2. According to Ulam’s procedure, in case of a partitioned map starting from cell \( C_1 \) at time \( t = 0 \), a possible sequence can be 121212.... In this case cells were related to states by their index \( (C_1 \to s_1, C_2 \to s_2) \). Notice that the chosen alphabet, \( \{1, 2\} \), labels the Markov chain states.

5 Topological Markov chains are examples of subshifts on finite type [8, 1], because they allow to update a sequence of symbols one by one. Interestingly, if a Markov chain is homogeneous and ergodic, its topological counterpart inherits the same properties.

6 Such connotation refers to continuous maps with first derivative \( |F'(x)| > 1 \), for every \( x \in J \), and piecewise constant on a given partition of \( J \).
for the transition probability from cell \( C_i \) to cell \( C_j \) becomes

\[
p_{ij} = \frac{1}{|C_j|} \frac{\mu_L(M_j)}{\mu_L(M_j)}.
\]

where \( c_j \) is the constant first derivative of \( F \) on \( M_j \), while \( \mu_L \) is the length of the interval \( M_j \).

On the interval \( I_{tent} = [0, 1] \), the tent map \( x_{t+1} = F_{tent}(x_t) = 1 - [2x_t - 1] \) is an example of a piecewise linear expansive map (with \( |F_{tent}| = 2 > 1 \)). It was originally introduced to study the logistic map, but became later a useful toy model to play with [10, 24]. \( I_{tent} \) admits a rather simple Markov partition [18, 8], with \( M_1 = [0, 1/2] \) and \( M_2 = [1/2, 1] \). By using Ulam’s approximation, linking the dynamics of \( x_t \) to a symbolic binary one becomes possible. The relative topological Markov chain has a \( 2 \times 2 \) transition matrix \( \Pi \) with all \( 1/2 \) entries. This means that, given the underlying dynamics of the tent map, a coarse description of it in terms of a Markov process exists.

Furthermore, given the probabilistic meaning of the entries of \( \Pi \), then the point mapped from \( x_i \) has always a probability \( 1/2 \) to transition from one state (cell) to the other. Therefore, for different initial configurations, after a certain transient phase, the binary sequence follows the same probability distribution of flipping a fair coin: at each time step, the next symbol can be either 1 or 0 with a probability \( 1/2 \) respectively. Intuitively, this scenario implies that such probability distribution \( \rho_{bin} \) is uniform. However, because of the preservation properties of Markov partitions [10, 8], then also the distribution \( \rho(x_t) \) of the original \( x_t \)'s is uniform and, after a transient phase, independent on time step \( t \). Such \( \rho(x_t) = \rho(x) \) is called stationary distribution and it is proportional to the probability \( \rho(x) dx \) of finding the dynamical system in a neighborhood of \( x \) after the transient phase [22]. This is an example of how probability allows to make “predictions” even for chaotic systems. In fact, the tent map is topologically conjugated (i.e. equivalent with a special type of change of variables) to the celebrated logistic map \( y_{t+1} = F_{log}(y_t) = ry_t(1 - y_t) \) on the onset of chaos, \( r = 4 \) [8]. In fact, inserting \( y = (1 - \cos(\pi x))/2 \), with \( x \in [0, 1] \), into \( F_{log} \) allows to retrieve \( F_{tent} \). Nonetheless, topological conjugacy has the peculiarity of preserving probabilities. This means that the uniform distribution \( \rho(x) \) of the tent map can be used to retrieve the stationary distribution of the logistic map:

\[
\rho(y) = \frac{1}{\pi \sqrt{y(1-y)}}.
\]

This result is quite remarkable: topological Markov chains can act as a coarse grained tracation of more complicated dynamical systems, while leaving unaltered many core features such as the stationary distribution [4, 8, 18]. However, it is not over yet.

Even if, in general, using together Markov partitions and Ulam’s prescription can be tricky, also the logistic map \( F_{log} \) admits a Markov partition with two cells only [18], immediately before the period-3 window (i.e. for \( r = r_0 = 1 + 2\sqrt{2} \)). In this point, the logistic map

\[
The tent map \( F_{tent}(x) = f(x) \) admits a Markov partition which is equal to the set of its defining \( \{A_1, A_2\} \) intervals. The plot shows also the straight line \( y = x \). From [8].

\[
\text{Histogram distribution for } 10^3 \text{ (top), } 10^4 \text{ (middle) and } 10^5 \text{ (bottom) } x_T \text{'s sampled after } T = 200 \text{ iterations of the logistic map with } r = 4 \text{. These plots give a vivid representation of the emergence of order from chaos. The } x_T \text{'s, in fact, have sensible dependence on initial conditions but their average behavior is well established.}

\[
\text{Figure 1.}
\]
admits a period-3 orbit (which remains the only stable one when \( r \) slightly increases from \( r_0 \)). Consequently, \( F_{\text{log}} \) is chaotic in the sense of Li and York [23]. Furthermore, its coarse grained formulation retrieves a homogeneous topological Markov chain with the following binary\(^7 \) transition matrix \([18]\),

\[
\tilde{\Pi} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.
\]

This matrix is a powerful tool: \( \tilde{p}_{11} = 0 \) means that it is impossible to “jump” from a state in cell \( M_1 \) to any other state in the same cell, in just one iteration step. But there are also “fixed points”, i.e. points that start and remain in cell \( M_2 \), since \( \tilde{p}_{22} = 1 \). Nevertheless, the Markov process is memoryless: the transition to the next state depends only on the actual one, so that a “jump” lasting two time steps can always be decomposed in two “jumps” of one time step each. This means that all the cells reachable in two “jumps” (two time steps) are those having non-zero entries in the matrix \( \tilde{\Pi} \cdot \tilde{\Pi} = \tilde{\Pi}^2 \).

And so on.

However, \( \tilde{\Pi} \) is such that all its powers \( \tilde{\Pi}^n \) for \( n \geq 2 \) have diagonal elements \( \tilde{p}_{ii}^{(n)} \) greater than zero: it is always possible to go out from a given cell and return to it in a countable number (\( n \)) of steps. In other words, there are infinite (coarse grained) periodic sequences of period \( n \geq 1 \) (including the fixed points). If the logistic map admits a period-3 orbit (i.e. it is chaotic in Li and Yorke’s sense [23]) then equivalently also its coarse grained description, via topological Markov chain, admits an infinite number of periodic sequences. This finding [18, 7] represents a revisitation of Sharkovskii’s theorem [33]. However, again, it is not over yet.

In fact, the maximum eigenvalue of \( \tilde{\Pi} \) is the celebrated golden ratio [18], \( \eta_0 = \varphi = (1 + \sqrt{5})/2 \), whose natural logarithm is the topological entropy \( h_{\text{top}} = \log \varphi \sim 0.4812 \) of the symbolic dynamics emerging from \( F_{\text{log}} \). The fact that \( h_{\text{top}} > 0 \) for the logistic map, before the period-3 window, not only implies that the number of periodic orbits is infinite but that they are also scrambled [1], analogously to Li and Yorke’s definition [23].

All in all, Markov chains can be efficiently used as markers of chaos.

**Discussion and Conclusions: Does God Play Dice?**

Determinism and predictability are different concepts. In agreement with Poincaré’s and Ruelle’s visions, chaotic systems evolution can greatly differ even with small changes in the initial conditions. This feature motivates the idea of chaos as a type of dynamics that does not preserve uncertainty. Knowing the initial state of a chaotic system with an infinitely small but nonzero approximation \( \Lambda \) voids any possibility of predicting the system behavior on one run, at a given time. This a brilliant example of Laplace’s epistemological uncertainty. Ignorance always diverges. Here is where probability and

\[ p_{ij}^{\text{bin}} = \text{sign}(p_{ij}) = \begin{cases} 1 & p_{ij} > 0 \\ 0 & p_{ij} = 0 \end{cases} \]

\[
\tilde{\Pi} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.
\]

Graph representation of \( \tilde{\Pi} \), from [18]. It implies that in \( r_0 \), the logistic map gives rise to a binary symbolic dynamics of 0s (if \( y_t \in M_1 \)) and 1s (if \( y_t \in M_2 \)) with all the sequences including the segment 00 being prohibited.

A representation of the logistic map periodic orbits at various \( r \). The period-3 window starts at \( r_0 \sim 3.824 \). Immediately before this window, chaotic behavior is evident.
statistical mechanics make their appearance. Even if predicting the single instance is exponentially difficult, a random approximation discloses precious information on the average behavior of a chaotic system, which is “wandering” in its own state space according to given patterns. There is structure in chaos. An order which is possible to foresee through randomness, as clearly shown within the context of Markov chains. At the same time, the above mentioned notions of Markov partitions, entropy and probability measures inevitably move the attention from the single state to its ensemble, in space (for different initial conditions) and time (for different time steps), reminiscently of Hegel’s accidents coalescing into superior and more substancial structures.

Randomness is not the only way of analysing chaos. Still, it is a fascinating one. And a successful one, relying on the many achievements of ergodic theory and statistical mechanics, even if still limited in scope to low-dimensional maps (fulfilling precise constraints). Even so, probability has already been useful in extending the “chaos paradigm” also to quantum mechanics [6], where the classical concept of orbit loses any meaning. Furthermore, also the coarse grained description of attractors via graph theory and Markov chains is increasingly receiving attention from the scientific community [29].

However, as experiments and theory on chaotic behavior progress, the idea of chaos arising only from an epistemeological obstacle seems rather disappointing. It is really difficult not to be enthralled by the elegance of the mathematics strictly entwining chaotic and stochastic systems, nevertheless it still remains mathematics. Sadly, models and reality are separate entities. At macroscopic levels it is currently impossible to disprove the conjecture of a deterministic Nature that might, nonetheless, be predictable for minds closer to perfection than the human one. However, it has to be stated that physics annoverates many experiments in which Nature seems to behave in a stochastic way even before being experienced and measured, at least at microscopic scales. Irregularity or “sensitive dependance on initial conditions” is only one property of chaos, as expressed by the Ruelle and Yorke dispute. Even if, on the one hand, the “unreasonable efficacy” of probability in interpreting the order behind chaos could come from a purely epistemeological obstacle, on the other hand reality itself could still be intrisincly chaotic and stochastic. Id est complex.
Bibliography


