Lett Mat Int
DOI 10.1007/s40329-015-0099-8

Andrey Andreyevich Markov: a furious mathematician and his chains

Angelo Vulpiani[1]

Abstract The Russian mathematician Andrey Andreyevich Markov (1856–1922) produced many very famous results, such as the law of large numbers and central limit theorem, and Markov chains. He was also known for his irascible nature. This article provides a brief overview of his life, and mathematical details of some of his results.

Keywords Andrey Andreyevich Markov · Markov chains · law of large numbers · central limit theorem · Monte Carlo method

1 (Restless) life and works

As a young man, Andrey Andreyevich Markov (1856–1922) was not the usual brilliant student: he got mediocre grades in all subjects (with mathematics as the unique exception) and had a rough and stubborn personality. As a boy, he was nicknamed *Andrey reistovy* (the Furious); later, he would become *the militant academician* when, as a respected scientist, he was committed against the Tsar's autocracy, the institutions' subservience, and the conservativeness of the Orthodox Church.

In 1874 he enrolled in Saint Petersburg University, where he studied mathematics under great teachers, such as Korkin and Zolotarev, and especially the most important Russian mathematician of the time, Pafnuty Lvovich Chebyshev (1821–1894), who was to be his mentor. Markov's career was very quick: in 1877, while still a student, he was awarded a gold medal for his research on the use of continued fractions to solve differential equations; in 1880 he defended his thesis and in 1884 he got his doctorate. At thirty he already was a professor at Saint Petersburg University; a few years later he was elected a fellow of the august Academy of Sciences.

Markov, with his unyielding character, had quite a few troubles; luckily, in his youthful years, he was protected by Chebyshev, who had realised his merit.

Here is a short list of Markov's antiauthoritarian activities.

In 1902 he protested against the cowardly behaviour of the Academy of Sciences which, under pressures from the Tsar, did not ratify the election of writer Maksim Gorky as an honorary fellow, while welcoming as new members some aristocrats completely devoid of cultural merits.

When the Minister of the Interior ordered that university professors were to be considered police officers, hence being obliged to report any anti-government activities by the students, Markov replied that he was a professor of probability theory, not a cop, that he did not approve of these decisions anyway, and that *he could not change his views by orders from his superiors*.

In 1905 he attacked the rules of Saint Petersburg University that fixed quotas on Jewish students.

After the decision by the synod of the Orthodox Church to excommunicate Tolstoy, Markov asked formally to be excommunicated too, since he shared the great writer's opinions. His request was accepted, but only partially, since its formulation was not correct. The formal answer of the synod was *Markov has seceded from God's Church and we expunged him from the lists of Orthodox believers*.

What is more interesting is that Markov's radicalism was not just a private fact of his life, but was important in his scientific activity and in the history of probability theory too.

✉ Angelo Vulpiani
angelo.vulpiani@roma1.infn.it

1 Dipartimento di Fisica, Università degli Studi di Roma "La Sapienza", Piazzale Aldo Moro 2, 00185 Rome, Italy

Published online: 23 November 2015

Since his birth, Markov was afflicted by impairments at his legs, underwent several operations and often had to use crutches. The last years of his life were saddened by the suicide of his colleague and friend Aleksandr Mikhailovich Lyapunov (1857–1918) and by health issues. Markov died in 1922 due to the consequences of an operation on his legs.

Markov was also an international level chess player. In 1892 he played four games against Mikhail Chigorin (the Russian champion, challenger of the world champion); Markov won one and tied another one. One of Markov's children, Andrey Andreyevich junior (1903–1979), followed in his father's footsteps and was a great mathematician; his contributions to algebra and logic are especially important.

Markov's mathematical activity (especially in probability theory) in Saint Petersburg fits in a well-established tradition that began in 1724, when Peter the Great founded the Academy of Sciences. Among the members of the Academy were such mathematicians as Euler and Daniel Bernoulli. The latter introduced probability theory in Russia in 1738, with a paper about risks. Since then, probability was one of Russian mathematicians' favourite branches. Chebyshev, Markov's guide, was the person who, around 1850, began giving full mathematical status to probability theory.

Let us recall, among Markov's contributions to probability theory, the way he improved Chebyshev's proof of the central limit theorem (CLT). The first general proof of the CLT for independent variables was given by another famous pupil of Chebyshev, A.M. Lyapunov, who used moments and characteristic functions.

After the Revolution of 1917, Markov's undisputed scientific prestige and his liberal opinions were instrumental in the new government's decision to strengthen the mathematical school of Saint Petersburg to the detriment of that of Moscow, which included several conservative mathematicians. Needless to say, Andrey the Furious found a way to clash with the new regime too.

Markov was an enthusiastic teacher, strongly convinced that the only way to learn was by solving problems; he was always available for his students, even for unofficial lectures during the holidays. He was beloved by his students; many of them, after passing their exam, followed again the lectures given by Markov.

In 1913, the Tsar ordered a one-year celebration for some anniversary of the Romanov dynasty. Markov, not quite interested in the event, tried to oppose the celebrations and organised a great international conference for the second centennial of the publication of *Ars Conjectandi* by Jacob Bernoulli, where the first proof of the law of large numbers appeared.

At the beginning of the twentieth century, Markov was involved in a lively debate with Moscow mathematician Nekrasov, who held political and religious opinions opposite to his own. The object of the controversy was the statistical regularity of social behaviours (see Appendix B). In Markov's opinion, this regularity was just a consequence of the law of large numbers and had nothing to do with such factors as the free will or political and religious beliefs. Nekrasov observed that the law of large numbers could not suffice to explain statistical regularities, since it only held under the hypothesis that the events are independent on each other. The remark by Nekrasov was somehow reasonable; indeed, at that time the only law of large numbers that had been proved was that by Jacob Bernoulli for independent events. In order to counter this objection, Markov had to create a theory for non-independent processes.

A letter to a friend shows how he was glad to have given his Moscow colleague a hard time:

The unique role of P.A. Nekrasov was, in my opinion, to have brought up the matter ... I have now constructed a system with properties so general as P.A. Nekrasov cannot even dream about. I have studied variables linked in a simple chain, hence the idea of the possibility of extending the limit theorems to these chains too.

This is the birth certificate of those stochastic processes now known as Markov chains (MC), which have since been used in a large class of problems in physics, chemistry, biology, economy and even in Google search engine.

In order to get an idea of Markov chains and of how it can be possible to have a law of large numbers even for non-independent events, let's consider the following game. We have three circles, numbered with 1, 2, and 3; within each circle there is a kind of roulette, each with different properties. A traveller starts from circle 1, where the roulette is divided into two sectors, labelled with the number 1 (in a sector 120 degrees wide) and 2 (in a sector 240° wide). The traveller spins the roulette; if the result is 1 he stays in the circle, while if the result is 2 the traveller jumps into circle 2. Hence, he stays in 1 with probability 1/3, or jumps into 2 with probability 2/3. Then the game repeats: so the jumping rules are defined by the roulettes or, in mathematical terms, by the probabilities $P_{i \to j}$ of jumping from circle i to circle j. The positions occupied successively are not independent. For instance, in the case shown in Fig. 1 we have that, from 1, 2 is more likely than 1, while 3 is impossible; from 2, 3 is more probable than 2, while 1 cannot be reached and so on.

By exploiting the symmetry of this problem, it is possible to show that the law of large numbers holds here: even though the events are not independent, if the game lasts long enough, the traveller spends one third of the time in each circle.

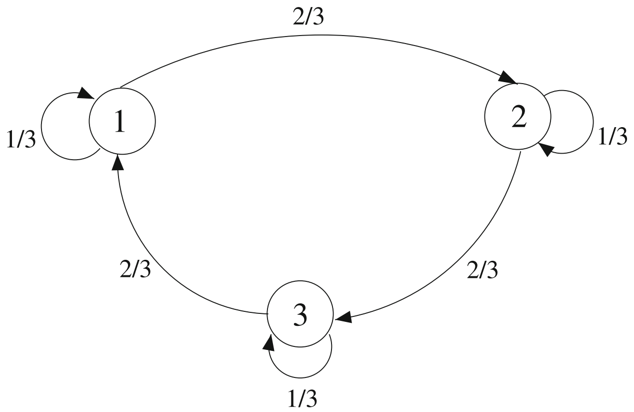


Fig. 1 Graph describing a 3-state Markov chain; each *arrow* is labelled with the transition probability

At first, Markov considered just the case with two-valued variables; later on, he covered the more general situation with a finite number of states, proving that the law of large numbers holds, under suitable conditions (see Appendix C), for non-independent variables too.

Few years later, Markov used MCs to perform a statistical analysis of some texts by Pushkin; from a linguistic viewpoint, his approach was rather elementary, since it considered the text as just a sequence of vowels and consonants. However, this work has been the starting point for the use of probabilistic techniques in linguistics: MCs are used still today to find the author of a text.

Markov chains describing stochastic processes that only assume discrete states and evolve in a discrete time are the simplest non-trivial case of a stochastic process. The development of this field was strongly motivated and driven by physics, and in particular by the study of Brownian motion at the beginning of the twentieth century by Einstein, Smoluchowski, and Langevin.

In the 1930s, Kolmogorov started formalising Markov stochastic processes. By this, we mean, besides the MCs with a finite number of states, the MCs with countably many states and moreover the processes with discrete states and continuous time (described by so-called master equations) and those with continuous states and continuous time (governed by Fokker-Planck equations). Now stochastic processes are applied in a number of different fields: in physics, chemistry, biology, as well as economy and finance.

2 Appendix A: The law of large numbers and the central limit theorem

Probability theory came out of the trifling (and quite limited) scope of card and dice games, where it was born in the seventeenth century, with *Ars Conjectandi* by Jacob Bernoulli (1654–1705), published posthumously in 1713. In

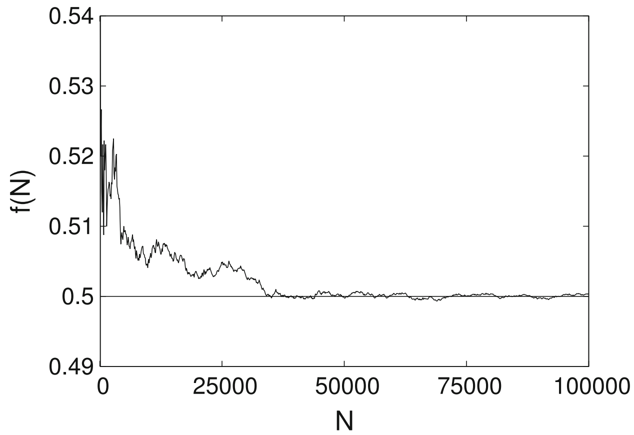


Fig. 2 A balanced coin is thrown N times. The graph shows the course of frequency $f(N)= N^*/N$, where N^* is the number of times head comes up

this book the *Law of large numbers* was proved: given an event that happens with probability p, the frequency $f(N)$ with which this event occurs in N independent trials, in the limit for large N, "tends" to p (Fig. 2):

$$f(N) \to p.$$

In more precise terms, for every $\epsilon > 0$ the probability that $f(N)$ differs by more than ϵ from p becomes arbitrarily small as N increases:

$$\lim_{N\to\infty} \text{Prob}\left(\left|\frac{f(N)}{N} - p\right| > \epsilon\right) = 0. \tag{1}$$

Analogously, we have that, given a sequence of independent variables x_1, x_2, ..., x_N having average value m and a finite variance, in the limit for large N the probability that the "empirical average" $(x_1 + x_2 + \cdots + x_N)/N$ differs by more than ϵ from m tends to zero when N approaches infinity:

$$\lim_{N\to\infty} \text{Prob}\left(\left|\frac{1}{N}\sum_{j=1}^{N} x_j - m\right| > \epsilon\right) = 0. \tag{2}$$

Result (1) suggests a possible way to connect the notion of probability to the real world: we may interpret the probability of an event as its frequency in the limit of a large number of trials. This is the essence of the frequentist interpretation which is usually accepted by physicists, but often not considered appropriate within the scope of social and economic sciences. This is not the right place to discuss the interpretations of probability, an interesting and delicate question which, however, belongs more to the philosophy of science than to the mathematics.

In 1716, in his book *The Doctrine of Changes*, Abraham De Moivre (1667–1754) gave a proof of the first case of the central limit theorem: consider N independent events, each occurring with probability p, and denote by $N^*(N)$ the number of times the event occurs. For large values of N we have

$$\mathrm{Prob}\left(a \leq \frac{N^*(N) - pN}{\sqrt{p(1-p)N}} \leq b\right) \simeq \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

This result is an improvement on the law of large numbers; indeed, we may not only state that $N^*(N)/N$ is close to p, but also have the probability of a given deviation from the average value.

As remarked by Kac and Ulam, to some purist's eye the result by De Moivre would not look especially important, since it could be seen as a quite simple application of elementary combinatorial formulas and of Stirling approximation $n! \simeq \sqrt{2\pi n}\, n^n e^{-n}$.

The result by De Moivre was extended by Laplace to the case of independent and identically distributed discrete variables $\{x_i\}$ with mean m and variance σ^2: in the limit for large N, we have

$$\mathrm{Prob}\left(a \leq \frac{\sum_{j=1}^N (x_j - m)}{\sigma\sqrt{N}} \leq b\right) \simeq \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx. \tag{A.1}$$

The first rigorous treatment of CLT for identically distributed variables, not necessarily discrete and having a finite variance, is due to Chebyshev, Markov and Lyapunov who, by using the method of characteristic functions and moments, proved (A.1) under very general hypotheses. The most general proof of the central limit theorem for independent variables was given by Swedish mathematician Lindeberg in the 1920s.

For non-independent variables, if the $\{x_n\}$s are "weakly dependent", we may expect the CLT to keep holding if the variance σ^2 is substituted with an "effective variance" σ^2_{eff} that takes the correlations into account. In the case of stationary processes, it can be shown that, if the correlation function

$$c(k) = \langle (x_t - m)(x_{t+k} - m) \rangle$$

decays to zero quickly enough when k increases (that is, more quickly than 1 / k), then the CLT holds by replacing σ^2 in (A.1) with the effective variance σ^2_{eff}:

$$\sigma^2_{\text{eff}} = \sigma^2 + 2 \sum_{k=1}^{\infty} c(k).$$

3 Appendix B: The strange birth of Markov chains

The story of how Markov chains were born is not well known. They were not introduced for technical reasons regarding probability theory, nor even to be put to some concrete use in science or technology, but within a debate about a philosophical topic: free will.

All began with a lively quarrel between Markov and Nekrasov, a Moscow mathematician, a conservative and a reactionary. The disagreement was about the interpretation of the statistical regularity of social behaviours.

Due to Quetelet's work, one of the fathers of modern statistics, in the predominant positivistic culture of the nineteenth century, the so-called *social physics* emerged. It interpreted the patterns observed on a large scale in the social phenomena as something analogous to the laws of physics: *If we bother to examine and collect accurate and sufficiently many observations, we shall find that what we believed to be an effect of chance is subject to stable principles ... Everything is predicted, everything is law: only our ignorance leads us to suppose it all to be due to the whims of chance*. Some reservation notwithstanding – for instance, Quetelet was accused by some of fatalism and of an attempt to apply uncritically Laplace's determinism to social phenomena—in the second half of the nineteenth century social physics had a great importance in sociology and philosophy, influencing people such as Marx and Durkheim.

Nekrasov regarded social physics as the first step towards materialism and, even worse, atheism and Marxism; he did not attribute to the statistical regularities of social phenomena the status of actual laws, since, in accordance with the tradition of Orthodox religion, human behaviour was supposed to be a consequence of free will.

Markov, on the other hand, claimed that such patterns are essentially an effect of the law of large numbers. Nekrasov remarked that the LLN could not suffice to explain statistical regularity, since it only held under the hypothesis of independent events. Nekrasov provided even a (modest) technical contribution: he showed that for the LLN the pairwise independence of the variables $\{x_i\}$ is sufficient, a condition weaker than actual independence. However, he concluded that the pairwise independence was a necessary condition too. From this mistaken conclusion he deduced that the presence of statistical regularities in social phenomena would prove the existence of the free will.

In his obscure connections between religion, philosophy, and mathematics, Nekrasov even tried to avail himself of the authority of late Chebyshev: for Andrey the Furious this was too much. Not only did a mediocre mathematician ranted about probability, but he dared even implicate his great master. In order to completely annihilate his opponent, Markov created a theory of non-independent processes, introducing what are now called Markov chains and proving that a LLN can be had even without independence.

The friction between Markov and Nekrasov was due only in part to their personal disagreement; it was part of a more general academic (and political in the wider sense) confrontation between the Moscow mathematical school and Saint Petersburg one, the former being religious and close to the government, the latter progressive and materialist.

Nekrasov did not leave a great mathematical legacy: in fact, his single really important contribution was having provoked Markov into creating his chains. Surprisingly, Nekrasov had a quite easy life after October Revolution and managed to find a compromise with the new government. When he died in 1924 the official newspaper *Izvestia* even published a laudatory obituary which exalted his contribution as a scientist in the proletariat's service. Only a few years later, in 1933, in the middle of the Stalin's great purges, he was branded as a reactionary scientist serving the bourgeoisie and the conservative Orthodox Church.

4 Appendix C: Markov chains in two words

A stochastic process x_t that at the discrete time t may assume one of M possible states (which we may label with integer numbers 1, 2,..., M) is a Markov chain if the future state only depends on the present one. In formulas:

$$\mathrm{Prob}(x_t = i_t | x_{t-1} = i_{t-1}, \ldots, x_{t-n} = i_{t-n}) \\ = \mathrm{Prob}(x_t = i_t | x_{t-1} = i_{t-1}),$$

where i_t can be 1,..., M. The essential property of MCs is that the transition to state $x_t = j$ under the condition that $x_{t-1} = i$, $x_{t-2} = k$,..., occurs with probability

$$\mathrm{Prob}(x_t = j | x_{t-1} = i) = P_{i \to j} = W_{ji},$$

independently on the state at time $t-2$, the state at time $t-3$ and so on.

One could wonder why we do not study even more general cases, in which the future at time t depends on the states at time $t-1$, at time $t-2$, up to time $t-r$. A moment's thought persuades us that this type of processes can be treated as one in which $r = 1$: it suffices to consider a new chain where state y_t is the vector $(x_t, x_{t-1}, \ldots, x_{t-r+1})$.

The simplest case is that of the time-homogeneous chains, where the transition probabilities do not depend on time. The matrix elements W_{ji} cannot be completely arbitrary and have to satisfy the following relations:

$$W_{ij} \geq 0, \quad \sum_{i=1}^{M} W_{ij} = 1.$$

The transition matrix $\{W_{ij}\}$ is, in manner of speaking, the DNA of Markov chain, since it determines all its properties. For instance, for the evolution in time of the vector $\mathbf{P}(t) = (P_1(t), P_2(t), \ldots, P_M(t))$, where $P_i(t)$ is the probability of being in state i at time t, we can immediately write

$$P_j(t+1) = \sum_{i=1}^{M} W_{ji} P_i(t) \tag{C.1}$$

and so

$$\mathbf{P}(t) = \hat{W}^t \mathbf{P}(0).$$

It is natural to ask whether, when t increases, $\mathbf{P}(t)$ converges to a limit vector $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_M)$; if this happens, then

$$\Pi_j = \sum_{i=1}^{M} W_{ji} \Pi_i, \tag{C.2}$$

and $\{\Pi_i\}$ are called invariant (or stationary) probabilities.

The following fundamental theorem holds:

> If there is an integer n such that for each pair (i, j) there is a non-zero probability of going from j to i in n steps, that is,
>
> $$[W^n]_{ji} > 0,$$
>
> *then there is a unique invariant probability* Π *and the convergence is exponential:*
>
> $$\mathbf{P}(t) = \hat{W}^t \mathbf{P}(0) = \Pi + O(e^{-t/\tau_c}) \rightarrow \Pi.$$

The characteristic time τ_c is

$$\tau_c = \frac{1}{|\ln(|\alpha_2|)|},$$

where α_2 is the second eigenvalue of the matrix $\hat{W}$. In this kind of matrix, by a theorem in linear algebra due to Perron and Frobenius, the first eigenvalue is never degenerate: $\alpha_1 = 1 > |\alpha_2|$, hence $|\alpha_2| < 1$ and τ_c is finite.

Such MCs are ergodic, that is, the time average along a random walk obtained with transition probabilities $\{P_{i \to j}\}$ is asymptotically equal to the average with the probabilities $\{\Pi_i\}$.

5 Appendix D: Two applications of Markov chains

Markov chains play a fundamental role in many applications in physics, astrophysics, chemistry, biomathematics, genetics, geophysics, engineering and communications; we shell discuss briefly two very important examples of the use of MCs.

5.1 Monte Carlo method

Consider the following problem: compute the average

$$\langle Q \rangle = \frac{1}{M} \sum_{j=1}^{M} Q_j \Pi_j, \qquad \text{(D.1)}$$

where Π_j is the probability of event j; if M is very large, then the computation may well be unfeasible. For instance, in statistical mechanics the value of M is huge even in small systems. To give an idea, consider the Ising model on a lattice (a very simplified description of magnetism) in which we have two possible values for each site; even with as few as 100 sites the number of admitted states is enormous: $M = 2^{100} \simeq 10^{30}$. The idea behind Monte Carlo method is to exploit ergodicity, and hence to substitute (D.1) with a time average obtained by following a "traveller" who jumps among the states following the rules of a Markov chain with invariant probabilities $\{\Pi_j\}$, and to compute the time average:

$$\frac{1}{T} \sum_{t=1}^{T} Q_{j_t}$$

where j_1, j_2, ..., j_T are the successive locations of the travellers. In the case when the Markov chain is ergodic (a condition that can be ascertained by considering the transition probabilities $\{P_{i \to j}\}$), for large T we have that the time average "tends" to $\langle Q \rangle$ (in the sense of Eq. 2).

Note that in Monte Carlo method there is a large arbitrariness: there are no specific constraints in the choice of the transition probabilities, as long as the Markov chain has as its invariant probabilities $\{\Pi_j\}$ and is ergodic. Monte Carlo method turns out to be hugely powerful and useful. For instance, it allows us to determine the thermodynamic properties of non-trivial systems (such as liquids) only knowing their macroscopic interactions (the potential between pairs of molecules).

Of course, when the number of possible states is very large (as almost always happens in problems in statistical mechanics), the trajectory cannot possibly be long enough to visit all of them. So we might wonder about the secret of the efficacy of this method. In few words, we may say that its strength lies in the fact that we typically compute averages of "not too strange" quantities and, moreover, the traveller does not waste time venturing out in states with too low a probability.

5.2 Google and Markov chains

Every time we use Google (or another web-based search engine), we use Markov chains without being aware of it. When we type some words in Google ("Markov chains applications", say), the search engine, by using an algorithm based on Markov chains, returns a list, ordered by relevance, of the web pages containing those words.

Here is a sketch of the procedure used to order the pages:

1. the N pages containing the words are singled out; each page is labelled with a number 1, 2,..., N;
2. the link number L_k is determined, that is, the number of pages linking to the kth page;
3. starting from N and from the link numbers $\{L_j\}$, following certain rules, the matrix of transition probabilities $\{P_{i \to j}\}$ is constructed, in a way giving an ergodic Markov chain;
4. the invariant probabilities Π_1, Π_2, ..., Π_N are computed;
5. the ranking is created: the first page is the one with the highest probability and so on.

The transition matrix is determined as follows: if there is a link from page i to page j, we have

$$P_{j \to i} = \frac{\alpha}{L_j} + \frac{(1-\alpha)}{N},$$

otherwise

$$P_{j \to i} = \frac{(1-\alpha)}{N},$$

where α has a value between 0 and 1 (a typical value is $\alpha = 0.85$).

This choice for the matrix can be easily and intuitively interpreted: for a fraction $1 - \alpha$ of the time an (hypothetical) web wanderer, starting from page j, picks a random website from the N possible ones, or he jumps, with probability α, to website i following the links. The technical reason to use a value $\alpha \neq 0$ is that, because of this, the MC is sure to be ergodic.

The invariant probabilities are determined by a system of linear equations:

$$\Pi_i = \sum_{j=1}^{N} \Pi_j P_{j \to i} \quad i = 1, 2, \ldots, N;$$

of course, if N is very large it is not possible to find the exact solution and generally an iterative method is followed. Starting, for instance, from $P_i(0) = 1/N$, we use repeatedly (C.1):

$$P_i(n+1) = \sum_{j=1}^{N} P_j(n) P_{j \to i};$$

since the MC is ergodic, when n increases we have a (exponentially fast) convergence to the invariant probabilities. Alternatively, we may use Monte Carlo method: for each i, the frequency $f_i(T)$ of visiting state i during a long time interval [0, T] will be close to Π_i.

In conclusion, we must point out that the method discussed here is the "pure", basic one; in practice, Google use a customised algorithm, with differences for each users, which takes into account the pages and websites already visited.

Translated from the Italian by Daniele A. Gewurz.

References

About Markov's life and scientific work:

- Langville, A.N., Stewart, W.J., (eds.): Markov Anniversary Meeting. (Boson Books, 2006)
- Basharin, G.P., Langville, A.N., Naumov, V.A.: The life and work of A.A. Markov. Linear Algebra and its Applications **386**:3 (2004)
- Seneta, E.: Markov and the Birth of Chain Dependence Theory. International Statistical Review **64**:255 (1966)
- Hayes, B.: First Links in the Markov Chain. American Scientist **101**:92 (2013)
- Sheynin, O.B.: "A.A. Markov's Work on Probability", Arch. Hist. Exact Sci. **39**:337 (1988).

About Nekrasov and statistics:

- Seneta, E.: "Statistical regularity and free will: L.A.J. Quetelet and P.A. Nekrasov", International Statistical Review **71**:319 (2003).

About statistics and probabilty in social sciences:

- SAddeo, F.: È normale la curva normale? (Bonanno Editore, 2008).

For an introduction to probability theory and to its connection to measure theory:

- Kac, M., Ulam, S.: Mathematics and Logic. Dover Publications (1992).

For an introduction to the different interpretations of probability:

- A. Hájek, "Interpretations of Probability", in *Stanford Encyclopedia of Phylosophy*, online at: http://plato.stanford.edu/entries/probability-interpret/

Two good books on Markov processes, which cover in particular their applications to physics, chemistry, and biology:

- Häggström, O.: Finite Markov Chains and Algorithmic Applications. Cambridge University Press (2002).
- Bharucha-Reid, A.T.: Elements of the Theory of Markov Processes and Their Applications. Dover Publications (2010).

About the use of Markov chains in linguistics:

- Khmelev, D.V., Tweedie, F.J.: "Using Markov chains for identification of writers", Literary an Linguistic Computing **16**:299 (2001).
- Khmelev, D.V.: Disputed authorship resolution through using relative empirical entropy for Markov chains of letters in human language texts. J. Quantit. Linguist. **7**:201 (2000).

For a general and historical survey on stochastic processes in physics (starting with Brownian motion):

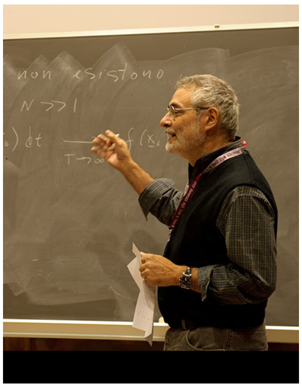
- Chandrasekhar, S.: Stochastic problems in physics and astronomy. Rev. Mod. Phys. **15**:1 (1943).

An interesting book about the (sometimes unsuspected) influence of religion on the development of mathematics:

- Graham, L., Kantor, J.-M.: Naming Infinity. Harvard University Press (2009).

About the history of the central limit theorem:

- Adams, W.J.: The life and times of the central limit theorem. American Mathematical Society (2008)
- Fischer, H.: A History of the Central Limit Theorem. Springer Verlag, Berlin (2010)



Angelo Vulpiani full professor of Theoretical Physics, teaches Statistical Mechanics and Non-equilibrium Statistical Mechanics courses at the Physics Department of the University of Rome "Sapienza". He was a visiting fellow in several research institutes and universities in France, Belgium, Sweden, Denmark, and United States. His scientific interests include chaos and complexity in dynamical systems, statistical mechanics of non-equilibrium and disordered systems, developed turbulence, phenomena of transport and diffusion. In addition to papers in international journals and academic texts, he published: Determinismo e Caos (Nuova Italia Scientifica 1994; Carocci 2004) and Caso, Probabilit'a e Complessit'a (Ediesse, 2014).