Statistical Science
2016, Vol. 31, No. 3, 399–414
DOI: 10.1214/16-STS568

Markov Chains as Models in Statistical Mechanics

Eugene Seneta

Abstract. The Bernoulli [*Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae* **14** (1769) 3–25]/Laplace [*Théorie Analytique des Probabilités* (1812) V. Courcier] urn model and the Ehrenfest and Ehrenfest [*Physikalische Zeitschrift* **8** (1907) 311–314] urn model for mixing are instances of simple Markov chain models called random walks. Both can be used to suggest a probabilistic resolution to the coexistence of irreversibility and recurrence in Boltzmann's H-Theorem. Marian von Smoluchowski [In *Sitzungsberichte der Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Klasse* (1914) 2381–2405 Hölder] also modelled by a simple Markov chain, with analogous properties, have fluctuations over time in the number of particles contained in a small element of volume in a solution.This paper explores the themes of entropy, recurrence and reversibility within the framework of such Markov chains.

A branching process with immigration, in this respect like Smoluchowski's model, is introduced to accentuate common features of the spectral theory of all models. This is related to their reversibility, a key issue.

Key words and phrases: Ehrenfest, Smoluchowski, entropy and recurrence, reversible Markov chain, stochastic matrix, Krawtchouk, Hahn, Charlier, Meixner polynomials, branching process with immigration.

1. INTRODUCTION

1.1 Structure

In their original formulation, the models with which we initially deal, the Bernoulli/Laplace model and the Ehrenfest model, are urn models. These are now commonly cast in the form of homogeneous finite Markov chains, a more general model, but are still studied through their (tridiagonal, random walk-type) transition matrices using difference equations. Transition matrices of finite Markov chains in general are stochastic matrices, which are in turn a class of nonnegative matrices. The theory of finite Markov chains is generally accepted as beginning with Markov in 1907, the year of his dating of the paper published as Markov (1906).

Markov's motivation in writing his chain papers was to show that the two classical theorems of probability theory, the weak law of large numbers and the central limit theorem, could be extended to sums of dependent random variables. Markov's methodology works well for strictly positive transition matrices, or at most for transition matrices having a strictly positive column. His (probabilistic) methodology was strongly focussed on the method of moments in the guise of conditional and absolute expectations, and double probability generating functions. The functions are, indeed, closely linked (Schneider, 1977, Seneta, 1998) to the determinant and hence spectral theory of stochastic matrices, and thus necessarily interact with the positioning of any zeros in the transition matrix. The underlying structural matrix properties of nonnegative stochastic matrices such as irreducibility, periodicity, stationary (invariant) vector, and asymptotic behavior of powers which determine the nature of the evolutionary probabilistic behavior, were not, however, clearly in evidence in Markov's work.

The theory of finite nonnegative matrices was beginning to emerge only contemporaneously with Markov's

School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia (e-mail: eseneta@maths.usyd.edu.au).

first papers on Markov chains, with the work of Perron (1907) and Frobenius (1908). The appearance of the Ehrenfest and Ehrenfest (1907) urn model, in the context of statistical mechanics, is also of that time. The eventual connection between the three directions, Markov, Perron–Frobenius and statistical mechanics is credited to Von Mises (1931). The analytical treatment of the long-term stochastic evolution of finite chains and the Perron–Frobenius theory of nonnegative matrices were not completely synthesized until the paper of Romanovsky (1936). Hawkins (2013), Section 17.3.2, gives an extensive background to these statements.

The primary aim of this paper is to illuminate the statistical mechanics direction, by focussing on several classical models in the setting of Markov chain formulation. Such a formulation enables addressing classical issues of statistical mechanics in a unified way.

We begin in Section 2 with historical background which details how Markov's work on chains finally came to the attention of western European mathematicians, not least because of the connection with statistical mechanics.

Sections3–6 are a technically light historical exploration of the physical features of entropy, recurrence and reversibility within the unifying framework of simple Markov chains as models. The explicit spectral structure of the specific models considered interacts with their "entropy analogue" behavior.

The necessary elements of Markov chain theory are deferred to Appendix A. B rounds out biographical/historical aspects. C describes the evolution of this paper..

1.2 Motivation

The author's study of interaction of Markov chains with classical models of statistical mechanics was stimulated by the well-known article, with its strong stochastic process coloration, of Chandrasekhar (1943), reprinted in the collection of Wax (1954). Chapter III of Chandrasekhar (1943) focusses on the recurrence and entropy paradoxes of thermodynamics, and in particular on the contributions of the physicist Marian Smoluchowski (1872–1917) of whom Chandrasekhar (1943) (pages 88, 89; pages 90, 91 of Wax, 1954) writes:

"The theory of density fluctuations as developed by Smoluchowski represents one of the most outstanding achievements of molecular physics.... The absence of references (in the more recent discussions of the laws of thermodynamics) in particular to Smoluchowski, is to be deplored since no-one has contributed so much as Smoluchowski to a real clarification of the fundamental issues involved."

Each of the four models studied has a now-familiar orthogonal polynomial system as its set of right eigenvectors corresponding to real distinct eigenvalues. These polynomial systems are the Krawtchouk, Hahn, Charlier and Meixner systems, which are orthogonal with respect to the simplest nonnegative integer-valued distributions very familiar to the mathematical statistician, respectively, the binomial, hypergeometric, Poisson and negative binomial, which in the four models occur, pleasingly, as stationary distributions.

The common spectral features go well beyond the commonality of structure expressed by Perron–Frobenius-type properties, which relate only to the dominant positive eigenvalue.

The simple spectral structure of the four models makes it possible to express powers of the transition matrix as an explicit spectral decomposition: that is, an expansion in powers of the eigenvalues. Such expansions were initiated in the setting of statistical physics by Kac (1947); and in the setting of branching processes are the focus of Karlin and McGregor (1966). Historically, expansions in terms of orthogonal polynomials played a central part in Markov's (1898) methodology, descended from the work on interpolation and expansion of probability densities of his supervisor Chebyshev. At least French-language writings of the work on polynomials by Chebyshev and Markov are well documented in Szegö (1939).

2. EVOLUTIONARY LINES OF MARKOV CHAIN MODELS

The best source on Markov's publications in number theory and probability has been Markov (1951), a Russian-language book of about 720 pages. The part entitled Probability Theory includes reprinting of 7 of Markov's papers on Markov chains, including Markov (1906, 1908, 1911). Sheynin (2004a) contains translations into English of the first two of these three. Sapogov (1951) contains commentary on the methodology of both Markov (1908) and Markov (1911), especially on the effect of presence of zero elements in the stochastic matrices of Markov's treatment. Sapogov's commentary is also available in English in Sheynin (2004b).

Our historical focus here is on the papers of Markov (1908, 1911).

Markov's (1898, 1910) papers appeared in French. The French language was standard for the times in the hope of international attention to Russian scientific endeavour period. The first paper was in a St. Petersburg Academy journal. The second, Markoff (1910), was in the prestigious *Acta Mathematica*, but likewise seems to have failed to attract attention. Markov (1910) is encompassed by two earlier Russian-language articles, Markov (1907, 1908).

In correspondence in late 1910 with the St. Petersburg statistician Chuprov, Markov was made aware of earlier work on special kinds of Markov chains by Ernst Heinrich Bruns, later called "Markov–Bruns chains" by Romanovsky (1949). Bruns's (1906), Lecture 18, methodology like Markov's is direct, that is: not matrix theory focussed. Bruns (1906) claimed that his book arose out of his lectures over the past 25 years. On becoming aware of Bruns's work, Markov (1911) immediately produced a paper on "Markov–Bruns" chains, presented to the St. Petersburg Academy of Science in January, 1911. Markov's paper begins by citing Bruns (1906), and saying that Bruns studies "notable cases of dependent trials which are not encompassed by the concept of a chain of trials as established by us, to which, however, one may successfully apply the method of mathematical expectation." The paper is again concerned with central limit theory, and uses generating functions. What is entailed is dependence of each outcome on the results of the previous two outcomes. By overlapping successive pairs of outcomes, and thus expanding the state space, an ordinary Markov chain, with zero entries in the transiition matrix, obtains.

A translation into German by Heinrich Liebmann of the second edition (of 1908) of Markov's textbook, *Ischislenie Veroiatnostei (The Calculus of Probabilities)* appeared as Markoff (1912). The translated book contained additionally, as three appendices, translations into German of Markoff (1898) and Markov (1908, 1911).

There is a Preface by Markov dated December, 1911. In relation to the 1908 Russian second edition of the book, he says that he has broadened, over the first edition, without attempting to produce a complete version, his bibliography on probability theory. This list of books, on page 17, consists of 12 items, of which the most recent is Bruns (1906). The 10 others by year (readily identified, so we do not include most in our own citations list) are: Laplace (1812), Poisson (1843), Lacroix (1816), Buniakovsky (1846), Bertrand (1889), Poincaré (1896), Kries (1886), Stumpf (1892), Goldschmidt (1897), and Czuber (1899). The twelfth item listed is undated: Czuber's entry,*Wahrscheinlichkeitsrechnung*, in the German Mathematical Encyclopedia. Of the 3 newly added appendices, Markov says only that these are examples of many to which the remarkable method of Bienaymé–Tschebyscheff (Chebyshev) is applicable in connection with mathematical expectation. Markov concludes that the aim of the work is not the derivation of approximative formulae for the calculation of probabilities, but to give rigorous proofs for the fundamental limit theorems of probability theory, and to provide the capability for extensive generalization.

The translator, Heinrich Liebmann, praises Markov as having presented his aims clearly; and moreover having related his work to the detailed study of probability by Czuber, and to the applied mathematics of Bruns. It is thus not unlikely that Markov was encouraged to include in the material for translation the item Markov (1911) for its German connection. Liebmann also recalls, as a companion item to Markoff (1912), the translation of 1896 into German as *Differenzenrechnung* Markov's book on difference calculus.

Liebmann's translation is clearly aimed at a German audience, to accentuate German-language eminence in probability. And the book was cited even internationally, in probabilistic monographs, as a matter of course. But the book is distant in nature from using Markov chains to model physical processes; and was hardly likely, in any case, to catch the deeper attention, with World War I and its immediate aftermath imminent, of a French, or even German, readership. [Notice that one of the appendices, Markov (1898), had originally been published in French.]

Of the three appendices to Markoff (1912), Markov (1898) is a showcase for Markov's methodology before its application to chains. Markov's methodology for chains is showcased in Markov (1908), which comes closest to the subsequent theory of nonnegative matrices. Von Mises (1931), still in a German context, recognized the significance of this appendix to both this theory and to models in statistical mechanics.

Von Mises cites Markoff (1912) in a footnote (on page 62) to his *6. Markoffsche Ketten (Markov Chains)*, among the "Aufgaben zum IV. Abschnitt." In Sections 3 to 5 inclusive of Section 16 of the Abschnitt mentioned, von Mises develops the ergodic theory under Perron–Frobenius structural assumptions on the stochastic matrix P. His main theorem on ergodicity assumes an irreducible ("unzerlegbar") matrix P. Thus, Von Mises (1931) studies Markov chains primarily through the structure of powers of their transition matrices, that is, from the then-new standpoint of nonnegative matrix theory.

Hawkins (2013), page 644, in his Section 17.3 *Markov Chains* 1908–1936 writes: "...von Mises was aware that the mathematics of his thought experiment (urn models for phenomena in statistical physics) was 'closely connected' to the 'problem of Markov chains'...," and could be used to justify certain assumptions in statistical physics. Von Mises (1931) does not actually formulate specific models which he studies in statistical physics, as Markov chains.

The link between Markov's (Russian) pre-World War I work on chains, and that of the French School of Poincaré founded on the concept of card-shuffling, came through Georg Pólya and the French-trained Sergei N. Bernstein, at the famous 1928 Bologna International Congress of Mathematicians, where chain dependence and the ergodic principle were hot topics. Bru (2003), page 145, writes: "The motivations of Markov were sufficiently different from those of Poincaré and Borel...It was not the barrier of language which prevented the French (and others)...his works had been presented in a widely read journal in French in 1910, and in German in 1912." The paper of Hadamard (1928) presented at the Conference, and written under the impetus of statistical physics, partially provided by Poincaré and Hostinsky, was later recognized as anticipating the method of Wolfgang Doeblin, Fréchet's student in the latter 1930s, on classifying chain structure focussed on sample paths.

The link between the Russian and French directions led to the booklet of Hostinsky (1931), with its extensive multinational bibliography. So the works of Hostinsky (1931) and of Von Mises (1931) mark the initial coalescence of all three directions, Markov's direct approach to classical probability limit laws in the presence of statistical dependence, the Perron–Frobenius matrix-theoretic approach to analysis, and the approach focussed on evolutionary behavior of chains as statistical models. The Markov chain contribution of Von Mises (1931), too, was soon appreciated by the French. Although its author appears not to have been at the Bologna Conference, in Von Mises (1932), pages 175–190, he presents (in French publication) his "statistical theory of successively chained events" within the context of statistical physics, using "certain results of algebra and of analysis." Further, Hadamard and Fréchet (1933) then praise von Mises fulsomely, in French, in von Mises's own German journal.

Fréchet's (1938) well-known monograph, with World War II imminent, marked the end of an era for finite Markov chain theory. It encompassed all directions, and writings in the interim, including those of the tragic Wolfgang Doeblin (1915–1940) on discrete chains (see Seneta, 2016).

For the contact dating from just after World War I, between Maurice Fréchet and the Czech mathematician Bohuslav Hostinsky (then in Brno), see Havlová, Mazliak and Šišma (2005). For contact between Doeblin and Hostinsky, see Mazliak (2007).

Hawkins (2013) has a Section, 17.3.2.2, on Romanovsky's role, culminating with Romanovsky (1936). We amplify on this to connect with our account. Vsevolod Ivanovich Romanovsky (1879–1954) was born in Verny (later Alma Ata, and now Almaty) in Kazakhstan. He received his secondary education in an academic high school ("Reelschule") in Tashkent, where he received an excellent grounding in languages, graduating in 1900. In 1906, he graduated from the St. Petersburg University and remained there to prepare for an academic career. After passing his Master's examinations in 1908, he returned to Tashkent as a teacher of mathematics and physics at his old high school. From 1911 to 1915, he was Privat-Docent and then Professor at Warsaw University. At that time, part of Poland was still in the Russian Empire. In 1912, after he had defended his Master's dissertation *On partial differential equations*, the degree of Master of Mathematics was conferred on him by St. Petersburg University. In 1916, Romanovsky completed his doctoral thesis, but its defence under wartime conditions proved impossible. Warsaw University, as a Russian institution, was closed down, and for a year or so he worked at Don University at Rostov-on-the-Don, and returned to Tashkent in 1917. From its beginning stages in 1918 till his death, he was heavily involved in teaching, research and administration at what became Tashkent State University (earlier called Central Asian University). In the early period of his research, he worked on differential equations, algebraic equations and (as expected from his student days in St. Petersburg), on number theory, Markov's other great sphere of interest and influence.

Romanovsky's research activities of the 1920s were largely devoted to mathematical statistics. He managed to keep in touch with, and publish in, the important western European statistical and mathematical journals, most notably *Biometrika*, where a number of his papers were devoted to polynomial expansions of probability densities corresponding to Pearson's curves.

His name is sometimes attached to one such polynomial system. Romanovsky's most important scientific work was on finite Markov chains, but this began only in 1928. His first publication on the topic, Romanovsky (1929), was in French, and it is from this point that Hawkins (2013) picks up his story in Section 17.3.2.2. It is not clear what motivated Romanovsky, although his good contacts with western European as well as Soviet scientists, and their possible connection with the Bologna Conference of 1928, may have encouraged him to give Markov's work its due. His publications in the Parisian *Comptes Rendus* in 1930s brought him into contact in particular with the Czechoslovak group of mathematicians working on finite Markov chains which was forming round Hostinsky (see Hostinsky, 1931). Hawkins (2013) writes that eventually Romanovsky (1936) "devoted 33 of the 105 pages of his memoir to Frobenius' theory and its application to stochastic matrices, thereby exposing his readers to all of Frobenius' significant results and making clear their relevance to the theory of stochastic matrices and Markov chains." Romanovsky had not been Markov's research student as is sometimes thought. Sarymsakov (1955) writes that Romanovsky perfected and adopted methods of the Chebyshev school for solving problems in mathematical statistics, and that this "can partly be explained by his... having attended the course in probability theory read by the celebrated Markov."

For more detail, the reader may wish to consult Seneta (2006), Section 5; Seneta (2009), Section 9; and the obituary, Sarymsakov (1955), of Romanovsky by his star student in Markov chain theory and applications, Sarymsakov (1915–1995). Sarymsakov was born in the same year as Doeblin, and was familiar with Doeblin's work on Markov chains, arising from contact between Doeblin and Kolmogorov (Doeblin, 2016).

3. THE BERNOULLI/LAPLACE AND THE EHRENFEST MODELS

These two-urn models for mixing when expressed as Markov chains have finite irreducible transition matrices which have zero entries outside of the three leading diagonals. They are examples of random walks with reflecting barriers. Historically the random walk structure, including our two irreducible special cases, has been, and is still, treated using difference equations.

All the eigenvalues of irreducible random walk transition matrices are real since their transition probabilities satisfy (25).

(a) The Bernoulli (1769)/Laplace (1812) model. This is a two-urn model. Label the Urns A and B. Each urn has N balls, so total number of balls is $2N$. The totality of $2N$ balls consists of N white, and N black.

An interchange consists of selecting a ball at random from Urn A, and a ball at random from Urn B, and exchanging them.

Let X_n be the number of black balls in Urn A after n interchanges.

Then the process $\{X_n\}$ is a finite Markov chain with irreducible transition matrix whose tridiagonal entries are given by

$$p_{i,i-1} = \left(\frac{i}{N}\right)^2, \quad p_{i,i+1} = \left(\frac{N-i}{N}\right)^2,$$
$$p_{i,i} = 2\frac{i}{N}\left(\frac{N-i}{N}\right), \quad i = 0, 1, \dots, N.$$

The Markov chain $\{X_n\}, n = 0, 1, \dots$ has stationary/limiting distribution $\pi^T = \{\pi_0, \pi_1, \dots, \pi_N\}$ given by

$$\pi_i = \frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}}.$$

This is the *hypergeometric distribution*, as one would expect from "good mixing." The condition (25) is satisfied, so that in its stationary regime, the Markov chain is *reversible*.

For the Bernoulli/Laplace model, the complete set of eigenvalues is

$$\lambda_n = 1 - \frac{n(2N+1-n)}{N^2}, \quad n = 0, 1, \dots, N$$

and the entries of the corresponding right eigenvector are the nth *Hahn polynomial* evaluated at $x = 0, 1, \dots, N$. These polynomials (Karlin and McGregor, 1961) are orthogonal with respect to the hypergeometric distribution. The spectral results may be found in broader context in Seneta (2001a) and earlier in Diaconis and Shahshahani (1987), and Donnelly, Lloyd and Sudbury (1994). An early partial investigation of eigenvalue structure is due to Hostinsky (1939).

(b) Ehrenfest (1907) model. Also a two-urn model, Urns A and B. Total number of balls is N. All N balls are black, and labeled 1 to N.

An interchange consists of selecting a number at random from the set $\{1, 2, \dots, N\}$, finding the ball with this number and placing it in the other urn.

Let X_n be the number of (black) balls in Urn A after n interchanges.

$$p_{i,i-1} = \left(\frac{i}{N}\right),$$

$$p_{i,i+1} = \left(\frac{N-i}{N}\right), \quad i = 0, 1, \ldots, N.$$

The Markov chain $\{X_n\}, n = 0, 1, \ldots,$ has stationary/limiting distribution $\pi^T = \{\pi_0, \pi_1, \ldots, \pi_N\}$ given by

$$\pi_i = \binom{N}{i}\left(\frac{1}{2}\right)^N. \tag{1}$$

This is the *symmetric binomial distribution*. The condition (25) is satisfied, so that in its stationary regime, the Markov chain is *reversible*.

For the Ehrenfest model, the complete set of eigenvalues is

$$\lambda_n = 1 - \frac{2n}{N}, \quad n = 0, 1, \ldots, N$$

and the corresponding right eigenvector is the nth *Krawtchouk (Kravchuk) polynomial*. These polynomials are orthogonal with respect to the symmetric binomial distribution. These spectral results are due to Kac (1947) in a classic paper, although he does not recognize that the polynomials are the Krawtchouk polynomials.

The Bernoulli/Laplace urn model had already been investigated by Daniel Bernoulli (1769) during his stay in St. Petersburg, and published in the journal of the Russian Imperial Academy. He obtained, in particular, the relation (which we express in modern notation)

$$E\left(X_k - \frac{N}{2}\right) = \left(1 - \frac{2}{N}\right)^k E\left(X_0 - \frac{N}{2}\right),$$

and the diffusion approximation (leading to "Newton's law of cooling") in the special case where $X_0 = N$. The very same two-urn model is treated in the celebrated treatise of Laplace (1812), pages 287ff., eventually with the same diffusion approximation, within his Chapitre III which begins on page 275. This is where Markov (1915) found it, after independently considering in 1912 a slightly more general model where the Urns A and B are permitted to contain different numbers of balls.

It is not surprising that Markov does not mention Daniel Bernoulli, since Laplace (1812) had a tendency not to cite (see Todhunter, 1865, pages 488–494) although he mentions "les Bernoullis." Bernstein (1934), pages 127–130, in the second edition of his book, takes up the model in Markov's version and Markov's difference equation treatment, without mentioning any of Bernoulli, Laplace, or even Markov, presumably because he is writing a "textbook." (The pagination of the material is the same in the celebrated fourth edition of 1946 of Bernstein's book.)

4. STATISTICAL MECHANICS

Gases were to be viewed as aggregates of particles undergoing movement at different velocities, and collisions between the particles were to accord with the principles of Newtonian mechanics. Hence, the term Statistical Mechanics.

In classical thermodynamics, the process of heat exchange between two isolated bodies at initially unequal temperatures is *irreversible*: the second law of thermodynamics says *entropy* is nondecreasing. And Boltzmann's H-Theorem asserts this. However, its derivation is based on classical kinetic considerations (of Newtonian mechanics), where essentially collisions between particles are *reversible.* In this kind of situation, any mechanical system constrained to move in a finite volume with fixed total energy must return to any specified initial configuration. Thus, "recurrence" must occur; and entropy defined in such a system cannot always increase with time, but must eventually decrease in order to return to its initial value.

Thus, one paradox was the apparent conflict in Boltzmann's theory between *irreversibility* (as manifested by increasing entropy) and *recurrence of states* as expected from the assumptions of Newtonian mechanics.

The Ehrenfest urn model (Ehrenfest and Ehrenfest, 1907) was created in response to such paradoxes which appeared in Boltzmann's (≈1872) theory. Parenthetically, on the page where the Ehrenfest article ends, one by von Mises (on an unrelated topic) begins, foreshadowing the role von Mises was to play in unifying statistical mechanics models with Markov chains.

On an elementary level with which the Ehrenfest model has come to be associated, if we regard the two urns as symbolizing two bodies, and the number of white balls in each as symbolizing their temperatures, we have a simple model for heat exchange between two bodies at unequal temperatures. The model represents the heat exchange as a random process, rather than an orderly one as in classical thermodynamics, and insofar as movement of particles is concerned, as a kinetic process.

It can be used to explain the apparent contradiction between irreversibility and recurrence as follows. Denoting by μ_i the mean recurrence time of state i, and using (24) and (1),

$$\mu_i = \frac{1}{\binom{N}{i}(\frac{1}{2})^N}.$$

If $N = 20{,}000$, and $i = 0$, $\mu_0 = 2^{20{,}000}$ time units. If the time units are seconds, this is approximately 10^{6000}

years. However, if $i \approx N/2$, $\mu_{N/2} \approx 175$ time units. So if one starts in a state with a long mean recurrence time, one will observe an essentially irreversible evolutionary process (this observation is due to Smoluchowski, although he used a different version of mean recurrence time) and one therefore has vindication of Boltzmann's assertion that "Poincaré Cycles" are so long compared to time intervals involved in ordinary experiences, that predictions based on classical thermodynamics are fully to be trusted.

The other aspect is *entropy.* The Ehrenfests chose as an analogue of the *negative entropy* of Boltzmann (which is supposed to be decreasing with time) the quantity:

(2) $$2|X_n - N/2|, \quad n = 0, 1, 2, \ldots.$$

This quantity is the absolute value of the difference in the numbers of balls in the two urns. It "jumps" by increments of 2 with increasing n, and does not have the same smooth behavior in the vicinity of $N/2$ as $2(X_n - N/2)$, which however will take negative as well as positive values.

In an important follow-up article which is surprisingly little-mentioned in the literature, Kohlrausch and Schrödinger (1926) focus on $2(X_n - N/2), n \geq 0$, to put the analysis of the Ehrenfest model on a proper probabilistic footing using difference equation techniques.

Additionally, they report (in their Sections 1 and 3) a simulation study with 5000 successive drawings, when $N = 100$, and $X_0 = 100$. By drawing no. 200, the plot of the quantity $2(X_n - N/2)$ against drawing number oscillates closely about 0. In Section 3, they use the data (from the last 4800 drawings) to plot $-\log|E(X_n - N/2)|$, and $-\log$ of absolute sample averages of $X_n - N/2$ at each of $n = 0, 1, 2, \ldots, 9$ for each of the starting values $|X_0 - N/2| = 5, 10, 15$. The averages are obtained from the number of available realizations for the starting value available within the data, respectively, 453, 106, 10. The agreement is generally very good, almost perfect for the starting value $|X_0 - N/2| = 5$ at $n = 0, 1, 2, 3, 4$. This is due to the fact that the sample size 435 is large, so agreement of the averages with $E|X_n - N/2|$ will be good, and because for $n = 0, 1, 2, 3, 4$ $(X_n - N/2)$ does not change sign on account of the random walk structure of transition probabilities, so that $|E(X_n - N/2)|$ and $E|X_n - N/2|$ coincide. Near-coincidence of $|E(X_n - N/2)|$ and $E|X_n - N/2|$ at $n = 0, 1, 2, \ldots, 9$ for each of the fixed starting values may be an attempt by the authors to justify $2|E(X_n - N/2)|$, an analytically tractable deterministic function of n as an appropriate analogue of the negative entropy (Boltzmann's *H-Kurve*), by arguing that it is essentially equivalent to $E|X_n - N/2|$, which is obtained by taking expectation of (2). In fact, more generally, $|E(X_n - N/2)| \leq E|X_n - N/2|$ by the triangle inequality.

Now, for both the Ehrenfest and the Bernoulli/Laplace models, it is easily shown, by first calculating conditional expectation $E(X_{n+1}|X_n)$ from the transition matrix, that

$$E\left(X_{n+1} - \frac{N}{2}\right) = \left(1 - \frac{2}{N}\right) E\left(X_n - \frac{N}{2}\right)$$

reflecting (23). So

(3) $$2\left|E\left(X_n - \frac{N}{2}\right)\right|$$

is a deterministic function which decreases as n increases providing $E(X_0) \neq N/2$, since

$$E\left(X_k - \frac{N}{2}\right) = \left(1 - \frac{2}{N}\right)^k E\left(X_0 - \frac{N}{2}\right).$$

The quantity (3) is reflected in all four Markov chain models with which we are concerned, and we take it as our analogue of negative entropy, motivated by our discussion above of Kohlrausch and Schrödinger (1926).

A referee is ambivalent of this particular use of expectations, since it reflects an inherent randomness in modelling, and writes: "…our understanding of the second law is quite insensitive to whether the underlying dynamics is stochastic or deterministic…. The probabilities and the corresponding expectations are indeed relevant to derivations of the second law, but only as a tool for establishing the typical behavior of individual systems via the law of large numbers." Our attempt to accommodate this reasoning is to consider a "universe" consisting of a large number of replications $\{X_n^{(r)}\}, r \geq 1$ of the Markov chain $\{X_n\}$, each starting with the same initial value. Then at any fixed time point, n, from the law of large numbers, $\lim_{n\to\infty} \sum_{r=1}^{R} X_n^{(r)}/R = E(X_n)$, so (3) at any fixed time n reflects the average entropy state of the "universe" at time n.

A physically desirable feature is that in the stationary (i.e., probabilistically stable) state, when the entropy remains at zero, the kinetic model should be (probabilistically) *reversible*, and we have noted this feature in both the Bernoulli–Laplace and Ehrenfest models.

So, transparently, the Bernoulli–Laplace model could have been used also to resolve by analogy the paradoxes in statistical mechanics and indeed has the more desirable feature of the Markov chain having the stationary distribution $\boldsymbol{\pi}^T$ as the limiting distribution as $n \to \infty$, since it is aperiodic. But in 1907 such a context was unlikely to be perceived; hence, the Ehrenfest model.

To return to the concept of reversibility, for irreducible finite Markov chains the connection with reversibility in statistical mechanics is simultaneously due to Kolmogorov (Kolmogoroff, 1935) and Hostinsky and Potoček (1935). Kolmogorov cites on page 155 as a specific example a paper of Schrödinger of 1931, titles his Section 4 *Die Umkehrung der Naturgesetze* (The reversal of natural laws), and gives a random walk example. He also cites Von Mises (1931), but in connection with the structure of transition matrices, rather than in connection with models of a system moving from state to state. Clearly excited by this idea of reversibility, perhaps partly by the distinction between a reverse Markov chain, and one that is reversible, Kolmogorov published a paper in the same journal the following year (Kolmogoroff, 1936) with the title now being *Zur Umkehrbarkeit der statistischen Naturgesetze* (The reversibility of statistical laws). In their tribute to the then-recently deceased Kolmogorov (1903–1987), Dobrushin, Sukhov and Frits (1988) explore Kolmogorov's legacy in this respect, both in stochastic process theory and in statistical mechanics. The idea of reversibility of Markov chains was, however, already briefly present in a paper of Markov (pages 171–186 of the same year, source and volume as Markov, 1911); and was also explored in the early papers of Bernstein, and Onicescu and Mihoc. A listing of such papers is given in Fréchet (1938).

We now pass to a model which receives much attention in Chandresekhar (1943). It has a similar wealth of features as the two above, but needs to be placed in a Markov chain modelling context to reveal this.

5. A BRANCHING PROCESS WITH IMMIGRATION

Let $X_n, n = 0, 1, 2, \ldots,$ denote the number of individuals at time n, where movement from time n to time $n+1$ is defined by

$$(4) \quad X_{n+1} = Z_1^{(n+1)} + Z_2^{(n+1)} + \cdots + Z_{X_n}^{(n+1)} + I_{n+1}.$$

Here, $Z_j^{(n+1)}$ is the number of offspring of the jth individual existing at time n, I_{n+1} is the number of immigrants coming into the population to supplement these offspring in forming the totality of the number of individuals X_{n+1} at time $n+1$.

All the random variables $Z_j^{(k)}, I_k, j, k \geq 1$ are *assumed independent.* All the $Z_j^{(k)}$'s are assumed to have the same probability distribution $\{p_j\}, j \geq 0$ with probability generating function (pgf) $f(s) = \sum_j p_j s^j, 0 \leq s \leq 1$; and all the I_k's are assumed to have the same probability distribution $\{b_j\}, j \geq 0$ and pgf $b(s) = \sum_j b_j s^j, 0 \leq s \leq 1$.

The process $\{X_n\}, n \geq 0$, with X_0 having some initial distribution, is clearly a Markov chain on the countably infinite state space $S = \{0, 1, 2, \ldots\}$, and if $H_n(s) = \sum_{j=0}^{\infty} \Pr(X_n = j) s^j$, from (4)

$$(5) \qquad H_{n+1}(s) = b(s) H_n\bigl(f(s)\bigr), \quad 0 \leq s \leq 1.$$

If the offspring mean $m = \sum_{j=0}^{\infty} j p_j < 1$, and if the immigration mean $\lambda = \sum_{j=0}^{\infty} j b_j < \infty$, a balance is set up between immigration and the tendency to extinction of the branching process without immigration. There is an approach to a limiting/stationary distribution as $n \to \infty$ whose pgf is $H(s)$, that is, $H_n(s) \to H(s)$. Thus, there is a strictly stationary (actually unique) regime for the process, with the pgf of the stationary X_n satisfying

$$(6) \qquad H(s) = b(s) H\bigl(f(s)\bigr), \quad 0 \leq s \leq 1.$$

Finally, from (4), using $E(X_{n+1} | X_n)$,

$$(7) \qquad E(X_{n+1}) = m E(X_n) + \lambda,$$

so that, if μ denotes the mean of the limiting/stationary distribution, $\mu = \lambda/(1-m)$, substituting for λ in (7) gives

$$E(X_{n+1} - \mu) = m E(X_n - \mu),$$

again reflecting (23).

5.1 Marian Smoluchowski's (1914) Model

Smoluchowski (von Smoluchowski, 1914) considers a simple model for the fluctuation in the number of particles contained in a geometrically well-defined small element of volume, v, in a much larger volume of solution containing particles exhibiting random motion. Observations $X_n, n \geq 1$ are made at points of time at equal intervals, τ, apart.

His model is a special case of a branching process with immigration if we take the intervals to be of unit length, the offspring distribution is Bernoulli(P), so that each individual replaces itself or "dies" and $m = 1 - P$; and the immigration distribution is Poisson(λ) so that

$$f(s) = P + (1-P)s, \quad b(s) = \exp\{\lambda(s-1)\}.$$

The transition probabilities of the Markov chain $\{X_n\}, n \geq 0$, are in this case clearly given, using (4) and the argument for the convolution of a binomial and a Poisson distribution, by

$$\Pr(X_{n+1} = j | X_n = i) \tag{8}$$
$$= p_{i,j}$$
$$= e^{-\lambda} \sum_{k=0}^{\min(i,j)} \binom{i}{k} (1-P)^k P^{i-k} \frac{\lambda^{j-k}}{(j-k)!},$$

and putting $r = i - k$:

$$\Pr(X_{n+1} = j | X_n = i)$$
$$= e^{-\lambda} \sum_{r=i}^{r=\max(0,i-j)} \binom{i}{r} \tag{9}$$
$$\cdot P^r (1-P)^{i-r} \frac{\lambda^{j-i+r}}{(j-i+r)!}.$$

If we take

$$H(s) = \exp\{\mu(s-1)\}, \tag{10}$$

where $\mu = \lambda/P$, it is readily seen that (6) is satisfied so the (unique) stationary distribution of $\{X_n\}$ is Poisson(μ).

Moreover, from (7),

$$E(X_{n+1} - \mu) = (1-P)E(X_n - \mu). \tag{11}$$

Thus, the decreasing negative "entropy" feature is common to all three models: Bernoulli–Laplace, Ehrenfest and Smoluchowski.

Further, the Smoluchowski model's transition matrix given by (9) and stationary distribution given by $\Pr(X_n = i) = \pi_i = e^{-\mu}\mu^i/i!$ satisfy (25), so when in a stationary regime this Markov chain is *reversible*. So another common feature of each of the three models is reversibility in their stationary regime.

The idea of a transition probability, a fundamental idea in Markov chain modelling, is present already in Smoluchowski (1914), where the expression for it on the right of our (9) occurs as equation (18), page 2392. Von Mises (1931) gives an account of Smoluchowski's theory, but makes no connection with Markov chains.

Smoluchowski (von Smoluchowski, 1914) did not use pgf's, nor did any of Fürth (1918), Von Mises (1931) or Chandrasekhar (1943) in their accounts of the same work. There was no need, because his model is one of the few cases of the general branching process with immigration where simple forms of expression are available.

We will address the question of spectral structure shortly, but now pass to a remarkable additional new aspect of Smoluchowski's (1914) model, of inference for a stochastic process.

5.2 Statistical Inference for Branching Processes

In Smoluchowski's theory, the number P, called the *probability after-effect*, is the probability that a particle somewhere inside v will have emerged from v during time τ. The exact value of $P = (1-m)$ (as also that of λ) depends on the precise circumstances of the problem. An explicit expression for it, in terms of the various physical parameters, is obtained by Smoluchowski (1914) when the motions are governed by the laws of Brownian movement.

On the other hand, P can be estimated *statistically* from observation of a trajectory of $\{X_n\}$, when the system is in *equilibrium* (that is when the Markov chain is in its stationary regime) and equilibrium is Smoluchowski's context.

A comparison of the predictions of the theory of colloid statistics with the data observed is therefore made possible, and was in fact carried out on data of Th. Svedberg by Smoluchowski himself (see Sredniawa, 1992 for an account of the collaboration). The striking advance on earlier fluctuation theory is the introduction of the probability after-effect ("Wahrscheinlichkeitsnachwirkung") P, which clearly incorporates a Markovian probabilistic structure of the assumed model, as well as being of great significance in physical contexts.

As regards the statistical estimation procedure, the underlying equation [since stationarity of regime is being assumed and the stationary distribution is Poisson(μ)] is the elegant expression:

$$E\big((X_{n+1} - X_n)^2\big) = 2\mu P, \tag{12}$$

where $\mu = \lambda/P$. Equation (12) is equation (23) of Smoluchowski (von Smoluchowski, 1914), page 2304.

The left-hand side of (12) was estimated by Smoluchowski using observations $X_1, X_2, \ldots, X_{N+1}$ by

$$\sum_{i=1}^{N} (X_{i+1} - X_i)^2 / N,$$

and μ, which is the variance as well as the mean of the stationary Poisson(μ) distribution, by

$$\hat{\mu} = \sum_{i=1}^{N} (X_i - \bar{X})^2 / N,$$

where $\bar{X} = \sum_{i=1}^{N} X_i / N$. Thus,

$$\hat{P} = \frac{\sum_{i=1}^{N} (X_{i+1} - X_i)^2}{2 \sum_{i=1}^{N} (X_i - \bar{X})^2}.$$

Verifications of the theory as reported by Chandrasekhar (1943), specifically in relation to data of Westgren, begin by first calculating the exact values of P and μ, *using physical constants and colloid theory.*

Then μ is used to give *expected frequencies* using the Poisson(μ) distribution, and the expected frequencies are compared with *observed frequencies*. This agreement appears very good.

Then the value of P is compared with $\hat{P}$ from statistical estimation using observations at times $n\tau_0, n \geq 1$, where τ_0 is the actual time gap initially used between observations. Again the agreement looks to be very good.

An asymptotic theory of estimation for subcritical branching processes with immigration was initiated by Heyde and Seneta (1972).

However, clearly what is actually needed is a large sample *test of the null hypothesis* that an observed nonnegative data sequence comes from

1. A branching process with immigration, in stationary regime;
2. More narrowly,*a Bernoulli–Poisson branching process with immigration.*

Such tests were finally developed, respectively for 1 and 2, by Mills and Seneta (1989, 1991) as analogues of Quenouille's test in times series analysis, using partial sample autocorrelations. The Bernoulli–Poisson null-hypothesis (i.e., Smoluchowski's model) was found to give a striking simplification of the general case, with sample partial autocorrelations at lag $k \geq 2$ asymptotically independent and Gaussian, as for classical time series models.

6. BRANCHING PROCESS SPECTRAL THEORY

For the branching process with immigration in general, we would like to prove

$$\sum_{j=0}^{\infty} p_{ij} p_r(j) = m^r p_r(i), \quad i = 0, 1, 2, \ldots, \tag{13}$$

where $m, 0 < m < 1$, is the mean of the offspring distribution, and $p_r(i)$ is the rth orthogonal polynomial, in $i, i = 0, 1, 2, \ldots$, with the polynomial system being orthogonal with respect to the stationary distribution on $\{0, 1, 2, \ldots\}$ of the process $\{X_n\}$. We proceed by forming the generating function

$$G(i, w) = \sum_{r=0}^{\infty} K(r) p_r(i) w^r \tag{14}$$

for a sequence $K(r), r = 0, 1, 2 \ldots$ of positive constants. Then (13) becomes

$$\sum_{j=0}^{\infty} p_{ij} G(j, w) = G(i, mw). \tag{15}$$

6.1 The Smoluchowski Model

From (10), the stationary distribution is Poisson(μ).

The Charlier polynomials are known to be orthogonal with respect to the Poisson(μ) distribution, where the rth Charlier polynomial evaluated at $j = 0, 1, 2, \ldots$ is given by

$$p_r(j) = \mu^{r/2}(r!)^{-1/2} \sum_{\nu=0}^{r} (-1)^{r-\nu} \binom{r}{\nu} \nu! \mu^{-\nu} \binom{j}{\nu},$$

and for small $|w|$ (Szegö, 1939, page 35)

$$G(j, w) = \sum_{r=0}^{\infty} \mu^{-r/2}(r!)^{-1/2} p_r(j) w^r \tag{16}$$

$$= e^{-w}(1 + \mu^{-1}w)^j, \quad j = 0, 1, 2 \ldots. \tag{17}$$

Then from the transition probabilities p_{ij} as given by (8):

$$\sum_{j=0}^{\infty} p_{ij} G(j, w)$$

$$= e^{-\lambda} e^{-w} \sum_{j=0}^{\infty} \Biggl(\sum_{k=0}^{\min(i,j)} \binom{i}{k} \bigl(m(1 + \mu^{-1}w)\bigr)^k \cdot (1 - m)^{i-k} \frac{\{(1 + \mu^{-1}w)\lambda\}^{j-k}}{(j - k)!} \Biggr). \tag{18}$$

Now, the inner summation is the coefficient of z^j in the product

$$\bigl(1 - m + m(1 + \mu^{-1}w)z\bigr)^i e^{(1+\mu^{-1}w)\lambda z},$$

so putting $z = 1$, and invoking the outer summation (over j) in (18) we obtain, since $\mu = \lambda/(1 - m)$, finally,

$$e^{-mw}(1 + m\mu^{-1}w)^i = G(i, mw),$$

which establishes (15), and hence (13).

Hence, $m^r, r = 0, 1, 2, \ldots$ is the rth eigenvalue of the infinite transition matrix, and the column vector $\mathbf{p}_r = \{p_r(j), j = 0, 1, 2, \ldots\}$, of values of the rth Charlier polynomial is the corresponding right eigenvector.

We have already shown that the transition matrix satisfies the reversibility condition, so the Smoluchowski model, an infinite Markov chain, parallels all the properties possessed by the two finite chain models, and can be used in the same way to explain physical paradoxes.

6.2 The Negative Binomial Model

A remaining familiar probability distribution, also on all the nonnegative integers, is the negative binomial, with probabilities specified by

$$\pi_i(n) = \binom{n+i-1}{i} p^n q^i = \binom{-n}{i} p^n (-q)^i, \quad i = 0, 1, 2, \ldots, \tag{19}$$

where $0 < p = 1 - q < 1$, and $n, n \geq 1$, is an integer. The pgf of this distribution is

$$H(s) = p^n(1-qs)^{-n} = \left(\frac{1-q}{1-qs}\right)^n. \tag{20}$$

The questions to be addressed are: is it the stationary distribution of a branching process with immigration, and if so what are appropriate offspring and immigration distributions? What is a system of polynomials orthogonal with respect to the negative binomial? If so, can they be regarded as right eigenvectors corresponding to a sequence of real eigenvalues?

And finally, if such a branching process with immigration can be found, does its infinite transition matrix satisfy the *reversibility condition* (25)?

A system of polynomials orthogonal with respect to the negative binomial distribution specified by (19) was found by Kulik (1953). They have a simple generating function:

$$G(j, w; n) = \sum_{r=0}^{\infty} (r!)^{-1} p_r(j; n) w^r = \frac{(1-w)^j}{(1-qw)^{n+j}}, \quad j = 0, 1, \ldots. \tag{21}$$

Next, we notice that if we take

$$b(s) = (1+q-qs)^{-n}, \qquad f(s) = (1+q-qs)^{-1}, \tag{22}$$

then (6), namely

$$H(s) = b(s)H\big(f(s)\big), \quad 0 \leq s \leq 1,$$

is satisfied with $H(s)$ given by (20). Thus, we have a branching process with immigration, with immigration and offspring distributions specified by the pgf's $b(s), f(s)$, respectively, for which the negative binomial distribution specified by the pgf $H(s)$ is the unique stationary distribution. The mean of the offspring distribution is given by $m = q$.

Next,

$$\sum_{j=0}^{\infty} p_{ij} G(j, w; n) = \frac{1}{(1-qw)^n} b\left(\frac{1-w}{1-qw}\right) f^i\left(\frac{1-w}{1-qw}\right),$$

and substituting from (22):

$$= \frac{(1-qw)^i}{(1-q^2w)^{n+i}} = G(i, qw; n), \quad i = 0, 1, 2, \ldots.$$

Thus, $p_r(j; n)$, $j = 0, 1, 2, \ldots$ forms the right eigenvector of P, corresponding to eigenvalue $q^r (= m^r)$, $r = 0, 1, 2, \ldots$.

Now, using (20), we find

$$\pi_j = (1-q)^n q^j \binom{n+j-1}{j}$$

and from $b(s) f^i(s)$, using (22)

$$p_{ij} = \binom{n+i+j-1}{j} \frac{q^j}{(1+q)^{n+i+j}},$$

so that the reversibility condition (25) is satisfied. Finally, differentiating $H(s), b(s) f^i(s)$, respectively, and evaluating at $s = 1$ we obtain respectively

$$\mu \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} j\pi_j = \frac{nq}{1-q}; \qquad \sum_{j=0}^{\infty} j p_{ij} = q(n+i)$$

so that

$$\sum_{j=0}^{\infty} (j-\mu) p_{ij} = q(i-\mu),$$

which is the decreasing negative entropy condition (23). Kulik (1953) was in fact generalizing to arbitrary $n \geq 1$ the case $n = 1$ of Gottlieb (1938). Gottlieb's paper is mentioned in passing in Szegö's (1939) treatise. We note that Szegö's book of 1959 is a almost a reprinting of a 1939 version, so papers dating from the middle 1930s, would receive little attention. Papers of Krawtchouk's associates such as Kulik and Smohorshevsky are not mentioned.

Further, we note that the Meixner (1934) orthogonal polynomials $M_r(x; b, a), r = 0, 1, 2, \ldots$ satisfy

$$\sum_{r=0}^{\infty} M_r(x; b, a)[b]_r \frac{s^r}{r!} = \frac{(1-\frac{s}{a})^x}{(1-s)^{b+x}},$$

where $0 < a < 1, b > 0, [b]_r = b(b+1)\cdots(b+r-1)$. Thus, Kulik's orthogonal polynomials $p_r(j; n)$ are essentially the Meixner polynomials, the precise relation being

$$p_r(j; n) = M_r(j; n, q) q^r.$$

Thus, the four familiar integer-valued distributions which we have considered each relate to a well-known orthogonal polynomial system relative to which distribution they are orthogonal. The rth orthogonal polynomial forms the right eigenvector corresponding to eigenvalue of form m^r.

6.3 The Generalized Negative Binomial Model

From (20), we are led to the obvious generalization of the negative binomial distribution with pgf

$$H(s) = \left(\frac{1-q}{1-qs}\right)^b,$$

which is the stationary distribution of a branching process with immigration, whose immigration and offspring distributions have respective pgf's

$$b(s) = (1+q-qs)^{-b}, \quad f(s) = (1+q-qs)^{-1}.$$

The set of orthogonal polynomials $p_r(j;b)$ orthogonal with respect to the stationary distribution are now given by

$$p_r(j;b) = M_r(j;b,q)q^r,$$

with the rth polynomial forming the right eigenvector corresponding to eigenvalue m^r.

Finally, as before with $b = n$, for any $b > 0$ this process is *reversible*, and so the left eigenvector is easily obtained.

6.4 The Karlin and McGregor Spectral Theory

In their concluding Section 9, Karlin and McGregor (1966) consider branching processes with immigration, with their brief Case II dedicated to the subcritical case $m < 1$, assuming also that $f(0) > 0$ and $f(s), b(s)$ are analytic in the neighbourhood of 1. Their Theorem 13 asserts that under these conditions the eigenvalues of P are $1, m, m^2, \ldots$ and the left eigenvector for m^r has generating function $H_r(s) = \sum_{i=0}^{\infty} U_i(r)s^i = H(s)(A(s))^r, r = 0, 1, 2, \ldots$. The right eigenvectors $V_j(r)$ for each fixed argument value j are generated by

$$\sum_{r=0}^{\infty} V_j(r)w^r = \frac{B^j(w)}{H(B(w))}.$$

Here,

$$A(s) = \lim_{n\to\infty} \frac{f_n(s)-1}{m^n},$$

$f_n(s)$ is the nth functional iterate of $f(s)$, and $B(s)$ is the inverse function of $A(s)$. In the Smoluchowski case of our Section 6.1 and in the negative binomial case of our Section 6.2 when $n = 1$, the functions $A(s)$, $B(s)$ are easily established as in fact Karlin and McGregor (1966) point out, and our results follow almost trivially from their exposition.

However, Karlin and McGregor's (1966) intention is to establish a spectral theory for *given* $f(s), b(s)$, with $A(s)$, $H(s)$ well defined but in general not explicitly known. Our aim, on the other hand, is to start with a familiar integer-valued distribution [with known pgf $H(s)$], with respect to which there is a well-known system of orthogonal polynomials, and then show that the orthogonal polynomials form the right eigenvectors, corresponding to eigenvalues m^r of a branching process with immigration, with $H(s)$ as the pgf of the stationary distribution. In the cases we have considered, reversibility of the process gives the left eigenvector.

The explicit results of our Sections 6.2, and 6.3, for general paramater $b > 0$ bypass the need to obtain the functions $A(s)$, $B(s)$ for the explicit construction of the right and left eigenvectors when $b \neq 1$ for Karlin and McGregor's (1966), Theorem 13.

APPENDIX A: ELEMENTS OF MARKOV CHAIN THEORY

The purpose of this section is to review those elements of Markov chain theory that are specifically reflected in the context of statistical mechanical models of our preceding account.

A Markov chain is a probability model which allows for simple statistical dependence between observations $X_0, X_1, X_2, \ldots$ on a sample space S at successive time points $n = 0, 1, 2, \ldots$.

In its aspect as a dynamic model over time, that is, as a stochastic process, it is said to describe the evolution over time of a "system" on a fixed "state space" S, where movement is from state to state at unit time intervals. This description may derive from Markov chains as models in statistical mechanics, and is in any case appropriate for this paper.

The homogeneous Markov property is expressed by

$$\begin{aligned}&\Pr(X_{m+1} = j | X_m = i, X_{m-1} = i_{m-1}, \ldots, X_0 = i_0)\\ &\quad = p_{ij}, \quad i, j \in S.\end{aligned}$$

When the p_{ij} are written as entries of a matrix P, then

$$P = \{p_{ij}\} \geq 0, \quad P\mathbf{1} = \mathbf{1}.$$

Nonnegative matrices with this property are called *stochastic*. P is the *transition matrix* of the Markov chain. Markov chains have the property that

$$P^n = \{p_{ij}^{(n)}\} \quad \text{where } p_{ij}^{(n)} = \Pr(X_{m+n} = j | X_m = i),$$

which allows their analysis by the tools of matrix theory, and particularly the theory of nonnegative matrices, specifically the Perron–Frobenius theory, if P is finite.

The matrix P is said to be *irreducible* if for every pair $i, j \in S$ there exists a positive integer $m \equiv m(i, j)$ such that $p_{i,j}^{(m)} > 0$. In modelling terminology: for every pair of states $i, j \in S$, it is possible with positive probability to pass from state i to state j in some number of steps which depends in general on i, j.

For the finite chains that we shall consider, it is natural to label the states as $\{0, 1, \ldots, N\}$. For our infinite state space chains, the labels will be $S = \{0, 1, 2, \ldots\}$.

If P is finite and irreducible, there is a unique solution vector π of

$$\pi^T P = \pi^T, \quad \pi^T \mathbf{1} = 1.$$

Its elements form a probability distribution

$$\boldsymbol{\pi}^T = \{\pi_0, \pi_1, \ldots, \pi_N\}.$$

with strictly positive entries. This π is called the *stationary distribution* vector. It is clearly a left eigenvector of P corresponding to eigenvalue 1.

Since for any stochastic $P, P\mathbf{1} = \mathbf{1}, \mathbf{1}$ is a right eigenvector corresponding to eigenvalue 1.

If $\sum_{j \geq 0} j p_{ij} = ai + b, i \geq 0$ or equivalently,

$$E(X_{n+1}|X_n) = aX_n + b, \quad |a| < 1$$

for constants a and b for irreducible P, and $\mu = \sum_j j\pi_j$, then $\sum_{j \geq 0}(j - \mu)p_{ij} = a(i - \mu), i \geq 0$. Equivalently: $E((X_{n+1} - \mu)|X_n) = a(X_n - \mu)$. Thus $\{(j - \mu), j = 0, 1, \ldots\}$ is a right eigenvector of P corresponding to eigenvalue a, where $|a| < 1$, and

$$\begin{aligned} E(X_{n+1} - \mu) &= aE(X_n - \mu) \\ &= a^{n+1}E(X_0 - \mu), \end{aligned} \tag{23}$$

so that $|E(X_n - \mu)|$ is decreasing as the chain evolves over time.

Irrespective of the distribution over S at time 0, a chain with irreducible finite transition matrix, is "positive recurrent," which means that *every state recurs with probability one*, and the mean time between recurrences is finite. For state i, the mean recurrence time is

$$\mu_i = \frac{1}{\pi_i}, \quad i = 0, 1, \ldots, N. \tag{24}$$

If the Markov chain with transition matrix P starts off at time 0 with the distribution vector $\boldsymbol{\pi}^T$ over its states, this is the distribution at all time points $n = 0, 1, 2, \ldots$:

$$\Pr(X_n = j) = \pi_j, \quad j = 0, 1, \ldots, N$$

and, more generally, the Markov chain is (strictly) stationary.

A positive recurrent stationary Markov chain with transition matrix $P = \{p_{ij}\}$ viewed backward in time is also a stationary Markov chain—called the reverse chain—with transition probability from state i to state j given by $\hat{p}_{ij} = \pi_j p_{ji}/\pi_i, i, j \in S$.

If the entries of the transition matrix P satisfy $p_{ij} = \hat{p}_{ij}$, that is, if

$$p_{ij} = \frac{\pi_j p_{ji}}{\pi_i}, \quad i, j \in S \tag{25}$$

a stationary Markov chain is *reversible* in time since the transition and stationary probabilities for the process are the same for the chain viewed backward in time as when viewed forward. For example,

$$P(X_n = i, X_{n+1} = j) = P(X_n = j, X_{n+1} = i).$$

Such a finite transition matrix P satisfying (25) has all its eigenvalues real, since the matrix $A = \{\pi_i^{1/2} p_{ij}/\pi_j^{1/2}\}$, a similarity transform of the matrix P, is symmetric.

If $\mathbf{w}^{(r)} = \{w_i^{(r)}\}, i = 0, 1, 2, \ldots, N$ is a right (column) eigenvector of a reversible irreducible P corresponding to eigenvalue λ_r, then a left (row) eigenvector $\mathbf{v}^T = \{v_i\}, i = 0, 1, 2, \ldots, N$ is given by $v_i = w_i \pi_i$. Thus, if all eigenvalues are distinct, the eigenvectors $\mathbf{w}^{(r)}, r = 0, 1, 2, \ldots, N$ form an orthonormal set (when properly standardized) with respect to the stationary distribution $\{\pi_0, \pi_1, \pi_2, \ldots, \pi_N\}$.

The property that $\{(j - \mu), j = 0, 1, \ldots\}$ is a right eigenvector of P corresponding to eigenvalue a, where $|a| < 1$, together with the property that *1* is a right eigenvector corresponding to eigenvalue 1, and reversibility suggest, then that the right eigenvector system of P is a system of polynomials orthogonal with respect to the stationary distribution.

APPENDIX B: SOME ADDITIONAL BIOGRAPHICAL NOTES

For additional detail on Markov's life and work, and his legacy, see Seneta (2006), Sections 1–6.

Some biographical detail on Marian Smoluchowski (1872–1917) and Mikhailo Kravchuk (Krawtchouk) (1892–1942) is given in Seneta (2001b). Both lives were cut short tragically by the times.

Stephen Kulik (Koulik) was a colleague and coauthor of Krawtchouk (better transliterated into English as Kravchuk) in Kyiv (Kiev) prior to World War 2, during the period of Ukrainianization. Kravchuk was eventually sentenced to the Siberian camps where he died in 1942 (see Seneta, 1997, 2001b). Kulik's (1943) paper is his last in a Soviet journal. It was received by the journal on 27 February, 1941. This was before the outbreak of hostilities between Germany and the Soviet Union later in 1941. Kravchuk had been sentenced by then, however, and became a nonperson; mention of his publications was prohibited. Thus, Koulik (1943) begins with the derivation of the polynomial system orthogonal to the binomial distribution, without mention of Kravchuk's (1929) well-known paper of 1929.

Kulik (1943) has a brief résumé in French titled "Fonctions génératrices de quelques polynomes orthogonaux." In the event, he deals only with polynomials orthogonal to a generalization of the binomial distribution, still on a finite number of points. One of his sources is a rare Ukrainian-language version of Bernstein (1934). Bernstein around this time was Commissar for Education in the Ukrainian SSR.

During the war, Kiev was for a time occupied by the Germans. Kulik managed to make his way to the West, first apparently to England where he was publishing from No. 1 Laboratory of Cortauld's Limited, Coventry, in 1948; and then to the US, where he was teaching and publishing from the-then Claremont Men's College in California, in 1953. He is very likely the Stephen Kulik born 6 January 1899, who died 12 October 1989, in Santa Ana, Orange County, California.

Kulik (1943, 1953) cites the work of the slightly older disciple of Kravchuk on orthogonal polynomial systems Aleksandr (Oleksander) Stepanovich Smohorshevsky, born 1896, who had been a schoolteacher. Smohorshevsky continued to publish until 1941 and then, having remained in Kiev, from 1945 until at least 1956 still in Soviet journals, but not on orthogonal polynomials.

The contributions of the Kravchuk School in Kiev of the later 1930s gained little traction in the West compared to the contributions on orthogonal polynomials by German and French authors. The relative isolation of Soviet authors from outside journals led to overlap.

The journal in which Kulik's (1953) paper [which does cite Krawtchouk (1929) as well as his own Kulik 1943] was published, was an organ of the Shevchenko Scientific Society of L'viv (Lemberg, Lwów, L'vov) until the Society was closed down by the Soviets in 1939. Kravchuk and Smohorshevsky had published in it, and eminent scientists such as Einstein had been Honorary Members of the Society, which had been founded when Lemberg was in the Austro-Hungarian Empire. The Society was incorporated in New York after the war, and its *Proceedings (Sitzungsberichte)* continued as a new series in 1953.

APPENDIX C: EVOLUTION OF THIS PAPER

My initial publication on themes of the present paper in the context of finite Markov chains was Seneta (1982).

An invited talk to the 4th World Congress of the Bernoulli Society, in Vienna, Austria, 26–31 August, 1996 (presented on my behalf by Professor Peter Jagers), introduced Smoluchowski's model as a special case of a branching process with immigration, and thus of a Markov chain with countably infinite state space which was *reversible* and, in this guise, could be also used to illuminate "the fundamental issues involved."

A unified consideration of the properties of three models (Bernoulli/Laplace, Ehrenfest and Smoluchowski) was presented on 3 November, 2006 as the Moyal Medal Lecture, at Macquarie University, Sydney (Seneta, 2014). Part of its focus was on the remarkably simple common features of the three models, their reversibility and the nature of their spectral theory in contrast to the dissimilar probabilistic structure of the Smoluchowski model from the other two.

Sections 3–6.1 of the sequel contain the material of these lectures in enhanced form, at another interval of 10 years.

Section 6 of the present paper was completed recently. It includes a fourth model, another special case of a branching process with immigration, to round out the picture of common features of the models.

ACKNOWLEDGMENTS

I am indebted to Dr. Bill Lloyd-Smith for encouragement to publish from the Bernoulli Society and Moyal Lectures, and to Professor Peter Green for encouragement to submit an article on history. I am especially grateful for the professional and extensive reports of an Associate Editor and three referees, who between them looked at all aspects of the subject matter: historical, statistical physics and probabilistic.

REFERENCES

BERNOULLI, D. (1769). Disquisitiones analyticae de novo problemate conjecturale. *Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae* **14** 3–25.

BERNSTEIN, S. N. (1934). *Teoriia Veroiatnostei*. [*Theory of Probabilities*], 2nd ed. Gostehizdat, Moskva-Leningrad.

BRU, B. (2003). Souvenirs de Bologne. *Journal de la Société Française de Statistique* **144** 135–226.

BRUNS, H. (1906). *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*. Teubner, Leipzig.

CHANDRESEKHAR, S. (1943). Stochastic problems in physics and astronomy. *Rev. Modern Phys.* **15** 1–89. MR0008130

DE LAPLACE, P. S. (1812). *Théorie Analytique des Probabilités*. V. Courcier, Paris.

DIACONIS, P. and SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J. Math. Anal.* **18** 208–218. MR0871832

DOBRUSHIN, R. L., SUKHOV, YU. M. and FRITTS, Ĭ. (1988). A. N. Kolmogorov—founder of the theory of reversible Markov processes. *Uspekhi Mat. Nauk* **43** 167–188. MR0983882

DOEBLIN, W. (2016). Oeuvres-Collected Works (M. Yor and B. Bru, eds.). Springer, Berlin.

DONNELLY, P., LLOYD, P. and SUDBURY, A. (1994). Approach to stationarity of the Bernoulli–Laplace diffusion model. *Adv. in Appl. Probab.* **26** 715–727. MR1285456

EHRENFEST, P. and EHRENFEST, T. (1907). Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. *Physikalische Zeitschrift* **8** 311–314.

FRÉCHET, M. (1938). *Recherches Théoriques Modernes sur Le Calcul des Probabilités. Second Livre. Méthode des Fonctions Arbitraires. Théorie des Événements en Chaîne dans Le Cas D'un Nombre Fini D'états Possibles*, 2nd ed. Gauthier-Villars, Paris.

FROBENIUS, G. (1908). Über Matrizen aus positiven Elementen. In *S.-B. Preuss. Akad. Wiss., (Berlin)* 471–476.

FÜRTH, R. (1918). Statistik und Wahrscheinlichkeitsnachwirkung. *Physikalische Zeitschrift* **19** 421–426.

GOTTLIEB, M. J. (1938). Concerning some polynomials orthogonal on a finite or enumerable set of points. *Amer. J. Math.* **60** 453–458. MR1507326

HADAMARD, J. (1928). Sur le battage des cartes et ses relations avec la mécanique statistique. *Atti del Congr. Inter. Mat., Bologna* **5** 133–139.

HADAMARD, J. and FRÉCHET, M. (1933). Sur les probabilités discontinues des événements en chaîne. *Zeitschrift Für Angewandte Mathematik und Mechanik* **13** 92–97.

HAVLOVÁ, V., MAZLIAK, L. and ŠIŠMA, P. (2005). Le début des relations mathématiques franco-tchécoslovaques vu à travers la correspondance Hostinský–Fréchet. *J. Électron. Hist. Probab. Stat.* **1** 18 pp. (electronic). MR2208346

HAWKINS, T. (2013). *The Mathematics of Frobenius in Context*. Springer, New York. MR3099749

HEYDE, C. C. and SENETA, E. (1972). Estimation theory for growth and immigration rates in a multiplicative process. *J. Appl. Probab.* **9** 235–256. MR0343385

HOSTINSKY, B. (1931). *Méthodes Générales du Calcul des Probabilités. Mémorial des Sciences Mathḿatiques* **52**. Gauthier-Villars, Paris.

HOSTINSKY, B. (1939). Probabilités relatives aux tirages de deux urnes avec l'échange des boules extraites. *Acta* [*Trudy*] *Univ. Asiae Mediae. Ser. V-a. Fasc.* **21** 4–10. MR0020226

HOSTINSKY, B. and POTOČEK, J. (1935). Chaînes de Markoff inverses. *Bulletin International de L'Acad. des Sciences de Bohême* **36** 64–67.

KAC, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54** 369–391. MR0021262

KARLIN, S. and MCGREGOR, J. L. (1961). The Hahn polynomials, formulas and an application. *Scripta Math.* **26** 33–46. MR0138806

KARLIN, S. and MCGREGOR, J. (1966). Spectral theory of branching processes. I. The case of discrete spectrum. *Z. Wahrsch. Verw. Gebiete* **5** 6–33. MR0205338

KOHLRAUSCH, K. W. F. and SCHRÖDINGER, E. (1926). Das Ehrenfestsche Modell der H-Kurve. *Physikalische Zeitschrift* **27** 306–313.

KOLMOGOROFF, A. (1935). Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112** 155–160.

KOLMOGOROFF, A. (1936). Zu Umkehrbarkeit der statistischen Naturgesetze. *Math. Ann.* **113** 766–772.

KOULIK, S. (1943). Proizvoizvodiashchie funktsii nekotorikh ortogonalnikh polynomov. [Generating functions of some orthogonal polynomials]. *Matematicheskii Sbornik* **12** 320–334.

KRAWTCHOUK, M. (1929). Sur un généralization des polynomes d'Hermite. *Compte Rendus de l'Académie des Sciences, Paris* **189** 620–621.

KULIK, S. (1953). Orthogonal polynomials associated with the binomial probability function with a negative integral index. *Shevchenko Scientific Society. Proceedings of the Mathematical, Physical and Medical Section* **1** 7–10.

MARKOFF, A. A. (1898). Sur les racines de l'équation $e^{x^2} d^m e^{-x^2}/dx^m = 0$. *Izvestiia Akad. Nauk* (*St. Petersburg*), (*5th Ser.*) **9** 435–446.

MARKOFF, A. (1910). Recherches sur un cas remarquable d'épreuves dépendantes. *Acta Math.* **33** 87–104. MR1555057

MARKOFF, A. A. (1912). *Wahrscheinlichkeitsrechnung*, 2nd ed. Teubner, Leipzig.

MARKOV, A. A. (1906). Extension of the law of large numbers to dependent quantities. *Izvestiia Fiz.-Matem. Obsch. Kazan Univ.*, (*2nd Ser.*) **15** 135–156. (In Russian). Also in Markov (1951), pp. 339–361.

MARKOV, A. A. (1907). Investigation of a notable instance of dependent trials. *Izvestiia Akad. Nauk* (*St. Petersburg*), (*6th Ser.*) **1** 61–80. (In Russian.)

MARKOV, A. A. (1908). Extension of limit theorems of the calculus of probabilities to sums of quantities associated into a chain. *Zapiski Akad. Nauk* (*St. Petersburg*), *Fiz.-Mat. Otd.*, (*7th Ser.*) **22** 363–397. (In Russian). Also in Markov (1951), pp. 363–397. English translation in Sheynin (2004a), pp. 159–178.

MARKOV, A. A. (1911). On dependent trials not forming a true chain. *Izvestiia Akad. Nauk* (*St. Petersburg*), (*6th Ser.*) **5** 113–126. (In Russian). Also in Markov (1951), pp. 399–416.

MARKOV, A. A. (1915). On a problem of Laplace. *Izvestiia Akad. Nauk* (*St. Petersburg*), (*6th Ser.*) **9** 87–104. (In Russian). Also in Markov (1951), pp. 549–571.

MARKOV, A. A. (1951). *Izbrannye Trudy*. Izdat. Akad. Nauk SSSR, Leningrad. MR0050525

MAZLIAK, L. (2007). On the exchanges between Wolfgang Doeblin and Bohuslav Hostinský. *Rev. Histoire Math.* **13** 155–180. MR2422946

MEIXNER, J. (1934) Orthogonale Polynomsysteme Mit Einer Besonderen Gestalt Der Erzeugenden Funktion. *J. Lond. Math. Soc.* (2) **S1-9** 6–13. MR1574715

MILLS, T. M. and SENETA, E. (1989). Goodness-of-fit for a branching process with immigration using sample partial autocorrelations. *Stochastic Process. Appl.* **33** 151–161. MR1027114

MILLS, T. M. and SENETA, E. (1991). Independence of partial autocorrelations for a classical immigration branching process. *Stochastic Process. Appl.* **37** 275–279. MR1102874

PERRON, O. (1907). Zur Theorie der Matrices. *Math. Ann.* **64** 248–263.

ROMANOVSKY, V. (1929). Sur les chaînes de Markoff. *Doklady Akademii Nauk SSR, Ser. A* 203–208.

ROMANOVSKY, V. (1936). Recherches sur les chaînes de Markoff. *Acta Math.* **66** 147–251. MR1555412

ROMANOVSKY, V. I. (1949). *Discretnie Tsepi Markova*. Moskva-Leningrad, Gostekhizdat. English translation 1970 by E. Seneta: *Discrete Markov Chains*, Wolters-Noordhoff, Groningen.

SAPOGOV, N. (1951). Commentary in Russian on a group of Markov's papers, including Markov (1908, 1911). In *Markov (1951)*, pp. 662–664.

SARYMSAKOV, T. A. (1955). Vsevolod Ivanovich Romanovsky. An obituary. *Uspekhi Mat. Nauk* **10** 79–88. (In Russian). English translation in Sheynin (2004b), Section 8g. MR0067810

SCHNEIDER, H. (1977). The concepts of irreducibility and full indecomposability of a matrix in the works of Frobenius, König and Markov. *Linear Algebra Appl.* **18** 139–162. MR0446850

SENETA, E. (1982). Entropy and martingales in Markov chain models. *J. Appl. Probab.* **19A** 367–381. MR0633206

SENETA, E. (1997). M. Krawtchouk (1892–1942): Professor of mathematical statistics. *Theory Stoch. Process.* **3** 388–392.

SENETA, E. (1998). Complementation in stochastic matrices and the GTH algorithm. *SIAM J. Matrix Anal. Appl.* **19** 556–563 (electronic). MR1614015

SENETA, E. (2001a). Characterization by orthogonal polynomial systems of finite Markov chains. *J. Appl. Probab.* **38A** 42–52. MR1915533

SENETA, E. (2001b). Marian Smoluchowski. In *Statisticians of the Centuries* (C. C. Heyde and E. Seneta, eds.). Springer, New York. MR1863206

SENETA, E. (2006). Markov and the creation of Markov chains. In *MAM* 2006: *Markov Anniversary Meeting* (A. N. Langville and W. J. Stewart, eds.) 1–20. Boson, Raleigh, NC.

SENETA, E. (2009). Karl Pearson in Russian contexts. *Int. Stat. Rev.* **77** 118–146.

SENETA, E. (2014). Markov Chains as Models in Statistical Mechanics. Available as www.maths.mq.edu.au/wp-content/uploads/2014/07/moyal-seneta.pdf.

SENETA, E. (2016). Doeblin on discrete Markov chains. In *Doeblin* (2016).

SHEYNIN, O. B., ed. (2004a). *Probability and Statistics. Russian Papers. Selected and Translated by Oscar Sheynin*. NG Verlag, Berlin.

SHEYNIN, O. B., ed. (2004b). *Probability and Statistics. Russian Papers of the Soviet Period. Selected and Translated by Oscar Sheynin*. NG Verlag, Berlin.

SREDNIAWA, B. (1992). The collaboration of Marian Smoluchowski and Theodor Svedberg on Brownian motion and density fluctuations. *Centaurus* **35** 325–355.

SZEGÖ, G. (1939). *Orthogonal Polynomials*. Amer. Math. Soc., New York.

TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability*. Macmillan, London. [Reprinting].

VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung*. Deuticke, Leipzig und Wien.

VON MISES, R. (1932). Théorie des Probabilités. Fondements et Applications. *Annales de l'Institut Henri Poincaré* **3** 137–190.

VON SMOLUCHOWSKI, M. (1914). Studien über Molekularstatistik von Emulsionen und deren Zusammenhang mit der Brown'schen Bewegung. In *Sitzungsberichte der Akademie der Wissenschaften*. Mathematisch-Naturwissenschaftliche Klasse **CXXIII**. Band X, Abteilung IIA. 2381–2405. Hölder, Wien.

WAX, N. (1954). *Selected Papers on Noise and Stochastic Processes*. Dover, New York. MR0062373