International Statistical Review (3), **1966**, 64, 255–263, Printed in Mexico

Markov and the Birth of Chain Dependence Theory

E. Seneta

School of Mathematics and Statistics, University of Sydney, N.S.W. 2006, Australia

Summary

Markov's work on chain dependence was motivated by his desire to refute a statement by Nekrasov that pairwise independence of random summands was a necessary condition for the Weak Law of Large Numbers. He did this by obtaining such a Law in 1906 for systems of dependent random variables, in particular for finite homogeneous 'Markov' chains. Nekrasov's incorrect assertion arose out of the theological doctrine of free will, with which some members of the Moscow School of Mathematics of the time were much concerned.

The first part of the paper presents the background to the above. The second part deals with the somewhat neglected techniques of Markov's 1906 paper, especially his use of what is now known as the ergodicity coefficient, to express the contractive effect of applying a stochastic matrix to a column vector. This coefficient underlies his ergodicity arguments, and his proof of the Weak Law.

Key words: Law of Large Numbers; Markov chain; Free will; Moscow Mathematical School; Contraction; Ergodicity coefficient.

1 Introduction

Markov chains, and, more generally, Markov processes, are named after the great Russian mathematician Andrei Andreevich Markov [1856–1922]. The paper in which "Markov chains" first make an appearance in his writings (Markov, 1906) concludes with the sentence

> "Thus, independence of quantities does not constitute a necessary condition for the existence of the law of large numbers."

This sentence encapsulates Markov's motivation for studying schemes of chain dependence. One of the purposes of our paper is to elucidate the little known circumstances of how his work on a general theory of such chain dependence was influenced by interaction with Pavel Alekseevich Nekrasov [1853–1924]. Material related to what is said here may be found within a broader context in Seneta (1984, §8 and Appendix 1); and in Sheynin (1989a).

A second aim is to give an account of some of the techniques and concepts, introduced by Markov within the context of chain dependence, which have persisted till the present day.

A (finite) Markov chain $\{X_n\}$ is a sequence of dependent variables with the following probabilistic structure. X_n denotes the 'state' at time n; this is one of a finite set, S, of possible values. The conditional probability

$$Pr\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots\}$$

Invited paper, 50th Session ISI, Beijing, 1995.

for any $j, i, i_{n-1}, i_{n-2}, \dots \quad \varepsilon S$

is independent of the sequence of past values $\{i_{n-1}, i_{n-2}, \dots\}$ whenever $\{i, i_{n-1}, i_{n-2}, \dots\}$ has positive probability, so that the probabilistic structure of the future is completely specified by a knowledge of the present state ($X_n = i$). If the above conditional probability is independent of time n (time-homogeneity) it can be denoted by p_{ij}, and the array of numbers $P = \{p_{ij}\}, i, j, \varepsilon S$ is the transition matrix of the Markov chain. Even then, the marginal distribution of X_n may differ for various n (more formally: the homogeneous chain may not be strictly stationary). If the conditional probability depends on n, there is a sequence of transition matrices, one for each $n \geq 0$ and the Markov chain is inhomogeneous. In either case, an additional specification of an initial distribution $\{\pi_i\}$, where $\pi_i = Pr(X_0 = i), i \varepsilon S$, completely determines the joint distribution of the X_n's. Markov dealt with inhomogeneous as well as homogeneous chains, but assumed that transition matrices were strictly positive.

Special instances of Markov chains were known before Markov: the gambler's ruin problem dates back at least to the correspondence between Pierre de Fermat and Blaise Pascal (*circa* 1654); there were various urn models—especially those due to Daniel Bernoulli in 1769 and Laplace in 1812. Fundamental results were obtained in 1845 by Irenée Jules Bienaymé [1796–1878] and then Galton & Watson (1873–4) for the simple branching process. Quetelet [1796–1874] applied two-state Markov chains to hydrology (see Stigler, 1975) in 1846. Contact with Chuprov in 1910 through correspondence led Markov to realize that by considering overlapping Bernoulli trials, Ernst Heinrich Bruns [1848–1919] had in fact studied in 1906 what came to be known as Markov–Bruns chains (Romanovsky, 1949, Chapter VI). Markov's outstanding contribution was a general formulation and basic general results. The general formulation occurred in response to specific stimuli from an unexpected and unscientific quarter.

2 Markov and Nekrasov

Markov belonged to the St. Petersburg Mathematical "School" founded by Chebyshev, and he and Liapunov were the most eminent of Chebyshev's disciples in probability. The work of Chebyshev focussed on the Weak Law of Large Numbers (WLLN) and the Central Limit problem for independent but not necessarily identically distributed summands. The probabilistic work of Markov initially consisted in making rigorous, and in generalizing, Chebyshev's version of the Central Limit Theorem (CLT), following his teacher in applying the method of moments (in contrast to Liapunov's exploitation of characteristic functions).

In regard to the WLLN, the standard approach of this time was via the Bienaymé–Chebyshev Inequality

$$Pr\{|\bar{X}_n - E\bar{X}_n| \geq \varepsilon\} \leq \text{ Var } \bar{X}_n/\varepsilon^2$$

where $\bar{X}_n = \sum_{i=1}^n X_i/n$ and $X_i, i \geq 1$ are random variables. Here, since independent X_i's were being considered with $\sigma_i^2 = \text{ Var } X_i$ well defined, $\text{Var } \bar{X}_n = n^{-2}\sum_{i=1}^n \sigma_i^2$, and thus $n^{-2}\sum_{i=1}^n \sigma_i^2 \to 0$ was sufficient to ensure the WLLN.

More generally,

$$\text{Var } \bar{X}_n = \frac{1}{n^2}\left\{\sum_{i=1}^n \text{ Var } X_i + 2\sum_{i<j} \text{ Cov } (X_i, X_j)\right\}$$

so the same sufficient condition would work if the X_i's were merely "pairwise independent", thus still resulting in the disappearance of the covariance terms. This fact was noticed by Nekrasov (1902) in examining the "logical underpinnings" of the celebrated Inequality. This fact, obvious to us now, but with its suggestion of orthogonality, is an important advance, and is so recognized by Chuprov

(1959, p.168) in his monograph of 1909.

However, Nekrasov wanted to use the *observed* stability of averages in everyday life to *infer* the pairwise independence of the X_i's (that is, to infer that pairwise independence is *necessary* for the WLLN). Behind this was his belief that observed statistical laws for observational data implied "free will", and that pairwise independence was an expression of free will. Chuprov (1959, pp. 222–224) rightly perceives Nekrasov's position as both philosophically and mathematically invalid. However Chuprov's mention of Nekrasov in a positive light (relating to the sufficiency of pairwise independence) brought an irate postcard (dated 2 November 1910) from Markov when he had "discovered" Chuprov's book. This postcard, which initiated their correspondence reads:

> "I notice with astonishment that . . . P.A. Nekrasov, whose work in recent years represents an abuse of mathematics, is mentioned next to Chebyshev."

The enmity between Markov and Nekrasov had its genesis in another long and bitter controversy between them (Seneta, 1984, §§6–7), initiated by a probabilistic paper of Nekrasov (1898), dedicated to Chebyshev (!) and containing no proofs. In this and later publications on this topic, Nekrasov — who was highly proficient in the use of complex variable theory in general and the Lagrange expansion in particular—attempted to use the method of saddlepoints, of Laplacian peaks and the Lagrange inversion formula to establish, for sums of non-identically distributed lattice random variables, what are now standard local and global limit theorems of Central Limit type for large deviations. The attempt was very many years ahead of its time, but was only partly successful and poorly presented. Its specific inaccuracies were criticised by Markov and Liapunov who never understood the general direction (nor was it noticed by their successors). Indeed, since Nekrasov, subsequent to 1898, wrote about 1000 pages on the topic, spread serially over volumes **21–23** of *Matematicheskii Sbornik* under the grandiose title "New Foundation of the Study of Probabilistic Sums and Mean Values", the task of penetrating the work was formidable.

It was, however, Nekrasov's (1902) attempt to use mathematics and statistics in support of "free will" that led Markov to construct a scheme of *dependent* random variables $\{X_n\}$ in his 1906 paper on the basis of which Nekrasov's claim for necessity of independence for the WLLN was convincingly refuted. The background to this therefore deserves some attention.

3 The Moscow Mathematical School

A basic Judeo-Christian doctrine is that every person is responsible for his/her actions. That is, every person has free will. In the times of which we speak this doctrine was of fundamental interest to certain mathematicians, who were firm adherents of the Russian Orthodox faith which permeated every aspect of life in the pre-revolutionary Russian Empire. We quote here Chuprov (1959, pp. 222–223)

> "Attempts to enlist statistics as a foundation of the theory of will have not abated, even so . . . . Of considerable interest in this connection is the reasoning of a group of Russian academics inclined to call itself the 'Moscow School'. In recent years their views have received publicity well beyond the limits of Russia thanks to the German-language articles of V.G. Alekseev, which exposit them in readily comprehensible form. The 'Moscow School' decidedly insists that free will is the *conditio sine qua non* of statistical laws governing everyday life."

These mathematicians were involved in leading roles within the Moscow Mathematical Society which published the *Matematicheskii Sbornik*. The most eminent of them was N.V. Bugaiev, who focussed on the study of discontinuous functions, since he perceived discontinuity as according with free will. Others were V.Ia. Zinger (or Tsinger) and P.A. Nekrasov. An account of the "idealist" activities of this group, and the fate after the revolution of their "offspring" Egorov and Luzin, is given in English by Ford (1991).

Needless to say, while there was clearly tension between the St. Petersburg School and the Moscow School in pre-revolutionary times, it took considerable time for the *mathematical* achievements of the Moscow School to come out of the severe decline engendered by its "idealist" philosophy. Eventually, one of the leaders of the "Initiative Group for the Reorganization of the Mathematical Society" in 1930 (see Ford, 1991), L.A. Lyusternik, in his memoirs (Lyusternik, 1967, §3) exonerates all the founders of the Moscow Mathematical Society of reactionary character, laying the blame squarely for the damage to their reputations on

> "... the pseudo-scientific articles of the MU [Moscow University] professor Nekrasov ... in which he tried to substantiate the indispensability of the regime prevailing at the time."

Lyusternik goes on to recollect that in:

> "... the first half of the 1920's Nekrasov still attended the meetings of the MMO [Moscow Mathematical Society] and sometimes even presented papers. A queer shadow of the past, he seemed decrepit—physically and mentally—and it was difficult to understand him ... This pitiful old man was like a shabby owl."

A substantially different view of P.A. Nekrasov is given in an obituary by his student S.P. Sluginov (1927; translated in Seneta, 1984) written at a time of detente and reconstruction in Russia before the onset of Stalinism.

It is not surprising that one of the brightest student members of the Moscow Mathematical School, Pavel Aleksandrovich Florensky, a friend of Luzin, became a Russian Orthodox priest (Ford, 1991). He acquired great eminence as a scientist and theologian, but the manner and date of his death were unknown or misreported (as 1943) until recently. In fact, Florensky was executed by shooting on December 8, 1937 (Khoruzhy, 1990, p. IX). He was born January 21 (new style), 1882.

4 Markov in Biography

Just as the personality of Nekrasov inflamed their various controversies, so did that of Markov, who, in contrast however, is often painted in glowing colours. This is deservedly so in respect of his mathematics, but is also due to the fact that his character and beliefs were very acceptable to the incoming political system following the October revolution. This made it politically acceptable, and indeed desirable, to put the St. Petersburg School into exclusive eminence in Soviet mathematical historiography. We recall from the brief biography by his son in Markov (1951) the protests of Markov in 1902 over the reversal by the Tsar of the election as Honorary Member of the Russian Academy of A.M. Gorky (Peshkov), the subsequent refusal to accept any awards ("orders") from the Academy, or to act as "agent of the government" in relation to students at St. Petersburg University. In 1912 when the Synod of the Russian Orthodox Church excommunicated Leo Tolstoy, Markov likewise requested excommunication.

On the other hand, Grodzensky's (1987) more recent biography, while it tries continuously to justify Markov's maverick behaviour in terms of a high-minded, uncompromising, liberal and emotional nature, shows nevertheless that Markov was also ready to dash off an aggrieved and strongly-worded letter in causes not as just as described above. In 1891 he implicitly accused his mentor Chebyshev of plagiarism (pp. 62–63). He had an extremely negative attitude to the person and work of Sofia Kovalevskaya (Sonia Kowalewski), initially his competitor for Academy membership (pp. 66–77), and this did not abate even after her death and had connections with Nekrasov.

There is another rather pathetic but illuminating connection with Gorky in the last year of Markov's life in the aftermath of the revolution (Grodzensky, p. 136):

> "On the 5th of March 1921 A.A. Markov communicated that on account of the absence of footwear he is not able to attend meetings of the Academy. A few weeks later the KUBU [Committee for Improvement of the Existence of Scientists], meeting under the chairmanship of M. Gorky, fulfilled this prosaic request of the famous mathematician. Time, however, provided a colourful sequel, of sorts, to this. At the meeting of the physico-mathematical section of the Academy of Science on the 25th May, Andrei Andreevich announced: Finally, I received footwear; not only, however, is it stupidly stitched together, it does not in essence accord with my measurements. Thus, as before, I cannot attend meetings of the Academy. I propose placing the footwear received by me in the Ethnographic Museum as an example of the material culture of the current time, to which end I am ready to sacrifice it."

5 Techniques of Markov's 1906 Paper

The very first paper on chain dependence of Markov (1906), contains incisive ideas for the future development of the theory, and is the basis of several of the early correspondence exchanges. N. Sapogov's commentary on this paper in Markov (1951, pp. 626–628, 660–662) does not, however, fully explore these.

The thrust of §§2 and 5 of Markov (1906) is to obtain a WLLN for a *general* scheme of dependent variables $\{X_n\}, n \geq 1$. The scheme he constructs is in fact a finite Markov chain, which Markov takes to be time-homogeneous with all transition probabilities p_{ij} strictly positive. He shows that $E(X_{n+k}|X_n = i), i \varepsilon S$, and $E(X_{n+k})$ all have the same limit as $k \to \infty$; denote this limit by a. Using the same technique he is able to show that

$$E\Big(\big(\sum_{i=1}^{n}(X_i - EX_i)\big)^2\Big) < Gn \tag{5.1}$$

for a positive number G. Then since

$$E\left\{\big(\sum_{i=1}^{n}(X_i - EX_i)\big)^2 - \big(\sum_{i=1}^{n}(X_i - a)\big)^2\right\} = -\left\{\sum_{i=1}^{n}(EX_i - a)\right\}^2$$

he deduces that $Pr\{|\bar{X}_n - a| \geq \varepsilon\} \to 0$ as $n \to \infty$,
from the Bienaymé–Chebyshev Inequality, since $\left\{\sum_{i=1}^{n}(EX_i - a)\right\}^2/n^2 \to 0$ (because $EX_i \to a$ as $i \to \infty$).

This achieves the WLLN construction, but let us focus, rather, on the technique. An examination of Markov's mathematics (Markov, 1951, pp. 358–359) shows that central to his argument is the derivation of a recurrence inequality $\Delta^{(k+1)} < H\Delta^{(k)}$, where $0 < H < 1$ (and we have substituted k for his $i - 1$). The derivation actually gives the following general result (c.f. Seneta, 1981, Theorem 3.1): Let $\mathbf{w} = \{w_i\}$ be an arbitrary vector and $P = \{p_{ij}\}, i, j \varepsilon S$ a stochastic matrix. If $\mathbf{z} = P\mathbf{w}, \mathbf{z} = \{z_i\}$, then for any two indices h, h':

$$z_h - z_{h'} \leq \frac{1}{2}\sum_j |p_{hj} - p_{h'j}| \quad \{\max_j w_j - \min_j w_j\}$$

whence (tantamount to the inequality obtained by Markov):

$$\max_{h,h'} |z_h - z_{h'}| \leq H \max_{j,j'} |w_j - w_{j'}|. \tag{5.2}$$

where

$$H = \frac{1}{2} \max_{i,j} \sum_{s \varepsilon S} |p_{is} - p_{js}|. \tag{5.3a}$$

From the equivalent form

$$H = 1 - \min_{i,j} \sum_{s \varepsilon S} \min(p_{is}, p_{js}) \tag{5.3b}$$

it is clear that $0 \leq H \leq 1$.

The contractivity property (5.2) of a stochastic matrix P acting on $\mathbf{w}$ to form $\mathbf{z}$ is a tighter form of the inequalities

$$\min_i w_i \leq \min_i z_i, \quad \max_i w_i \geq \max_i z_i \tag{5.4}$$

which follow obviously from $z_i = \sum_j p_{ij} w_j, i \varepsilon S$, and are also implicit in Markov's reasoning.

Markov works with the specific choices $w_i = E(X_{n+k} | X_n = i)$, so that

$$\begin{aligned} z_i &= \textstyle\sum_j P(X_n = j | X_{n-1} = i) E(X_{n+k} | X_n = j) = E(X_{n+k} | X_{n-1} = i) \\ &= E(X_{n-1+k+1} | X_{n-1} = i). \end{aligned}$$

From (5.4) and time-homogeneity, $\min_i E(X_{n+k} | X_n = i)$ is non-decreasing with k, $\max_i E(X_{n+k} | X_n = i)$ is non-increasing, EX_{n+k} is sandwiched between, and from (5.2)

$$\Delta^{(k+1)} = \max_i E(X_{n+k+1} | X_n = i) - \min_i E(X_{n+k+1} | X_n = i) \leq K H^k$$

where

$$K = \max_i E(X_1 | X_0 = i) - \min_i E(X_1 | X_0 = i).$$

It is clear from (5.3b) that if the elements of P are all positive, then $H < 1$, yielding the limiting results for expectations required.

Equally striking is the justification for (5.1), which Markov says follows from

$$E((X_k - EX_k)\Big(\sum_{i=1}^{k-1}(X_i - EX_i)\Big)) < D(H + H^2 + \cdots + H^{k-1}) \tag{5.5}$$

for a positive constant D. He justifies (5.5) and (5.1) thus:

"... arguing entirely as in §2, we easily obtain the inequalities ... ".

Now, §2 is concerned with the case of a Markov chain on the state space $S = \{0, 1\}$ with (2×2) transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} q'' & p'' \\ q' & p' \end{bmatrix} \end{array}$$

corresponding to dependent trials (state 0 = failure, state 1 = success), in the *stationary regime* of the chain. This is a rather specific situation in which $H = |p' - p''|$. In §2, Markov certainly is close to (5.5), but actually argues from an exact expression for the left-hand side as a sum of powers of $p' - p''$, according as $p' < p''$ or $p' > p''$, in a situation where strict stationarity of the homogeneous chain is being assumed, whence neither Var X_n nor $\text{Cov}(X_n, X_{n+k})$ depend on n. In §5, on the other hand, Markov takes care not to assume stationarity of the homogeneous chains $\{X_n\}$. Note that, in fact, modulus signs are required on the left of (5.5).

He returns to the themes of his §2 repeatedly in subsequent work, but as far as we have been able to determine, the use of H as a tool in a general setting seems not to reappear. Markov's foci of WLLN and CLT for Markov chains (especially the inhomogeneous case) were taken up by S.N. Bernstein, beginning with a notable study, as one of his very earliest, in 1918 of the case of dependent trials (see Bernstein, 1964, pp. 61–65). Bernstein used in place of H for general inhomogeneous chains, a simpler tool; we take this up in the next section, when we discuss briefly the validity of (5.5) and (5.1) in a general setting.

On the other hand, the contraction effect (5.2) of H underlies various proofs of *ergodicity* (in which Markov seems to have been interested only in passing, as expressed above), in cases (such as that of a strictly positive P) where $H < 1$. Thus consider the matrix of k-step transition probabilities $P^k = \{p_{js}^{(k)}\}$ for a homogeneous chain, and take in (5.2) $w_j = p_{js}^{(k)}$; then $z_i = p_{is}^{(k+1)}$ so that

$$\max_{h,h'} |p_{hs}^{(k+1)} - p_{h's}^{(k+1)}| \leq H \max_{j,j'} |p_{js}^{(k)} - p_{j's}^{(k)}|.$$

Thus, using (5.4), we see that as $k \to \infty$, $p_{js}^{(k)}$ approaches a limit independent of j. For reasons such as this, H can be thought of as an *ergodicity coefficient*, and is denoted $\tau_1(P)$ in Seneta (1981). We need to use this alternative notation for the next section.

6 The Ergodicity Coefficient

In his commentary Sapogov notes (Markov, 1951, pp. 662) that for $\tau_1(P)(\equiv H) < 1$ it is sufficient that P have just one column of strictly positive entries. Indeed from (5.3b), if $\max_j \min_i p_{ij} \geq \lambda > 0$, then $\tau_1(P) \leq 1 - \lambda < 1$. Such stochastic matrices, sometimes called Markov matrices, became the basis of Bernstein's studies of homogeneous Markov chains, where $P_{n+1}, n \geq 0$, denotes the matrix of transition probabilities $Pr\{X_{n+1} = j | X_n = i\}, i, j \varepsilon S$ (assume constant finite state space S). Of particular interest here is a WLLN for such chains of Bernstein in 1946, following on the work of Markov (1910). This paper of Markov deals with the case where $S = \{0, 1\}$, where the entries in the P_n's are each uniformly bounded from 0 and 1, again by methods specific to that case and similar to Markov (1906, §2). Bernstein's treatment (see Bernstein, 1964, pp. 455–460) of the WLLN for the general case repeats some of the logical steps of Markov (1906, §5) leading to an expression of form (5.2), and then invokes the assumption that all the transition matrices are uniformly Markov (that is, the same $\lambda > 0$ may be used for each P_n) to obtain in essence (5.5) and (5.1) with $(1 - \lambda)$ in place of H.

The derivation in the setting of inhomogeneous chains and in terms of $\tau_1(P_n)$'s, assuming $\tau_1(P_n) < \beta < 1$ (where β is independent of n), in the manner of (5.5) and (5.1) is, finally, contained in Hartfiel & Seneta (1994, Section 4). This derivation is heavily dependent on the fact that

$$\tau_1(P) = \sup\{\delta^T, \|\delta^T\|_1 = 1, \delta^T \mathbf{1} = 0 : \|\delta^T P\|_1\} \tag{6.1}$$

which is apparently due to Dobrushin in 1956.

The equivalence of the forms (5.3a) and (6.1) is a non-trivial exercise, and it is not clear to the author how (even in the general homogeneous case) Markov could have derived (5.5), especially in the absence of stationarity. The apparent disappearance of H in the general case from his writings after its only appearance in a "provincial journal" (Ondar, 1981, Letter No. 4), and the fact that Bernstein does not incorporate it into his proof, suggest an inspired guess on Markov's part, by analogy with the stationary dependent trials case.

Dobrushin's (1956) use of the ergodicity coefficient τ_1 was in the classical setting of the CLT for non-homogeneous Markov chains. Its use in connection with problems of ergodicity was popularized at about the same time by Hajnal & Sarymsakov (see Iosifescu, 1980; Seneta, 1981).

A stochastic matrix P with the crucial property $\tau_1(P) < 1$ is called "scrambling". From (5.3b) we see that a stochastic matrix is scrambling iff any two rows have at least one position in which positive entries occur. This is a richer and more convenient class to work with than positive or Markov matrices. It is clear that τ_1 as a tool may be used also for stochastic matrices with a countable infinite index set S.

For completeness as regards the WLLN, we remark that the uniformly scrambling condition $\tau_1(P_n) < \beta < 1, n \geq 1$, above, may be replaced by the condition $n\{1 - \tau_1(P_n)\} \to \infty$ an $n \to \infty$, adapted from the 1918 analysis of the two-state case by Bernstein (1964, pp. 61–65). For that setting, Bernstein shows that this condition is close to necessary for the WLLN.

To bring the role of $H = \tau_1(P)$ in ergodicity glancingly right up to the present, we first note the inequality:

$$\|\pi_0^T P^k - \pi^T\|_1 \leq \|\pi_0^T - \pi^T\|_1 \tau_1^k(P) \tag{6.2}$$

if P is the transition matrix of a finite Markov chain whose states contain a single closed aperiodic class, π^T is the unique limiting-stationary distribution vector, and π_0^T is any (initial) distribution vector. The inequality (6.2) may be obtained in a few steps from (6.1) (see e.g. the proof of Theorem 2.1 in Seneta, 1993; and Tan, 1983). Thus (6.2) actually measures the variation distance between $\pi_0^T P^k$ and its limit π^T in terms of the initial variation distance between $\pi_0^T = \{\pi_i^0\}$ and $\pi^T = \{\pi_i\}$, and of the maximal variation distance between the rows of P as measured by $\tau_1(P)$. If P is scrambling (6.2) provides an explicitly calculable geometric convergence rate $\tau_1(P)$, and thus provides an ultimate form (albeit implicit) to Markov's ergodicity approach for finite homogeneous chains.

For the subcase of finite irreducible aperiodic P (so $\pi_i > 0$, each $i \varepsilon S$) Fill (1991, Theorem 2.7) gives the bound, obtained in rather more involved fashion:

$$\|\pi_0^T P^k - \pi^T\|_1 \leq \left\{\sum_i (\pi_i^0 - \pi_i)^2/\pi_i\right\}^{1/2} \{\beta_1^{1/2}(M)\}^k \tag{6.3}$$

where $\beta_1(M)$ is the second-largest eigenvalue of the "reversibilized" matrix $M = P\tilde{P}$ where $\tilde{P} = \{\tilde{p}_{ij}\}$ and $\tilde{p}_{ij} = \pi_j p_{ji}/\pi_i$. We note with Fill that $\beta_1(M)$, being an eigenvalue, is not in general explicitly available, and bounds for it (involving, for example, the Poincaré constant or Cheeger constant) must be employed, thus weakening (6.3).

There appears to be little to choose in fact even between (6.2) and (6.3) as it stands for irreducible aperiodic P, since matrices P may be found giving $\tau_1(P) < \beta_1^{1/2}(M) < 1$. An example (taken from Tan, 1983) of such a P, with associated quantities π^T, $\tau_1(P)$ and $\beta_1(M)$ is

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 5/16 & 5/16 & 3/16 & 3/16 \\ 5/16 & 5/16 & 3/16 & 3/16 \\ 5/16 & 5/16 & 3/16 & 3/16 \end{bmatrix}$$

$\pi^T = (5/13,\ 5/13,\ 3/26,\ 3/26)$

Eigenvalues of P, $\tilde{P}$, and $M = P\tilde{P}$ are: 1, 3/16, 0, 0.

Thus $\beta_1^{1/2}(M) = \sqrt{3/16} = 0.4330, \quad \tau_1(P) = 3/8 = 0.375.$

It may not be inappropriate, therefore to call the constant $H = \tau_1(P)$, the Markov–Dobrushin constant because of its remarkable and pervasive properties. It is remarkable that this constant is already central in Markov's first paper on chain dependence theory.

References

Bernstein, S.N. (1964). *Sobranie Sochinenii: Tom IV [Collected Works, Vol. IV], Teoriia Veroiatnostei i Matematicheskaia Statistika [1911–1946]* Izd. Nauka, Moscow.

Bruns, H. (1906). *Wahrscheinlichkeitsrechnung und Kollektivmasslehre.* Teubner, Leipzig. [See esp. Achtzehnte Vorlesung].

Chuprov, A.A. (1959). *Ocherki po Teorii Statistiki.* [Essays in the Theory of Statistics]. Moscow. [Reprinting of 1910 edition; 1st edition: 1909].

Dobrushin, R.L. (1956). Central limit theorem for non-stationary Markov chains, I and II. *Theory Prob. Appl.* **1**. 65–80, 329–83 [English translation].

Fill, J.A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Applied Prob.* **1**, 62–87.

Florensky, P.A. (1990). *Stolp i Utverzhdenie Istini [The Pillar and Foundation of Truth.]*, Vol. 1, Part 1. Pravda, Moscow. [Facsimile of 1914 edition].

Ford, C.E. (1991). Dmitrii Egorov: Mathematics and religion in Moscow. *Mathematical Intelligencer*, **13**, 24–30.

Grodzensky, S.Ia. (1987). *Andrei Andreevich Markov 1856–1922*. Nauka, Moscow.

Hartfiel, D.J. & Seneta, E. (1994). On the theory of Markov set-chains. *Adv. Appl. Prob.* **26**, 947–964.

Iosifescu, M. (1980). *Finite Markov Processes and Their Applications*. Wiley, Chichester; Editura Tehnica, Bucharest.

Khoruzy, S.S. (1990). On the philosophy of the priest Pavel Florensky [in Russian] pp. VI–XVI of Florensky (1990).

Lyusternik, L.A. (1967). The early years of the Moscow Mathematical School. *Russian Mathematical Surveys*, **22**, No. 2, 171–211. [Part 2 of this article. English translation].

Markov, A.A. (1906). Extension of the law of large numbers to dependent quantities [in Russian]. *Izv. Fiz.-Matem. Obsch. Kazan Univ.*, (2nd Ser.) **15**, 135–156 [Also in Markov (1951), pp. 339–61].

Markov, A.A. (1910). Investigation of the general case of trials associated into a chain [in Russian]. *Zapiski. Akad. Nauk (St. Petersburg)* Fiz.-Matem. Otd., (8th ser.) **25**, No. 3, 33 pp. [Also in Markov (1951), pp. 465–507].

Markov, A.A. (1951). *Izbrannie Trudy [Selected Works]*. A.N.S.S.S.R., Leningrad.

Nekrasov, P.A. (1898). General properties of numerous independent events in connection with approximative calculation of functions of very large numbers [in Russian]. *Mat. Sbornik* **20**, 431–442.

Nekrasov, P.A. (1902). *Filosofia i Logika Nauki o Massovikh Proiavleniakh Chelovecheskoi Deiatelnosti. [The Philosophy and Logic of the Science of Mass Phenomena in Human Activity]*. Moscow.

Ondar, Kh.O. (1981). *The Correspondence Between A.A. Markov and A.A. Chuprov on the Theory of Probability and Mathematical Statistics*. Springer, New York. [Translated by C. & M. Stein].

Romanovsky, V.I. (1949). *Diskretnie Tsepi Markova [Discrete Markov Chains]*. Moscow. [English translation by E. Seneta, published in 1970 by Wolters-Noordhoff, Groningen].

Seneta, E. (1981). *Non-negative Matrices and Markov Chains*. (2nd Edn.). Springer, New York.

Seneta, E. (1984). The central limit problem and linear least squares in pre-revolutionary Russia: The background. *Mathematical Scientist*, **9**, 37–77.

Seneta, E. (1993). Applications of ergodicity coefficients to homogeneous Markov chains. *Contemporary Mathematics*, **149**, 189–199.

Sheynin, O.B. (1989a). A.A. Markov's work on probability. *Archive Hist. Exact Sci.*, **39**, 337–77.

Sheynin, O.B. (1989b). *Materiali o A.A. Markove v Arkhive A.A. Chuprova. [Materials on A.A. Markov in the Archive of A.A. Chuprov.]* Institut Istorii Estestvoznania i Tekhniki ANSSSR, Moscow. Preprint No. 26, 1989.

Sluginov, S.P. (1927). P.A. Nekrasov [in Russian]. *Trudy Matem. Seminariia, Gosud. Universitet, Perm.* **1**, 37–38.

Stigler, S.M. (1975). The transition from point to distribution estimation. *Bull. International Statistical Institute*, **46**, Bk. 2, 332–340.

Tan, C.P. (1983). Coefficients of ergodicity with respect to vector norms. *J. Appl. Prob.* **20**, 277–287.

Résumé

L'oeuvre de Markov sur la dépendance en chaîne eut pour motivation son désir de réfuter l'affirmation de Nekrasov que l'indépendance en paires de variables aléatoires dans une somme était une condition nécessaire de la Loi Faible des Grands Nombres. Markov acheva son résultat en obtenant une telle Loi en 1906 pour un système de variables aléatoires dépendantes, et en particulier pour des chaînes de Markov finies homogènes. L'affirmation incorrecte de Nekrasov était basée sur la doctrine théologique de libre arbitre, que certains membres de l'Ecole de Mathématiques de Moscou entretenaient à cette époque.

La première partie de l'article présente l'arrière-plan de la situation. La deuxième traite des techniques plutôt négligées de l'article de Markov en 1906, et surtout de son emploi de ce qu'on a par la suite nommé le coefficient d'ergodicité. Ce coefficient exprime l'effet contractoire que l'application d'une matrice stochastique a sur un vecteur colonne; il est à la base de ses arguments d'ergodicité et de sa preuve de la Loi Faible des Grands Nombres.

[Received March 1996, accepted July 1996]