As an example of a nice computational project, suppose that upon reaching the level $\mathbf{0}$, say in the state $(0, j)$, the Markov chain is moved after one unit of time to the state $(N, j)$ with $N \geq 1$, and is allowed to continue. We wish to determine the smallest value of $N$ which guarantees that the stationary version of the Markov chain spends a fraction of time at least $\theta$ above the level $K$. The quantities $\theta$, $0 < \theta < 1$, and $K \geq 0$, are given.

In this problem, the matrices $B_\nu$, $\nu \geq 0$, are given by $B_N = I$, $B_\nu = 0$, for $\nu \neq N$, and we wish to determine the smallest value of $N$ for which

$$\sum_{\nu=K+1}^{\infty} \mathbf{x}_\nu \mathbf{e} \geq \theta.$$

By implementing the algorithm we have just described and by planning its steps with some care, the desired value of $N$ may be efficiently computed. More involved problems of this type are common in applications, such as in the design of message storage buffers in communications engineering or in inventory models.

It is also possible to give a formally different analysis of the stationary probability $\mathbf{x}$ of $\hat{P}$, which is based on the matrix $G^*(z)$ of Section 2. That approach is more involved both in its mathematical analysis and its algorithmic implementation and we shall not present it here. The matrix $G^*(z)$ and various associated first-passage-time distributions play an essential role, however, in the study of Markov chains of the type $P_2$.

## References


[1] KEILSON, J. AND WISHART, D. M. G. (1964) A central limit theorem for processes defined on a finite Markov chain. *Proc. Camb. Phil. Soc.* **60**, 547–567.

[2] KEMENY, J. AND SNELL, J. L. (1960) *Finite Markov Chains.* Van Nostrand, Princeton, NJ.

[3] NEUTS, M. F. (1976) Moment formulas for the Markov renewal branching process. *Adv. Appl. Prob.* **8**, 690–711.

[4] NEUTS, M. F. (1977) Some explicit formulas for the steady-state behavior of the queue with semi-Markovian service times. *Adv. Appl. Prob.* **9**, 141–157.

[5] NEUTS, M. F. (1981) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* The Johns Hopkins University Press, Baltimore, MD.


*Math. Scientist*, 1984, **9**, 37–77

# The Central Limit Problem and Linear Least Squares in Pre-Revolutionary Russia: The Background


E. Seneta

*University of Sydney*


(Received: 11 November 1983)


**Abstract**

The primary purpose of this article is to give some insight into work in probability other than that of the St. Petersburg School within the Russian Empire, specifically on the central limit problem, and the linear model framework with which it was associated. It focusses in particular on the contributions and personalities of I. V. Sleshinsky (who completed a fully rigorous proof of the central limit theorem before Markov) and P. A. Nekrasov who used the method of saddlepoint, of Laplacian peaks, and of the Lagrange inversion formula to establish for sums of lattice variables, the standard local and global limit theorems of central limit type for large deviations; and their relations with Chebyshev, Markov and Liapunov. The controversy between Markov and Liapunov on one hand, and Nekrasov on the other, led Markov to ideas resulting in the appearance of Markov chains.


## 1. Introduction

The commonly accepted beginning of rigorous proofs of the central limit theorem is taken to be a paper of Chebyshev (Tchébichef (1887)) in which (§1) he asserts for independently but not necessarily identically distributed random variables $U_1, U_2, \ldots$, each described by a density, that if

(i) $EU_i = 0$, each $i$;
(ii) $|E\{U_i^k\}| \leq C$ for all $i$, and all integers $k \geq 2$,

where $C$ is a constant independent of $i$ and $k$; then as $n \to \infty$

$$\Pr[t < S_n/B_n < t'] \to (2\pi)^{-1/2} \int_t^{t'} \exp(-x^2/2)dx \qquad (1.1)$$

where

$$S_n = \sum_{i=1}^{n} U_i, \qquad B_n^2 = \sum_{i=1}^{n} \operatorname{Var} U_i.$$

(We have changed Chebyshev's notation and statement to accord with more modern usage.) The technique used in the proof, which is incomplete, is the method of moments, which dates back to Chebyshev's interaction with Bienaymé.

Standard accounts of subsequent early history focus on two of Chebyshev's illustrious students within the 'St. Petersburg School', A. A. Markov and A. M. Liapunov. In published letters to A. V. Vasiliev, Markov (1898a) states that a further condition needs to be added to make the statement of the theorem correct, and suggests that this can be taken to be:

(iii) $B_n^2/n$ is uniformly bounded away from 0.

In the very last part of his correction (Markoff (1898b)) of Chebyshev's theorem, he replaces (iii) by:

(iiia) $EU_n^2$ is bounded from 0 as $n \to \infty$.

Finally, in 1900–1901, Liapunov proves it in the version known as Liapunov's theorem, whose final version (Liapounoff (1901a), §2) states (for $U_i$ discrete rather than continuous) that (1.1) obtains, with convergence uniform with respect to $t$ and $t'$, if

(a) $EU_i = 0$ all $i$;
(b) there exists a $\delta > 0$ such that

$$\left(\sum_{i=1}^{n} E|U_i|^{2+\delta}\right)^2 \Big/ \left(\sum_{i=1}^{n} E(U_i^2)\right)^{2+\delta} \to 0$$

as $n \to \infty$. The proof is carried out by characteristic-function methods and what are now known as Liapunov's inequalities. In §11, Liapunov alludes to Chebyshev's and Markov's work, in relation to deducing as a consequence of his result that his conclusion holds if

(i)′ $EU_i = 0$;
(ii)′ $E|U_i|^k \le C$ for all $i$ and positive integer $k \ge 2$ where $C$ is a constant; and
(iii)′ there exists a $\beta$, $0 < \beta < 1$, such that, as $n \to \infty$

$$n^\beta / B_n^2 \to 0.$$

The relation of (i)′–(iii)′ to (i)–(iii) is evident.

There is often allusion to later work of Markov (1913) who reproves Liapunov's result using the method of moments and truncated variables, but there is rarely in the literature even passing mention of Russian contributions to the central limit problem other than of these three members of the St. Petersburg School (see e.g. Adams (1974), who devotes his last two chapters, 7 and 8, to this work; Maistrov (1967), Chapter 4; Gnedenko and Sheynin (1978)).

Yet, the central limit problem in pre-revolutionary Russia does not have such a monolithic structure, and it is the intention of this essay to give some insight into work and workers other than that of the St. Petersburg School. Indications of such are not explicitly given by Markov in the above sequence of works, but in the first of his two major papers, Liapunov (Liapounoff (1900)), speaking of resolution of the central limit problem based on the notion of the 'discontinuity factor' (which is essentially the characteristic function approach), says†

> 'I was at first unaware that an attempt in this direction had been made by Sleshinsky. Nevertheless, when I subsequently acquainted myself with it, I saw that new investigations would not be superfluous, since Sleshinsky, in using the ideas of Cauchy, made assumptions which are too restrictive, and his analysis does not appear to extend to more general situations.'

† All quotations from the Russian are given in free English translation.



The reference is to Sleshinsky (1892). Another English rendering of this passage is by Adams (1974), pp. 86–87; he transliterates Sleshinsky as 'Slechinskii', who is not otherwise mentioned (nor do the relevant papers of Cauchy attract comment). A few paragraphs further on, Liapunov (Liapounoff (1900)), says

> 'Yet I should mention that the question under investigation was also the subject of investigations of Nekrasov, who recently published a note containing a brief exposition of results obtained by him on it. Insofar as one may judge by the note, the assumptions made by Nekrasov are essentially different from mine. In regard to his methodology, this is still unknown, since Nekrasov has not published his analysis. In any case it is permissible to suppose that this methodology is far from elementary, for, to quote Nekrasov, it is based on general investigations relating to the Lagrange series, which were published by him earlier.'

The (implicit) reference is to Nekrasov (1898a). There is another allusion to Nekrasov (in a footnote given as being to 'Novie osnovania . . .', 1901, *Mat. Sbornik* **21–22**) in Liapounoff (1901a) at the end of §11—after the allusion mentioned above to the conditions of Chebyshev and Markov:

> 'Nekrasov, in his work on probabilistic sums which he has recently published, claims that Markov's supplementary condition [i.e. (iii) above] may be replaced by one even more general, which requires only that the quantity $B_n^2$ should go to infinity as $n \to \infty$.
>
> As we shall show in the next section, in some cases this condition is indeed sufficient. But in general it is not sufficient, as may be seen from examples, as was already noted by Markov.'

There is no explicit allusion to an article of Markov here, and the existence of such published examples due to him is doubtful. The only examples in the central limit connection published by Markov occur in the Markov–Vasiliev correspondence (Markov (1898a)) and in his (1899b) reply to Nekrasov. All these examples concern sequences of independent random variables on $[-1, 1]$, presumably to satisfy condition (ii); and for uniformly bounded random variables we now know it to be true that $B_n^2 \to \infty$ is necessary and sufficient for the central limit theorem to hold.

Indeed, assuming conditions (i)′ and (ii)′, the additional assumption $B_n^2 \to \infty$ *is* sufficient for the central limit theorem to hold so in essence an example of this kind alluded to by Liapunov as due to Markov could not have been constructed. To see this, note that for any integer $m \ge 0$, and fixed $\eta > 0$, we may choose $n_0(\eta)$ so that $\eta B_n \ge 2$ for $n \ge n_0(\eta)$ since $B_n \to \infty$ in which case using (i)′ and (ii)′ for all $i \ge 1$,

$$C \ge \int_{|y| > \eta B_n} |y|^{2+m} dP[U_i \le y] \ge (\eta B_n)^m \int_{|y| > \eta B_n} y^2 dP[U_i \le y]$$

so that

$$\sum_{i=1}^{n} \frac{1}{B_n^2} \int_{|y| > \eta B_n} y^2 dP[U_i \le y] \le \frac{Cn}{(\eta B_n)^m B_n^2} \le \frac{Cn}{2^m B_n^2}$$

and it is clearly possible to determine $m \equiv m(n)$ so that the right-hand side approaches 0 as $n \to \infty$, so that the Lindeberg condition is satisfied and the central limit theorem holds. This reasoning depends heavily on the bound $C$ in (ii)′ being uniform in $k$. Liapunov's reasoning, if not explicit statement, permits this constant to depend on $k$; it has been claimed, beginning with Liapunov (1901b), p. 57, that Chebyshev's condition (ii) might be so interpreted, but it was certainly not so interpreted by Markov.

There is an obviously understandable temptation to dismiss the apparently minor contributions of Sleshinsky and Nekrasov engendered by a natural desire to obtain a simplistic picture, and this is in fact what has occurred. The attitude is reinforced by the following information concerning P. A. Nekrasov given in a biography of A. A. Markov by his son (also A. A. Markov) in Markov (1951), pp. 610–611†:

> 'My father paid very great attention to the method of teaching mathematics in high school. He protested energetically against various harmful experiments in this area. In particular, such experiments were attempted by a professor of Moscow university, P. A. Nekrasov, a member of the Black Hundred and mystic, who sought to make out of mathematics a bulwark for Orthodox Christianity and autocracy. In 1915, Nekrasov, associated with the administration of the Ministry of National Education, and formerly in charge of an educational area, proposed with P. S. Florov a scheme for the introduction of probability theory into the high school curriculum. In essence this scheme amounted to inculcation into the minds of students the confused pseudo-scientific views of its authors as regards probability theory, mathematical statistics and mathematics in general.'

The scheme came to nothing on the initiative of A. A. Markov. A similar characterization of Nekrasov, in generally greater detail, is given by Maistrov (1967) in Chapter V, §1. We have for example:

> 'In his numerous works Nekrasov adopted idealistic positions.'
>
> 'Speaking of social problems, Nekrasov sharply opposes political changes in which the masses participate. He considers private property a prime principle, which it is the czarist regime's province to protect.'

Presumably, one must understand words such as 'mystic' and 'idealistic' according to their usage in political terminology. Maistrov mentions a number of books of Nekrasov, purportedly of sociological, philosophical and pedagogical nature, expressing reactionary views, as well as a probability text book based on his lectures at Moscow university (Nekrasov (1896)) where he had been professor. We shall not pursue any of these publications further except to note, with Maistrov, a strong attack in 1916 on Markov by Nekrasov in apparent retaliation for Markov's opposition to the high-school project. We shall see in the sequel that animosity between the two had much earlier, technical rather than ideological, origins which have been forgotten.

† The Black Hundred was a name applied by their adversaries to extreme right-wing elements in Russia of the time. They supported anti-Semitism, absolutism, and nationalism, and carried out pogroms against Jews and students.



It is well known that Markov's ('progressive materialistic') ideologies were acceptable, and, indeed, welcome in post-revolutionary Russia; it is clear that Nekrasov's reactionary ones were not, ostensibly with good reason. Yet it is curious that Sluginov's (1927) obituary† is very positive.

Although I. V. Sleshinsky was ethnically Polish, his published work in probability was done in the city of Odessa where he studied and taught at the university up to his retirement in 1909. He subsequently moved to Kraków. We have been able to trace a good obituary in Polish of Sleshinsky by Hoborski (1931), and this has been supplemented by personal information from various sources, especially from his Kraków period. In addition, the article of Leibman (1961) contains much material on Sleshinsky during his Odessa period (on p. 400, there is a photograph of him), and on his colleagues there, within the framework of a general biographical and evolutionary account of their academic activity, to which we shall allude in the sequel.

We include as appendices, synthesized biographical sketches of both Nekrasov and Sleshinsky, to aid in the understanding of their roles, and their eclipse. Neither had any lasting effect on the subsequent development of the central limit problem for reasons which do not relate to the significance of their work. This significance has been briefly mentioned elsewhere (Seneta (1979)), and we shall develop it at greater length below, but it is in place to summarize it here. Nekrasov in his 1898a paper attempted to use what we now know as the method of saddlepoint, of Laplacian peaks, and of the Lagrange inversion formula to establish, for sums of lattice variables, the standard local and global limit theorems of central limit type for large deviations. The attempt was many years ahead of its time. The controversy between Markov, Liapunov and Nekrasov, led Markov to ideas resulting in the appearance of 'Markov chains' in his work; and certainly led to a significant interaction between Markov and A. A. Chuprov (or Tschuprow), which in turn shaped the significant role that Chuprov was to play in the development of statistical theory.

Sleshinsky's (1892) work on the central limit theorem already manifests the notion of 'triangular arrays' in terms of which general central-limit-type results have latterly been formulated. To Cauchy and Sleshinsky (see Heyde and Seneta (1977), §§4.6–4.7) we may attribute the first rigorous proof of the central limit theorem (albeit under restrictive conditions), and the first really successful use of characteristic functions in this connection. With its rigour, and its difficult and involved estimates, after earlier authors, the paper of 1892 is modern probability theory, in anticipation of Markov and Liapunov.

Although we shall devote most attention to Nekrasov and Sleshinsky, this work arose out of a broader desire: to gain some insight into work in probability other than that of the St. Petersburg School within the Russian Empire, in consequence of Chapter 4 (on linear least squares) of the book of C. C. Heyde and E. Seneta (1977) with which the reader will find a little overlap. Linear least squares, in its probabilistic aspect, has had a long historic connection with the central limit problem; we shall develop this theme beginning in the next section, and use least squares as a framework for the general character of this account. Significant work in our general subject area was being done at centres other than St. Petersburg in the empire, mainly under the

† See Appendix 1.

indirect stimulus of Chebyshev (1821–1894), from about 1890. We may consider, as examples, Moscow (P. A. Nekrasov); the three Ukrainian† cities of Odessa (I. V. Sleshinsky, S. P. Yaroshenko), Kharkov (V. G. Imshenetsky, M. A. Tikhomandritsky, and later Liapunov), and Kiev (V. P. Ermakov); Kazan (A. V. Vasiliev); and the Polish city of Warsaw (P. S. Nazimov, W. Gosiewski). The account presented here would doubtless have been more complete with access to materials which a more appropriately-based worker might have had (although political considerations may have constituted a hindrance); yet it is hoped that it will make more complete, in regard to language, period and location, even the recent valuable survey of least squares by Harter (1974–1976). In his list of 408 items, Merriman (1877–1882) lists 16 emanating from Russia, but includes no Russian-language sources.

## 2. Linear least squares

The basic problem of the classical linear model is to estimate an $r \times 1$ vector, $\boldsymbol{\beta}$ of unknowns from a number $N$ of observations $\boldsymbol{Y}$, related linearly to $\boldsymbol{\beta}$ but subject to error $\boldsymbol{\varepsilon} = \{\varepsilon_i\}$:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.1}$$

Here $X = \{x_{ij}\}$ is a known fixed $N \times r$ matrix with $N \geq r$, which we shall assume to be of full column rank $r$. In the nineteenth century this problem was generally viewed as one of finding an $r \times N$ constant matrix $K = \{k_{ij}\}$ (or, in the usage of that time, a system of 'multipliers' $k_{ij}$) such that

$$KX = I \tag{2.2}$$

where $I$ is the unit matrix. Then $\boldsymbol{\beta}$ is, accordingly, estimated by $\bar{\boldsymbol{\beta}} = \{\bar{\beta}_i\}$

$$\bar{\boldsymbol{\beta}} = K\boldsymbol{Y} = \boldsymbol{\beta} + K\boldsymbol{\varepsilon} \tag{2.3}$$

the matrix $K$ being chosen (under the constraint (2.2)) in some optimal manner. Such a procedure is linear, insofar as according to (2.3), the elements of $\boldsymbol{\beta}$ are each estimated by a fixed linear combination of the elements of $\boldsymbol{Y}$, irrespective of the value of $\boldsymbol{Y}$.

The situation may be regarded from two viewpoints. The first is non-statistical and regards it as a problem of *interpolation*, associated with the overdetermined set of linear equations $\boldsymbol{Y} = X\boldsymbol{\beta}$. In this case $K$ is determined in accordance with a direct requirement on $\boldsymbol{\varepsilon}$ itself. Legendre in 1805 determined $K$ according to the criterion that the sum of squares of errors, $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$, be minimized: this leads to the 'least-squares' estimate of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{Y}, \tag{2.4}$$

and corresponding 'least-squares' choice of $K$ (which clearly satisfies (2.2)):

$$K = (X'X)^{-1}X'. \tag{2.5}$$

† See Gnedenko and Gikhman's (1956) account of the history of Ukrainian probability; for the Odessa scene in particular. Curiously, in a general mathematical setting, Gnedenko (1965) does not mention Sleshinsky with his Odessa colleagues Yaroshenko, S. O. Shatunovsky, V. F. Kagan and I. Yu. Timchenko. Ermakov, Imshenetsky and Tikhomandritsky are mentioned.



The second viewpoint is *probabilistic* (statistical), $K$ being chosen (under constraint (2.2)) optimally in accordance with some *distributional requirement* on $\boldsymbol{\varepsilon}$. Gauss's contributions to this problem are essentially well known, and a historical discussion of them (see e.g. Plackett (1949), (1972), Seal (1967), Maistrov (1967), van der Waerden (1977), Heyde and Seneta (1977)) is not within the scope of the present work. Suffice it to say, for purposes of convenient reference, that in his second probabilistic justification of least squares, Gauss in 1821 continued to assume that the $\varepsilon_i$, $i = 1, \ldots, N$ are i.i.d. with common mean 0 and common variance $\sigma^2$ finite. He then showed that among the class of estimates of the elements of $\boldsymbol{\beta}$ of the kind (2.3), satisfying (2.2), those provided by (2.4) have minimal variance. Since it is readily seen that $E(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta})(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta})' = \sigma^2 KK'$, it follows that the choice (2.5) minimizes $\sum_{h=1}^{N} k_{ih}^2$, for each $i = 1, \ldots, r$.

The long historic connection between linear least squares, in its probabilistic aspect, and the central limit problem arises as follows. Laplace (1812), seeking to justify the least-squares choice of $K$ on *large sample* grounds (Gauss's justifications had been for fixed sample size $N$), assumed (much like Gauss) the $\varepsilon_i$'s to be i.i.d. with unspecified symmetric density confined to a finite interval, and considered a general $K$ satisfying (2.2). He showed (at least in the case $r = 1$) that, in effect and under unspecified further conditions on $K$, the standardized random variable

$$\frac{\bar{\beta}_i - \beta_i}{\sqrt{\operatorname{Var} \bar{\beta}_i}} = \sum_{h=1}^{N} k_{ih}\varepsilon_h \Bigg/ \left\{\sigma^2 \sum_{h=1}^{N} k_{ih}^2\right\}^{1/2} \tag{2.6}$$

has, approximately, as $N \to \infty$, an N(0, 1) distribution, where $\sigma^2 = \operatorname{Var} \varepsilon_i$. (He then argued that, for large $N$, the optimal choice of $K$ would be that minimizing the denominator for every $i$). The first part is clearly an early version of the central limit theorem for independent, but not identically distributed, random variables $\{k_{ih}\varepsilon_h\}$, $h = 1, 2, \ldots$. Since the first general (i.e. for general random variables), if non-rigorous, versions of the central limit theorem are usually associated with Laplace, the connection of the central limit problem with linear least squares clearly has its origins at this point.

This connection persisted strongly throughout the nineteenth century within the subsequent dominant streams of probabilistic evolution, first the French and then the Russian; we shall in fact have much more to say about aspects of the central limit problem than about least squares.

Apart from the early contributions of Euler and Daniel Bernoulli associated with the beginnings of the St. Petersburg Academy, probability within the Russian empire began to develop only in the 1820's. From this point, Russian historians characteristically consider its pre-revolutionary history in two stages, that preceding the St. Petersburg School, and that of this school itself. To the earlier period notably belong B. Ya. Buniakovsky and M. V. Ostrogradsky in St. Petersburg, and A. Yu. Davydov in Moscow, with contributions made by a number of others. This early period is described in some detail in the book of Maistrov (1967), Chapter III, §12, and we shall not dwell on it, except to note for completeness the contributions generally relevant to our subject matter by Buniakovsky (1846), Savich (1857) and Zinger (1862) on least squares. According to Chebotarev (1961), Savich's was the first Russian-language text on least squares. The era of the St. Petersburg School may be taken to begin in about 1860 when Chebyshev began to teach probability at the St. Petersburg university on the departure of Buniakovsky, although he had done some work on probability before this time. It must be mentioned that in the general Russian tradition Chebyshev, Markov and Liapunov all made many contributions to branches of mathematics other than probability and therefore cannot exclusively be described as probabilists; nor did the St. Petersburg School, founded by Chebyshev, consist only of them, though in regard to probability, they were the outstanding figures.

We now pass onto a study of the later of the two pre-revolutionary periods.

## 3. Interpolational aspects of least squares

The interpolational aspect of least squares devolves to methods of solution of the system of *normal equations*

$$(X'X)\hat{\boldsymbol{\beta}} = X'Y \tag{3.1}$$

for the least-squares estimator of (2.4)—the solution is unique in view of the assumed linear independence of the columns of $X$, which renders $X'X$ non-singular.

Gauss (circa 1809) gave an algorithm for solving (3.1) which has become the standard method for solving any set of linear equations

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{b} \neq \boldsymbol{0} \tag{3.2}$$

for $\boldsymbol{x}$ where $A = \{a_{ij}\}$ is $r \times r$ and non-singular and $A$ and $\boldsymbol{b}$ are known; namely the method of successive elimination of unknowns. This Gaussian elimination process is a *direct* solution procedure, that is, one in which the solution is attained in a finite number of operations. In general Gaussian elimination for solution of (3.2) will require rearrangement of equation order to avoid a pivot becoming 0 on the triangularization, but in the special case $A = X'X$ of (3.1) this will not occur (e.g. Wendroff (1966), pp. 124–125) since this matrix is symmetric and positive definite.

Opposed to *direct* solution methods are iterative procedures. Stationary linear iterative methods generate a sequence of approximative vectors $\{\boldsymbol{x}_k\}$ by a scheme of the form

$$\boldsymbol{x}_{k+1} = B\boldsymbol{x}_k + \boldsymbol{c}, \qquad k \geq 0, \tag{3.3}$$

where $B$ and $\boldsymbol{c}$ are specified, independent of $k$, and related to $A$ and $\boldsymbol{b}$; $\boldsymbol{x}_0$ is a starting vector. One such procedure is commonly known as the point-Gauss–Seidel procedure after related work by Gauss in 1823 and Seidel in 1874 (see Bodewig (1956), p. 126 *et seq.*; Ostrowski (1955)). This defines $B$ and $\boldsymbol{c}$ by writing $A = L + U$ where $L$ has the same elements on the diagonal and below as $A$, zeroes elsewhere. Then (providing $A$ has no zeroes on the diagonal),

$$B = -L^{-1}U, \qquad \boldsymbol{c} = L^{-1}\boldsymbol{b}. \tag{3.4}$$

The expression (3.3), for practical application, is then rewritten

$$L\boldsymbol{x}_{k+1} + U\boldsymbol{x}_k = \boldsymbol{b}. \tag{3.5}$$

This reveals that the procedure has a *cyclic* character, inasmuch as the approximating



vector is modified one element at a time, in strict order, a single cycle constituting modification of all elements: that is, replacement of $\boldsymbol{x}_k$ by $\boldsymbol{x}_{k+1}$. From (3.3) and (3.4) it is clear that the sequence $\{\boldsymbol{x}_k\}$ converges for the process defined by (3.5) if $\rho(-L^{-1}U) < 1$ where $\rho(-L^{-1}U)$ is the spectral radius of $-L^{-1}U$; that this spectral radius is then the convergence rate; and, from (3.5), that the limit vector is the unique solution of (3.2). It is also clear that if $\rho(-L^{-1}U) \geq 1$, the sequence $\{\boldsymbol{x}_k\}$ generally diverges.

The original context of application of Gauss–Seidel-type iterative procedures was that of the system (3.1). {This context is thus concerned essentially with the situation of a *symmetric positive–definite* coefficient matrix $A$ in (3.1), since any such matrix may be written in the form $A = X'X$.} The predicted readings according to the least squares fit are given by $\hat{\boldsymbol{Y}} = X\hat{\boldsymbol{\beta}}$ so the residual sum of squares

$$(\hat{\boldsymbol{Y}} - \boldsymbol{Y})'(\hat{\boldsymbol{Y}} - \boldsymbol{Y}) \tag{3.6}$$

may be used as a measure of goodness of fit. Indeed, for any approximation to $\hat{\boldsymbol{\beta}}$, a predicted set of readings and a corresponding residual sum of squares can be calculated, and it was the behaviour of this with successive approximations which was used as a criterion for the convergence of such a process of approximations. Indeed, if the residual sum of squares was non-increasing, and presumably approaching the limit (3.6), then the process was deemed to converge. Such processes $\{\boldsymbol{x}_k\}$ where some object function (the 'entropy') is progressively reduced are called 'relaxation processes'. In the context of (3.1) such a relaxation process had already been studied by Seidel (1874), where, in contrast to the point-Gauss–Seidel method as described above, at each stage the choice of next element of the present vector for modification is not already determined, but optimal. We shall return to this point shortly. Such optimal-choice variants of relaxation methods are sometimes described as 'free-steering' (Ostrowski (1955)).

Nekrasov (1885b) gives as his motivation for studying the strictly cyclic point-Gauss–Seidel procedure (he calls it Seidel's method) the instigation of the noted astronomer, V. K. Tserassky, in relation to a system (3.1) when a large number of observations and unknowns are involved, with a view to determining the speed of convergence (a problem not considered by Seidel).

If we denote by $\boldsymbol{x}_1$ the first column of the design matrix $X$, then, in terms of (2.1), Nekrasov first notes that we may write

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \{(\boldsymbol{Y} - X\boldsymbol{\beta})'\boldsymbol{x}_1\}^2/(\boldsymbol{x}_1'\boldsymbol{x}_1) + P \tag{3.7}$$

where $P$ is an expression not involving $\beta_1$. On the grounds of successive expressions of this kind he shows that successive iterative approximations (in the Gauss–Seidel procedure) to $\hat{\boldsymbol{\beta}}$ reduce the corresponding sum of squares of (estimated) residuals, since upon modification of successive elements a component of the form of the first summand in (3.7) becomes 0, while the other summand is unchanged. Seidel's free-steering procedure involved choosing at each stage that element for modification (in place of $\beta_1$ above) which reduces the current residual sum of squares by the greatest amount.

Nekrasov concludes that the Gauss–Seidel method for the setting (3.1) always converges; it is not, however, clear from his argument that the limit to which the progressively modified sum of squares of residuals is decreasing is the required minimum, or that the sequence of iterates is converging to $\boldsymbol{\beta}$. In his §3 he says the method 'doubtless converges', but in a later paper (Nekrasov (1892)) again claims that convergence has been proved in the setting of the present paper. He does deduce that the quantities of interest as regards convergence are the solutions $\lambda$ to the determinantal equation $|\lambda L + U| = 0$ (i.e. the eigenvalues of $(-L^{-1}U)$) but cannot prove that $\rho(-L^{-1}U) \equiv \max|\lambda| < 1$ always, as is required. This is a very early attempt to establish convergence of the Gauss–Seidel iterative method for a system (3.2) with symmetric positive definite matrix $A$; indeed, a complete proof seems not to have been given until Mises and Geiringer (1929).

As regards the purported main problem, that of speed of convergence, he investigates it by means of examples and difference equations (essentially matrix spectral methods), showing that it is fast if $\rho(-L^{-1}U)$ is small, and slow if $\rho(-L^{-1}U)$ is close to unity; and then obtains a lower bound on $\rho(-L^{-1}U)$, the argument leading to which Ostrowski ((1955), p. 179, footnote) describes as very useful ('sehr nützlich'). Nekrasov (1885b) passes onto Seidel's free-steering variant of the procedure and uses examples to show that the sum of squares of estimated residuals does seem to decrease faster, but does not consider this variant in his subsequent papers in the area. Nekrasov's further work on Gauss–Seidel iteration, in conjunction with Mehmke, is discussed in Seneta (1981).

The contributions of Chebyshev to interpolational least squares are well known in broad outline if not in details, and we shall only sketch them here. Chebyshev (Tchébichef (1855)) was apparently the first author to make use of an orthogonalized design matrix, produced from a given matrix $X = \{x_{ij}\}$ in a rather specialized setting in that $x_{ij} = x_i^{j-1}, i = 1, \ldots, N, j = 1, \ldots, r$. In other words, he was concerned with fitting, for fixed $r$, a polynomial

$$y = \sum_{j=1}^{r} \beta_j x^{j-1} \tag{3.8}$$

on the basis of $N$ pairs of observations $(x_i, Y_i)$, $i = 1, \ldots, N$, and does so by producing from the $r$ powers $1, x, \ldots, x^{r-1}$ a set of $r$ polynomials $T_0(x) = 1, T_1(x), \ldots, T_{r-1}(x)$ which are orthogonal in respect of the points $x_1, \ldots, x_N$ in that

$$\sum_{i=1}^{N} T_k(x_i)T_j(x_i) = 0, \qquad k \neq j.$$

The coefficients fitted to the readjusted polynomial system: $\sum_{j=1}^{r} \tilde{\beta}_j T_{j-1}(x)$ are those arising out of a least-squares fit to the data, as he makes clear at the outset of his paper. Apart from its important connections with orthogonal polynomials the paper is devoted largely to the application of continued-fractions theory† to this specific problem, as is reflected in its title. Chebyshev's treatment is also more general than we have indicated, in that he minimizes a 'weighted sum of squares' of residuals; that is, he effectively considers the 'generalized' least-squares approach to (2.1), allowing heteroscedasticity.

The full significance of the orthogonalization is not made apparent until a follow-up

† See Appendix 2 for Chebyshev's possible influence on Sleshinsky from this standpoint.



paper (Tchébichef (1859)). Its appearance is largely due to I. J. Bienaymé as catalyst, who had translated Chebyshev's 1855 paper into French in 1858 with a long prefatory footnote (Heyde and Seneta (1977)). This refers to the problem that, say, in the original setting of (3.8), one might wish to continue fitting with progressively higher $r$, until an 'adequate fit' (as measured, say, by (3.6)) obtains (this problem was first raised by Cauchy—see Heyde and Seneta (1977), Chapter 4). Translated to the context of (2.1), one wishes to add, in general, further columns to the design matrix $X$. If one proceeds directly by least-squares fit to the system (3.8), and goes from stage $r$ to stage $r + 1$, the new fitted vector of coefficients $\hat{\boldsymbol{\beta}}_{r+1}$ is different in general, in every element, specifically in the first $r$, to $\hat{\boldsymbol{\beta}}_r$, and most of the work done at stage $r$ is wasted. However, if one progressively orthogonalizes the columns of the design matrix as one increases $r$, the estimate of $\tilde{\boldsymbol{\beta}}_r$ forms the first $r$ elements of the estimate for $\tilde{\boldsymbol{\beta}}_{r+1}$ and only the last element of the latter needs to be calculated at stage $r + 1$. The *decrease* in residual sum of squares in going from stage $r$ to $r+1$ is $\tilde{\boldsymbol{x}}'_{r+1}\tilde{\boldsymbol{x}}_{r+1}\hat{\tilde{\beta}}^2_{r+1}$, where $\tilde{\boldsymbol{x}}_{r+1}$ is the $(r+1)$th (orthogonalized) column, and $\hat{\tilde{\beta}}_{r+1}$ is the estimate of $\tilde{\beta}_{r+1}$, and the simplicity of this is understood by Chebyshev. The central role of the estimated residual sum of squares in early least-squares work is again clear.

It is worth mentioning also three additional papers of Chebyshev in the area: those of 1858, 1864, and 1875, which explore the consequences of assuming the points $x_i, i = 1, \ldots, N$ at equidistant intervals in the context of ordinary least squares. The orthogonal polynomials so produced by Chebyshev are the finite difference analogues of the Legendre polynomials; they are widely tabulated and often associated with the name of Chebyshev. In this connection the reader should consult pp. 878–881 of Chebyshev (1955) for a commentary by N. I. Achiezer.

It should be noted, however, that in all his work in the area, Chebyshev was preoccupied with polynomial (or, as it is sometimes known, parabolic) regression, and approached it via continued-fractions theory. There is a less-known note in the stream of Chebyshev's work, viz. Tchébichef (1870), stimulated by practical considerations. In this he deals with a more general system than hitherto, viz.

$$y = F(x) \sum_{j=1}^{\infty} \beta_j x^{j-1}$$

where $F$ is a known function. The polynomials now produced are orthogonal with respect to the weight function $F(x)$:

$$\sum_{i=1}^{N} T_k(x_i)T_j(x_i)F(x_i) = 0, \qquad k \neq j,$$

assuming homoscedasticity.

It is appropriate to conclude this section by mentioning that subsequently 'Chebyshev's methods' were a stimulus for, and indeed tended to permeate, all Russian-language writings on least squares. (This statement applies not only to writings of interpolational nature, but those with a probabilistic viewpoint of linear least squares, where we shall see shortly that attempts to justify optimality of the procedure on the grounds of Chebyshev's (more accurately, the Bienaymé–Chebyshev) inequality (Tchébichef (1867)) were not infrequent.) As examples we cite here Maievsky

(1881), Grave (1889), Kleiber (1890), Yaroshenko (1893a,b), Veselovsky (1897), Tikhomandritsky (1898), Zabudsky (1898), Avrinsky (1904), Ermakov (1905) and Iveronov (1912). In early post-revolutionary times, the booklet of Khotimsky (1925) is even appropriately subtitled as 'Chebyshev's Method'.

We shall have more to say shortly about the writings of Maievsky, Yaroshenko and Tikhomandritsky. The latter two are representatives of the Odessa–Kharkov groups, and therefore of particular interest to us.

## 4. Probabilistic aspects of least squares (Odessa and Kharkov)

In his probabilistic study of the linear model, Sleshinsky (1892) assumes the residual random variables $\varepsilon_i, i = 1, \ldots, N$ in (2.1) are i.i.d. with symmetric positive density $f(\cdot)$ on the finite interval $[-\tau, \tau]$, and is first concerned with proving rigorously the convergence as $N \to \infty$

$$\sum_{i=1}^{N} \lambda_i \varepsilon_i \Big/ \left\{\sigma^2 \sum_{i=1}^{N} \lambda_i^2\right\}^{1/2} \to \mathrm{N}(0, 1) \tag{4.1}$$

in line with investigations along these lines (see Section 2) initiated by Laplace. Notice that it may be assumed without loss of generality that $\boldsymbol{\lambda} = \{\lambda_i\} \geq \mathbf{0}$. Sleshinsky's line of investigation in this respect is not original, but is actually a carefully detailed and rigorous reworking and supplementation of a proof sketched by Cauchy in 1853 (see Cauchy (1853) and Heyde and Seneta ((1977), Chapter 4). He obtains by characteristic function methods an estimate of closeness of two probabilities:

$$\left| \Pr\left\{ -t \leq \frac{\sum_{i=1}^{N} \lambda_i \varepsilon_i}{\left\{\sigma^2 \sum_{i=1}^{N} \lambda_i^2\right\}^{1/2}} < t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} \exp(-w^2/2) dw \right| \leq I_1 + I_2 + I_3 \tag{4.2}$$

where $I_1, I_2, I_3$ are certain integrals satisfying:

$$I_1 \leq \frac{1}{\pi M} \exp\{-M\}$$

with

$$M = \tfrac{1}{2}\{\sigma^2 \Lambda \Theta^2 / (1 + \sigma^2 \lambda^2 \Theta^2)\}, \qquad \lambda = \max(\lambda_1, \ldots, \lambda_N), \qquad \Lambda = \sum_{i=1}^{N} \lambda_i^2,$$

and $\Theta$ is essentially arbitrary;

$$I_2 \leq \frac{2h\sqrt{3}}{\pi} \log\left\{ \frac{t\Theta\sqrt{\sigma^2 \Lambda}}{\sqrt{3}} + \sqrt{1 + \frac{t^2 \Theta^2 \sigma^2 \Lambda}{3}} \right\},$$

$h$ being the larger of

$$\exp\left\{\frac{\sigma^2 \Lambda \lambda^2 \Theta^4 \tau^2}{8}\right\} - 1, \qquad 1 - \exp\frac{1}{4}\left\{\frac{-\sigma^4 \Lambda \lambda^2 \Theta^4}{1 - (\sigma^2 \lambda^2 \Theta^2)/2}\right\}; \qquad I_3 \leq \frac{\exp - \{\sigma^2 \Lambda \Theta^2 / 2\}}{\pi \sigma^2 \Lambda \Theta^2 / 2}.$$



The various peripheral conditions required for the validity of the inequalities are stated beginning at the bottom of his p. 251, and reformulated on p. 253. If we first take two positive numbers $l, L$ such that $l \leq N\lambda_0 \leq N\lambda \leq L$, where $\lambda_0 = \min(\lambda_1, \ldots, \lambda_N)$, these are

$$N \geq \max\left\{8, \frac{4L^2}{l^2}, \frac{8\tau^2 L^2}{\sigma^2 l^2}\right\}$$

and

$$\frac{2^{3/2}\sqrt{N}}{l\sigma} < \Theta < \frac{N}{\tau L}.$$

Hence taking $l$ and $L$ as fixed, and letting $\Theta$ grow as a power law of $N$, between $N^{1/2}$ and $N^{3/4}$, each of $I_1, I_2, I_3$ approaches 0, and Sleshinsky obtains the limit result (4.1) under the condition that the numbers

$$|N\lambda_1|, |N\lambda_2|, \ldots, |N\lambda_N|, \tag{4.3}$$

be uniformly bounded away from 0 and $\infty$ as $N \to \infty$.

With its rigour and difficult and involved estimates, some of which are not contained in Cauchy's work (neither is the conclusion), Sleshinsky achieves for the first time, as he has intended, full rigour of mathematical analysis in a proof of the central limit theorem.

The bounds on $I_1$ and $I_3$ are independent of $t$ and go down exponentially fast as $N \to \infty$. The $I_2$ bound unfortunately involves $t$, so overall Sleshinsky's bound is inhomogeneous. Asymptotics for this show that $h \leq \text{const.}\, N^{4\gamma - 1}$ and the log expression $\leq \text{const.} \log(N^\gamma t)$, if we put $\Theta = N^{1/2+\gamma}, 0 < \gamma < \frac{1}{4}$. Thus overall we have

$$I_1 + I_2 + I_3 \leq \text{const.} \frac{\log N}{N^{1-4\gamma}}, \qquad 0 < \gamma < \tfrac{1}{4}$$

which can be seen to be a quite good asymptotic estimate if $\gamma$ is chosen close to 0 {the 'const.', independent of $N$, will in general depend on $t$ and $\gamma$}. In fact we know from the later work of Liapunov (Liapounoff (1901a)), since the $|\lambda_i N|$ are uniformly bounded away from 0 and $\infty$ and $\varepsilon_i \in [-\tau, \tau]$, that the left-hand side of (4.2) can be bounded by the homogeneous bound const. $(\log N)/N^{1/2}$.

It is also an immediate deduction that Sleshinsky's results actually hold under conditions more general than he believes. For suppose for all $N$,

$$0 < l \leq |\phi(N)\lambda_i| \leq L < \infty, \qquad i = 1, \ldots, N$$

($l$ and $L$ independent of $N$). Then

$$\sum_{i=1}^{N} \lambda_i \varepsilon_i \Big/ \left\{\sigma^2 \sum_{i=1}^{N} \lambda_i^2\right\}^{1/2} = \sum_{i=1}^{N} \tilde{\lambda}_i \varepsilon_i \Big/ \left\{\sigma^2 \sum_{i=1}^{N} \tilde{\lambda}_i^2\right\}^{1/2},$$

where $\tilde{\lambda}_i = \lambda_i \phi(N)/N$, so that $l \leq |N\tilde{\lambda}_i| \leq L, i = 1, \ldots, N$, so the results hold for such $\lambda_i$ also. Conversely, *without loss of generality* we may assume that for $i = 1, \ldots, N$, and all $N, l \leq |\lambda_i| \leq L$ where $\lambda_i$ may still depend on $N$. Now, we know that if an infinite sequence independent of $N$, $\{\lambda_i\}$, $i = 1, 2, \ldots$ satisfies $|\lambda_i| \leq L, i \geq 1$, then necessary† and sufficient in this situation for convergence to normality is $\sum_{i=1}^{\infty} \lambda_i^2 = \infty$; in Sleshinsky's formulation this divergence is replaced by the stronger assumption: $|\lambda_i| \geq l > 0, i \geq 1$.

It is evident and remarkable, as mentioned in our Section 1, that Sleshinsky's boundedness condition (4.3) on the $\lambda_i \equiv \lambda_i(N), i = 1, \ldots, N$ already manifests the notion of 'triangular' arrays in terms of which general central-limit-type results have since been formulated. Indeed, he is clearly dealing with the triangular array of variables (with obvious independence within rows) where

$$X_{N,i} = \lambda_i(N)\varepsilon_{N,i}$$

the $\varepsilon_{N,i}, i = 1, \ldots, N$ being i.i.d. with density $f$.

Gnedenko and Gikhman (1956), p. 490, remark that Sleshinsky's paper:

> '... is passed over in silence in the literature. Only A. M. Liapunov... refers to the work of Sleshinsky, and, in the event, does not characterize it with complete accuracy.'

but do relatively little (*loc. cit.*, p. 491–492) to remedy the situation, and conclude as follows:

> 'Even so, in the work of Sleshinsky the method of characteristic functions, from a conceptual and technical standpoint, is not yet sufficiently developed. This circumstance, and possibly the narrowness of his purpose, associated with the specific problem of elucidating the method of least squares, resulted in unnecessary restrictions in the formulation and proof of the theorem.'

A little later, in an analysis of Liapunov's work in probability, Gnedenko (1959), pp. 145–146, does no more in relation to Sleshinsky than to give Liapunov's assessment, quoted in our Section 1.

Gnedenko and Gikhman have to some extent in mind the last part of Sleshinsky's paper. In his last section (§11, p. 261 *et seq.*) Sleshinsky reverts to the linear least-squares setting of his motivating problem, that is, the framework represented by our (2.1)–(2.3), in attempting to bring his theory to bear to justify Laplace's conclusion (2.6). He effectively makes the following assumptions on the design matrix $X = \{x_{ij}\}$:

(1) *Every one* of the $\binom{N}{r}$ matrices formed from selecting $r$ rows of $X$ is non-singular [this is rather stronger than the assumption that $X$ is of full rank $r$].
(2) The absolute values of the determinants of these matrices are bounded away from 0 as $N \to \infty$.
(3) $|x_{ij}| < \alpha < \infty, j = 1, \ldots, r; i = 1, 2, \ldots$.

He deduces in regard to the 'multipliers' matrix $K = \{k_{ij}\}$ that

$$|k_{ij}| = O(N^{-1}) \qquad \text{as } N \to \infty \tag{4.4}$$

† Necessity follows from e.g. Cramér's theorem (Lukacs (1960), p. 173), and is not a consequence of the Lindeberg–Feller Theorem, as Fisz ((1963), pp. 206–207) suggests.



$i = 1, \ldots, r$, *uniformly in* $j \geq 1$. It is clear that he had hoped to get a two-sided condition $0 < l \leq |Nk_{ij}| \leq L < \infty$, so that the preceding central limit work would be applicable in regard to (2.6), but it seems that the lower bound eluded him; and the paper terminates at this point in rather unresolved fashion.

However, (4.4) is sufficient for *consistency* of the estimator $\bar{\boldsymbol{\beta}}$, given by (2.3) for $\boldsymbol{\beta}$, and this point is immediately noticed by Sleshinsky's Odessa colleague, Yaroshenko, in a paper with identical title. He considers the full-rank system (2.1), with the $\varepsilon_i$ independently and symmetrically distributed, *but not necessarily identically distributed*, $\varepsilon_i$ taking on a set of *discrete* values on $[-\tau_i, \tau_i]$. The reason for the discreteness seems to be in relation to his forthcoming use of the 'Chebyshev' inequality, referenced from the Russian version of 'Chebyshev's paper (Tchébichef (1867)), where it is proved for discrete variables. Indeed, since the variance of $\bar{\beta}_i = \sum_{j=1}^{N} k_{ij} Y_j$ is $\sum_{j=1}^{N} k_{ij}^2 \sigma_j^2$, where $\sigma_j^2 = \operatorname{Var} \varepsilon_j$, and $\bar{\beta}_i$ is unbiased for $\beta_i$, application of that inequality to the $i$th element of (2.3), yields

$$\Pr\{|\bar{\beta}_i - \beta_i| \geq \varepsilon\} \leq \sum_{j=1}^{N} k_{ij}^2 \sigma_j^2 / \varepsilon^2 \tag{4.5}$$

which will $\to 0$ as $N \to \infty$, *if* $|k_{ij}\sigma_j| \leq L/N$, *uniformly* in $j \geq 1$. Now, this is precisely Sleshinsky's conclusion (4.4), once Yaroshenko's heteroscedastic situation (possibly unequal $\sigma_j^2$) is transformed to Sleshinsky's homoscedastic one in the manner

$$\Sigma^{-1/2}\boldsymbol{Y} = \Sigma^{-1/2}X\boldsymbol{\beta} + \Sigma^{-1/2}\boldsymbol{\varepsilon}$$

where $\Sigma = \operatorname{Var} \boldsymbol{\varepsilon} = \operatorname{diag}(\sigma_i^2)$, with $K \to K\Sigma^{1/2}$ (Sleshinsky's *conditions* (1), (2), (3) then being applied to the matrix $\Sigma^{-1/2}X$).

Yaroshenko also shows that the requirement that what is now known as the generalized least-squares estimate for $\boldsymbol{\beta}$, viz.

$$\hat{\boldsymbol{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\boldsymbol{Y}$$

obtained by Yaroshenko through use of Lagrange multipliers (a technique used by Sleshinsky), also emerges from requiring the 'tightest' interval in the bound provided in (4.5) for $i = 1, \ldots, r$. This use of Chebyshev's inequality is of course quite irrelevant, since the conclusion emerges out of the *minimality of the* $\operatorname{Var} \bar{\beta}_i$, $i = 1, \ldots, r$, and that result had already been obtained by Gauss, in his 1821 work. He goes on to consider in the same manner a slightly more general situation of his model still with $\boldsymbol{w} = E\boldsymbol{\varepsilon}$, where $\boldsymbol{w}$ is not necessarily $\boldsymbol{0}$; this situation also requires only a simple transformation to revert to the simpler one, so needs no separate discussion.

There is a French version of the paper (Yarochenko (1893)) which otherwise differs substantially only in now citing the French version of Chebyshev (Tchébichef (1867)), and omitting all reference to Sleshinsky in connection with the situation leading to (4.4), glossing over that whole central and difficult issue. It would appear from the dual publication that Yaroshenko thought the paper important, but for the wrong reasons; only the consistency is of interest, not, as it were, the justification of generalized least squares on the basis of Chebyshev's inequality.

Sleshinsky (1893–1894), hot on the heels of Yaroshenko, begins by noting that Chebyshev's inequality leads to least squares, and that both Chebyshev and Yaroshenko deal with the discrete case. With a view to extending Yaroshenko's conclusions, he modifies Chebyshev's proof to obtain the inequality for the case of a single random variable described by a probability *density* on a finite interval $[A, B]$, and then to a linear function $\lambda'(\varepsilon - E\varepsilon)$ where the $\varepsilon_i$ are independent, and $\varepsilon_i$ is distributed on $[A_i, B_i]$. It is quite surprising that Sleshinsky, who was very familiar with the work of Bienaymé (1853), discussing it on p. 210 of his own 1892 paper, did not perceive that in the paper cited Bienaymé had already proved the inequality in its general form and Chebyshev (Tchébichef (1874)) had more or less conceded priority. Even more surprising is the fact that Sleshinsky did not notice that the use of the inequality is totally irrelevant, since the first part of the 1892 paper is much concerned with the history and nature of least squares. This follow-up note of Sleshinsky is thus only interesting inasmuch as it is the second and apparently the last of his writings in probability and statistics. The anomaly mentioned above is not pointed out by Gnedenko and Gikhman ((1956), p. 487)† who are presumably aware of this note of Sleshinsky from Tikhomandritsky's book, which we mention shortly.

It is clear that Sleshinsky and Yaroshenko share a somewhat nationalistic desire to emphasize the contributions to probability of Chebyshev, particularly in the matter of the discovery and importance of the celebrated inequality. This has ever since been the case with Russian authors; we find it even in the writings (as we shall see) of Nekrasov, let alone in those of Chebyshev's students. In the probability textbook of one such student, M. A. Tikhomandritsky (1898), we find very definite evidence of this. While manifesting considerable influence by Chebyshev's exposition, it dwells extensively on the work of Yaroshenko, and, in part, on that of Sleshinsky. The tone of the book is set by its preface, where Tikhomandritsky emphasizes least squares as the most important practical application of probability. He states (!?):

> '...least squares does not give the most probable estimates, as is acknowledged by everybody, but only those giving tightest bounds [in the sense of Chebyshev's inequality]...'

and refers to Yaroshenko in this connection. His actual exposition on least squares, in Chapter VIII, follows Yaroshenko's closely, but he allows his residuals to have a probability density, in view of the fact that in his §50 he had reported Sleshinsky (1893–1894). In the preceding section, §49, he follows Chebyshev's derivation of the inequality in the discrete case. Chapter VIII itself contains no explicit mention of Yaroshenko or Sleshinsky, although elementary textbooks of the time, of Ermakov and Nekrasov, are mentioned.

There is some confusion in Tikhomandritsky's exposition: the residuals are allowed to have densities on $(-\infty, \infty)$ without apparent justification; and random variables and their expectations are used more or less interchangeably. This last is particularly unfortunate in relation to one of the few features of the exposition not following previous Russian writing, when he arrives (§70, p. 101 *et seq.*) at what is tantamount to the formula

$$\sigma^2 = \frac{E\{(Y - X\hat{\beta})' A^{-1} (Y - X\hat{\beta})\}}{N - r} \tag{4.6}$$

† Sleshinsky (1893–1894) is again presumably alluded to as 'Sleshinsky [2]' on p. 488 of Gnedenko and Gikhman; however, the allusion there is invalid, and 'Sleshinsky [2]' is missing from their bibliography.



where $A = \operatorname{diag} k_i = \sigma^{-2}\Sigma$—having written $\sigma_i^2 = k_i\sigma^2$—is assumed known, and $\sigma$, a 'measure of precision' is not. He credits this formula to Gauss, in whose writings it indeed is well known to occur. Its importance lies in the fact that it implies

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})' A^{-1} (Y - X\hat{\beta})}{N - r} \tag{4.7}$$

is an unbiased estimator of $\sigma^2$.

## 5. Markov, Liapunov and Slutsky on probabilistic least squares

Markov expresses himself on linear least squares within a letter to Vasiliev (published within Markov (1898a), which itself is reprinted within Markov (1951), of which the relevant pages are pp. 246–251). His view is that the only rational justification for (generalized) least-squares estimation is that it yields minimum variance estimators: that is, the early justification of Gauss of 1821, that he mentions, which has come to be known as the Gauss–Markov theorem. He is less impressed with Gauss's justification of least-squares estimators as being the 'most probable' (or, as we should now say, maximum likelihood). In the matter of the name 'Gauss–Markov theorem', we have found no reason to differ from the opinion of Seal (1967), who writes concerning the proof of this proposition to be found in editions of Markov's textbook beginning in 1900:

> 'The Markov proof, essentially the same as Helmert's, was thought by Neyman...to be an original theorem and the Russian writer's name became attached to it.'

We have examined the last edition (Markov (1924)) which presumably differs little in regard to Chapters I–VIII from earlier editions. Chapter VII ('The method of least squares'), on its opening page (p. 323) carries a footnote referring to Markov (1898a):

> 'My view on various attempts at theoretical justification of the method of least squares is presented in *The law of large numbers and the method of least squares*...'.

In the course of the chapter he considers independently distributed residuals with possibly different variances. In §50 he treats of Chebyshev's parabolic interpolation, referring specifically to Tchébichef (1875). In §51 he considers the situation where there is linear dependence between the columns of the design matrix, but does so in an unremarkable manner, reducing the situation to one of independence. At the end of the chapter is a short bibliography which includes Sleshinsky (1892), who does not appear to be mentioned in the text. The only other Russian-language works mentioned in the bibliography are Maievsky (1881) and the theoretical part of an astronomy text of N. Zinger.

Only faint echoes of the work discussed in our Section 4 are to be found in the writings of Markov and Liapunov, generally.

Markov (1898a); (1951), p. 249, for example, refers to Russian work which has attempted to justify least squares on the basis of results found in Tchébichef (1867). He criticizes explicitly only the work of Maievsky ((1881), §31)†; but then on p. 250 mentions Russian mathematicians who have attempted justification on the basis of Chebyshev's inequality, making clear, as we have noted, the non-relevance of this approach. It seems obvious that he has Yaroshenko and Tikhomandritsky in mind. Tikhomandritsky's book is mentioned in Markov's textbook, but only in connection with Markov's Chapter VI ('Probabilities of hypotheses and future events').

There is also in existence a manuscript by Liapunov whose date is unknown but which is closely related to Liapunov's other papers (Liapounoff (1900), (1901a)) in theme and technique, which bears on linear least squares in the heteroscedastic situation discussed in the preceding section, in particular on (4.6) and (4.7). This has now been published (Liapunov (1975)) with a commentary by O. B. Sheynin (1975). Liapunov claims that (4.6) is as yet to be rigorously justified, and can be based on a deduction that he makes in his note: that if $(E\varepsilon_i^4)/(E\varepsilon_i^2)^2$ remains uniformly bounded above for all $i$, then $\hat{\sigma}^2 \to \sigma^2$ in probability as $N \to \infty$. Thus Liapunov is concerned with what we now call consistency of $\hat{\sigma}^2$ as an estimator of $\sigma^2$. To achieve his end, Liapunov uses the inequality for a non-negative random variable $\xi$:

$$\Pr\{\xi > a\} \le E(\xi)/a$$

which can be used to deduce Chebyshev's inequality; it is now known as 'Markov's inequality' though Liapunov refers to the 'considerations of P. L. Chebyshev'. Sheynin (1975) notices the absence of direct references to Gauss, and also that in his §§39–40, Gauss had in 1823 already demonstrated an inequality whose right-hand side reads

$$\operatorname{Var}\hat{\sigma}^2 \le \frac{2\sigma^4}{N-r}.$$

Clearly the right-hand part of this alone is enough to yield the consistency of $\hat{\sigma}^2$ as $N \to \infty$, and Liapunov's condition is redundant. It is easy to conjecture, with Sheynin, reasons why Liapunov discontinued work on this project.

E. E. Slutsky's (1914) paper needs to be mentioned in this section, although it treats a slightly different problem. In general this relates to fitting a function $f(x; \beta_1, \ldots, \beta_r)$ of one variable $x$, depending on $r$ parameters $\beta_1, \ldots, \beta_r$, for example the polynomial (3.8), when there are *repeated readings* on response $Y$ for each value of $x$ considered, the system additionally being normal and possibly heteroscedastic. If there are $n_i$ responses for the value $x_i$, and their average is $\bar{Y}_i$, $i = 1, \ldots, N$, then

$$\bar{Y}_i = f(x_i; \beta_1, \ldots, \beta_r) + \varepsilon_i, \qquad i = 1, \ldots, N$$

where the $\varepsilon_i$ are independent and $\varepsilon_i \sim \mathrm{N}(0, \sigma_i^2/n_i)$, where $\sigma_i^2$ is the variance of a normal response corresponding to setting $x_i$. Slutsky proposes to estimate $\boldsymbol{\beta}$ by minimizing

$$\chi^2 = \sum_{i=1}^{N} n_i(\bar{Y}_i - f(x_i; \beta_1, \ldots, \beta_r))^2/\sigma_i^2 \tag{5.1}$$

which he recognizes as a chi-square variable, so we may regard this procedure as an

† Maievsky was a student of Chebyshev's, and an artillery theoretician; in this connection the nature of Tchébichef (1870) is interesting.



early instance of minimum-$\chi^2$ estimation (if we assume $\sigma_i^2$ known). Of course it is also obviously maximum-likelihood estimation and (without the normality assumption) an instance of weighted least squares, and, in the particular case of (3.8), of the linear model with repeated observations.

Slutsky, in this paper much influenced by K. Pearson on the topic of chi-square, proposes to use the estimated value of (5.1); that is, the residual sum of squares, (estimating $\sigma_i^2$ in the usual way), as a measure of goodness-of-fit, by comparison with tabulated values of $\chi^2$, and uses $N$ degrees of freedom following K. Pearson.

## 6. Central limit theorems for large deviations and Nekrasov

Although Nekrasov's first paper in probability may have been written in about 1890, it is his paper of 1898 (Nekrasov (1898a)), dedicated to the memory of Chebyshev, which is the beginning of a long and bitter controversy between himself and Markov (and also Liapunov) which lasted till 1915. The first controversy thus spanned the crucial years 1898–1901 of the central limit theorem.

The 1898a paper, as Liapunov comments † in 1900, contains no proofs. Assessing it now, we may say in general that in this and later publications on the same topic Nekrasov, highly proficient in the use of complex-variable theory in general and knowledgeable about the Lagrange expansion in particular (Nekrasov (1885a)), attempted to use what we now call the method of saddlepoints, of Laplacian peaks, and of the Lagrange inversion formula, to establish, for sums of non-identically distributed lattice 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. A standard reference for this type of theory is Petrov ((1972), Chapter 8); the earliest references there are to Khinchin in 1929 in connection with treatment of the Bernoulli scheme by using Stirling's formula. The direct use of saddlepoint methods in problems of this kind is generally attributed to Daniels (1954). Nekrasov's attempt was only partly successful, poorly presented, badly defended, and never understood by Markov and Liapunov, nor noticed by their successors. Indeed, Nekrasov's enormous outpouring of material on the subject subsequent to 1898, in the form of a monograph of some 1000 pages spread serially over volumes **21**, **22** and **23** of *Matematicheskii Sbornik* under the grandiose title 'Novie osnovania . . .' which in translation reads 'New foundations of the study of probabilistic sums and mean values', made the task of penetrating to the essence of this work formidable. Markov and Liapunov, consequently, reject Nekrasov's work on the grounds of a few of the more obvious points only. We shall come to these later (one has been mentioned in our Section 1).

To understand the general nature and components of Nekrasov's working, it seems unavoidable to *outline* first the mathematical content of the kind of theorem which he attempted to prove. W. Richter (1957) seems to have been the first in modern times to take up consideration of such general central limit problems by methods of the same kind; and we begin by stating a result of his. We follow the notation of our Section 1: that is, $U_1, U_2, \ldots$ are independently but not necessarily identically distributed random variables.

† Cited in our Section 1.

We assume each $U_i$ has zero mean, and is defined on a lattice $a_i + rh$, $r = 0, \pm 1, \pm 2, \ldots$ where $h$ is the (common) period. We put as before $S_n = \sum_{i=1}^{n} U_i$ and $B_n^2 = \sum_{i=1}^{n} \text{Var } U_i$. We further make the following assumptions (which are not as weak as possible, but serve our historical purpose best), where $M_j(z) = E(\exp(zU_j))$.

*Assumption* A. There are positive numbers $K$, $k$ and $A$ independent of $j$ such that for complex $z$ in the circle $|z| < A$

$$k \le |M_j(z)| \le K, \qquad j = 1, 2, \ldots.$$

*Assumption* B†. For all $n$, and some positive constant $\delta$

$$B_n^2/n \ge \delta > 0.$$

Then, under an additional assumption ('Assumption C') whose precise form need not concern us, putting

$$P_n(r) = \Pr\{S_n/B_n = x\}, \qquad x \equiv x_{n,r} = \frac{\left(\sum_{i=1}^{n} a_i\right) + rh}{B_n},$$

*it can be shown that if* $x > 1$ *and* $x = o(\sqrt{n})$ *as* $n \to \infty$,

$$P_n(r) = \frac{h}{B_n\sqrt{2\pi}} \exp(-x^2/2) \exp((x^3/\sqrt{n})\lambda_n(x/\sqrt{n})) \left\{1 + O\left(\frac{x}{\sqrt{n}}\right)\right\} \tag{6.1}$$

*where* $\lambda_n(t)$ *is a power series in t convergent for sufficiently small* $|t|$, *uniformly for all n.* (If $x < -1$ always, then the '$O$' term is replaced by $O(|x|/\sqrt{n})$.)

This kind of theorem is a *local* limit theorem of central limit type for large deviations for non-identically distributed lattice random variables.

Assumption A implies that $M_j(z)$ is analytic along the strip $|\text{Re } z| < A$ and the cumulant generating functions $K_j(z) = \ln M_j(z)$ where 'ln' denotes the principal branch of the logarithm, are each analytic in $|z| < A$, and all the $K_j(z)$, $K_j'(z)$ and $K_j''(z)$ are uniformly bounded in modulus in some circle $|z| < A_1 < A$. Richter's subsequent argument focusses on the function $K_{(n)}(z) = \sum_{j=1}^{n} K_j(z)/n$. He seeks a *saddlepoint*, that is, a real stationary point $w_0$, at which the real function $K_{(n)}(w) - B_n xw/n$, $-A_1 < w < A_1$, is minimal in this interval. Its existence for sufficiently large *n follows* from an inversion of the equation

$$K_{(n)}'(z) = B_n \tau/\sqrt{n} \tag{6.2}$$

(where $\tau = x/\sqrt{n}$) by *Lagrange's inversion formula* (Lagrange series) to give an expansion for $z$ in strictly positive powers of $\tau$ using Assumption B. Its uniqueness follows from convexity of the function considered. The existence and uniqueness of the requisite $w_0$, in general dependent on $n$, in $(-A_1, A_1)$ leads to an asymptotic treatment

† Notice that this is the same as Markov's condition (iii) of our Section 1.



of integrals by *Laplace's method of peaks*, to give first the expression

$$P_n(r) = \frac{h}{\sqrt{2\pi}} \frac{\exp\left\{n\left\{K_{(n)}(w_0) - w_0\tau \frac{B_n}{\sqrt{n}}\right\}\right\}}{\sqrt{nK_{(n)}''(w_0)}} [1 + O(\tau)] + \frac{h}{2\pi i} I_2 \tag{6.3}$$

to which we shall need to refer to shortly, where $I_2$ is a complex expression negligible (under 'Condition C') relative to the first term stated; and then to give (6.1). Mathematical details of this argument may be found in Petrov (1972), esp. p. 274.

A corresponding *global* limit theorem given by Richter needs only Conditions A and B (and does not require different treatment for lattice and non-lattice variables) and states *that if* $x > 1$ *and* $x = o(\sqrt{n})$

$$\left.\begin{aligned} \frac{1 - F_n(x)}{1 - \Phi(x)} &= \exp\left\{\frac{x^3}{\sqrt{n}} \lambda_n\left(\frac{x}{\sqrt{n}}\right)\right\}\left[1 + O\left(\frac{x}{\sqrt{n}}\right)\right] \\ \frac{F_n(-x)}{\Phi(-x)} &= \exp\left\{-\frac{x^3}{\sqrt{n}} \lambda_n\left(-\frac{x}{\sqrt{n}}\right)\right\}\left[1 + O\left(\frac{x}{\sqrt{n}}\right)\right] \end{aligned}\right\} \tag{6.4}$$

where

$$F_n(x) = \Pr\{S_n/B_n \le x\}, \qquad \Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-u^2/2) du.$$

Most important for the sequel is to notice that if we put $x, x' = o(n^{1/6})$ (assuming $x, x' > x_0 > 0$) then from (6.4)

$$1 - F_n(x) = (1 - \Phi(x))\{1 + o(1)\}, \qquad F_n(-x') = \Phi(-x')\{1 + o(1)\}$$

which suggests, for large $n$, the reasonableness *under the stated conditions*, of the *approximation*

$$\Pr\{-x' < S_n/B_n \le x\} \cong \Phi(x) - \Phi(-x'). \tag{6.5}$$

In particular we may take *fixed* $x, x' > 0$. We then have by an indirect route, and under rather restrictive conditions, what in fact follows more easily from the central limit theorem.

However, the basic importance of (6.5) is its plausibility in the case of $x$ and $x'$ *increasing with n*, which is what is meant by *large deviation* results. Nekrasov (1898a), Theorems 2 and 3, has this kind of result in mind, when he requires that $x', x \le n^{(1/2)-v}$ where $1/3 < v < 1/2$; *indeed then, certainly* $n^{(1/2)-v} = o(n^{1/6})$. At a later time, within his grandiose 'Novie osnovania...', in reply to Liapunov (1901b) he simply requires $v > 1/3$, which is also correct.

We have tried to indicate the crucial role played by Assumption B, in the above argument. We have noted this assumption in Markov's work, albeit in the context of a very conventional global limit theorem (see our Section 1); it is lacking in Nekrasov's work, and is the point on which both Markov and Liapunov focus in the controversy, from the standpoint of conventional global limit theorems, which are in fact particular cases (as we have noted) albeit under excessive conditions, of such theorems for large deviations (as Nekrasov is aware).

The controversy does not emphasize the local limit theorems; these do not figure in

Markov's and Liapunov's work even in conventional form. If we let $x = o(n^{1/6})$ in (6.1) we obtain as $n \to \infty$

$$P_n(r) = \frac{h \exp(-x^2/2)}{B_n\sqrt{2\pi}} \{1 + o(1)\}.$$

The conclusion of Nekrasov's (1898a) Theorem 1 under the assumption (*inter alia*) that $1/3 < v < 1/2$ is that this conclusion holds providing $|x| \le n^{(1/2)-v}$.

The preceding exposition enables us to describe the procedure which was the general thrust of Nekrasov's argument; we paraphrase this from §3 of his (1898a) paper. Suppose $\phi_i(s) = E(s^{U_i})$ where the $U_i$ are assumed to have zero means. The generating functions exist at least on the unit circle. Now put

$$\psi(s) = \left\{\prod_{i=1}^{n} \phi_i(s) s^{-xB_n}\right\}^{1/n}$$

where in place of $s$ Nekrasov wishes to use the 'positive root of $\psi'(s) = 0$'. If we assume that Richter's Assumptions A and B are satisfied, we note that if we take principal values

$$\log \psi(e^z) = K_{(n)}(z) - xzB_n/n = K_{(n)}(z) - \tau z B_n/\sqrt{n}.$$

Treating $z$ as a real variable as in our previous exposition, $w$, we find that $d\log\psi(e^w)/dw = \psi'(e^w)e^w/\psi(e^w) = 0$ yields the saddlepoint $w_0$, so the 'positive root of $\psi'(s) = 0$' is $\alpha = e^{w_0}$. He adds that the definition of the root $\alpha$ presents no problems, for a given $rh \equiv xB_n$, since the equation $\psi'(s) = 0$ can be written in the form $s - 1 = \tau(B_n/\sqrt{n})F(s)$, (this is obviously analogous to (6.2), and would follow from an expansion of $\psi'(s)$ in the vicinity of $s = 1$, taking into account that $\psi'(1) = (-\tau B_n/\sqrt{n})$), at which point $s$ may be expanded in terms of $\tau(B_n/\sqrt{n})$ in a rapidly convergent Lagrange expansion.

He assumes the analyticity for all $i$ of $\phi_i(e^z) = M_i(z)$ within a *fixed* circle. This more or less corresponds to Assumption A, but is, it would seem, far from adequate.

There is a final assumption which is qualitatively important for the sequel. If $w_0$ denotes the saddlepoint, let

$$R_{(n)}(y) = |\exp\{K_{(n)}(w_0 + iy) - \tau(w_0 + iy)B_n/\sqrt{n}\}|.$$

As $y$ varies from $-\infty$ to $\infty$, this modulus clearly achieves its maximum at $y = 0$ (that is, at the saddlepoint), and is periodic in $y$ with period $2\pi/h$. He says: 'We imagine the remaining *maxima* of the function $R_{(n)}(y)$ and denote by $R_1(n)$ the largest of these. If $R_{(n)}(y)$ has no other maxima, we shall take for $R_1$ the *minimum* of $R_{(n)}(y)$.' That is, $R_1(n)$ is the largest local (but not global) maximum of $R_{(n)}(y)$. He then appears to require (Nekrasov (1898a), p. 437)

$$1/(n\{K_{(n)}(w_0) - \tau w_0 B_n/\sqrt{n} - \log R_1(n)\}) = O(1/n^{\sigma}),$$

where $\sigma > 0$, although this is not altogether clear. This condition, unclearly expressed though it is, is likely envisaged by Nekrasov as playing a role in what we have presented



as the estimation of $I_2$. In any case, his stated theorem immediately following his statement of assumption, viz. Theorem 4, has the general form of the local limit expression (6.3) expressed in terms of $\psi(s)$; Theorem 5 is a global limit theorem for large deviations, and Theorem 6 attempts to give a more explicit (asymptotic-expansion type) correction term of Laplacian type, in the global limit theorem. Theorems 4 to 6 do not appear to require the assumption $|x| \le n^{(1/2)-v}$, $1/3 < v < 1/2$, but are intended as more general propositions of the kind we have exposited under the general assumption $(x/\sqrt{n}) = o(1)$ from which Theorems 1, 2, 3 would follow under the more specific assumption $x = o(n^{1/6})$.

Nekrasov (1898a) applies his Theorem 6 to the case of Bernoulli trials with constant success probability $p = 1 - q$. First we need to say that if we denote by $\{\bar{S}_n\}$ the partial sums of i.i.d. random variables $X_i$, $i = 1, 2, \ldots$ where $\Pr(X_1 = 1) = p = 1 - \Pr(X_1 = 0)$ and specialize his (local limit theorem for large deviations) Theorem 1 to this case, we deduce, after considerable manipulation, that for any $r \equiv r(n)$ satisfying $|r - np|/(npq)^{1/2} \le n^{(1/2)-v}$, where $1/3 < v < 1/2$, that

$$\Pr\{\bar{S}_n = r\} = \Pr\{S_n = -np + r\} \sim (2\pi npq)^{-1/2} \exp\{-(r - np)^2/(2npq)\}$$

as $n \to \infty$. Recourse to Khinchin (Khintchine (1929)) shows this conclusion is correct, certainly if $|r - np|/(npq)^{1/2} > 1$, although the restriction $v > 1/3$ is unnecessary. Theorem 6 is imprecise in its formulation; nevertheless for the Bernoulli case we may establish from its statement by tedious asymptotics, using Nekrasov's notes, the expression

$$\Pr\{\bar{S}_n \le np + t(npq)^{1/2}\} = \Pr\{S_n \le t(npq)^{1/2}\}$$
$$= (2\pi)^{-1/2} \int_{-\infty}^{t} \exp(-y^2/2)dy + \frac{\exp(-t^2/2)(1 - t^2)(p - q)}{6\sqrt{(2\pi npq)}}$$
$$+ \frac{\exp(-t^2/2)}{2\sqrt{(2\pi npq)}} + O\left(\frac{1}{n}\right)$$

under the assumption that $n'' \equiv np + t(npq)^{1/2}$ is an integer, and $t$ $(>0)$ is constant. Recourse to the detailed treatment of Uspensky (1931), p. 130, shows that this expression is also correct; in that more general result, the third term on the right is multiplied by $(1 - 2\theta)$, $\theta$ being the fractional part of $n''$, and Uspensky actually gives an error bound instead of an '$O$' term. The idea of expansions of this kind for the binomial distribution go back to Laplace in 1812, and they have their genesis for general distributions in Chebyshev (Tchébichef (1887)); Markov later makes much of the application of Chebyshev's ideas to the binomial case in Markov (1914a), but makes no mention of Nekrasov and his more general setting. This would seem to be a consequence of his own and Liapunov's refutations of Nekrasov's work in the interim, which, it is clear from the above, were not totally justified.

There is one more thing to note: in relation to his Theorems 1 to 3, Nekrasov's definition of $R_1(n)$ by taking $w_0 = 0$ in the expression for $R_{(n)}(y)$ as given above; let us call the corresponding value $\tilde{R}_1(n)$ $(<1)$. Then he accordingly requires that $\{\tilde{R}_1(n)\}^n \to 0$ as $n \to \infty$. It is not easy to see what Nekrasov has in mind with such a condition, in modern terms. A plausible modern condition of similar product-appearance, if we write $M_{(n)}(s) = E(s^{S_n})$, is

$$\limsup_{n\to\infty} |M_{(n)}(\exp(iy_n))| \equiv \limsup_{n\to\infty} \prod_{j=1}^{n} |\phi_i(\exp(iy_n))| < 1 \tag{6.6}$$

as $n \to \infty$, for any sequence $\{y_n\}$, $|y_n| \to \infty$. This, of course, cannot be satisfied for *lattice* random variables.

## 7. Nekrasov, Markov and Liapunov

We have noted in our Section 5 that Nekrasov's full treatment of the topic of his 1898 paper appeared in volumes **21** (1900–1901), **22** (1901–1902) and **23** (1902) of *Matematicheskii Sbornik*; and in our Section 1 that at least the first two sets of writings are cautiously mentioned by Liapunov in his ultimate paper (Liapounoff (1901a)). We recall also that Nekrasov's (1898a) summary paper was dedicated to the memory of Chebyshev; and mention in passing that it was followed by a paper by Nekrasov (1898b) dealing exclusively with the aspect of Bernoulli trials. We need to recall also, from our Section 1, Markov's (1898a) first rigorization, in correspondence with Vasiliev, of Chebyshev's version of the central limit theorem, and his follow-up paper (Markoff (1898b)).

Nekrasov (1899), writing in the same Kazan mathematical journal with which Vasiliev was closely associated, claims that his (1898a) publication preceded both of Markov's contributions, that he sent Markov a copy of it, and states that since Markov does not mention the similarities, he must needs indicate them himself. He claims also that Markov's main results can be deduced from his own (local limit) Theorem 1, and its conditions. He states that his condition $(\tilde{R}_1(n))^n \to 0$ as $n \to \infty$ *implies* Markov's condition ((iiia) in our Section 1), and supports this with some excitably written mathematics. He also responds to criticism by 'one critic' that in his (1898a) paper Nekrasov mentions Chebyshev's (Tchébichef (1867)) paper, but not Tchébichef (1887) on the central limit problem, that he had done so for brevity and because he used methods which he regards as more productive than Chebyshev's. He also explicitly states that a global limit theorem (such as in Chebyshev's paper) does not suffice for all statistical applications, and a local limit theorem is necessary. There is a footnote which points out that his results give probability $\Pr\{9800 \le X \le 10200\}$—where $X$ is the number of successes in 20000 tosses of a fair coin, as 0·995330 with an error which in absolute value is smaller than 0·0001, and that nobody had hitherto succeeded in obtaining such accuracy. This reference is to pp. 585–586 of Part I of 'Novie osnovania...'.

There are really two notes by Markov in the same journal. The first, Markov (1899a) notes (it would seem in ignorance of Nekrasov (1899)), that in his follow-up papers on Bernoulli trials, Nekrasov (1898b) obtained the numerical value of the above probability as 0·9953301 with an error of less than 0·000085; but his own method of attacking such problems by *continued fractions* and thereby obtaining actual *bounds* on the required probability gives the interval (0·995424, 0·995428). He also notes that Laplace's formula (with correction term) gives 0·9954256. The paper (Markov (1899a)) is interesting in that his approach has to some extent passed into the folklore-methodology of the subject of the binomial distribution (see Uspensky (1937), p. 52, for example), but its motivation has been forgotten. At the time of the second response, Markov (1899b) has obviously read Nekrasov's (1899a) comments, and the note is written in a very sharp tone. He gives as a counterexample to Nekrasov's claim a probability distribution for $U_k$ concentrated on the points $(+1, -1, 2^{-k}, -2^{-k})$ with probabilities, respectively, $\frac{1}{2}(1-p, 1-p, p, p)$ where $p$ is independent of $k$ and $0 < p < \frac{1}{2}$, for which $EU_n^2$ is bounded away from 0, while

$$M_{(n)}(e^{iy}) = \prod_{j=1}^{n} \{(1-p)\cos y + p\cos(y/2^j)\}$$

which attains a local maximum of $1-2p$ at $y = 2^n\pi$, so $(\tilde{R}_1(n))^n \ge 1-2p$. The example would not do to contradict (6.6) since $M_{(n)}(\exp(i2^{n+1}\pi)) = 1$, where we use the sequence $\{y_n\}$, $y_n = 2^{n+1}\pi$; nevertheless it is clear that this kind of condition has little to do with moment conditions, since these relate to the behaviour of characteristic functions at the *origin*. There is a response to this by Nekrasov (1900), in *Matematicheskii Sbornik* this time, which asserts that Markov's counterexample is irrelevant, since $\tilde{R}_1(n)$ must be $<1$ even in the limit as $n \to \infty$. This would seem to accord with a condition such as (6.6), which relates to Cramér's condition for similar contexts in the case where all $\phi_i(\cdot) = \phi(\cdot)$ (i.i.d. random variables):

$$\limsup_{|y|\to\infty} |\phi(e^{iy})| < 1.$$

At this stage we need to say a little about Nekrasov's full exposition of his technique 'Novie osnovania...' to which we have alluded in particular in our Section 6. The motivation is set out in the Introduction, pp. 579–586 of *Matematicheskii Sbornik* **21** (1900–1901). This is to *refine* existing results in the study of sums of independent random variables—coarse results such as Bernoulli's, Poisson's and Chebyshev's laws of large numbers; and theorems such as Chebyshev's central limit theorem, which (with oblique reference to Markov) are sometimes proved under conditions which are too restrictive†. To this end, new methods of proof are needed to yield greater precision, and complexity of technique is unavoidable. The work is to be in three parts, of increasing level of sensitivity of analysis, whose general scope is explained in terms which make it clear that a central role is to be played by Nekrasov's specialty developed in other publications listed in our bibliography: the theory of the Lagrange series and complex-variable methods. We have tried to convey the essence of what this might involve in our Section 6. A point on which much of the justification hangs is the excellent approximation to the binomial which we have mentioned above. Chapter VI of Part I (pp. 99–111 of **22** (1901–1902), entitled 'Historical and Critical Remarks') again attempts to set out his philosophy. It mounts a direct attack on the St. Petersburg school in general: Chebyshev and Markov explicitly, others implicitly, in particular on the grounds of 'Petersburg Methods' and that it was his own (1898a) paper that led to the realization of inadequacies in Chebyshev's (Tchébichef (1887)) conditions and proof of the central limit theorem, and Markov's attempt to save it

† There is no reference in any of his writing to Sleshinsky (1892).

through introduction of condition (iii) (see our Section 1). The question, for us, still exists: was Sleshinsky in 1892 aware of defects (in either aspect) in Chebyshev's treatment? He only says that the proof of Chebyshev's theorem '...cannot be called simple...'. This would seem to be a characteristically gentlemanly stance in full awareness of the problems, and rather preempts Nekrasov's claims, although not the value of his contribution in concept and methodology. The issue on which Liapunov later seizes as exemplifying the weakness of Nekrasov's argument, viz. that Markov's condition (iii) can be replaced by the weaker: $B_n^2 \to \infty$, occurs within this chapter (pp. 108–109); we have noted in our Section 1 that this claim is not, in the setting, false, although it is unlikely that Nekrasov could have established it. As a whole, Nekrasov's presentation is rambling and unclear, and we leave it to the interested reader to pursue mathematical detail.

Nekrasov's (1901) direct confrontation with Liapunov actually interrupts the presentation of his own 'Novie osnovania...' in **22** (1901) 225–238 of *Matematicheskii Sbornik*, occurring between the second and third instalments; the first and second instalments, as we have noted, are alluded to by Liapunov in his (Liapounoff (1901a)) paper. This paper of Nekrasov is an immediate attack on Liapunov's methodology. There is a reply by Liapunov (1901b) (which reference Adams ((1974), p. 110) mentions with the comment: 'Lyapunov's reply to a misdirected attack on his probabilistic investigations by P. A. Nekrasov'—the only time Nekrasov is mentioned in this work); and a further response directed at Liapunov's work on the central limit problem in the conclusion of Nekrasov's 'Novie osnovania...' in *Matematicheskii Sbornik* **23** (1902) 41–462, specifically within pp. 413–455. B. V. Gnedenko (1959), in his assessment of Liapunov's probabilistic *oeuvre*, discusses the overall exchange in §5, understatedly entitled 'Discussion with P. A. Nekrasov'. Gnedenko comments concerning Nekrasov (1901):

> 'The critical remarks of P. A. Nekrasov were so vague in terms of mathematical content and so dogmatic in character that even today they elicit only wonder and irritation.'

In essence Nekrasov asserts that Liapunov's results:

> '...contain all the main inadequacies of proof of his predecessors, indicated in detail in my cited investigation.'

He asserts that Liapunov, using in his proofs the *discontinuous Dirichlet multiplier*, has overlooked the well-known difficulties associated with the use of this multiplier in relation to the problems raised (and cites in support Markov's *Ischislenie Veroiatnostei* as well as the historico-critical section of his own 'Novie osnovania...'). Nekrasov's various *criticisms*, cited at some length by Gnedenko, are indeed unfounded; but his article (1901) contains what appear to be several clarifications of his own attempts—in particular: (p. 227) that he is indeed dealing with sums of independent lattice variables of common period (this concept would appear to be new for its time); and (p. 237) that whereas Chebyshev, Markov and Liapunov all deal with what is now called convergence in distribution, what is needed in the case of large deviations is a



consideration of asymptotic expressions. Liapunov's (1901b) reply is, according to Gnedenko,

> '...restrained in form, but very sharp in content...'.

It is, like Nekrasov's attack, largely polemical, but much to the point; and in his response at the conclusion of his 'Novie osnovania...', Nekrasov rather unwillingly, partially withdraws his remarks (pp. 441–446), while trying to save face and give the impression that his earlier work stimulated Liapunov. At one point (§168) he states that in Liapounoff (1901a), equation (8) is an assumption which Liapunov uses but does not state (and is enough to ensure that Nekrasov's own conditions are satisfied); but this claim appears not to be true.

Much of the controversy between Nekrasov and Markov/Liapunov involves aspects of analytical *approach* to the probabilistic problems of the time. Some of these are no longer familiar, at least in the form pertinent to the time, and we shall speak of a central one here which is effectively crucial as regards motivation for both Sleshinsky and Nekrasov.

This central notion is that of a 'factor of discontinuity' sometimes associated with the name of Dirichlet ('Dirichlet multiplier'), who appears to have introduced use of such factors, though he did not concern himself with problems of probability. The simplest version of a factor of discontinuity is a specific function of two variables (generally involving trigonometric functions) which is unity over portion of the range of definition of the variables and 0 elsewhere.

The Dirichlet discontinuity factor as used by Glaisher and Liapunov (see Gnedenko (1959)) is

$$I = \frac{2}{\pi}\int_0^\infty \frac{\sin ht}{t} \cos st\, dt,$$

which is unity if $h > s \geq 0$. Various modern 'inversion formulae' for characteristic functions of probability distributions may be regarded as applications of the notion. For example, for a lattice distribution on the points $a + rh$, $r = 0, \pm 1, \ldots$, the characteristic function is

$$\phi(t) = \sum_{r=-\infty}^{\infty} \Pr\{X = a + rh\} \exp(it(a + rh)), \qquad t \text{ real}.$$

If we multiply this by $\exp(-it(a + kh))$ and integrate from $-\pi/h$ to $\pi/h$ we obtain the inversion formula

$$\Pr\{X = a + kh\} = \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} \phi(t) \exp(-it(a + kh))\, dt.$$

Here the operative factor of discontinuity is

$$\frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} \exp(it(r - k)h)\, dt = \begin{cases} 0, & r \neq k \\ 1, & r = k. \end{cases}$$

Inversion formulae for passing from a characteristic function to a distribution function, or to a probability density if one exists, essentially necessitate integration over an infinite range of $t$, an objection to Liapunov's work raised by Nekrasov. A frequent objection to the use of factors of discontinuity has ever been

> 'that the result is given in the form of some definite integral or definite sum whose value is often unknown.'
>
> (Jordan (1972), p. 71)

This is a statement about the practical applicability of inversion formulae, which, however, does not detract from their use, after introduction of transforms, in the proofs of limit theorems such as the central limit theorem. Clearly, also, the present widespread use of the notion of an indicator random variable in probability theory may derive in part from the notion of a factor of discontinuity, even though such random variables are used as an analytical aid in contexts where explicit functional form for them is not required.

It is to be noted again that much of the early development was in terms of discrete (lattice) distributions. Thus, as we have already mentioned, Chebyshev's (Tchébichef (1867)) treatment of the Bienaymé–Chebyshev inequality occurs in this setting. Earlier, the characteristic function was introduced as a generating function for discrete random variables by Laplace (1812), and the specific simple 'factor of discontinuity' mentioned above is due to him. Continuing preoccupation with discrete distributions seems temporarily to have caused a drift from characteristic functions; thus, Laurent (1873), p. 5, commenting† on this work of Laplace:

> '.... le calcul des fonctions génératrices est aujourd'hui à peu pres oublié, parce que le calcul des résidus de Cauchy conduit plus sûrement au même but.'

A footnote beginning on the same page explains this in reference to a generating function

$$f(z) = \sum_{r=0}^{\infty} a_r z^r$$

with Cauchy's inversion formula

$$a_n = \frac{1}{2\pi i} \oint_C \frac{f(z)dz}{z^{n+1}} \tag{7.1}$$

the contour integral being taken along $C$ about the origin. The reader will note that this is not generally appropriate even for a lattice distribution characterized by the numbers $r = 0, \pm 1, \pm 2, \ldots$, although (7.1) continues to hold for all $n = 0, \pm 1, \pm 2, \ldots$ providing

$$f(z) = \sum_{-\infty}^{\infty} a_r z^r$$

† '...the calculus of generating functions is today almost forgotten, because the calculus of residues of Cauchy leads more surely to the same objective.'



is regular within an annulus about the origin, if $C$ lies within the annulus. Laurent's† analytical abilities and inclinations are manifested in this remarkable book of 1873, whose entire first chapter is concerned with analytical methods appropriate in probability. He uses Laplace's characteristic functions approach when he sets out to prove on pp. 98 *et seq.* Poisson's law of large numbers (that is, for independent trials, where the probability of success in the $i$th trial depends on $i$); though Sleshinsky (1892) notes that this proof contains an error on p. 103, line 13 *et seq.* which substantially alters the conclusion, and himself aims to improve, consequently, the discontinuity factor approach. One of the uses to which Laurent puts the generating function method in his introduction is the deduction (pp. 8–10) of Stirling's asymptotic formula for $n!$ through consideration of the power series for $e^x$, and inversion, with appropriate estimation procedures *en route*. It is well known that this is akin to the method of Laplacian peaks for integrals (see Bruijn, (1961), Chapter 4). Inasmuch as Nekrasov was familiar with Laurent's book (citing it in a number of places), it is not unlikely that it was an influencing factor on his own approach, in view of his own mathematical capabilities, in avoiding a discontinuity factor.

## 8. The aftermath: The law of large numbers and Markov chains

We mention in our Appendix 1 the probably tenuous association dating to before 1896 between Nekrasov and A. A. Chuprov, and Chuprov's citation of Nekrasov twice in his 1909 book *A Precis of Statistical Theory*, a fundamental influence on statistics in the Russian Empire. On pp. 167–168 (Chuprov (1959)), Nekrasov is mentioned in connection with the weak law of large numbers (WLLN), where Chuprov speaks‡ of all its mathematical formulations, beginning with Bernoulli and Laplace, and

> '. . . finishing with the more general "law of large numbers" of Poisson, the even more general theorem of Chebyshev and, most general of all, the constructions of P. A. Nekrasov and Bruns . . .'.

This sentence provoked Markov into the first item of his correspondence with Chuprov (Ondar (1977), p. 10)—a postcard which reads:

> 'I note with astonishment that . . . together with Chebyshev is mentioned P. A. Nekrasov, whose works of recent years represent an abuse of mathematics.'

In a long editorial comment on this Ondar (*loc. cit.*) is of the opinion that both Chuprov and Markov have in mind the paper Nekrasov (1898a), the source of the controversy we have discussed in Section 7. This opinion is too restrictive in regard to

† Mathieu Paul Hermann Laurent, also author of *Théorie des résidus* (1865), Gauthier-Villars, Paris, although the Laurent expansion is named after P. A. Laurent.

‡ In Nekrasov's and Chuprov's writings, as often among Russian authors of the period, the WLLN in the modern sense is often called Chebyshev's theorem, without further qualification.

Markov, who probably has in mind more the mystical connotations (see below) which Nekrasov was wont to ascribe to mathematics. It is incorrect as regards Chuprov as, firstly, consultation of the reference list of Chuprov's book reveals that the only work of Nekrasov was wont to ascribe to mathematics. It is incorrect as regards Chuprov as, mention of Nekrasov is, further, made clear at the second point (pp. 222–223) where Nekrasov is mentioned†:

> 'Attempts to enlist statistics as a foundation for the theory of will have not abated, even so. But the weaponry is now directed against those who first attempted to use it; from the fact of stability of statistical numbers deduction is made of free will.... 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 on the tenet that free will is the *conditio sine qua non* of statistical laws governing everyday life. The proof of this proposition, put forward by *P. A. Nekrasov*, is based on his analysis of the logical underpinnings of Chebyshev's theorem, which is a fundamental entity in the theoretical elucidation of the stability of arithmetic means. The proof of Chebyshev's theorem, *Nekrasov* (Filosofia i Logika Nauki o Massovikh Proiavleniakh, p. 29) points out, assumes mutual pairwise independence of those separate events considered as a whole for the calculation of arithmetic means.'

Indeed, the Bienaymé–Chebyshev inequality, in the version of Chebyshev for independent but not identically distributed random variables $X_i, i = 1, \ldots, n$, with $S_n = \sum_{i=1}^n X_i$, takes the form

$$\Pr(|S_n - E(S_n)|/n \geq \varepsilon) \leq \sum_{i=1}^n \operatorname{Var} X_i/(n\varepsilon)^2,$$

so a WLLN holds if $(\sum_{i=1}^n \operatorname{Var} X_i)/n^2 \to 0$, as $n \to \infty$. By examining the 'logical underpinnings' of the inequality, Nekrasov (1902) notices that 'pairwise independence' of the $X_i$ is *sufficient* for the same conclusions. This (with its suggestion of orthogonality) is an important advance on 'Chebyshev's theorem', and is so recognized by Chuprov (1959), p. 168, in 1909. However, Nekrasov also wishes to use observed stability of large numbers to deduce pairwise independence of the $X_i$'s (i.e. that pairwise independence is *necessary* for the WLLN), as proof of 'free will'; and Chuprov (1959), pp. 222–224, rightly perceives this position as both mathematically and philosophically invalid. It is Nekrasov's propensity to use such ideas which were sometimes even not mathematically well-founded, in mystical connotations, that partly earned him the contempt of Markov.

Markov's postcard evoked a spirited, if appropriately deferential, reply from Chuprov (Ondar (1977), p. 11, Communication No. 2), that, nonetheless, Nekrasov deserves for various reasons to be mentioned in the above connection, irrespective of

† The work of Nekrasov mentioned in this quotation is Nekrasov (1902).



his mathematical failings. Markov's reply (Ondar, pp. 11–12, Communication No. 3) is illuminating in a number of respects:

> 'Certainly, I was astonished also by your citation of Bruns, whom I regard as a nonentity.
>
> I can evaluate writings only from a purely mathematical standpoint, and from this standpoint it is clear to me, that neither Bruns, nor Nekrasov, nor Pearson have done anything worthy of attention. You speak of some extremely general formulations, but I do not find them in the writings of these authors.
>
> On the other hand, I do find really quite general theorems by authors whom you have essentially overlooked: A. M. Liapunov and A. A. Markov. The only credit P. A. Nekrasov deserves, to my way of thinking, consists in the fact that he sharply revealed his error, shared, I claim, by many, to the effect that independence is a necessary condition for the law of large numbers. This state of affairs stimulated me to elucidate in a sequence of articles that the law of large numbers and Laplace's formula may hold for dependent variables. In this manner a formulation of very wide applicability, of which P. A. Nekrasov could not even dream, was indeed achieved.
>
> I investigated quantities, associated into a simple chain, but this leads to the notion of the possibility of extending limit theorems of the probability calculus also to a complex chain.'

Nekrasov's error in the above specific connection thus stimulated Markov to give thought to the WLLN for dependent random variables (noted also by Ondar (1977), p. 12). This resulted in the paper: Markov (1906), whose last sentence reads

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

though Nekrasov is not mentioned. It is, indeed, in §5 of this paper, as N. Sapogov's commentary in Markov (1951), p. 662, points out, that 'Markov chains' first make their appearance in Markov's writings. This is probably the paper (and *not* the 1910 one of Markov mentioned as likely candidate by Ondar (1977), p. 13) an offprint of which Markov sent to Chuprov with this letter. By the time of Communication 7, (Ondar, p. 12) (from Markov to Chuprov, a few days later) Markov's attitude seems to have softened somewhat:

> 'Bohlmann's work caused me to remember, that, as was remarked by me earlier in lectures, and as was noted by P. A. Nekrasov, in the investigation of the expectation of a known squared expression, pairwise independence suffices.'

It is clear that, albeit indirectly, or even as a catalyst, Nekrasov had a significant effect on the development of probability theory, and does not altogether deserve the usual picture which is painted of him. Markov was certainly influenced by Nekrasov's writings.

We shall pursue elsewhere the focussing of Chuprov's researches onto a theoretical direction embodied in mathematical statistics, and away from practical aspects, as a result of the intense exchange of correspondence with Markov. It seems important to do this since it is, unfortunately, only the evolution of Markov's thinking as a result of this correspondence that interests reviewers as eminent as Neyman (1978), (1981). As a conclusion to this work it is relevant to quote several lines† from this review of Neyman, who had indirect acquaintance with Markov:

> 'During the early decades of the present century, Markov was somewhat conspicuous by the sharpness of his polemics and was frequently referred to as 'Andrew the Terrible' (=Neistoviy Andrey). This trait of Markov is reflected in the correspondence.... In fact, this correspondence was initiated by Markov scolding Tshuprov for mentioning the name of Necrasov next to that of Chebyshev.'

### Appendix 1. P. A. Nekrasov (1853–1924)

The information we have been able to obtain is sketchy, possibly due to his political and ideological inclinations already alluded to in the text. Sluginov's (1927) short obituary generally lacks detail. It does not even give Nekrasov's full name (nor have we been able to determine from other sources what 'P. A.' stands for) nor the years of his birth and death (which we have from Maistrov (1967)) let alone the dates and places, nor a bibliography. It may be more revealing to translate most of it rather than to give a commentary almost as long.

> 'In recent years, science has lost a whole sequence of famous savants. It was therefore all the more difficult to learn of the death of P. A. Nekrasov, professor of the 1st Moscow State University, one of the most eminent representatives of mathematical knowledge. The name of P. A. Nekrasov is widely known not only in Russia, but also beyond. Possessing a rare erudition, P. A. Nekrasov distinguished himself in various areas of pure and applied mathematics. The deceased's work included writings on the most delicate and difficult problems. He worked out with completeness the theory of the Lagrange series..., P. A. Nekrasov is well known not only in specialist circles, but also to the general public. His fine work on the theory of probability enjoys deservedly general fame. The larger part of the works of P. A. Nekrasov is contained in the Moscow journal *Matematicheskii Sbornik*, in relation to which P. A. played a most active role. Taking lively and vociferous part in the life and activity of the Moscow Mathematical Society, in which the deceased assumed in turn almost all administrative positions, P. A. through his numerous works affected very significantly the development of the printed organ of this Society on the one hand, and, on the other, the development of the scientific capacity of young researchers, connected in one way or another with the Society.

† Using Neyman's transliterations of Russian names.



> Far from restricting himself to purely academic activity... he was very interested in pedagogics, to which he devoted considerable energy and time. He wrote several pedagogical works, and talks which were presented by him at all-Russian meetings of lecturers. The name of P. A. was widely popular in pedagogical circles, ....
>
> P. A. Nekrasov belonged to that far from large number of savants for whom science represents the main aim and meaning in life.... Till his very death he continued to enrich learning by his numerous and valuable contributions, some of which have considerable general significance, such as the yet unpublished memoir "Anthropological Precis".... He helped many.... The author of the present note is not a little indebted to P. A. as his unforgettable guide for a period of a number of years, and this academic guidance did not cease to the end.
>
> The deceased stood out for his very wide outlook, and his extremely high scientific objectivity with which, regrettably, far from all savants are endowed... P. A. Nekrasov's work has deep significance, some of his original mathematical ideas being well ahead of their time....'

These comments make an interesting contrast to those of Markov's son and of Maistrov already mentioned. It is perhaps notable that the obituary was written at a time of reconstruction and detente in post-revolutionary Russia.

The following is additional to information already given in the present text and this appendix.

Maistrov (1967), Chapter V, §1, mentions that Nekrasov was rector of Moscow university, as of 1873. An examination of the review journal *Jahrbuch über die Fortschritte der Mathematik* (Berlin–Leipzig) reveals that Nekrasov's last-listed and only post-revolutionary publication, appears to be a six-page booklet in 1923, entitled (in English translation) 'New Periodic Functions' with place of origin given as Samara, which came to be known as Kuybishev in 1936. It may be significant that this town, on the Volga and the main Moscow–Siberia railway, was in 1918 the seat of the anti-Bolshevik Committee of Members of the Constituent Assembly; and that a university, founded here in 1918, was abolished in 1927.

Unlike Sleshinsky, Nekrasov published (as Sluginov remarks) prodigious amounts of material, beginning, according to the *Fortschritte*, in 1883. We have therefore not listed all his publications, which may be found in that source, and the companion review source of the times which we have used in our investigations, *Revue semestrielle des publications mathématiques* (redigée sous les auspices de la Société mathématique d'Amsterdam). He was particularly proficient in complex-variable theory and his dissertation, partly published as Nekrasov (1885a), is mentioned in our discussion of his probabilistic work. The same year sees the publication of his paper (1885b) on least squares, though in 'interpolational' setting, on convergence of the Gauss–Seidel iterative method of solution to the 'normal equations'

$$(X'X)\boldsymbol{\beta} = X'Y$$

(where $X$ is the design matrix and $Y$ the observation vector in the classical linear model) giving as solution the least-squares estimator $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. This important line of work, in collaboration with R. Mehmke, which has its conclusion in Nekrasov (1892) is discussed in our Section 3 and Seneta (1981). It represents the only favourable context within our subject area in which Nekrasov is remembered today: the work is mentioned, for example, by Ostrowski (1955), Faddeeva (1959), and, consequently, Varga (1962); and Faddeev and Faddeeva (1963).

Nekrasov was elected to the 'Moskovskoe Matematicheskoe Obschestvo' (Moscow Mathematical Society) in 1883 and became very active from about 1887; all such information may be traced from the protocols of the Society published from time to time in its organ till 1935, the journal *Matematicheskii Sbornik*. He ultimately became vice-president in 1891 and president in 1903, resigning the presidency on moving to St. Petersburg in 1905. It may be assumed, therefore, that he could publish in the *Sbornik* virtually at will. We note from the protocols that Chebyshev was a foundation member of the Society, that Liapunov was elected in about 1893, and Sleshinsky, Yaroshenko, and Tikhomandritsky in 1894; and that there were disputes between the Society and Markov on certain matters, specifically in 1895, Markov being supported by Liapunov at the time. It is unclear whether Markov was ever a member; he was not a subscriber to the *Sbornik* in 1897. An obvious further factor which may be important in the nature of the later enmity between Markov and Nekrasov is the fact that while Chebyshev, Markov and Liapunov were all ultimately elected to the St. Petersburg Academy, doubtless deservedly though perhaps in the often nepotic manner of such things, Nekrasov, in spite of his power and influence, apparently never was. It may be that partly out of resentment he may have attempted consequently to sustain (perhaps with the cooperation of the eminent number-theorist, N. V. Bugaiev) an opposing polarity (the 'Moscow School'; see Section 8) to the St. Petersburg School within the Society. There is evidence of this even in the title of the grandiose philosophical article of Nekrasov (1904–1906). Relations, at any rate, seemed cordial enough at an earlier time, for according to the selected works of Markov (1951), p. 686, in 1892, Nekrasov read a paper of Markov's to the Society.

Although Nekrasov's first writings on probability were published earlier it is his paper of 1898a, dedicated to the memory of Chebyshev, which is the beginning of a long and bitter controversy between himself and Markov (and also Liapunov) which lasted till 1915, thus spanning the crucial years 1898–1901 of the central limit problem. Some details have been considered in Section 7.

Nekrasov seemed to enjoy controversies; he was also involved in one of a philosophical nature within the ramifications of the Moscow Psychological Society. The relevant polemics appear in *Voprosy Filosofii i Psikhologii* (**68–70**, 1903, Moscow).

The contact of the eminent statistician A. A. Chuprov† with Nekrasov seems to date to before 1896, when Chuprov finished his baccalaureate studies at the faculty of mathematics at the University of Moscow (Tschetwerikoff (1926)). His baccalaureate thesis, The Mathematical Bases of Statistics, was examined by Nekrasov, then in charge of the course in the theory of probabilities. Nekrasov, writes Chetverikov (Tschetwerikoff, p. 315), was only interested

> '. . . in the calculations within the work insofar as Chuprov brought to the forefront logic as one of the precise bases for the adoption of probability theory to statistical methodology.'

† Whose contributions to statistics will be treated elsewhere.



but implies that Nekrasov had no influence on Chuprov. In 1909, Chuprov's magisterial dissertation (A Precis of Statistical Theory, for which the University of Moscow awarded the grade of doctor) was published (Chuprov (1959)). In this Nekrasov is mentioned twice (pp. 167–168, 222–223); we have discussed this in Section 8. Chuprov's contact with Markov only began as a result of Markov's reaction to these citations, and his own assessment of Nekrasov's work seems a more rational one.

## Appendix 2. I. V. Sleshinsky (1854–1931)

The correct (latinized Polish) version of Sleshinsky's name is Jan Sleszyński; an English transliteration of the Russian version (the middle name being the patronymic) is Ivan Vladislavovich Sleshinsky. During his life in Odessa (Novorossisk) his two probabilistic articles (Sleshinsky (1892), (1893–4)) appeared, and since during his Odessa period he published exclusively in Russian, we use this transliteration. He used the German transliteration, Sleschinsky.

He was born in Lysianka, a small town in the Kiev region of Ukraine on 11 July (old style) 1854 and died in Kraków on 9 March 1931. He finished high school (gymnasium) in Odessa in 1871 with a silver medal, and subsequently, in 1871–1875, studied mathematics at the university there, and completed his course by winning a gold medal for a treatise on trigonometric series. This treatise, a subsequent one on continued fractions (not published until 1889), and finally an examination for a master's degree in pure mathematics, revealed his considerable potential. In 1880 he received a stipend of the Russian Ministry of Education and the directive (according to Russian practice) to continue his studies for two years outside Russia. He travelled to Berlin where he attended lectures by Kronecker, Kummer and Weierstrass. Here also he worked on a dissertation on variational calculus in the sense of Weierstrass, and on the basis of this gained the teaching post of 'privat-dozent' on his return to Odessa in 1883.

His published writings of the Odessa period begin in 1885, and the early ones can be seen, from his list of publications which we include in our bibliography, to be in continued fractions and analytic functions. Some bare trace of him in this connection remains: his papers Sleshinsky (1889a,b); (1890)† are mentioned by Khovansky (1956), from p. 35 of which we learn that these were unjustly criticized by Pringsheim (1898).

The substantial and important paper Sleshinsky (1892), ostensibly on linear least squares, but actually giving a rigorous proof of the central limit theorem in the manner of connection described in our Section 2, is essentially his doctoral dissertation. We have discussed it in our Section 4. This paper is mentioned briefly in the survey of Czuber (1899) without due recognition; only mentioned in a bibliography at the end of Chapter 7 (on the method of least squares) in Markov (1924), as we have noted in our Section 5, and possibly in the earlier editions, which begin in 1900, of this celebrated book; mentioned, as we have noted in our Section 1, by Liapunov (1900); and then forgotten till a brief description by Gnedenko and Gikhman (1956), pp. 490–491, and the more recent discussion, in the context of Chapter 4, in Heyde and Seneta (1977) and Seneta (1979), §5. It should be noted that this 1892 paper, like a number of his others of

† Note that the first two of these are in *Matematicheskii Sbornik*.

the Odessa period, was published in a provincial (Odessa–Novorossisk) journal. This, combined with its being written in Russian, then an even greater obstacle for the foreign reader than now, doubtless contributed in no small measure (perhaps in conjunction with Sleshinsky's personality) to its being neglected.

The work of this 1892 paper was motivated by the Bienaymé–Cauchy controversy on least squares entailing the central limit problem (Heyde and Seneta (1977), Chapter 4). We have noted Sleshinsky's early interest in continued fractions; and Chebyshev, at the time a prevailing influence in Russian mathematics, contributed directly to this area, and also used continued fractions as a tool in his interpolational writings. These in turn were connected to least squares and the controversy (Heyde and Seneta (1977), §4.5), so bearing in mind Chebyshev's (Tchébichef (1887)) work on the central limit theorem, it is evident how Sleshinsky came to choose his subject.

Sleshinsky was made ordinary professor in 1898 and until his retirement from the university in Odessa in 1909 he taught analysis continuously, and otherwise mainly calculus, number theory and probability calculus. His students included the later eminent Odessa mathematicians V. F. Kagan and S. O. Shatunovsky (see Leibman (1961)).

From 1888 Sleshinsky was very active, within the mathematical section of the Novorossisk (Odessa) Society of Natural Scientists, in the pedagogical and methodological aspects of elementary mathematics (Leibman (1961), p. 413 *et seq.*). This has the consequence that he became strongly associated with the well-known 'mathematical-educational' journal whose proper name was *Vestnik Opitnoi Fiziki i Elementarnoi Matematiki* (i.e. 'Messenger of Experimental Physics and Elementary Mathematics'), though this was often, confusingly, in the mathematical literature such as the *Fortschritte* known by a German contraction involving the name of its current editor, e.g. as '*Spaczinski's Bote*' and '*Kagan's Bote*'. The detailed history of this journal is given by Dakhia (1956). This journal, in turn, was published from 1904 by a mathematical publishing association in Odessa called 'Mathesis', which became influential in the Russian empire of the time. It also published a series of general mathematical monographs, and translations from western languages. Sleshinsky remained associated with 'Mathesis' till his departure for the Polish city of Kraków in October 1911, as his listed publications testify. It is interesting, and perhaps indicates some kind of belated connection between Markov and Sleshinsky, that the second edition (1910) of Markov's celebrated book on finite differences was published by 'Mathesis', as well as the later important article Markov (1914b).

His move to Kraków was at the invitation of the Polish Academy of Sciences, which on the occasion of a large bequest for the specific purpose of improving the quality of lectures in mathematics in this city, chose Sleshinsky as a suitably eminent mathematician. He immediately began to teach at the Jagiellonian University in much the same areas as in Odessa, and in addition in mathematical logic, a new area in Kraków, which gained wide popularity there. Sleshinsky's publications show his interest in this area from his Odessa days.

His Kraków period is notable for his role as a teacher; his lectures were meticulously prepared and continuously reworked, so that the proofs he gave, unlike those of his sources, were absolutely complete. This left him little time for additional research activity, and he was not, in any case, in the habit of publishing hastily. In fact, claiming that people were publishing too much and reading too little, at about the time of his arrival in Kraków he decided to stop publishing altogether. The article Sleshinsky (1921) was an exception which he made under strong persuasion; it may have been associated with his election to the Polish Academy as correspondent that year. Also under strong persuasion he agreed to the publication of the book *Teoria Dowodu* on the condition that he would not look at the text till it was printed; his books listed in our bibliography from this period were prepared from his lectures by his students. His pedanticism and meticulousness explain his progressive inclination towards mathematical logic and the foundations of mathematics.

Sleshinsky finally retired, at his own request, in 1924 at the age of 70, and after a short rest began to work on his favourite area, number theory.

His personality, honesty and mathematical and general culture greatly impressed people, but he does not seem to have had any appreciable influence in Poland outside Kraków. Within Kraków, he created no mathematical school, and in particular, does not seem to have inspired any disciples in probability/mathematical statistics.



## Acknowledgement

The author thanks B. L. van der Waerden for a careful reading of the manuscript, and the Australian Research Grants Committee for its support.

## Bibliography

ADAMS, W. J. (1974) *The Life and Times of the Central Limit Theorem*. Kaedmon, New York.

ALEXEJEFF (ALEKSEEV), W. G. (1904) Ueber die Entwicklung des Begriffes der höheren arithmologischen Gesetzmässigkeit in Natur- und Geisteswissenschaften. *Vierteljschr. Wiss. Philosophie u. Sociologie* Bd. **28** (N.F., Bd. **3**), Leipzig.

ALEXEJEFF (ALEKSSEV), W. G. (1905) N. W. Bugajew und die idealistischen Probleme der Moskauer mathematischen Schule. *Vierteljschr. Wiss. Philosophie u. Sociologie* Bd. **29** (N.F. Bd. **4**). [No further bibliographical information available.]

AVRINSKY, N. A. (1904) *Textbook on the Theory of Probability and its Application to the Examination of the Results from Measurements and Observations* (in Russian). St. Petersburg, 134 pp. [Information from *Fortschritte der Mathematik*. Precise Russian title unknown.]

BIENAYMÉ, I. J. (1853) Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C.R. Acad. Sci. Paris* **37**, 309–324. [Also: *Liouville's J. Math. Pures Appl.* (2) **12** (1867), 158–176.]

BODEWIG, E. (1956) *Matrix Calculus*. North-Holland, Amsterdam.

BOLZANO, B. (1910?) *Paradoksi Bezkonechnago*. Mathesis, Odessa. [Translation from German into Russian edited by I. V. Sleshinsky, in preparation in 1910.]

BRUIJN, N. G. DE (1961) *Asymptotic Methods in Analysis*, 2nd edn. North-Holland, Amsterdam.

BUNIAKOVSKY, V. IA. (1846) *Osnovania Matematicheskoi Teorii Veroiatnostei*. St. Petersburg.

CAUCHY, A. L. (1853) Mémoire sur les résultats moyens d'un très-grand nombre des observations. *C.R. Acad. Sci. Paris* **37**, 381–385.

CHEBOTAREV, A. S. (1961) From the history of the development of the method of least squares (in Russian). *Voprosy Istor. Estestvoznan. i Tekhn.* **11**, 20–28.

CHEBYSHEV, P. L. (no date) *Oeuvres*. Ed. A. Markov and N. Sonin, 1899–1907, 2 vols. Chelsea, New York.

CHEBYSHEV, P. L. (1955) *Izbrannie Trudy*. AN SSSR, Moscow.

CHUPROV, A. A. (1959) *Ocherki po Teorii Statistiki*. Moscow. [Reprinting of 1910 edition; first edition: 1909.]

COUTURAT, L. (1909) *Algebra Logiki*. Mathesis, Odessa. [Translation from French into Russian by I. V. Sleshinsky, of *L'Algèbre de la Logique*. Reprint of second edition of this published by Librairie Scientifique et Technique, Albert Blanchard, Paris, 1980.]

CZUBER, E. (1899) Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. *J. Deutschen Math. Verein.* **7** (2nd part), 1–279.

DAKHIA, S. A. (1956) "The Journal of Elementary Mathematics" and "The Messenger of Experimental Physics and Elementary Mathematics" (in Russian). *Istor.-Mat. Issled.* **9**, 537–612.

DANIELS, H. E. (1954) Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–650.

ERMAKOV, V. P. (1879) *Elementarniy Kurs Teorii Veroiatnostei.* Kiev. (First published in *Univ. Izv.* (*Kiev*), **9–12**; also published St. Petersburg 1902.)

ERMAKOV, V. P. (1905) The method of least squares (in Russian). Fiziko-matematicheskoe obschestvo. Otchet i protokoly. *Univ. Izv.* (*Kiev*) **3**, 1–22.

FADDEEVA, V. N. (1959) *Computational Methods of Linear Algebra.* Dover, New York.

FADDEEV, D. K. AND FADDEEVA, V. N. (1963) *Vichislitelnie Metodi Lineynoy Algebri*, 2nd augmented edn. Gos. Izd. Fiz.-Mat. Lit., Moscow–Leningrad.

FISZ, M. (1963) *Probability Theory and Mathematical Statistics*, 3rd edn. Wiley, New York.

GNEDENKO, B. V. (1958) On certain problems in the history of mathematics (in Russian). *Istor.-Mat. Issled.* **11**, 47–62.

GNEDENKO, B. V. (1959) On A. M. Liapunov's work in probability theory (in Russian). *Istor.-Mat. Issled.* **12**, 135–160.

GNEDENKO, B. V. (1965) Mathematics (in Ukrainian). *Ukrainska Radianska Sotsialistychna Respublika.* Red. Ukr. Rad. Entsyklopedii, AN URSR, Kiev, pp. 465–468.

GNEDENKO, B. V. AND GIKHMAN, I. I. (1956) The development of probability theory in Ukraine (in Russian). *Istor.-Mat. Issled.* **9**, 477–536.

GNEDENKO, B. V. AND SHEYNIN, O. B. (1978) Probability theory. In *Matematika XIX Veka* [Mathematics of the 19th Century], Nauka, Moscow, pp. 184–240.

GRAVE, P. P. (1889) On parabolic interpolation according to the method of least squares (in Russian). *Obschestvo Estestvoispitatelei, Sekt. Fiz.-Mat. Nauk, Univ. Kazan* **7**. [No more information available. From *Fortschritte der Mathematik*.]

HARTER, H. L. (1974–1976) The method of least squares and some alternatives. *Internat. Statist. Rev.* **42**, 147–174, 235–264; **43**, 1–44, 125–190, 269–278; **44**, 113–159.

HEYDE, C. C. AND SENETA, E. (1977) *I. J. Bienaymé: Statistical Theory Anticipated.* Springer-Verlag, New York.

HOBORSKI, H. (1931) Jan Sleszyński (in Polish). *Wiadom. Mat.* (Warszawa) **36**, 71–76.

IVERONOV, I. (1912) *The Method of Least Squares. Lectures held at the Konstantine Institute of Surveying and at the Moscow Institute of Agriculture* (in Russian). Moscow, 2nd edn, 136 pp. [Precise Russian title not known. Information from *Fortschritte der Mathematik*.]

JORDAN, K. (1972) *Chapters on the Classical Calculus of Probabilities.* Akadémiai Kiado, Budapest.

KHINTCHINE (KHINCHIN), A. (1929) Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Ann.* **101**, 745–752.

KHOTIMSKY, V. (1925) *Vyravnivanie Statisticheskikh Riadnov po Metodu Naymenshikh Kvadratov* (*Sposob Chebysheva*). Moscow–Leningrad.

KHOVANSKY, A. N. (1956) *Prilozhenie Tsepnikh Drobei i ikh Obobscheniy k Voprosam Priblizhennoho Analiza.* Gos. Izd. Tehn.-Teoret. Lit., Moscow.

KLEIBER, J. A. (1890) Empirical formulae on the best ordinates for interpolation according to the method of least squares (in Russian). *Obschestvo Estestvoispitatelei, Sekt. Fiz.-Mat. Nauk, Univ. Kazan* **8**, 232–245.

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

LAURENT, H. (1873) *Traité du Calcul des Probabilités.* Gauthier-Villars, Paris.

LEIBMAN, E. B. (1961) The mathematical section of the Novorossisk society of scientists (1876–1928) (in Russian). *Istor.-Mat. Issled.* **14**, 393–440.

LIAPOUNOFF (LIAPUNOV), A. M. (1900) Sur une proposition de la théorie des probabilités. *Bull. Acad. Imp. Sci. St. Pétersbourg* **13**, 359–386. [Russian translation in Liapunov (1954), 125–151.]

LIAPOUNOFF (LIAPUNOV), A. M. (1901a) Nouvelle forme du théorème sur la limite des probabilités. *Mém. Acad. Imp. Sci. St. Pétersbourg* **12**, No. 5, 1–24. [Russian translation in Liapunov (1954), 157–176.]

LIAPUNOV, A. M. (1901b) Reply to P. A. Nekrasov (in Russian). *Zap. Imp. Kharkov Univ.* **3**, 51–63.

LIAPUNOV, A. M. (1954) *Sobranie Sochineniy*, Vol. I. AN SSSR, Moscow.

LIAPUNOV, A. M. (1975) On the formula of Gauss for estimation of the precision of observations (in Russian). *Istor.-Mat. Issled.* **20**, 319–322.

LUKACS, E. (1960) *Characteristic Functions.* Griffin, London.

MAIEVSKY, N. (1881) *Izlozhenie Sposoba Naymenshikh Kvadratov.* St. Petersburg. [With applications to the results of gunnery.]

MAISTROV, L. E. (1967) *Teoriya Veroiatnostei: Istoricheskii Ocherk.* [Translated and edited by S. Kotz as *Probability Theory. A Historical Sketch.* Academic Press, New York, 1974.]

MARKOV, A. A. (1898a) The law of large numbers and the method of least squares (in Russian). *Izv. Fiz.-Mat. Obschestva Kazan Univ.* (2) **8**, 110–128. [Also in Markov (1951), pp. 231–251.]

MARKOFF (MARKOV), A. A. (1898b) Sur les racines de l'équation $e^{x^2} d^i e^{-x^2}/dx^i = 0$. *Bull. Acad. Imp. Sci. St. Pétersbourg* **9**, 435–446. [Also in Markov (1951), pp. 253–269.]



MARKOV, A. A. (1899a) Application of continued fractions to the calculation of probabilities (in Russian). *Izv. Fiz.-Mat. Obschestva Kazan Univ.* (2) **9**, 29–34.

MARKOV, A. A. (1899b) Reply [to P. A. Nekrasov] (in Russian). *Izv. Fiz.-Mat. Obschestva Kazan Univ.* (2) **9**, 41–43.

MARKOV, A. A. (1906) Extension of the law of large numbers to dependent quantities (in Russian). *Izv. Fiz.-Mat. Obschestva Kazan Univ.* (2) **15**, 135–156. [Also in Markov (1951), pp. 339–361.]

MARKOV, A. A. (1910) *Ischislenie Konechnikh Raznostei*, 2nd edn. Mathesis, Odessa.

MARKOV, A. A. (1912) Answer to P. A. Nekrasov (in Russian). *Mat. Sbornik* **28**, 215–227.

MARKOV, A. A. (1913) The theorem on the limit of probability in the cases of academician A. M. Liapunov (in Russian). [In Markov (1951), pp. 319–338. Extracted from later editions of Markov's *Ischislenie Veroiatnostei*.]

MARKOV, A. A. (1914a) On the problem of Jacob Bernoulli (in Russian). *Izv. Akad. Nauk St. Petersburg* (6) **8**, 237–246. [Also in Markov (1951), pp. 509–521.]

MARKOV, A. A. (1914b) The bicentenary of the law of large numbers (in Russian). *Vestnik opitnoi fiziki i elementarnoi matematiki* (*Odessa*) **603**, 59–64. [Also in Ondar (1977).]

MARKOV, A. A. (1924) *Ischislenie Veroiatnostei*, 4th, posthumous, edn. Gosizdat, Moscow. [1st edn 1900; 2nd end 1908; 3rd edn 1913.]

MARKOV, A. A. (1951) *Izbrannie Trudy.* AN SSSR, Leningrad.

MEHMKE, R. (1892) On the method of Seidel used for the solution of a system of linear equations with a very large number of unknowns by means of successive approximations. (Extract from a letter of Prof. Mehmke to Prof. Nekrasov (in Russian). *Mat. Sbornik* **16**, 342–345.

MEHMKE, R. AND NEKRASOV, P. A. (1892) Solution of a linear system of equations by means of successive approximations. (From correspondence of Prof. P. A. Nekrasov with Prof. R. Mehmke.) *Mat. Sbornik* **16**, 437–459. [Two letters of Mehmke in German, two of Nekrasov in Russian.]

MERRIMAN, M. (1877–1882) A list of writings relating to the method of least squares, with historical and critical notes. *Trans. Connecticut Acad. Arts. Sci.* **4**, 151–232.

MISES, R. VON AND GEIRINGER, H. (1929) Praktische Verfahren der Gleichungsauflösung. *Z. Angew. Math. Mech.* **9**, 58–77, 152–164.

NEKRASOV, P. A. (1885a) The Lagrange series and approximative expressions for functions of very large numbers. Part I. The Lagrange series (in Russian). *Mat. Sbornik* **12**, 49–189, 315–377, 481–579, 643–725.

NEKRASOV, P. A. (1885b) The specification of unknowns by the method of least squares when there is a very large number of unknowns (in Russian). *Mat. Sbornik* **12**, 189–204.

NEKRASOV, P. A. (1885c) Cyclic equations, their relation with the method of least squares, and their application to astronomy (in Russian). *Mat. Sbornik* **12**, 377–420.

NEKRASOV, P. A. (1892) On the question of solution of a linear equation system with a large number of unknowns by means of successive approximations (in Russian). *Zap. Imp. Akad. Nauk* (Addendum) **69** (No. 5), 1–18.

NEKRASOV, P. A. (1898a) 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. (1898b) Limits for the error in approximate expressions for the probability *P* in Jacob Bernoulli's theorem (in Russian). *Mat. Sbornik* **20**, 485–534, 535–548.

NEKRASOV, P. A. (1899) On A. A. Markov's paper "The law of large numbers and the method of least squares" and my communication presented to the mathematical section of the 10th meeting of Russian scientists and physicians in Kiev (in Russian). *Izv. Fiz.-Mat. Obschestva Kazan Univ.* (2) **9**, 18–26.

NEKRASOV, P. A. (1900) On academician A. A. Markov's reply (in Russian). *Mat. Sbornik* **21**, 379–386.

NEKRASOV, P. A. (1900–1902) New foundations of the study of probabilistic sums and mean values (in Russian). [Novie osnovania uchenia....] *Mat. Sbornik* **21**, 579–764; **22**, 1–142, 323–498; **23**, 41–462.

NEKRASOV, P. A. (1901) On a simple theorem on probabilistic sums and mean values (in Russian). *Mat. Sbornik* **22**, 225–238.

NEKRASOV, P. A. (1902) *Filosofia i Logika Nauki o Massovikh Proiavleniakh Chelovecheskoy Deiatelnosti.* Moscow.

NEKRASOV, P. A. (1904–1906) The Moscow philosophical-mathematical school and its founders (in Russian). *Mat. Sbornik* **25**, 3–249.

NEKRASOV, P. A. (1911) On the foundations of the law of large numbers, the method of least squares and statistics (in Russian). *Mat. Sbornik* **27**, 433–451.

NEKRASOV, P. A. (1912) Laplace's theory of the method of least squares simplified with the aid of a theorem of Chebyshev (in Russian). *Mat. Sbornik* **28**, 228–234.

NEYMAN, J. (1978) *Historia Math.* **5**, 485–486. [Review of Ondar (1977).]

NEYMAN, J. (1981) Introduction to the English translation of the Markov/Chuprov correspondence edited by H. O. Ondar. Scientific Report ONR 81-01, Statistical Laboratory, University of California, Berkeley. [See Ondar (1977).]

ONDAR, KH. O. (ED.) (1977) *O Teorii Veroiatnostei i Matematicheskoi Statistike.* (Perepiska A. A. Markov'a i A. A. Chuprov'a.) Nauka, Moscow. [English translation by Charles and Margaret Stein: *The Correspondence between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics.* Springer-Verlag, New York, 1981.]

Ostrowski, A. (1955) Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen. *Comment. Math. Helv.* **30**, 175–210.

Petrov, V. V. (1972) *Summi Nezavisimikh Sluchainikh Velichin.* Nauka, Moscow.

Plackett, R. L. (1949) A historical note on the method of least squares. *Biometrika* **36**, 458–460.

Plackett, R. L. (1972) The discovery of the method of least squares. *Biometrika* **59**, 239–251.

Pringsheim, A. (1898) Ueber die Convergenz unendlicher Kettenbruche. *Sitzungsber. math.-phys. Kl. Kgl. Bayer. Akad. Wiss., München* **28**, 295–324.

Richter, W. (1957) Local limit theorems for large deviations (in Russian). *Teor. Veroiatn. Primenen.* **2**, 214–229.

Savich, A. N. (1857) *Prilozhenia Teorii Veroiatnostei k Vichisleniu Nabliudeniy i Geodezicheskikh Vichisleniy.* St. Petersburg. [In German translation: A. Sawitsch: *Die Anwendungen der Wahrscheinlichkeitstheorie auf die Berechnung der Beobachtungen und geodätischer Messungen oder die Methode der Kleinsten Quadrate.* Mitau, 1863.]

Schneider, H. (1977) Olga Taussky-Todd's influence on matrix theory and matrix theorists. *Linear and Multilinear Algebra* **5**, 197–224.

Seal, H. L. (1967) The historical development of the Gauss linear model. *Biometrika* **54**, 1–24.

Seidel, L. (1874) Über ein Verfahren, die Gleichungen . . . durch successive Annärung aufzulösen. *Abh. Bayer. Akad. Wiss. Math.-Nat. Kl.* **11**, 81–108.

Seneta, E. (1979) Round the historical work on Bienaymé. *Austral. J. Statist.* **21**, 209–220.

Seneta, E. (1981) Least squares in pre-revolutionary Russia. In *History of Mathematics. Proceedings of the First Australian Conference*, ed. J. N. Crossley, Monash University, Victoria, pp. 168–174.

Sheynin, O. B. (1975) Observations on the article of A. M. Liapunov 'On the formula of Gauss for estimation of the precision of observations' (in Russian). *Istor.-Mat. Issled.* **20**, 323–328.

Sleshinsky, I. V. (1886) On the question of the development of continued fractions of analytical functions (in Russian). *Zap. Mat. Otdel. Novorossiyskago Obschestva Estestvoispitatelei (Odessa)* **6**, 33–104.

Sleshinsky, I. V. (1888a) On the convergence of continued fractions (in Russian). *Zap. Mat. Otdel. Novorossiyskago Obschestva Estestvoispitatelei (Odessa)* **8**, 97–127.

Sleshinsky, I. V. (1888b) Proof of the existence of some limits (in Russian). *Zap. Mat. Otdel. Novorossiyskago Obschestva Estestvoispitatelei (Odessa)* **8**, 129–137.

Sleshinsky, I. V. (1889a) On the question of convergence of continued fractions (in Russian). *Mat. Sbornik* **14**, 337–343.

Sleshinsky, I. V. (1889b) Appendix to the note on the convergence of continued fractions (in Russian). *Mat. Sbornik* **14**, 436–438.

Sleshinsky, I. V. (1890) On the convergence of continued fractions (in Russian). *Zap. Mat. Otdel. Novorossiyskago Obschestva Estestvoispitatelei (Odessa)* **10**, 201–256.

Sleshinsky, I. V. (1892) On the theory of the method of least squares (in Russian). *Zap. Mat. Otdel. Novorossiyskago Obschestva Estestvoispitatelei (Odessa)* **14**, 201–264.

Sleshinsky, I. V. (1893–1894) On Chebyshev's theorem (in Russian). *Zap. Novorossiyskago Univ. (Odessa)* **59**, 503–506.

Sleshinsky, I. V. (1893a) On the theory of analytical functions (in Russian). *Mat. Sbornik* **17**, 223–228.

Sleshinsky, I. V. (1893b) *Logicheskaya Mashina Jevons'a.* Edns. of *Vestnik Opitnoi Fiziki i Elementarnoi Matem.* Mathesis, Odessa.

Sleshinsky, I. V. (1897) Obituary of Prof. Weierstrass (in Russian). *Vestnik Opitnoi Fiziki i Elementarnoi Matematiki (Odessa)* ['*Spaczinski's Bote*'] **255**, 59–62.

Sleshinsky, I. V. (1899) Report on the following thesis by Mr Zimmermann (in Russian). *Zap. Novorossiyskago Univ. (Odessa)* **77**, 13–18.

Sleshinsky, I. V. (1900) Report on the thesis of Mr Timchenko: Foundations of the Theory of Analytical Functions (in Russian). *Zap. Novorossiyskago Univ. (Odessa)* **79**, 27–32.

Sleshinsky, I. V. (1903) The life and works of N. Abel (in Russian). *Vestnik Opitnoi Fiziki i Elementarnoi Matematiki (Odessa)* ['Spaczinski's Bote'] **344**, 169–176; **345**, 193–205.

Sleshinsky, I. V. (1909) In memory of P. S. Poretsky. (Obituary) (in Russian). *Vestnik Opitnoi Fiziki i Elementarnoi Matematiki (Odessa)* ['Kagan's Bote'] **487**, 145–148.

Ślеszyński (Sleshinsky), J. (1912) Sur l'interversion de l'ordre des variables dans le passage à la limite. *Bull. Internat. Acad. Sci. Cracovie* **1–4**, 380–384.

Ślеszyński (Sleshinsky), J. (1921) Sur le raissonnement dans les sciences déductives. *Rozprawy Polskiego Towazystwa Matematicznego* **1**, 102–109. [After the first volume the title of this journal changed to *Annales de la Société Polonaise de Mathématique.*]

Ślеszyński (Sleshinsky), J. (1925) *Teoria Dowodu Podlug Wykladow Universiteckich*, Vol. 1 (ed. S. K. Zaremba). Nakladem Kolka Matematyczyno-Fizycznego U.U.J. Kraków. [Also published in 1929. Volume 2, entitled *Teoria Wyznacznikow*, also in Polish, appeared in 1926.]

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

Slutsky, E. E. (1914) On the criterion of goodness of fit of the regression line and on the best method of fitting them to the data. *J.R. Statist. Soc.* **77**, 78–84.



Tchébichef (Chebyshev), P. L. (1855) Sur les fractions continues. [Translation into French, and footnote, by I. J. Bienaymé, in *J. Math. Pures Appl.* **3** (1858), 289–323. Also in (1) his *Oeuvres* **1**, 201–230; (2) Chebyshev (1955).]

Tchébichef (Chebyshev), P. L. (1858) Sur une nouvelle série. *Bull. Phys.-Mat. Acad. Imp. Sci. St.-Pétersbourg* **17**, 257–261. [Also in his *Oeuvres* **1**, 379–384.]

Tchébichef (Chebyshev), P. L. (1859) Sur l'interpolation par la méthode des moindres carrés. *Mém. Acad. Sci. St. Pétersbourg* (8) **1**, No. 15, 1–24. [Also in (1) his *Oeuvres* **1**, 471–498; (2) Chebyshev (1955).]

Tchébichef (Chebyshev), P. L. (1864) Sur l'interpolation. [In French translation in his *Oeuvres* **1**, 539–560.]

Tchébichef (Chebyshev), P. L. (1867) Des valeurs moyennes. *Liouville's J. Math. Pures Appl.* (2) **12**, 177–184. [In Russian in *Mat. Sbornik* (2) **2**, 1–9. Also in (1) his *Oeuvres*; (2) Chebyshev (1955).]

Tchébichef (Chebyshev), P. L. (1870) Formule d'interpolation par la méthode des moindres carrés. *Mémoires couronnés et mémoires des savants étrangers publiès par l'Académie Royale des Sciences de Belgique, Bruxelles* **21**, 9 pp. [Appendix to: N. Majewski (Maievsky): Mém. sur les expériences faites à l'établissement de M. Krupp à Essen au mois de novembre 1867, pour déterminer les pressions des gaz de la poudre dans l'ame des bouches à feu.]

Tchébichef (Chebyshev), P. L. (1874) Sur les valeurs limites des intégrales. *Liouville's J. Math Pures Appl.* (2) **19**, 157–160. [Also in his *Oeuvres.*]

Tchébichef (Chebyshev), P. L. (1875) Sur l'interpolation des valeurs équidistants. [In French translation (by C. A. Possé) in his *Oeuvres* **2**, 217–242.]

Tchébichef (Chebyshev), P. L. (1887) Sur deux théorèmes relatifs aux probabilités. Supplement to *Zap. Imp. Akad. Nauk St. Petersburg* **55**, No. 6. [Also in: (1) *Acta Math.* **14** (1890–1891), 305–315; (2) his *Oeuvres*; (3) Chebyshev (1955), pp. 211–224.]

Tikhomandritsky, M. A. (1898) *Kurs Teorii Veroiatnostei.* Kharkov Imp. Universitet, Kharkov.

Tschetwerikoff (Chetverikov), N. S. (1926) Al. A. Tschouproff. *Metron* **6**, 314–320.

Uspensky, J. V. (1937) *Introduction to Mathematical Probability.* McGraw-Hill, New York.

Varga, R. S. (1962) *Matrix Iterative Analysis.* Prentice Hall, Englewood Cliffs, NJ.

Vashchenko-Zakharchenko, M. E. (1864) A brief survey of probability theory (in Russian). *Univ. Izv. (Kiev)* **1**, 59–77.

Veselovsky, N. N. (1897) *Elementary Theory of Errors of Observation and the Method of Least Squares with Application to Questions of Low Geodesy* (in Russian). Moscow, 239 pp. [Precise Russian title unknown. Information from *Fortschritte der Mathematik.*]

Waerden, B. L. van der (1977) Über die Methode der kleinsten Quadrate. *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II* **8**, 75–87.

Wendroff, B. (1966) *Theoretical Numerical Analysis.* Academic Press, New York.

Yaroshenko, S. P. (1893–1894) On the theory of the method of least squares (in Russian). *Zap. Novorossiyskago Univ. (Odessa)* **58**, 193–208.

Yarochenko (Yaroshenko), S. P. (1893) Sur la méthode des moindres carrés. *Darboux Bull. Sci. Math.* (2) **17**, 113–125.

Zabudsky, N. (1898) *Teoria Veroiatnostei i Primenenie Eia k Strelbe i Pristrelke.* St. Petersburg.

Zaremba, S. K. (1976) Private communication (in response to an inquiry about Sleshinsky). (May 13).

Zinger, V. Ya. (1862) *Sposob Naymenshikh Kvadratov.* Moscow.