Translations of

MATHEMATICAL MONOGRAPHS

Volume 47

Stochastic Approximation and Recursive Estimation

M. B. Nevel'son
R. Z. Has'minskiĭ

American Mathematical Society

10.1090/mmono/047

TRANSLATIONS OF
MATHEMATICAL
MONOGRAPHS
VOLUME 47

M. B. Nevel'son
R. Z. Has'minskiĭ

Stochastic Approximation and Recursive Estimation

American Mathematical Society
Providence, Rhode Island

СТОХАСТИЧЕСКАЯ АППРОКСИМАЦИЯ
И РЕКУРРЕНТНОЕ ОЦЕНИВАНИЕ

М. Б. НЕВЕЛЬСОН, Р. З. ХАСЬМИНСКИЙ

Издательство "Иаука" Москва 1972

Translated from the Russian by
Israel Program for Scientific Translations

Translation edited by B. Silver

2000 *Mathematics Subject Classification.* Primary 62L20;
Secondary 60J05, 93E10.

Abstract. This book is devoted to sequential methods of solving a class of problems to which belongs, for example, the problem of finding a maximum point of a function if each measured value of this function contains a random error. Some basic procedures of stochastic approximation are investigated from a single point of view, namely the theory of Markov processes and martingales. Examples are considered of applications of the theorems to some problems of estimation theory, educational theory and control theory, and also to some problems of information transmission in the presence of inverse feedback.

Library of Congress Cataloging in Publication Data CIP

Nevel'son, Mikhail Borisovich.
Stochastic approximation and recursive estimation.

(Translations of mathematical monographs ; v. 47)
Translation of Stokhasticheskaîa approksimatsiîa i rekurrentnoe otsenivanie.
Bibliography: p.
Includes index.
1. Stochastic approximation. 2. Estimation theory. I. Khas'minskiĭ, Rafail Zalmanovich, joint author. II. Title. III. Series.
QA274.N4813 519.2 76-48298
ISBN 0-8218-1597-0

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/

10 9 8 7 6 5 4 3 04 03 02 01 00

TABLE OF CONTENTS

CONTENTS

This page intentionally left blank

NOTES ON THE LITERATURE

Chapter 1

The proof of all theorems cited in §§1–4 may be found in Kolmogorov [1], Halmos [1], and Kolmorogov and Fomin [1]. The martingale and supermartingale concepts are due to Doob. Theorems 5.1 and 5.1′ and their proofs are taken from Doob [1].

Chapter 2

§1. For more details about properties of Markov processes and their transition functions, see Feller [1], Dynkin [1], Loève [1], and others.

§2. Theorem 2.1 is essentially a special case of Doob's theorem (Doob [1]) on the stopping of a supermartingale. Kolomogorov's inequality for supermartingales may also be found in Doob [1].

§3. Theorems 3.1 and 3.2 are well known. See, for example, Gihman and Skorohod [1].

§5. For related results, see, for example, ABR [1], Chapter IV.

§6. Theorems 6.1 and 6.2 were proved by Kolomogorov [1] in a more general setting.

§7. Theorem 7.1 is conceptually similar to Theorem 8 in Chapter 4 of ABR [1].

§8. For a detailed discussion of the questions related to this formulation of the problem, see ABR [1], Chapter 5.

Chapter 3

§1. For more details about the definition and properties of a continuous Markov process, see Dynkin [1].

§2. The limit passage from equation (2.1) to a continuous Markov process was considered by Bernstein [1], who, among other things, proved a theorem on the limit behavior of the transition probability as $h \to 0$.

§§3 and 4. This definition of the stochastic integral is due to Itô. For a proof of Theorem 4.1, see, for example, Gihman and Skorohod [2].

§7. A proof of Theorem 7.1 may be found in Has'minskiĭ [1].

§8. Similar results are presented in the authors' paper [1].

Chapter 4

§1. As mentioned in the text, s.a. procedures for location of the roots of regression equations were first suggested in the 1951 paper of Robbins and Monro [1], who proved convergence in mean square of the procedure (5.1) under certain assumptions. This result was generalized by Wolfowitz [1], Blum [1], Chung [1], and others. Blum [1], in particular, presented the first proof of convergence with probability 1. Theorem 1.1 and the method of proof used in the text are due to Gladyšev [1]. Multidimensional RM procedures were introduced by Blum [2]. Blum was the first author to apply the theory of supermartingales to investigate the convergence of s.a. procedures.

§2. The procedure (2.5) was proposed by Kiefer and Wolfowitz [1] in 1952. Blum [2], Derman [1], Burkholder [1] and others studied the procedure in detail and, in particular, proved that it is convergent with probability 1. No account is given here of more general s.a. procedures such as those proposed in Burkholder [1], Dvoretzky [1], etc. More detailed bibliographies may be found in the review articles of Dvoretzky [1], Fabian [3], Sakrison [3], Schmetterer [1], Loginov [1], and others.

§3. The continuous analog of the RM procedure was introduced by Driml and Nedoma [1]. They proved the convergence of the procedure under fairly general assumptions on the stochastic processes appearing on the right of the equation. Under certain assumptions, Sakrison [1] proved the convergence of the continuous analog of the Kiefer-Wolfowitz procedure. The continuous procedures considered here were defined in Has'minskiĭ [1] and Nevel'son and Has'minskiĭ [1].

§4. The first convergence theorems for multidimensional RM procedures were proved by Blum [2]; see also ABR [1], Schmetterer [1] and further bibliography cited there. Theorem 4.4 is similar to the above-mentioned result of Gladyšev [1]; Theorem 4.5 is due to Braverman and Rozonoèr (see ABR [1]). The formulation of the last theorem raises the question as to whether one can also weaken the condition $\Sigma a^2(t) < \infty$ in the proof that $X(t) \to x_0$ a.s. It turns out that this is indeed possible. For example, in some cases it is sufficient to demand that $\Sigma a^n(t) < \infty$ for some $n > 0$ (although one must then impose more stringent conditions on the "noise"). The theorems proved here for continuous RM procedures are due to the authors [1].

§5. Convergence theorems for discrete multidimensional KW procedures were proved by Blum [2], Dupač [1] and others. For the continuous case, see Nevel'son and Has'minskiĭ [1].

Chapter 5

The conjecture that a KW procedure cannot converge with positive probability to minimum points of the regression function was advanced by Fabian [1], [2]. The results presented here are based on Has'minskiĭ [1] and Nevel'son [1], [2]. The two last-mentioned papers also prove more general theorems. See also Krasulina [1].

Chapter 6

§1. Theorem 1.1 for the one-dimensional case is proved, e.g., in Loève [1]; for the multidimensional case, see Sacks [1].

§2. Conditions for the validity of the estimate $\mathsf{E}X^2(t) = O(1/t)$ as $t \to \infty$, where $X(t)$ is a discrete one-dimensional RM process, were first obtained by Chung [1]. An analog of Lemma 2.1 for discrete multidimensional RM processes was proved by Sacks [1] under the additional assumption $\mathsf{E}|G(t, x, \omega)|^2 < c < \infty$. The fact that (2.9) is a consequence of (2.8) follows from a well-known lemma of Chung [1].

§§3 and 4. Lemma 3.1 was proved by Has'minskiĭ [2]. Lemma 4.3 is due to Sacks [1].

§5. Theorem 5.3 for $\mu > 0$ is proved in Has'minskiĭ [2]. The asymptotic normality of a RM procedure was established, under certain assumptions, by Holevo [1].

§6. Theorem 6.1 was proved by Sacks under the following additional assumptions: a) the matrix B may be reduced to diagonal form by a similarity transformation involving an orthogonal matrix; b) the norm of the matrix $A(t, x)$ is bounded for $t \geqslant 1$ and $x \in E_l$. Theorem 6.3 is new. Asymptotic normality theorems for other s.a. procedures were proved by Fabian [5], Derman [1], Burkholder [1], Dupač [1] and others. The "truncation" method in asymptotic normality proofs for s.a. procedures was first applied by Hodges and Lehman [1].

§7. The first convergence theorems for moments of s.a. processes were proved by Chung [1] in the one-dimensional case. For the multidimensional case, the result of Theorem 7.2 is well known (see, for example, Schmetterer [1]) for $\lambda > \frac{1}{2}a$. But if Condition (α) of §7 holds with $\lambda < \frac{1}{2}a$, the only result known to us on convergence of moments is that of Sakrison [2]. Constructing a certain modification of the RM procedure, Sakrison proved that the matrix of second moments of the process $\sqrt{t}(X(t) - x_0)$ converges to the corresponding covariance matrix of a normal law, provided all moments $\mathsf{E}|G(t, x, \omega)|^p$, $p = 1, 2, \ldots,$ are bounded.

Chapter 7

§§1–4. Albert and Gardner [1] established convergence and asymptotic normality for the procedure (1.1) and a more general one in which the factor $a(t)$ is allowed to depend on past observations. These authors also considered a multidimensional modification of (1.1). Recently, Csibi [1], [2] has proved convergence of truncated procedures for a broad range of truncations.

§5. There is a considerable literature on the optimal choice of parameters for RM procedures and their adaptive modifications. Among others, we mention Chung [1], Dvoretzky [1], Kesten [1], Cypkin [1], [2], and Stratonovič [2]. Our approach is based on the work of Venter [1], who proved Theorem 5.1 under slightly more restrictive assumptions.

§6. Theorem 6.1 slightly generalizes a result of Venter [1]; Theorem 6.2 is new.

Chapter 8

§1. Theorem 1.1 follows from results of Chapman and Robbins [1] and Kagan [1]. Theorem 1.2 was proved by Ibragimov and Has'minskiĭ [2]. These papers also derive inequalities for biased estimates.

§2. Theorem 2.1 and its analog for biased estimates were proved by Ibragimov and Has'minskiĭ [2].

§3. The use of RM procedures for parametric estimation is discussed in Albert and Gardner [1], where a larger class of estimates, not necessarily forming Markov processes, is considered.

§§4 and 5. Theorems 4.1 and 4.5 were proved by Sakrison [2], [3] under stronger conditions (and without the convergence of distributions). Other optimality criteria were considered by Albert and Gardner [1], Cypkin [1], and Stratonovič [2], but their procedures generally depend on past observations (see, for example, the procedure (0.9) in the Introduction).

Chapter 9

§1. For inequality (1.7), see, for example, Van Trees [1].

§2. A procedure equivalent to (2.4) for m linear in the parameter ($m(x_0) = mx_0$, where m is a nonsingular matrix) was considered by Holevo [1]. He also considered the case that m is not a square matrix but mm^* is nonsingular. For the nonlinear case, Holevo proposed a procedure more complicated than (2.4). Under certain assumptions, he established asymptotic normality of the estimation procedure.

§3. Recursive procedures similar to (3.1) may be derived from the Bayesian approach, using linearization. For linear systems, similar estimation procedures may be found in Kalman and Bucy [1], Lipcer and Širjaev [1], and elsewhere.

§4. The problem of estimating x_0 based on observations of type (4.7) (estimation of linear regression coefficients) has been considered from other points of view in many papers.

§5. With suitable initial conditions, the procedure (5.3), (5.4) for a linear function $m(t, x) = m(t)x$ coincides with the Kalman-Bucy optimal linear filtering equations. In other words, this system is satisfied, in particular, by the conditional expectation of x_0 given $Y(s)$, $s < t$, provided the a priori distribution of x_0 is Gaussian.

BIBLIOGRAPHY

M. A. AĬZERMAN, È. M. BRAVERMAN AND L. I. ROZONOER (ABR)

1. *The method of potential functions in machine learning theory*, "Nauka", Moscow, 1970. (Russian)

ARTHUR E. ALBERT AND LELAND A. GARNDER, JR.

1. *Stochastic approximation and nonlinear regression*, MIT Press, Cambridge, Mass., 1967. MR 36 #7261.

SERGE BERNSTEIN

1. *Principes de la théorie des équations différentielles stochastiques*, Trudy Fiz.-Mat. Inst. Steklov. Otd. Mat. 5 (1934), 95–124.

DAVID BLACKWELL AND M. A. GIRSHICK

1. *Theory of games and statistical decisions*, Wiley, New York; Chapman and Hall, London, 1954. MR 16, 1135.

JULIUS R. BLUM

1. *Approximation methods which converge with probability one*, Ann. Math. Statist. 25 (1954), 382–386. MR 15, 973.
2. *Multidimensional stochastic approximation methods*, Ann. Math. Statist. 25 (1954), 737–744. MR 16, 382.

D. L. BURKHOLDER

1. *On a class of stochastic approximation processes*, Ann. Math. Statist. 27 (1956), 1044–1059. MR 19, 71.

DOUGLAS G. CHAPMAN AND HERBERT ROBBINS

1. *Minimum variance estimation without regularity assumptions*, Ann. Math. Statist. 22 (1951), 581–586. MR 13, 367.

K. L. CHUNG

1. *On a stochastic approximation method*, Ann. Math. Statist. 25 (1954), 463–483. MR 16, 272.

HARALD CRAMÉR

1. *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, N. J., 1946. MR 8, 39.

S. CSIBI

1. *Simple and compound processes in iterative machine learning*, Lectures delivered at the International Summer School, Trieste, 1971.
2. *Learning optimal decision functions recursively*, Second Internat. Sympos. Information Theory (Tsahkadsor, 1971), Abstracts of papers, Izdat. Akad. Nauk SSSR, Moscow, 1971, pp. 292–294. CCA 7 #12990.

JA. Z. CYPKIN

1. *Adaptation and learning in automatic systems*, "Nauka", Moscow, 1968; English transl., Academic Press, New York, 1971.
2. *Foundations of the theory of learning systems*, "Nauka", Moscow, 1970; English transl., Academic Press, New York, 1973.

CYRUS DERMAN

1. *An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure*, Ann. Math. Statist. 27 (1956), 532–536. MR 17, 1218.

A. G. D'JAČKOV AND M. S. PINSKER

1. *The optimal linear method of transmission over a memoryless Gaussian stationary channel with complete feedback*, Problemy Peredači Informacii 7 (1971), vyp. 2, 38–46 = Problems of Information Transmission 7 (1971), 123–129. MR 46 #8741.

J. L. Doob

1. *Stochastic processes*, Wiley, New York; Chapman and Hall, London, 1953. MR 15, 445.

Miloslav Driml and Jiří Nedoma

1. *Stochastic approximations for continuous random functions*, Trans. Second Prague Conf. Information Theory, Publ. House Czechoslovak Acad. Sci., Prague, 1960; Academic Press, New York, 1961, pp. 145–158. MR 23 #A4171.

Václav Dupač

1. *On the Kiefer-Wolfowitz approximation method*, Časopis Pěst. Mat. 82 (1957), 47–75. (Czech; Russian and English summaries) MR 19, 693.
2. *A dynamic stochastic approximation method*, Ann. Math. Statist. 36 (1965), 1695–1702. MR 33 #1939.

Aryeh Dvoretzky

1. *On stochastic approximation*, Proc. Third Berkeley Sympos. Math. Statist. and Probability (1954–55), vol. I, Univ. of California Press, Berkeley, Cal., 1956, pp. 39–55. MR 18, 946.

E. B. Dynkin

1. *Markov processes*, Fizmatgiz, Moscow, 1963; English transl., Springer-Verlag, Berlin, Academic Press, New York, 1965. MR 33 #1886, 1887.

Vaclav Fabian

1. *Stochastic approximation methods*, Czechoslovak Math. J. 10 (85) (1960), 123–159. MR 22 #4091.
2. *A new one-dimensional stochastic approximation method for finding a local minimum of a function*, Trans. Third Prague Conf. Information Theory, Statistical Decision Functions, Random Processes (Liblice, 1962), Publ. House Czech. Acad. Sci., Prague, 1964, pp. 85–105. MR 29 #6569.
3. *A survey of deterministic and stochastic approximation methods for the minimization of functions*, Kybernetika (Prague) 1 (1965), 499–523. (Czech; English summary) MR 32 #5378.
4. *On the choice of design in stochastic approximation methods*, Ann. Math. Statist. 39 (1968), 457–465. MR 37 #1038.
5. *On asymptotic normality in stochastic approximation*, Ann. Math. Statist 39 (1968), 1327–1332. MR 37 #6984.

A. A. Fel′dbaum

1. *Optimal control systems*, 2nd rev. ed., "Nauka", Moscow, 1966; English transl. of 1st ed., Academic Press, New York, 1965. MR 35 #2656; 32 #9075.

William Feller

1. *An introduction to probability theory and its applications*. Vol. II, Wiley, New York, 1966. MR 35 #1048.

G. M. Fihtengol′c

1. *A course of differential and integral calculus*. Vol. II, 4th ed., Fizmatgiz, Moscow, 1959; German transl., VEB Deutscher Verlag der Wissenschaften, Berlin, 1966. MR 39 #1b.

I. I. Gihman and A. V. Skorohod

1. *Introduction to the theory of random processes*, "Nauka", Moscow, 1965; English transl., Saunders, Philadelphia, Pa., 1969. MR 33 #6689; 40 #923.
2. *Stochastic differential equations*, "Naukova Dumka", Kiev, 1968; English transl., Springer-Verlag, New York, 1972. MR 41 #7777; 49 #11625.

E. G. Gladyšev

1. *On stochastic approximation*, Teor. Verojatnost. i Primenen. 10 (1965), 297–300 = Theor. Probability Appl. 10 (1965), 275–278. MR 32 #3184.

Paul R. Halmos

1. *Measure theory*, Van Nostrand, Princeton, N. J., 1950. MR 11, 504.

R. Z. HAS'MINSKIĬ

1. *Stability of systems of differential equations under random perturbations of their parameters*, "Nauka", Moscow, 1969. (Russian) MR 41 #3925.
2. *The behavior of the processes of stochastic approximation for large time-variables*, Problemy Peredači Informacii 8 (1972), vyp. 1, 81–91 = Problems of Information Transmission 8 (1972), 62–70. MR 48 #1321.

J. L. HODGES, JR., AND E. L. LEHMANN

1. *Two approximations to the Robbins-Monro process*, Proc. Third Berkeley Sympos. Math. Statist. and Probability (1954–1955), vol. I, Univ. of California Press, Berkeley, Cal., 1956, pp. 95–104. MR 18, 947.

A. S. HOLEVO

1. *Estimates for the drift parameters for a diffusion process by the method of stochastic approximation*, Studies in the Theory of Self-adjusting Systems (V. G. Sragovič, editor), Vyčisl. Centr Akad. Nauk SSSR, Moscow, 1967, pp. 179–200. (Russian) MR 38 #1800.

I. A. IBRAGIMOV AND R. Z. HAS'MINSKIĬ

1. *The asymptotic behavior of certain statistical estimates in the smooth case*. I: *Investigation of the likelihood ratio*, Teor. Verojatnost. i Primenen. 17 (1972), 469–486 = Theor. Probability Appl. 17 (1972), 445–462. MR 46 #10106.
2. *Information inequalities and superefficient methods*, Dokl. Akad. Nauk SSSR 204 (1972), 1300–1302 = Soviet Math. Dokl. 13 (1972), 821–824. MR 47 #2733.
3. *On the approximation of statistical estimators by sums of independent variables*, Dokl. Akad. Nauk SSSR 210 (1973), 1273–1276 = Soviet Math. Dokl. 14 (1973), 883–887. MR 49 #10020.

KIYOSI ITÔ

1. *On a formula concerning stochastic differentials*, Nagoya Math. J. 3 (1951), 55–65. MR 13, 363.

KIYOSI ITÔ AND MAKIKO NISIO

1. *On stationary solutions of a stochastic differential equation*, J. Math. Kyoto Univ. 4 (1964), 1–75. MR 31 #1719.

A. M. KAGAN

1. *On the theory of Fisher's information quantity*, Dokl. Akad. Nauk SSSR 151 (1963), 277–278 = Soviet Math. Dokl. 4 (1963), 991–993. MR 27 #2032.

R. E. KALMAN AND R. S. BUCY

1. *New results in linear filtering and prediction theory*, Trans. ASME Ser. D. J. Basic Engrg. 83 (1961), 95–108. MR 38 #3076.

HARRY KESTEN

1. *Accelerated stochastic approximation*, Ann. Math. Statist. 29 (1958), 41–59. MR 20 #371.

J. KIEFER AND J. WOLFOWITZ

1. *Stochastic estimation of the maximum of a regression function*, Ann. Math. Statist. 23 (1952), 462–466. MR 14, 299; erratum, MR 14, 1278.

A. N. KOLMOGOROV

1. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, 1933; English transl., Chelsea, New York, 1950; 2nd ed., 1956. MR 11, 374; 18, 155.

A. N. KOLMOGOROV AND S. V. FOMIN

1. *Elements of the theory of functions and of functional analysis*, 2nd rev. ed., "Nauka", Moscow, 1968; English transl. of 1st ed., vols. 1, 2, Graylock Press, Albany, N. Y., 1957, 1961. MR 38 #2559; 19, 44; 22 #9565a.

N. N. KRASOVSKIĬ

1. *Certain problems in the theory of stability of motion*, Fizmatgiz, Moscow, 1959; English transl., *Stability of motion. Application of Lyapunov's second method to differential systems and equations with delay*, Stanford Univ. Press, Stanford, Cal., 1963. MR 21 #5047; 26 #5258.

T. P. KRASULINA

1. *The Robbins-Monro process in the case of several roots*, Teor. Verojatnost. i Primenen. 12 (1967), 386–390 = Theor. Probability Appl. 12 (1967), 333–337. MR 35 #5085.

L. M. LIBKIND

1. *On the probability of large deviations for a quadratic functional of a stationary Gaussian sequence*, Problemy Peredači Informacii 5 (1969), vyp. 2, 72–78 = Problems of Information Transmission 5 (1969), no. 2, 56–60.

R. S. LIPCER AND A. N. ŠIRJAEV

1. *Nonlinear filtering of Markov diffusion processes*, Trudy Mat. Inst. Steklov. 104 (1968), 135–180 = Proc. Steklov Inst. Math. 104 (1968), 163–218 (1971). MR 41 #6342.
2. *The absolute continuity with respect to Wiener measure of measures that correspond to processes of diffusion type*, Izv. Akad. Nauk SSSR 36 (1972), 847–889 = Math. USSR Izv. 6 (1972), 839–882. MR 47 #1119.

MICHEL LOÈVE

1. *Probability theory*, 2nd ed., Van Nostrand, Princeton, N. J., 1960. MR 23 #A670.

N. V. LOGINOV

1. *Methods of stochastic approximation*, Avtomat. i Telemeh. 27 (1966), no. 4, 185–204 = Automat. Remote Control 27 (1966), 706–728.

I. G. MALKIN

1. *Theory of stability of motion*, 2nd rev. ed., "Nauka", Moscow, 1966; English transl. of 1st ed., AEC Transl. #3352, U. S. Atomic Energy Commission, Oak Ridge, Tenn., 1958. MR 34 #6228; erratum, 43, p. 1695.

M. B. NEVEL'SON

1. *Certain properties of continuous procedures of stochastic approximation*, Teor. Verojatnost. i Primenen. 17 (1972), 310–319 = Theor. Probability Appl. 17 (1972), 298–308. MR 45 #6124.
2. *The convergence of continuous and discrete Robbins-Monro procedures in the case of several roots of the regression equation*, Problemy Peredači Informacii 8 (1972), vyp. 3, 48–57 = Problems of Information Transmission 8 (1972), 215–223. MR 48 #1320.

M. B. NEVEL'SON AND R. Z. HAS'MINSKIĬ

1. *Continuous procedures of stochastic approximation*, Problemy Peredači Informacii 7 (1971), vyp. 2, 58–69 = Problems of Information Transmission 7 (1971), 139–148. MR 48 #1319.
2. *On the convergence of the moments of the Robbins-Monro procedure*, Avtomat. i Telemeh. 1973, no. 1, 96–100 = Automat. Remote Control 34 (1973), 85–90.

M. S. PINSKER

1. *The probability of error in block transmission in a memoryless Gaussian channel with feedback*, Problemy Peredači Informacii 4 (1968), vyp. 4, 3–19 = Problems of Information Transmission 4 (1968), no. 4, 1–14.

C. RADHAKRISHNA RAO

1. *Linear statistical inference and its applications*, Wiley, New York, 1965. MR 36 #4668.

HERBERT ROBBINS AND SUTTON MONRO

1. *A stochastic approximation method*, Ann. Math. Statist. 22 (1951), 400–407. MR 13, 144.

JEROME SACKS

1. *Asymptotic distribution of stochastic approximation procedures*, Ann. Math. Statist. 29 (1958), 373–405. MR 20 #4886.

DAVID J. SAKRISON

1. *A continuous Kiefer-Wolfowitz procedure for random processes*, Ann. Math. Statist. 35 (1964), 590–599. MR 28 #4650.

2. *Efficient recursive estimation; application to estimating the parameters of a covariance function*, Internat. J. Engrg. Sci. 3 (1965), 461–483. MR 31 #6306.
3. *Stochastic approximation: A recursive method for solving regression problems*, Advances in Communication Systems: Theory and Applications (A. V. Balakrishnan, editor), Vol. 2, Academic Press, New York, 1966, pp. 51–106.

J. PIETER M. SCHALKWIJK
1. *Recent developments in feedback communication*, Proc. IEEE 57 (1969), 1242–1249.

J. PIETER M. SCHALKWIJK AND T. KAILATH
1. *A coding scheme for additive noise channels with feedback. I: No bandwidth constraint*, IEEE Trans. Information Theory IT-12 (1966), 172–182. MR 40 #3991.

L. SCHMETTERER
1. *Multidimensional stochastic approximation*, Multivariate Analysis, II (Proc. Second Internat. Sympos., Dayton, Ohio, 1968) (P. R. Krishnaian, editor), Academic Press, New York, 1969, pp. 443–460. MR 41 #2844.

C. E. SHANNON
1. *A mathematical theory of communication*, Bell System Tech. J. 27, (1948), 397–423, 623–656. MR 10, 133.

A. V. SKOROHOD
1. *Studies in the theory of random processes*, Izdat. Kiev. Univ. Kiev, 1961; English transl., Addison-Wesley, Reading, Mass., 1965. MR 32 #3082a,b.

R. L. STRATONOVIČ
1. *Conditional Markov processes and their application to the theory of optimal control*, Izdat. Moskov. Univ., Moscow, 1966; English transl., Amer. Elsevier, New York, 1968. MR 33 #5391; 36 #4912.
2. *Does there exist a theory of synthesis of optimal adaptive, self-learning, and self-adaptive systems?* Automat. i Telemeh. **1968**, no. 1, 96–107 = Automat. Remote Control 29 (1968), 83–92. MR 39 #5210.

HARRY L. VAN TREES
1. *Detection, estimation and modulation theory*. Part I, Wiley, New York, 1968.

J. H. VENTER
1. *An extension of the Robbins-Monro procedure*, Ann. Math. Statist. 38 (1967), 181–190. MR 34 #5225.

M. T. WASAN
1. *Stochastic approximation*, Cambridge Univ. Press, New York, 1969. MR 40 #975.

J. WOLFOWITZ
1. *On the stochastic approximation method of Robbins and Monro*, Ann. Math. Statist. 23 (1952), 457–461. MR 14, 299.

JOHN M. WOZENCRAFT AND IRWIN MARK JACOBS
1. *Principles of communication engineering*, Wiley, New York, 1965.

K. S. ZIGANGIROV
1. *Transmission of messages in a binary Gaussian channel with feedback*, Problemy Peredači Informacii 3 (1967), vyp. 2, 98–101 = Problems of Information Transmission 3 (1967), no. 2, 74–78.

SUBJECT INDEX

MAIN NOTATION

$\|B\| = \sqrt{\Sigma_{ij} b_{ij}^2}$: norm of matrix $B = ((b_{ij}))$, p. 125

$\mathbf{C}_2$: class of functions $V(t, x)$ continuously differentiable with respect to t and twice continuously differentiable with respect to x, p. 57

$\mathbf{C}_2^0$: set of functions $V(x) \in \mathbf{C}_2$ with bounded second-order partial derivatives, p. 92

D_L: domain of definition of generating operator, p. 31

$D_L^{(t,x)}$: domain of definition of generating operator L at point (t, x), p. 31

E_l: euclidean l-space, p. 12

$\mathfrak{F}_t$: monotone family of σ-algebras, pp. 36, 59

$I(x)$: Fisher information matrix, pp. 182, 186

J: identity matrix, p. 124

L: generating operator, pp. 31, 67

$\mathbf{L}_2[a, b]$: set of measurable random functions $f(t, \omega)$, $t \in [a, b]$, $\omega \in \Omega$, $\mathfrak{F}_t$-measurable for every t, such that $\int_a^b f^2(t, \omega) dt < \infty$ with probability 1, p. 59

$\mathfrak{N}_t$: σ-algebra of events generated by random variables $X(u)$, $u \leqslant t$, pp. 30, 73

$\tilde{U}$: closure of set U, p. 162

$U_\epsilon(B)$: ϵ-neighborhood of set B, p. 39

$U_{\epsilon,R}(B) = V_\epsilon(B) \cap \{x: |x| < R\}$, p. 40

$V_\epsilon(B) = E_l \setminus U_\epsilon(B)$, p. 39

$\mathfrak{X}$: domain of values of unknown parameter x, p. 181

$X(t) \sim \mathfrak{N}(m, S)$: asymptotic normality of $X(t)$, p. 123

$\mathfrak{y}^{(n)} = (y_1, \ldots, y_n)$, p. 202

$\mathfrak{Z}$: domain of values of control parameter z, p. 221

$\nabla_c f(x)$: vector with coordinates $[f(x + ce_i) - f(x - ce_i)]/2c$, p. 86

$\partial\varphi/\partial x$: matrix with elements $\partial\varphi_i/\partial x_j$, p. 107

τ_G: first exit time from domain G, pp. 32, 73

$\Phi(B)$: class of functions defined on p. 40

ISBN 978-0-8218-0906-8
9 780821 809068
MMONO/47.S