Stochastic Models in Systems Analysis

Gabriel V. Orman and Irinel Radomir

Department of Mathematical Analysis and Probability,
"Transilvania" University of Braşov,
500091 Brasov, Romania
(e-mail: ogabriel@unitbv.ro)

Abstract. From some time past our interest was focused to find new possibilities for characterizing the process of generation of the words by generative systems. In our previous papers Orman[8] and Orman[9] we have introduced some numerical functions able to characterize classes of derivations according to a given generative system up to an equivalence. They are referred to as *derivational functions.* In this paper, firstly we consider equivalence classes of derivations and we establish a property of symmetry. Secondly, we shall refer to some problems concerning the reliable systems. Many and very important results have been obtained especially by A.D. Solovyev and B.V. Gnedenko. In this sense we refer to some aspects regarding to the problem of the increase of the effectiveness of stand-by systems as a way in which the stochastic-approximation techniques can be applied in practice.
Keywords: random variables, Markov chains, transition matrices, stochastic differential equations, stochastic approximation procedures.
2000 MS Classification: 60J20; 60J10: 60K10, 60K20, 68M15, 68Q45

1 Systems of transmission of information

1.1 Introduction

In the process of transmission of information a very important aspect is that of generation of the words by a generative system. In our tentative for finding new possibilities to characterize the process of generation of the words by sequences of intermediate words we have adopted a stochastic point of view involving Markov chains. Because such sequences of intermediate words (called *derivations*) by which the words are generated are finite, it results that finite Markov chains will be connected to the process. In order that our discussion should be as general as possible, the derivations are considered according to the most general class of formal grammars from the so-called *Chomsky hierarchy*, namely those that are free of any restrictions and are called *phrase-structure grammars.*

The novelty that we have introduced consists in the fact that the process of generation of the words is organized by considering the set of all the derivations according to such a grammar split into equivalence classes, each of them containing derivations of the same length (here we are not interested in the internal structure of the intermediate words of a derivation but only in its length). We remind some basic definitions and notations.

A finite nonempty set is called an *alphabet* and is denoted by Σ. A *word* over Σ is a finite sequence $u = u_1 \cdots u_k$ of elements in Σ. The integer $k \geq 0$ is *the length* of u and is denoted by $|u|$. The word of length zero is called *the empty word* and is denoted by ε. If Σ is an alphabet, let us denote by Σ^* *the free semigroup*, with identity, generated by Σ (Σ^* is considered in relation to the usual operation of concatenation).

Definition 1. *A* phrase-structure grammar *is a system* $G = (V, \Sigma, P, \sigma)$ *where*
i V *is an alphabet called* the total alphabet*;*
ii $\Sigma \subseteq V$ *is an alphabet the elements of which are called* terminal symbols *(or letters);*
iii P *is a finite subset of the Cartesian product* $[(V \setminus \Sigma)^* \setminus \{\varepsilon\}] \times V^*$. *Its elements are called productions;*
iv $\sigma \in (V \setminus \Sigma)$ *is referred to as* the initial symbol.
The elements of $V \setminus \Sigma$ *are called variables (or nonterminals).*

For y and z in V^* it is said that y *directly generates* z, and one writes $y \Rightarrow z$ if there exist the words t_1, t_2, u and v such that $y = t_1ut_2,\ z = t_1vt_2$ and $(u, v) \in P$. Then, y is said *to generate* z and one writes $y \overset{*}{\Rightarrow} z$ if either $y = z$ or there exists a sequence $(w_0, w_1, \cdots, w_j)$ of words in V^* such that $y = w_0, z = w_j$ and $w_i \Rightarrow w_{i+1}$ for each i (we write $\overset{*}{\Rightarrow}$ for the reflexive-transitive closure of $\Rightarrow$). The sequence $(w_0, w_1, \cdots, w_j)$ is called a *derivation* of *length* j and from now on will be denoted by $D(j)$. Because a derivation of length 1 is just a production we shall suppose that the length of any derivation is ≥ 2.

Now we consider the family $\mathcal{D}$ of all derivations according to our generative system. Let D_x be the class of derivations of length x in $\mathcal{D}$.

1.2 The Markov dependence case

Now we consider that a word is in a random process of generation, the equivalence classes of derivations being connected into a simple Markov chain. Obviously, it can or cannot be generated into the equivalence class D_x. Now we take into consideration only the case when a word cannot be generated by an equivalence class of derivations. Thus, if it is not generated by the class $D_x,\ x \geq 2$, then it will be generated by the class D_{x-1} with probability q and by the class D_{x+1} with probability $p = 1 - q$. Relating to the first and the last classes we suppose that it can or cannot be generated by them.

But for the case when it is not generated we put the following supplemental conditions:

1. If it is not generated by the first class D_2 then, it will be certainly generated by the next class.
2. If it is not generated by the last class D_n then, it will be certainly generated by the last but one.

We refer to such a way for generating words as being a *fork-join generation procedure.* For the other classes $D_x,\ 2 < x < n$, we suppose that a word, being in each of them, is subject to a fork-join generation procedure.

Four cases arise:
i The word will be generated by the first class and the last;
ii it will be generated by the first class but it will be not generated by the last;
iii it will be not generated by the first class but it will be generated by the last;
iv it will be not generated both by the first class and the last class.

For each of these we determine the two-step transition matrix and we come to the following result:

I. The rows of rank $i = 3, 4, \cdots, n-3$ contain, each of them, the triplet of elements $q^2, 2pq, p^2$ disposed with q^2 and p^2 on two diagonals to the left and respective to the right of the main diagonal which contains the element $2pq$.

II. The first two and the last two rows are different from a case to another.

Thus, for these rows we have:

- In the first case: $p_{11} = p_{n-1\,n-1} = 1,\ p_{21} = q,\ p_{22} = p_{n-2\,n-2} = pq,\ p_{24} = p^2,\ p_{n-2\,n-4} = q^2,\ p_{n-2\,n-1} = p$.
- In the second case: $p_{11} = 1,\ p_{21} = p_{n-1\,n-3} = q,\ p_{n-1\,n-1} = p,\ p_{22} = pq,\ p_{24} = p^2,\ p_{n-2\,n-4} = q^2,\ p_{n-2\,n-2} = p + qp$.
- In the third case: $p_{11} = q,\ p_{13} = p_{n-2\,n-1} = p,\ p_{n-1\,n-1} = 1,\ p_{22} = q + pq,\ p_{24} = p^2,\ p_{n-2\,n-4} = q^2,\ p_{n-2\,n-2} = qp$.

• In the fourth case: $p_{11} = p_{n-1\,n-3} = q,\ p_{13} = p_{n-1\,n-1} = p,\ p_{22} = q + pq,\ p_{24} = = p^2,\ p_{n-2\,n-4} = q^2,\ p_{n-2\,n-2} = p + qp.$

Thus, we obtain a common property of these four matrices that is a *specific property of symmetry* and that can be stated as follows

Theorem 1. (***Symmetry Property***). *If a word is in a random process of generation by a fork-join generation procedure, then in all cases of generation, the two-step transition matrix has* $n-5$ *successive rows each of them containing the triplet of elements* q^2, $2pq$, p^2 *symmetrically disposed as against the first two and the last two rows. Furthermore* q^2 *and* p^2 *are elements of two distinct diagonals symmetrically disposed as against the main diagonal which contains the element* $2pq$.

1.4 Absorbing and reflecting barriers

Let now be again the situation when a word is generated by a fork-join procedure and let us consider the first and the fourth cases. They will conduct us to a very interesting result. We consider only the equivalence classes of derivations by which a word is not generated.

Let us denote by A_1 the event consisting in the word being generated by the class D_2, by A_2 being generated by the class $D_3, \cdots$, by A_{n-1} being generated by the class D_n.

• In the first case the two-step transition matrix is the same with the two-step transition matrix for a particle in a random walk between two absorbing barriers known in the theory of Markov chains.

For example let us consider that a particle located on a straight line moves along the line via random impacts occurring at times $t_1, t_2, t_3, \cdots$. The particle can be at points with integral coordinates $a, a+1, a+2, \cdots, b$. At points a and b there are absorbing barriers. Each impact displaces the particle to the right with probability p and to the left with probability $q = 1 - p$ so long as the particle is not located at a barrier. If the particle is at a barrier then, it remains in the states A_1 and A_{n-1} with probability 1.

• As regards the fourth case of generation of a word by a fork-join generation procedure, the two-step transition matrix is the same with the two-step transition matrix for a particle in random walk between two reflecting barriers also known in the theory of Markov chains. The conditions remain the same as in the former case, the only difference being that if the particle is at a barrier, any impact will transfer it one unit inside the gap between the barriers.

Therefore, these cases of generation of the words by a fork-join generation procedure become of a special interest. The practical character of these cases must be also emphasized and we believe that they will be very useful in some studies concerning the generative systems.

2 Reliable systems: the increase of the effectiveness of stand-by systems

The start point is the idea that dependind on the state of the stand-by equipment, can be distinguished loaded, nonloaded and partially loaded relief. In the case of loaded relief, the stand-by unit is in the same state as the operating unit and for this reason has the same intensity of breakdowns. In the partially loaded case, the stand-by device is loaded, but not so fully as the main equipment and for this reason has a different breakdown intensity. A stand-by unit that is not loaded does not, naturally, suffer breakdown. Quite naturally, loaded and nonloaded relief are special cases of partially loaded relief.

In this sense, we shall discuss, in short, some problems and results concerning *the increase of the effectiveness of stand-by systems*, due especially to A. D. Solovyev and B. V. Gnedenko (see Solovyev[13], Gnedenko[4], Gnedenko[5], Gnedenko[6]).

As Gnedenko himself said *this problem is a basic part of the theory of stand-by systems.*

And for this reason it is to be expected to offer a specific application of the theory of stochastic processes.

There are enough situations when it is possible to have an entire device in reserve as, for example, a generator at a power station. Also it is possible to have in reserve a component of a system or even a single element. A question arises: *what is preferable, to have large units or single elements in reserve ?* An answer is given in the following theorem

Theorem 2. *If the switching of stand-by devices (units, elements, a.s.o.) is flawless, then both in the case of loaded and nonloaded relief, an increase in the scale of the stand-by system reduces non-breakdown operation of the whole system.*

2.1 The probability that the system will operate flawlessly

Now to increase the effectiveness of stand-by systems, devices that have failed are repaired. Hence it is interesting to investigate the effect of repair on increasing the reliability. It is confined ourselves to the case of one basic and one reserve system.

Will be supposed that the following conditions are fulfilled:
i on breakdown of the basic device, the stand-by unit immediately takes up the load;
ii the device that has failed undergoes repair immediately;
iii the repairs fully restore the properties of the basic device that failed;
iv the repair time is a random variable with a distribution function $G(x)$;
v the repaired device becomes a stand-by unit;
vi the period of faultless operation of the device is random and is distributed in accord with the law $F(x) = 1 - e^{-\lambda x},\ \lambda > 0$, for the basic device and in accord with the law $F_1(x) = 1 - e^{-\lambda_1 x},\ \lambda_1 \geq 0$, for the stand-by device. In particular, if the stand-by unit is nonloaded then, $\lambda_1 = 0$ and if it is loaded then, $\lambda_1 = \lambda$.

Definition 2. *It is said that the system (basic unit plus stand-by unit) breaks down if both devices go out of commission at the same time.*

Let us denote by $P(x)$ the probability that the system will operate flawlessly for a time greater than x. Also the Laplace transforms is introduced

$$g(s) = \int_0^\infty e^{-sx} dG(x), \qquad \varphi(x) = -\int_0^\infty e^{-sx} dP(x).$$

Thus the following result is found

Theorem 3. *Under the conditions i-vi before, the probability P(x) satisfies the following integral equation*

$$P(x) = e^{-(\lambda+\lambda_1)x} + (\lambda+\lambda_1)e^{-\lambda x}\int_0^x e^{-\lambda_1 z}[1 - G(x-z)]dz +$$
$$+(\lambda+\lambda_1)\int_0^x \int_0^{x-y} e^{-(\lambda+\lambda_1)y - \lambda z} P(x-y-z)dG(z)dy. \qquad (1)$$

Proof. The event we are interesting in is decomposable into three mutually independent events (flawless operation of the system during time from 0 to x):

1. During the time $(0,\ x)$ neither the basic nor the stand-by element fails. The probability of this event is

$$P_1(x) = e^{-(\lambda+\lambda_1)x}. \tag{2}$$

2. The first breakdown occurs prior to time x. The remaining element operates flawlessly up to time x. Repair of the element which has failed is not completed prior to time x. In this case, the probability of the event is as follows

$$\begin{aligned} P_2(x) &= \int_0^x (\lambda+\lambda_1)e^{-(\lambda+\lambda_1)z}e^{-\lambda(x-z)}[1-G(x-z)]dz = \\ &= (\lambda+\lambda_1)e^{-\lambda x}\int_0^x e^{-\lambda_1 z}[1-G(x-z)]dz. \end{aligned} \tag{3}$$

3. The first breakdown occurs prior to time x, the repair of this element is completed also prior to time x, during the repair period, the remaining element was functional. From the time of repair to time x, the system functioned normally. Now, the probability of the event, in this case, is

$$\begin{aligned} P_3(x) &= \int_0^x\int_0^{x-y}(\lambda+\lambda_1)e^{-(\lambda+\lambda_1)y}e^{-\lambda x}P(x-y-z)dG(z)dy = \\ &= (\lambda+\lambda_1)\int_0^x\int_0^{x-y}e^{-(\lambda+\lambda_1)y-\lambda z}P(x-y-z)dG(z)dy. \end{aligned} \tag{4}$$

But $P(x) = P_1(x) + P_2(x) + P_3(x)$ so that (1) results.■

Now one can observe that the solution of (1) is as follows

Proposition 1. *In terms of Laplace transforms, the solution of the equation (1) is given by the formula*

$$\varphi(s) = \frac{\lambda(\lambda+\lambda_1)[1-g(\lambda+s)]}{(\lambda+s)[s+(\lambda+\lambda_1)(1-g(\lambda+s))]}. \tag{5}$$

Note 1. *By virtue of the properties of the exponential distribution, the result obtained can be immediately extended to the case when there are n operating devices and one stand-by unit. All devices have the same properties namely, they have the same distribution functions for operating time and repairs. It is necessary only to replace λ by $n\lambda$ in (1) and (5).*

2.2 The expectation of the time of flawless operation

Now to calculate the expectation of the time of flawless operation of the system will be considered

$$\left[\frac{d\varphi(s)}{ds}\right]_{s=0}.$$

One gets successively

$$\begin{aligned} \left[\frac{d\varphi(s)}{ds}\right]_{s=0} &= \frac{[-\lambda(\lambda+\lambda_1)g(\lambda)][\lambda(\lambda+\lambda_1)(1-g(\lambda))]}{[\lambda(\lambda+\lambda_1)(1-g(\lambda))]^2} - \\ &- \frac{[\lambda(\lambda+\lambda_1)(1-g(\lambda))][(\lambda+\lambda_1)(1-g(\lambda))+\lambda(1-(\lambda+\lambda_1)g(\lambda))]}{[\lambda(\lambda+\lambda_1)(1-g(\lambda))]^2} = \\ &= \frac{-\lambda(\lambda+\lambda_1)^2(1-g(\lambda))^2-\lambda^2(\lambda+\lambda_1)(1-g(\lambda))}{[\lambda(\lambda+\lambda_1)(1-g(\lambda))]^2} = \\ &= -\frac{[\lambda(\lambda+\lambda_1)(1-g(\lambda))][\lambda+(\lambda+\lambda_1)(1-g(\lambda))]}{[\lambda(\lambda+\lambda_1)(1-g(\lambda))]^2}. \end{aligned}$$

Therefore

$$m = -\left[\frac{d\varphi(s)}{ds}\right]_{s=0} = \frac{\lambda + (\lambda+\lambda_1)(1-g(\lambda))}{\lambda(\lambda+\lambda_1)(1-g(\lambda))}. \tag{6}$$

Now for a nonloaded stand-by system we have $\lambda_1 = 0$, so that it results

$$m_1 = \frac{\lambda + \lambda(1-g(\lambda))}{\lambda^2(1-g(\lambda))} = \frac{2-g(\lambda)}{\lambda(1-g(\lambda))} \tag{7}$$

while for a loaded stand-by system $\lambda_1 = \lambda$ and one gets

$$m_2 = \frac{\lambda + 2\lambda(1-g(\lambda))}{2\lambda^2(1-g(\lambda))} = \frac{3-2g(\lambda)}{2\lambda(1-g(\lambda))}. \tag{8}$$

2.3 Limit theorems

But, in the most practical cases, the mean duration of repairs is considerably less than the mean time of flawless operation of the device. For this reason it was observed that some limit theorems are necessary just to give a precise and rigorous meaning to the results obtained in these situations. In the sequel we shall refer, in short, to these.

Let us suppose that the function $G(x)$ depends on a certain parameter ν and for any $\varepsilon > 0$,

$$1 - G_\nu(\varepsilon) \to 0 \tag{9}$$

as $\nu \to \infty$.

On the other hand, from (6), it is obtained that

$$g_\nu(\lambda) \to 1 \tag{10}$$

as $\nu \to \infty$.

The converse is also true because, if for any $s > 0$ we have the relation $g_\nu(s) \to 1$, as $\nu \to \infty$, then for any $x > 0$,

$$G_\nu(x) \to 1$$

as $\nu \to \infty$.

Let us denote

$$\alpha_\nu = \left(1 + \frac{\lambda_1}{\lambda}\right)(1 - g_\nu(\lambda))$$

or

$$\lambda + \lambda_1 = \frac{\lambda\alpha_\nu}{1 - g_\nu(\lambda)}. \tag{11}$$

Now, by (5) one gets

$$\varphi_\nu(\alpha_\nu s) = \frac{\lambda(\lambda+\lambda_1)[1 - g_\nu(\lambda+\alpha_\nu s)]}{(\lambda+\alpha_\nu s)[\alpha_\nu s + (\lambda+\lambda_1)(1 - g_\nu(\lambda+\alpha_\nu s))]}.$$

Then, replacing $\lambda + \lambda_1$ from (11) it follows

$$\begin{aligned}\varphi_\nu(\alpha_\nu s) &= \frac{\lambda\dfrac{\lambda\alpha_\nu}{1-g_\nu(\lambda)}[1 - g_\nu(\lambda+\alpha_\nu s)]}{(\lambda+\alpha_\nu s)\left[\alpha_\nu s + \dfrac{\lambda\alpha_\nu[1 - g_\nu(\lambda+\alpha_\nu s)]}{1 - g_{\nu(\lambda)}}\right]} = \\ &= \frac{\dfrac{\lambda^2[1 - g_\nu(\lambda+\alpha_\nu s)]}{1 - g_\nu(\lambda)}}{(\lambda+\alpha_\nu s)\left(s + \dfrac{\lambda[1 - g_\nu(\lambda+\alpha_\nu s)]}{1 - g_\nu(\lambda)}\right)}\end{aligned}$$

Therefore

$$\varphi_\nu(\alpha_\nu s) = \frac{\lambda^2 \dfrac{1 - g_\nu(\lambda + \alpha_\nu s)}{1 - g_\nu(\lambda)}}{(\lambda + \alpha_\nu s)\left(s + \lambda \dfrac{1 - g_\nu(\lambda + \alpha_\nu s)}{1 - g_\nu(\lambda)}\right)}. \tag{12}$$

Thus, the following theorem results

Theorem 4. *If the conditions (1), (5) and (6) to (12) hold then, by the condition (9), the flow of failures of a reduplicated system tends to the elementary case, given the choice of a proper unit of time.*

The effect of repair on the operational effectiveness of a system can be estimated. In this case it is natural to consider the ratio of the mean operational time of a system with repair to that without repair. From the formula (6) the former can be calculated, and from the formula

$$a_0 = \frac{2\lambda + \lambda_1}{\lambda(\lambda + \lambda_1)}$$

the latter.

The effectiveness of repair is now given by the equality

$$\begin{aligned} e_\nu &= \frac{\lambda + (\lambda + \lambda_1)(1 - g_\nu(\lambda))}{\lambda(\lambda + \lambda_1)(1 - g_\nu(\lambda))} \cdot \frac{\lambda(\lambda + \lambda_1)}{2\lambda + \lambda_1} = \\ &= \frac{\lambda + (\lambda + \lambda_1)(1 - g_\nu(\lambda))}{(2\lambda + \lambda_1)(1 - g_\nu(\lambda))}. \end{aligned} \tag{13}$$

Now let us suppose that

$$\begin{aligned} m_1(\nu) &= \int_0^\infty x dG_\nu(x) = \frac{1}{\nu} \\ m_2(\nu) &= \int_0^\infty x^2 dG_\nu(x) < +\infty \end{aligned}$$

and

$$\frac{m_2(\nu)}{m_1(\nu)} \to 0 \tag{14}$$

as $\nu \to \infty$.

It is also useful to be retained that the following theorem holds

Theorem 5. *Let us suppose that the conditions (1), (5), (6) to (13) and (14) are satisfied. Then for ν sufficiently large, the mean time of flawless operation of a system with stand-by relief is asymptotically equal to the mean time of the system under the assumption that*

$$G_\nu(x) = 1 - e^{-\nu x}.$$

Remarks. *In recent years, algorithms of the stochastic approximation type have found applications in new and diverse areas, and new techniques have been developed for proofs of convergence and rate of convergence. The actual and potential applications in signal processing have exploded. Indeed, whether or not they are called stochastic approximations, such algorithms occur frequently in practical systems for the purposes of noise or interface cancellation, the optimization of* post processing *or equalization filters in time varying communication channels, adaptive antenna systems, and many related applications.*

In such applications, the underlying processes are often nonstationary, the optimal value of the parameter of the system changes with time, and we keeps the step size *strictly away from zero in order to allow tracking. Such tracking applications lead to new problems in the asymptotic analysis ; one wishes to estimate the tracking errors and their dependence on the structure of the algorithm.*

Let us return to the condition vi in Section 2.1 above and, for an unknown parameter $\lambda > 0$, *let us observe that a distribution function* $F(x)$, *which can be the distribution function of a "system" or "item", with a "life time" for which inspections are made at time* $t_1, t_2, t_3, \cdots$, *can be defined. If the conclusion of the inspections is that the system is inoperative, then it will be repared or replaced. In any other case nothing is done. Thus, the problem is to choose the inspection plan, that is to choose the sequence* $t_1, t_2, t_3, \cdots$ *in an optimal way in a suitable sense. Such problems were discussed, among other, by J.H. Venter and J.L. Gastwirth[14], M.T. Wasan[15], P. Clément and G. Da Prato[2].*

We shall come back to these aspects with a new occasion.

Acknowledgement. This work was supported by grant CNCSIS 431.

References

[1] A.T. Bharucha-Reid. Elements Of The Theory Of Markov Processes And Their Applications. *Dover Publications, Inc.*, Mineola, New York, 1997.

[2] P. Clément and G. Da Prato. White noise perturbation of the heat equation in materials with memory. *Dynam. Systems Appl.*, 6, 441-460, 1997.

[3] S. Ginsburg. Algebraic and Automata-Theoretic Properties of Formal Languages. North Holland Pub. Co., Amsterdam, 1975.

[4] B. V. Gnedenko. On Non-loaded Duplication. *Technical Cybernetics*, No. 4, 3-12, 1964.

[5] B. V. Gnedenko. On Duplication with Renewal. *Technical Cybernetics*, No. 5, 111-118, 1964.

[6] B. V. Gnedenko. The Theory of Probability. Mir Publisher, Moscow, 1976.

[7] B. Øksendal. Stochastic Differential Equations: An Introduction with Applications, Fifth Edition, Springer-Verlag, 1998.

[8] G. V. Orman. On the numerical functions utilized in the study of a class of discrete systems. *Journal of Systems Science*, Vol. 19, No. 1, 27-37, 1993.

[9] G. V. Orman. On the Characterization of some Generative Grammars by Numerical Functions. In C. Martin-Vide, editor, *Current Issues in Mathematical Linguistics*, pages 151-160, Elsevier Science B.V., North Holland, Amsterdam, 1994.

[10] G. V. Orman. Random Phenomena and some Systems Generating Words. *Yugoslav Journal of Operation Research*, Vol. 6, No. 2, 245-256, 1996.

[11] G. V. Orman. Stochastic Methods for Generative Systems Analysis. In A. Rizzi, M. Vichi, editors, *Advances in Data Science and Classification*, pages 205-210, Springer, Berlin, 1998.

[12] G. V. Orman. Handbook Of Limit Theorems And Sochastic Approximation. Transilvania University Press, Braşov, 2003.

[13] A. D. Solovyev. On Stand-by Systems Without Renewal. In *Sbornik "Cybernetics at the Service of Comm."*, vol. 2, 83-122, 1964.

[14] J. H. Venter and J. L. Gastwirth. Adaptive statistical procedures in reliability and maintenance problems. *Technical Report* No. 99, Department of Statistics, Stanford University, 1964.

[15] M. T. Wasan. Stochastic Approximation. Cambridge University Press, 1969.