Chaotic Modeling and Simulation (CMSIM) 1: 33–47, 2019

On stochastic approximation techniques in the study of a class of systems

Gabriel V. Orman[1] Irinel Radomir[1] Sorina-Mihaela Stoian[2]

[1] Department of Mathematics and Computer Science "Transilvania" University of Braşov, 500091 Braşov, Romania
(E-mail: ogabriel@unitbv.ro)
[2] "Excelsior" Excellency Centre, Braşov, Romania

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 as, for example, *the phrase-structure grammars* known in formal languages, up to an equivalence. Some aspects regarding the complexity of discrete time systems are discussed here. Also, we shall refer, in short, to models of the Brownian motion useful in many studies regarding the *chaotic and complex systems*.
Keywords: equivalence classes, Markov processes, stochastic approximation procedures, stochastic differential equations, Brownian motion..

1 Introduction

Brownian motion, used especially in Physics, is of ever increasing importance not only in Probability theory but also in classical Analysis. Its fascinating properties and its far-reaching extension of the simplest normal limit theorems to functional limit distributions acted, and continue to act, as a catalyst in random Analysis. It is probable the most important stochastic process. As some authors remarks too, the Brownian motion reflects a perfection that seems closer to a law of nature than to a human invention.

In Physics, the ceaseless and extremely erratic dance of microscopic particles suspended in a liquid or gas, is called *Brownian motion*. It was systematically investigated by Robert Brown (1828, 1829), an English botanist, from movement of grains of pollen in water to a drop of water in oil. He was not the first to mention this phenomenon and had many predecessors, starting with Leeuwenhoek in the 17^th century. However, Brown's investigation brought it to the attention of the scientific community, hence *Brownian*.

Brownian motion was frequently explained as due to the fact that particles were alive. It is only in 1905 that kinetic molecular theory led Einstein to the first mathematical model of Brownian motion. He began by deriving its possible existence and then only learned that it had been observed.

Received: 9 July 2018 / Accepted: 16 December 2018

ISSN 2241-0503

A completely different origin of mathematical Brownian motion is a game theoretic model for fluctuations of stock prices due to L. Bachélier from 1900.

In his doctoral thesis, *Théorie de la spéculation*, Ann. Sci. École Norm. Sup., 17, 1900, 21-86, he hinted that it could apply to physical Brownian motion. Therein, and in his subsequent works, he used the heat equation and, proceeding by analogy with *heat propagation* he found, albeit formally, distributions of various functionals of mathematical Brownian motion. Heat equations and related parabolic type equations were used rigorously by Kolmogorov, Petrovsky, Khintchine.

But Bachélier was unable to obtain a clear picture of the Brownian motion and his ideas were unappreciated at the time. This because a precise definition of the Brownian motion involves a measure on the path space, and it was not until 1908-1909 when É. Borel published his classical memoir on Bernoulli trials: *Les probabilités dénombrables et leurs applications arithmétique*, Rend. Circ. Math. Palermo 27, 247-271, 1909.

But as soon as the ideas of Borel, Lebesgue and Daniell appeared, it was possible to put the Brownian motion on a firm mathematical foundation. And this was achived in 1923 by N. Wiener, in his work: *Differential space*, J. Math. Phys. 2, 131-174, 1923.

Many researchers were fascinated by the great beauty of the theory of Brownian motion and many results have been obtained in the last decades. As for example, among other things, in *Diffusion processes and their sample paths* by K. Itô and H.P. McKean, Jr., in *Theory and applications of stochastic differential equations* by Z. Schuss, or in *Stochastic approximation* by M.T. Wasan as in *Stochastic calculus and its applications to some problems in finance* by J.M. Steele.

Itô's integral and other details and related topics in Stochastic Calculus are developed among other by B. Øksendal & A. Sulem, J. M. Steele, P. Malliavin, P. Protter, D. W. Stroock.

Some topics, in this sense, will be discussed below.

2 About some results involving the Itô's formula

Firstly, is considered an example due to Z. Schuss ([20]), involving the Itô's formula.

Definition 21 *A function $f(t)$ which is independent of the increment $w(t+s)-w(t)$ for all $s>0$ is called a nonanticipating function.*

We observe that such a function depends stochastically on $w(u)$ for $u \leq t$, that is, on the *past* only. For a nonanticipating step function $f(t)$, the integral

$$\int_0^t f(s) ds\, dw(s)$$

is also a nonanticipating function.

Let us denote by $H_2[0,T]$ the class of all nonanticipating functions $f(t)$ such that

$$\int_0^T E\, f^2(t)dt < \infty.$$

It is known (according to [20]) that for any function $f(t)$ in $H^2[0,T]$, there exists a sequence $\{g_n(t)\}$ of step functions such that

$$\int_0^T |f(t) - g_n(t)|^2 dt \to 0 \ \ a.s. \tag{1}$$

as $n \to \infty$, and, also,

$$\int_0^T g_n(s)dw(s) \to (limit) = L(t) \ \ a.s. \tag{2}$$

as $n \to \infty$, uniformly for t in $[0,T]$.

If $f(t)$ is a deterministic smooth function, the integral

$$\int_0^T f(t)\, dw(t)$$

is the Stieltjes integral and, hence,

$$\begin{aligned}\int_a^b f(t)\, dw(t) = f(b)w(b) - f(a)w(a) = \\ = \int_a^b w(t)\, f'(t)dt \ \ a.s.\end{aligned} \tag{3}$$

Now, let $X(t)$ be a stochastic process satisfying the condition

$$X(t_2) - X(t_1) = \int_{t_1}^{t_2} a(t)dt + \int_{t_1}^{t_2} b(t)dt$$

for all $0 \le t_1 < t_2 < T$, where $a(t)$ and $b(t)$ are functions in $H_2[0,T]$. Then, we say that $X(t)$ has a *stochastic differential*

$$dX(t) = a(t)dt + b(t)dw(t). \tag{4}$$

Now we shall refer to the following example due to Z. Schuss ([20]).

Example 21 *Let us consider the integral*

$$\int_a^b w(t)dw(t). \tag{5}$$

But $w(t)$ is continuous a.s., so that the step functions $w_n(t)$, defined by

$$w_n(t) = \sum_{i=0}^{n-1} w(t_i)\chi_{(t_{i+1}t_i)}(t)$$

where $t_i = a + i\,\dfrac{b-a}{n}$, $i = 0, 1, 2, \cdots, n$, *converge uniformly to* $w(t)$ *in* $[a, b]$.

The integrals of $w_n(t)$ *are as follows*

$$I_n = \int_a^b w_n(t)dw(t) = \sum_{i=0}^{n-1} w(t_i)[w(t_{i+1}) - w(t_i)].$$

But

$$\begin{aligned} & w(t_i)[w(w(t_{i+1}) - w(t_i))] = \\ & = \frac{1}{2}[w^2(t_{i+1}) - w^2(t_i) - (w(t_{i+1}) - w(t_i))^2] \end{aligned}$$

so that one gets

$$\begin{aligned} I_n &= \frac{1}{2}\sum_{i=0}^{n-1}[w^2(t_{i+1}) - w^2(t_i) - (w(t_{i+1}) - w(t_i))^2] = \\ &= \frac{1}{2}[w^2(b) - w^2(a)] - \frac{1}{2}\sum_{i=0}^{n-1}(\delta_i w)^2, \end{aligned}$$

where $\delta_i w = w(t_{i+1}) - w(t_i)$.

Let us denote now

$$\eta_n = \sum_{i=1}^{n-1}(\delta_i w)^2.$$

Then, the expectation and the variance of η_n *are as follows*

$$E\,\eta_n = \sum_{i=0}^{n-1} E\,(\delta_i w)^2 = \sum_{i=0}^{n-1}(t_{i+1} - t_i) = b - a$$

and

$$D^2\,\eta_n = \sum_{i=1}^{n-1} D^2\,(\delta_i w)^2$$

for $(\delta_i w)^2$ *being independent random variables. It is shown that*

$$\eta_n \to E\,\eta_n = b - a \quad \textit{in probability as } \; n \to \infty.$$

Therefore it follows that

$$\int_a^b w_n(t)dw(t) \to \frac{1}{2}[w^2(b) - w^2(a)] - \frac{1}{2}(b-a),$$

so that

$$\int_a^b w(t)dw(t) = \frac{1}{2}[w^2(b) - w^2(a)] - \frac{1}{2}(b-a). \tag{6}$$

Now we comeback to the stochastic differential (4) and consider $X(t) = w^2(t)$. From (6) we have

$$w^2(t_2) - w^2(t_1) = 2\int_{t_1}^{t_2} w(t)dw(t) + \int_{t_1}^{t_2} 1\,dt \tag{7}$$

so that one finds

$$dw^2(t) = 1\,dt + 2w(t)dw(t). \tag{8}$$

Therefore, $a(t) \equiv 1$ and $b(t) = 2w(t)$. If $f(t)$ is a deterministic smooth function then, by (3), it results

$$f(t_2)w(t_2) - f(t_1)w(t_1) = \int_{t_1}^{t_2} f(t)dw(t) + \int_{t_1}^{t_2} w(t)f'(t)dt$$

and finally one gets

$$d[f(t)w(t)] = f(t)dw(t) + w(t)df(t),$$

where $df(t) = f'(t)$.

Let us observe that Itô's formula can be also written for systems. To this end, let now be

$$\mathbf{w}(t) = \begin{pmatrix} w_1(t) \\ \cdots \\ w_n(t) \end{pmatrix}$$

a vector of independent Brownian motions. Also, let $\mathbf{v}(\mathbf{x}, t)$ be an $n \times n$ matrix, and we consider a vector $\mathbf{b}(\mathbf{x}, t)$

$$\mathbf{b}(\mathbf{x}, t) = \begin{pmatrix} b_1(x,t) \\ \cdots \\ b_n(x,t) \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ \cdots \\ x_n \end{pmatrix}.$$

The system of stochastic differential equations

$$d\mathbf{x} = \mathbf{b}dt + \mathbf{v}d\mathbf{w} \tag{9}$$

leads to the following Itô's formula

$$df(\mathbf{x}(t), t) = Lf dt + \nabla_x f^T \mathbf{v} d\mathbf{w} \tag{10}$$

where

$$\begin{aligned} Lf &= \frac{\partial f}{\partial t} + \mathbf{b} \cdot \nabla_x f + \frac{1}{2} \sum_{i,j=1} a_{ij} \cdot \frac{\partial^2 f}{\partial x_i \, \partial x_j} \equiv \\ &\equiv \frac{\partial f}{\partial t} + Mf \end{aligned} \tag{11}$$

and

$$a_{ij} = (\mathbf{v}\,\mathbf{v}^T)_{ij}. \tag{12}$$

The backward Kolmogorov equation will take the form

$$\frac{\partial p}{\partial t} + Mp = 0;$$

while the forward Kolmogorov equation will be as follows

$$\frac{\partial p}{\partial s} + \nabla_y \cdot (\mathbf{b}p) - \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 (a_{ij}p)}{\partial y_i \partial y_j} = 0.$$

3 Some considerations regarding the Brownian motion in connection with chaotic and complex systems

Let us imagine a chaotic motion of a particle of colloidal size immersed in a fluid. Such a chaotic motion of a particle is called, usually, *Brownian motion* and the particle which performs such a motion is referred to as a *Brownian particle*. Such a chaotic perpetual motion of a Brownian particle is the result of the collisions of particle with the molecules of the fluid in which there is.

But this particle is much bigger and also heavier than the molecules of the fluid which it collide, and then each collision has a negligible effect, while the superposition of many small interactions will produce an observable effect.

On the other hand, for a Brownian particle such molecular collisions appear in a very rapid succession, their number being enormous. For a so high frequency, evidently, the small changes in the particle's path, caused by each single impact, are too fine to be observable. For this reason the exact path of the particle can be described only by statistical methods.

Thus, the influence of the fluid on the motion of a Brownian particle can be described by the combination of two forces in the following way.

1. The considered particle is much larger than the particle of the fluid so that the cumulated effect of the interaction between the Brownian particle and the fluid may be taken as having a hydrodynamical character. Thus, the first of the forces acting on the Brownian particle may be considered to be the forces of *dynamical friction*. It is known that the frictional force exerted by the fluid on a small sphere immersed in it is determined from the Stockes's law: *the drag force per unit mass acting on a spherical particle of radius a is given by $-\beta\mathbf{v}$, with $\beta = \dfrac{6\pi a\eta}{m}$*, where m is the mass of the particle, η is the coefficient of dynamical viscosity of the fluid, and $\mathbf{v}$ is the velocity of particle.
2. The other force acting on the Brownian particle is caused by the individual collisions with the particles of the fluid in which there is. This force produces instantaneous changes in the acceleration of the particle. Furthermore, this force is *random both in direction and in magnitude*, and one can say that it is a *fluctuating force*. It will be denoted by $\mathbf{f(t)}$. For $\mathbf{f(t)}$ the following assumptions are made:

i The function $\mathbf{f(t)}$ is statistically independent of $\mathbf{v}(t)$.
ii $\mathbf{f(t)}$ has variations much more frequent than the variations in $\mathbf{v}(t)$.
iii $\mathbf{f(t)}$ has the average equal to zero.

In these conditions, the Newton's equations of motion are given by the following stochastic differential equation

$$\frac{dbfv(t)}{dt} = -\beta\mathbf{v}(t) + \mathbf{f}(t) \tag{13}$$

which is called the *Langevin's equation.*

From the Langevin's equation, the statistical properties of the function $\mathbf{f(t)}$ can be obtained if its solution will be in correspondence with known physical laws. One can observe that the solution of (13) determines the *transition probability density* (in brief *the transition density*) $\rho(\mathbf{v}, t\mathbf{v}_0)$ of the random process $\mathbf{v}(t)$, which verifies the equation

$$P(\mathbf{v}(t) \in A) \,|\, \mathbf{v}(0) = \mathbf{v}_0) = \int\limits_A \rho(\mathbf{v}, t, \mathbf{v}_0) d\mathbf{v}. \tag{14}$$

Now, the initial velocity $\mathbf{v}_0$ can be supposed to be given. Then, one gets

$$\rho(\mathbf{v}, t, \mathbf{v}_0) \to \delta(\mathbf{v} - \mathbf{v}_0)$$

as $t \to 0$ where δ is the $Dirac's$ δ-$function$. On the other hand, from the statistical physics it is known that the transition density $\rho(\mathbf{v}, t, \mathbf{v}_0)$ must approach the Maxwell's density for the temperature T of the surrounding medium and this, independently of $\mathbf{v}_0$ as $t \to \infty$. We come to the limit

$$\rho(\mathbf{v}, t, \mathbf{v}_0) \to \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{m|\nu|^2}{2kT}} \tag{15}$$

as $t \to \infty$. This means, in other words, that the fluctuating force $\mathbf{f}(t)$ has certain statistical properties. For the formal solution is as follows (according to (13))

$$\mathbf{v}(t) = \mathbf{v}_0 \, e^{-\beta t} + \int\limits_0^t e^{-\beta(t-s)} \, \mathbf{f}(z) \, dz. \tag{16}$$

Therefore, the integral and the difference $\mathbf{v}(t) - \mathbf{v}_0 \, e^{-\beta t}$ must have the same statistical properties. Since

$$\mathbf{v}(t) - \mathbf{v}_0 \, e^{-\beta t} \approx \mathbf{v}(t)$$

for latge values of t, it results that the integral must have in the limit a normal density. But the integral can be written as a finite Riemann sum in the following way

$$\int\limits_0^t e^{-\beta(t-s)} \, \mathbf{f}(z) \, dz \approx$$
$$\approx e^{-\beta t} \sum_n e^{\beta n \,\Delta t} \, \mathbf{f}(n \,\Delta t) \Delta t \approx e^{-\beta t} \sum_n e^{\beta n \,\Delta t} \, \Delta \mathbf{g}_n$$

where was denoted $\Delta\mathbf{g}_n = \mathbf{f}(n\,\Delta t)\Delta t$. Hence, for large values of t, the following approximation is found

$$\mathbf{v} \approx \sum_n e^{\,\beta(n\,\Delta t - t)}\,\Delta\mathbf{g}_n. \tag{17}$$

Here $\Delta\mathbf{g}_n$ is a random variable which gives the random accelerations transmited to a Brownian particle in an inteval of time $(n\,\Delta t), (n+1)\Delta t$. Therefore, the random variables $\Delta\mathbf{g}_n$ can be assumed to be statistically independent of each other, the successive collisions being completely chaotic.

One can assume that, in comparison with the average period of a single fluctuation of the function $\mathbf{g}_n$, the time intervals Δt are enough large. The function $\mathbf{g}_n$ has a period of fluctuation of the order of the time between successive collisions which appear between the Brownian particle and the molecules of the fluid.

Thus, if $\Delta\mathbf{g}_n$ is choosen to be a normal random variable with mean zero, it follows that $\nu(t)$ will be also a normal random variable, as it is desired. By means of 17, and setting $D^2(\Delta\mathbf{g}_n) = 2q\Delta t$ one gets

$$\begin{aligned} E|\mathbf{v}|^2 &= \sum_n 2q\Delta t\, e^{\,2\beta(n\,\Delta t - t)} \to \\ &\to 2q\int_0^t e^{\,2\beta(z-t)}\,dz = \frac{q}{\beta}\,(1 - e^{-\,2\beta t}) \end{aligned} \tag{18}$$

as $\Delta t \to 0$.

But, at the same time, one has

$$E|\mathbf{v}|^2 \to \frac{kT}{m}$$

as $t \to \infty$, so that q is given by the equality below

$$q = \frac{\beta kT}{m}. \tag{19}$$

If $\mathbf{x}(t)$ is the notation for the displacement of the Brownian particle then, we have

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{v}(z)dz. \tag{20}$$

Now substituting (16) in (20) one gets

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \left(\mathbf{v}_0\,e^{-\beta z} + e^{-\beta z}\int_0^z e^{\beta y} f(y)dy\right) dz.$$

If the order of integration is changed the following estimation follows

$$\begin{aligned} &\mathbf{x}(t) - \mathbf{x}_0 - \frac{\mathbf{v}_0(1 - e^{-\beta t})}{\beta} = \\ &= -e^{-\beta t}\int_0^t \frac{e^{\beta z}\mathbf{f}(z)dz}{\beta} + \int_0^t \frac{\mathbf{f}(z)dz}{\beta} \equiv \int_0^t g(z)\mathbf{f}(z)dz, \end{aligned} \tag{21}$$

where $g(z) = \dfrac{1-e^{\beta(z-t)}}{\beta}$. If a finite sum approximation to the integral is used again then, we come to the conclusion that

$$\mathbf{x}(t) - \mathbf{x}_0 - \frac{\mathbf{v}_0(1-e^{-\beta t})}{\beta}$$

is a normal random variable with the mean equal to zero and the variance given by the equality

$$\sigma^2 = 2q\int\limits_0^t g^2(z)dz = \frac{q}{\beta^3}\,(2\beta t + 4e^{-\beta t} - e^{-2\beta t} - 3). \tag{22}$$

Regarding to the probability density of the displacement $\mathbf{x}(t)$, it is given by the follwoing equality

$$p(\mathbf{x}, t, \mathbf{x}_0, \mathbf{v}_0) = \left[\frac{m\beta^2}{2kT\,(2\beta t + 4e^{-\beta t} - e^{-2\beta t} - 3)}\right]^{\frac{3}{2}} \times$$
$$\times\, e^{-\dfrac{m\beta^2\left|\mathbf{x}-\mathbf{x}_0\,\frac{1-e^{-\beta t}}{\beta}\right|^2}{2kT\,(2\beta t + 4e^{-\beta t} - e^{-2\beta t} - 3)}}. \tag{23}$$

Finally, for sufficiently large values of t it results

$$p(\mathbf{x}, t, \mathbf{x}_0, \mathbf{v}_0) \approx \frac{1}{(4\pi D t)^{\frac{3}{2}}}\, e^{-\frac{|\mathbf{x}-\mathbf{x}_0|^2}{4Dt}} \tag{24}$$

where D is

$$D = \frac{kT}{m\beta} = \frac{kT}{6\pi a\eta}. \tag{25}$$

and is referred to as the *diffusion coefficient*.

Therefore, it results that $p(\mathbf{x}, t, \mathbf{x}_0, \mathbf{v}_0)$ satisfies the diffusion equation given below

$$\frac{\partial p(\mathbf{x}, t, \mathbf{x}_0, \mathbf{v}_0)}{\partial t} = D\Delta p(\mathbf{x}, t, \mathbf{x}_0, \mathbf{v}_0). \tag{26}$$

The expression of D in (25) was obtained by A. Einstein.

Observation 31 *From physics it is known the following result due to Maxwell: Let us suppose that the energy is proportional to the number of particles in a gas and let us denoted $E = \gamma n$, where γ is a constant independent of n. Then,*

$$P\{a < v_i^1 < b\} = \frac{\int\limits_a^b \left(1 - \dfrac{x^2 m}{2\gamma n}\right)^{\frac{3n-3}{2}} dx}{\int\limits_{-\left(\frac{2\gamma n}{m}\right)^{\frac{1}{2}}}^{+\left(\frac{2\gamma n}{m}\right)^{\frac{1}{2}}} \left(\dfrac{1 - x^2 m}{2\gamma n}\right)^{\frac{3n-3}{2}} dx} \to$$
$$\to \left(\frac{3m}{4\pi\gamma}\right)^{\frac{1}{2}} \int\limits_a^b e^{-\dfrac{3mx^2}{4\gamma}}\, dx.$$

Now, for $\gamma = \dfrac{3kT}{2}$ *the following Maxwell's result is found*

$$\lim_{n\to\infty} P\{a < v_i^1 < b\} = \left(\frac{m}{2\pi kT}\right)^{\frac{1}{2}} \int_a^b e^{-\dfrac{mx^2}{2kT}}\, dx.$$

T is called the "absolute temperature", while k is the "Boltzmann's constant".

Conclusion 31 *We think that when, in various problems, we say "chaos" or "chaotic and complex systems" or we use another similar expression to define the comportment of some natural phenomena, in fact we imagine phenomena similarly to a Brownian motion which is a more realistic model of such phenomena.*

4 In short about the complexity of discrete time systems

4.1 Generative systems

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 41 *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 the derivations according to our generative system. Let D_x be the class of all the derivations of length x in $\mathcal{D}$. We set

$$n = \max\{x |\, D(x) \in D_x \text{ and } x \geq 2\}.$$

Evidently, $\mathcal{D}$ split into equivalence classes each of them being represented by one of its elements arbitrarily chosen. Let us consider $D_x = (w_0, w_1, \cdots, w_x)$. Then, let K be the following set of sequences:

$$\begin{aligned} K \;=\; & \{(w_h, \cdots, w_k)|(w_h, \cdots, w_k) \subset D(x) \\ & \text{and } h > 0, \;\text{ or }\; k < x \;\text{ or the both}\}. \end{aligned}$$

Now, if N is the set of all natural numbers then, for $k \in N$, the set of first k natural numbers is denoted by $[k]$. A function $p : [\alpha] \to K$, where $\alpha \in [x-1]$, is called a *partial derivation of length* α of $D(x)$. The length of a partial derivation $p \subset D(x)$ will be denoted by $|p|$ if another specification is not made. Evidently for each $p(\alpha) \subset D(x)$ we have $\alpha \in [x - 1]$, so that the condition $x \geq 2$ is justified. For $p_1, p_2 \subset D(x)$ we write $p_1 \leq p_2$ to mean that $|p_1| \leq |p_2|$ and $p_1(j) \leq p_2(j)$ for all $j \in [|p_1|]$. It is easy to see that $\leq$ is a partial order in the set of all partial derivations of $D(x)$.

4.2 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. Thus, if it is, then the probability that it should be also generated into the class D_{x-1} is denoted by γ; but given that it is not generated into D_x, the probability that it should be generated into D_{x+1} is denoted by β. 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: $p_{11} = p_{n-1\,n-1} = 1$; $p_{21} = q$; $p_{22} = p_{n-2\,n-2} = pq$; $p_{24} = p^2$; $p_{n-2\,n-4} = q^2$ and $p_{n-2\,n-1} = p$.
- In the second case: $p_{11} = 1$; $p_{21} = p_{n-1\,n-3} = q$; $p_{n-1\,n-1} = p$; $p_{22} = pq$; $p_{24} = p^2$; $p_{n-2\,n-4} = q^2$ and $p_{n-2\,n-2} = p + qp$.
- In the third case: $p_{11} = q$; $p_{13} = p_{n-2\,n-1} = p$; $p_{n-1\,n-1} = 1$; $p_{22} = q+pq$; $p_{24} = p^2$; $p_{n-2\,n-4} = q^2$ and $p_{n-2\,n-2} = qp$.
- In the fourth case: $p_{11} = p_{n-1\,n-3} = q$; $p_{13} = p_{n-1\,n-1} = p$; $p_{22} = q + pq$; $p_{24} = p^2$; $p_{n-2\,n-4} = q^2$ and $p_{n-2\,n-2} = p + qp$.

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 41 (***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 $2pq$.*

Let us remain in the case when a word is generated by more derivations according to a given generative system. This is a specific propriety of the so-called *ambiguous languages*, that is inte-resting to be characterized.

To this end let ν_x be the number of derivations into the equivalence class D_x, $x \geq 2$, by which the word w is generated. Obviously ν_x is a random variable that takes the values 1 and 0 with the probabilities p_x and $q_x = 1 - p_x$ respectively. Then, the number of derivations in $n-1$ equivalence classes by which w is generated is the following

$$\nu = \sum_{x=2}^{n} \nu_x$$

Now, because the equivalence classes of the derivations are connected into a homogeneous Markov chain, the expectation and the variance of ν are as follows

$$E\nu = \sum_{x=2}^{n} E\nu_x = (n-1)p + \sum_{x=2}^{n} (p_1 - p)\delta^{x-1} = (n-1)p + (p_1 - p)\frac{\delta - \delta^n}{1-\delta} \quad (27)$$

and

$$D\nu = E\left[\sum_{x=2}^{n} (\nu_x - p_x)\right]^2 = \sum_{x=2}^{n} E(\nu_x - p_x)^2 + 2 \sum_{i<j, i\geq 2} E(\nu_i - p_i)(\nu_j - p_j) \quad (28)$$

Now regarding the expectation of ν, excepting $(n-1)p$ the other term is bounded as n increases, such that it results $E\nu = (n-1)p + u_n$, while regarding its variance excepting $(n-1)pq$ and $npq\frac{\delta}{1-\delta}$, the other all terms are bounded as n increases, so that we get $D\nu = (n-1)pq + 2npq\frac{\delta}{1-\delta} + v_n$, where u_n and v_n are certain quantities that remain bounded as n increases. Thus, the following main result is obtained

Theorem 42 *If among the equivalence classes of the derivations according to a generative system G, a Markov dependence exists then, $L(G)$ tends to become an ambiguous language of order n if there exists a word $w \in L(G)$ such that the expectation and the variance of the random variable giving the number of derivations by which w is generated verify the following relations*

$$E\nu = (n-1)p + u_n$$
$$D\nu = pq\left[n\,\frac{1+\delta}{1-\delta} - 1\right] + v_n$$

4.3 A limit theorem

Now we consider the special case when a word can be generated into the equivalence class of a derivation on the following conditions:

1 It can be generated into the class $D_x, x \geq 2$, by more of its elements.
2 If it is not generated into the class $D_x, x \geq 2$, then it is generated into the preceding and the next class.

We refer to such a way for generating words as being *an alternating generation procedure*. We shall use the notation w for the case when this word is generated into an equivalence class and the notation $\overline{w}$ otherwise.

We propose to determine the probability $P_n(k)$ that a word w should be generated by $m\,(m < n)$ derivations in the following ways:

i It will be generated by the first class and the last and there is a direct rule (σ, w);
ii It will be generated by the first class but it will be not generated by the last and there is a direct rule (σ, w);
iii It will be not generated by the first class but it will be generated by the last and there is not a direct rule (σ, w);
iv It will be not generated both by the first class and the last and there is not a direct rule (σ, w).

Then $P_n(k)$ is given by the following equality

$$P_n(k) = P_n(k, ww) + P_n(k, w\overline{w}) + P_n(k, \overline{w}w) + P_n(k, \overline{ww}) \tag{29}$$

where $P_n(k, ww)$ is the notation for i, and so on.

Now, computing the terms in (29), it results:

$$P_n(k, ww) \approx \frac{p_1\beta}{\sqrt{2\pi[k\gamma(1-\gamma) + (n-k)\beta(1-\beta)]}}\, e^{-\frac{z^2}{2}},$$
$$P_n(k, w\overline{w}) \approx \frac{p_1(1-\gamma)}{\sqrt{2\pi[k\gamma(1-\gamma) + (n-k)\beta(1-\beta)]}}\, e^{-\frac{z^2}{2}},$$
$$P_n(k, \overline{w}w) \approx \frac{q_1\beta}{\sqrt{2\pi[k\gamma(1-\gamma) + (n-k)\beta(1-\beta)]}}\, e^{-\frac{z^2}{2}},$$
$$P_n(k, \overline{ww}) \approx \frac{q_1(1-\gamma)}{\sqrt{2\pi[k\gamma(1-\gamma) + (n-k)\beta(1-\beta)]}}\, e^{-\frac{z^2}{2}}.$$

$P_n(k)$ will be obtained by adding these probabilities, and we get

$$P_n(k) \approx \frac{p_1\beta + p_1(1-\gamma) + q_1\beta + q_1(1-\gamma)}{\sqrt{2\pi[k\gamma(1-\gamma) + (n-k)\beta(1-\beta)]}}\, e^{-\frac{z^2}{2}}.$$

or, after some transformations

$$P_n(k) \approx \frac{1-\gamma+\beta}{\sqrt{2\pi npq(1+\gamma-\beta)(1-\gamma+\beta)}}\, e^{-\frac{z^2}{2}}.$$

Thus, the following main result is obtained

Theorem 43 *If a word is generated by an alternating generation procedure, according to a generative system just considered, the derivations of which belonging to n equivalence classes then, the probability that it should be generated by k classes out of n is given by the following relation*

$$P_n(k) \approx \frac{1}{\sqrt{2\pi npq}} \sqrt{\frac{1-\gamma+\beta}{1+\gamma-\beta}}\, e^{-\frac{z^2}{2}}$$

.

References

1. Bharucha-Reid A.T., Elements Of The Theory Of Markov Processes And Their Applications. Dover Publications, Inc., Mineola, New York, 1997.
2. Feller, W., An Introduction to Probability Theory and its Applications, vol. I, II. John Wiley & Sons, Inc., New York, London, 1960.
3. Gihman, I.I. and Skorohod, A.V., Stochastic Differential Equations. Springer-Verlag, Berlin, 1972.
4. Itô, K. and McKean, H.P. Jr., Diffusion Processes and their Sample Paths. Springer-Verlag, Berlin Heidelberg, 1996.
5. Itô, K., Stochastic Processes. Eds. Ole E. Barndorff-Nielsen and Ken-iti Sato. Springer, 2004.
6. Kushner, H.J., Yin, G.G., Stochastic Approximation Algorithms and Applications, Springer-Verlag New York, Inc., 1997.
7. Malliavin P., Integration and Probability. Springer-Verlag New York, Inc., 1995.
8. Øksendal, B. and Sulem, A., Applied Stochastic Control of Jump Diffusions. Springer, 2007.
9. Orman, G.V., Limbaje formale, Ed. Tehnică, Bucureşti, 1982.
10. Orman, G.V., *On the numerical functions utilized in the study of a class of discrete systems*, Journal of Systems Science, Vol. 19, No. 1, 1993, pp. 27-37.
11. Orman, G.V., *On the Characterization of some Generative Grammars by Numerical Functions*, In: Current Issues in Mathematical Linguistics, ed. C. Martin-Vide, Elsevier Science B.V., North Holland, Amsterdam, 1994, pp. 151-160.
12. Orman, G.V., *Random Phenomena and some Systems Generating Words*, In: Yugoslav Journal of Operation Research, Vol. 6, No. 2, 1996, pp. 245-256.
13. Orman, G.V., *Derivational Operators in Generative Systems*, In: Proceedings of 13^{th} International Conference on SYSTEMS SCIENCE, Vol. I, September 15-18, Wroclaw, 1998, pp. 129-136.
14. Orman, G.V., Lectures on *Stochastic Approximation Methods and Related Topics.* Preprint. "Gerhard Mercator" University, Duisburg, Germany, 2001.
15. Orman, G.V., Handbook of Limit Theorems and Stochastic Approximation. "Transilvania" University Press, Brasov, 2003.
16. Orman, G.V., *On a problem of approximation of Markov chains by a solution of a stochastic differential equation*. In: Christos H. Skiadas, Ioannis Dimotikalis and Charilaos Skiadas (eds). Chaos Theory: Modeling, Simulation and Applications. World Scientific Publishing Co Pte Ltd, pp. 30-40, 2011.
17. Orman, G.V., Aspects of convergence and approximation in random systems analysis. LAP Lambert Academic Publishing, 2012.
18. Orman, G.V., Basic Probability Theory, Convergence, Stochastic Processes and Applications, *Concepts and Stochastic Models useful in the Study of the Systems.* Transilvania University Press, Braşov, 2016.
19. Protter P., Stochastic Integration and Differential Equations: a New Approach. Springer-Verlag, 1990.
20. Schuss Z., Theory and Application of Stochastic Differential Equations. John Wiley & Sons, New York, 1980.
21. Steele J.M., Stochastic calculus and financial applications. Springer-Verlag New York, Inc. 2001.
22. Stroock D. W., Markov Processes from K. Itô Perspective. Princeton Univ. Press, Princeton, 2003.
23. Wasan, M.T., Stochastic Approximation. Cambridge University Press, 1969.