On a Problem of Approximation of Markov Chains by a Solution of a Stochastic Differential Equation

Gabriel V. Orman

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

Abstract. Much scientific works has been done on the applications of the Brownian motion in such diverse areas as molecular and atomic physics, chemical kinetics, solid-state theory, stability of structures, population genetics, communications, and many other branches of the natural and social sciences and engineering. We shall refer below to some aspects concerning the approximation of Markov chains by a solution of a stochastic differential equation to determine the probability of extinction of a genotype. Thus, the Markovian nature of the problem will be pointed out.

Keywords: Brownian motion, stochastic differential equations, Markov chains, transition probabilities, binomial distribution.

2000 MS Classification: 60H10, 60H30, 60J65, 60J20, 60J70

1 Introduction

It is known that a precise definition of the Brownian motion involves a measure on the path space, such that it is possible to put the Brownian motion on a firm mathematical foundation. Much scientific works has been done on its applications in such diverse areas as molecular and atomic physics, chemical kinetics, solid-state theory, stability of structures, population genetics, communications, and many other branches of the natural and social sciences and engineering. In this sense, many contributions have been done by P. Lévy, K. Itô, H.P. McKean, Jr., S. Kakutani, H.J. Kushner, A.T. Bharucha-Reid and other. Also some models based on Brownian motion are successfully applied to nucleotide strings analysis.

We shall refer here only to some aspects concerning the approximation of Markov chains by a solution of a stochastic differential equation to determine the probability of extinction of a genotype. Thus, the Markovian nature of the problem will be pointed out again, and we think that this is a very important aspect.

Obviously, the interaction of a population can have a great complexity, which lead to the enhancement of the interdisciplinary coordination in these studies.

When a differential equation is considered if it is allowed for some randomness in some of its coefficients, it will be often obtained a so-called stochastic differential equation which is a more realistic mathematical model of the considered situation.

For example, let us consider the simple population growth model (according to [8])

\[
\frac{dN}{dt} = a(t)N(t)
\]

\[N(0) = k\]

(1)
where \( N(t) \) is the size of the population at time \( t \), and \( a(t) \) is the relative rate of growth at time \( t \). Obviously, it might happen that \( a(t) \) is not completely known, but subject to some random environmental effects, so that one gets

\[ a(t) = r(t) + \text{"noise"}, \]

where the exact behaviour of the noise term is unknown, but only its probability distribution.

The function \( r(t) \) is assumed to be nonrandom, and one put the problem to solve (1) in this case. Such a problem conduct us to the following notion: the equation obtained by allowing randomness in the coefficients of a differential equation is called a stochastic differential equation. Therefore, it is clear that any solution of a stochastic differential equation must involve some randomness. In other words one can hope to be able to say something about the probability distribution of the solutions.

## 2 Brownian motion

Regarding the Brownian motion we think that it is the most important stochastic process. As a practical tool, it has had profound impact on almost every branch of physical science, as well as several branches of social sciences. As a creation of pure mathematics, it is an entity of uncommon beauty. It reflects a perfection that seems closer to a law of nature than to a human invention.

In 1828 the English botanist Robert Brown observed that pollen grains suspended in water perform a continual swarming motion. The chaotic motion of such a particle is called Brownian motion and a particle performing such a motion is called a Brownian particle.

The first important applications of Brownian motion were made by L. Bachelier and A. Einstein. L. Bachelier derived (1900) the law governing the position of a single grain performing a 1-dimensional Brownian motion starting at \( a \) at time \( t = 0 \)

\[ P_a[x(t) \in db] = g(t,a,b)db \quad (t > 0), \tag{2} \]

where \( (t, a, b) \in (0, +\infty) \times \mathbb{R}^2 \) and \( g \) is the Green (or the source) function

\[ g(t,a,b) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(b-a)^2}{2t}} \]

of the problem of heat flow

\[ \frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial a^2}, \quad (t > 0). \]

A. Einstein (in 1905) also derived (2) from statistical mechanical considerations and applied it to the determination of molecular diameters.

Bachelier also pointed out the Markovian nature of the Brownian path expressed in

\[ P_a[a_1 \leq x(t_1) < b_1, a_2 \leq x(t_2) < b_2, \cdots, a_n \leq x(t_n) < b_n] = \]

\[ \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} g(t_1-a_1, \xi_1) g(t_2-t_1, \xi_2) \cdots g(t_n-t_{n-1}, \xi_n) \xi_1 \xi_2 \cdots \xi_n db_1 \cdots db_n \]

where \( 0 < t_1 < t_2 < \cdots < t_n \), and used it to establish the law of maximum displacement

\[ P_a[\max_{s \leq t} x(s) \leq b] = 2 \int_0^b e^{-\frac{s^2}{2t}} ds, \quad t > 0, \quad b \geq 0. \]

But he was unable to obtain a clear picture of the Brownian motion and his ideas were unappreciated at that 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 the works of É. Borel and H. Lebesgue have been appeared. Beginning with this moment was possible to put the Brownian motion on a firm mathematical foundation and this was achived by N. Wiener in 1923.
Definition 1. A continuous-time stochastic process \( \{B_t \mid 0 \leq t \leq T\} \) is called a standard Brownian motion on \([0, T)\) if it has the following four properties: i. \( B_0 = 0 \); ii. The increments of \( B_t \) are independent; that is, for any finite set of times \( 0 \leq t_1 < t_2 < \cdots < t_n < T \), the random variables 
\[
B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_n} - B_{t_{n-1}}
\]
are independent; iii. For any \( 0 \leq s \leq t < T \) the increment \( B_t - B_s \) has the normal distribution with mean 0 and variance \( t - s \); iv. For all \( \omega \) in a set of probability one, \( B_t(\omega) \) is a continuous function of \( t \).

The Brownian motion can be represented as a random sum of integrals of orthogonal functions. Such a representation satisfies the theoretician’s need to prove the existence of a process with the four defining properties of Brownian motion, but it also serves more concrete demands. Especially, the series representation can be used to derive almost all of the most important analytical properties of Brownian motion. It can also give a powerful numerical method for generating the Brownian motion paths that are required in computer simulation.

3 Stochastic differential equations

To describe the motion of a particle driven by a white noise type of force (due to the collision with the smaller molecules of the fluid) the Langevin equation
\[
\frac{dv(t)}{dt} = -\beta v(t) + f(t) \tag{3}
\]
is used, where \( f(t) \) is the white noise term. Its solution is the following
\[
y(t) = y_0 e^{-\beta t} + e^{-\beta t} \int_0^t e^{\beta s} f(s) ds. \tag{4}
\]
If we denote by \( w(t) \) the Brownian motion, then it is given by
\[
w(t) = \frac{1}{q} \int_0^t f(s) ds, \tag{5}
\]
so that \( f(s) = q \frac{dw(s)}{ds} \). But \( w(t) \) is nowhere differentiable, such that \( f(s) \) is not a function. Therefore, the solution (4), of Langevin’s equation, is not a well-defined function. This difficulty can be overcome, in the simple case, as follows. Integrating (4) by parts, and using (5), it results
\[
y(t) = y_0 e^{-\beta t} + q(w(t) - \beta \int_0^t e^{-\beta (t-s)} w(s) ds). \tag{6}
\]
But all functions in (6) are well defined and continuous, such that the solution (5) can be interpreted by giving it the meaning of (6). Now, such a procedure can be generalized in the following way. The functions \( f(t) \) and \( g(t) \) are considered to be defined for \( a \leq t \leq b \). For any partition \( P : a \leq t_0 < t_1 < \cdots < t_n \), we denote
\[
S_P = \sum_{i=1}^n f(\xi_i)[g(t_i) - g(t_{i-1})],
\]
where \( t_{i-1} \leq \xi_i \leq t_i \). If a limit exists
\[
\lim_{|P| \to 0} S_P = I
\]
where \( |P| = \max_{1 \leq i \leq n} (t_i - t_{i-1}) \), then it is said that \( I \) is the \textit{Stieltjes integral} of \( f(t) \) with respect to \( g(t) \). It is denoted

\[
I = \int_a^b f(t)dg(t).
\]

Now the stochastic differential equation

\[
\begin{align*}
\frac{dx}{dt} &= a(x(t), t)dt + b(x(t), t)dw(t) \\
x(0) &= x_0
\end{align*}
\]

is defined by the Itô integral equation

\[
\begin{align*}
x(t) &= x_0 + \int_0^t a(x(s), s)ds + \int_0^t b(x(s), s)dw(s).
\end{align*}
\]

The simplest example of a stochastic differential equation is the following equation

\[
\begin{align*}
\frac{dx}{dt} &= a(t)dt + b(t)dw(t) \\
x(0) &= x_0
\end{align*}
\]

which has the solution

\[
\begin{align*}
x(t) &= x_0 + \int_0^t a(s)ds + \int_0^t b(s)dw(s).
\end{align*}
\]

The transition probability density of \( x(t) \) is a function \( p(x, s; y, t) \) satisfying the condition

\[
P(x(t) \in A \mid x(s) = x) = \int_A p(x, s; y, t)dy,
\]

for \( t > s \), where \( A \) is any set in \( R \). It is supposed that \( a(t) \) and \( b(t) \) are deterministic functions.

## 4 Markov property

Some aspects regarding the Markov property, in a vision of Kiyosi Itô, are discussed below.

**Definition 2.** Let \( S \) be a state space. A Markov process is a system of stochastic processes \( \{X_t(\omega), t \in T, \omega \in (\Omega, K, P_a)\}_{a \in S} \).

For each \( a \in S \), \( \{X_t\}_{t \in S} \) is a stochastic process defined on the probability space \( (\Omega, K, P_a) \). It is considered that \( C = C(S) \) is the space of all continuous functions (it is a separable Banach space with a supremum norm).

The transition probabilities of a Markov process are \( \{p(t, a, B)\} \).

Let us denote by \( \{H_t\} \) the transition semigroup and let \( R_\alpha \) be the resolvent operator of \( \{H_t\} \). Now \( p(t, a, B) \), \( H_t \) and \( R_\alpha \) can be expressed in terms of the process as follows.

**Theorem 1.** Let \( f \) be a function in \( C(S) \). Then

1. \( p(t, a, B) = P_a(X_t \in B) \).
2. \( \text{For } E_a(\cdot) = \int_\Omega \cdot P_a(d\omega) \text{ one has } H_t f(a) = E_{a}(f(X_t)) \).
3. \( R_\alpha f(a) = E_{a} \left( \int_0^\infty e^{-\alpha t} f(X_t)dt \right) \).
Proof. One can observe that 1° and 2° are immediately. Now, for 3°, it is considered the following equality:

\[ R_a f(a) = \int_0^\infty e^{-at} H_t f(a) dt = \int_0^\infty e^{-at} E_a(f(H_t)) dt. \]

But \( f(X_t(\omega)) \) is right continuous in \( t \) for \( \omega \) fixed and measurable in \( \omega \) for \( t \) fixed and therefore it is measurable in the pair \((t, \omega)\). Thus, the Fubini's theorem can be used and one gets

\[ R_a f(a) = E_a \left( \int_0^\infty e^{-at} f(X_t) dt \right), \]

that is just 3°.

**Notation.** For \( \mathcal{C} \) a \( \sigma \)-field on \( \Omega \), the space of all bounded \( \mathcal{C} \)-measurable functions will be denoted by \( B(\Omega, \mathcal{C}) \), or simple \( B(\mathcal{C}) \).

Now the Markov property is given in the following theorem.

**Theorem 2.** Let be given \( \Gamma \in K \). The following is true

\[ P_a(\theta_t \omega \in \Gamma | K_t) = P_{X_t(\omega)}(\Gamma) \quad a.s.(P_a); \]

that is to say

\[ P_a(\theta_t^{-1} \Gamma | K_t) = P_{X_t(\omega)}(\Gamma). \]

**Note 1.** The following notation can be used

\[ P_{X_t(\omega)}(\Gamma) = P_b(\Gamma) |_{b = X_t(\omega)}. \]

Proof. It will be suffice to show that

\[ P_a(\theta_t^{-1} \Gamma \cap D) = E_a(P_{X_t(\Gamma)}, D) \] (10)

for \( \Gamma \in K \) and \( D \in K_t \).

One can distinguish three cases.

**[I.]** Let us consider \( \Gamma \) and \( D \) as follows:

\[ \Gamma = \{ X_{s_1} \in B_1 \} \cap \{ X_{s_2} \in B_2 \} \cap \cdots \cap \{ X_{s_n} \in B_n \}, \]

and

\[ D = \{ X_{t_1} \in A_1 \} \cap \{ X_{t_2} \in A_2 \} \cap \cdots \cap \{ X_{t_m} \in A_m \} \]

with \( 0 \leq s_1 < s_2 < \cdots < s_n \), \( 0 \leq t_1 < t_2 < \cdots < t_m \leq t \) and \( B_i, A_j \in K(S) \).

Now it will be observed that the both sides in (10) are expressed as integrals on \( S^{n+m} \) in terms of transition probabilities. Thus, one can see that they are equal.

**[II.]** Let now be \( \Gamma \) as in the case **[I.]** and let us denote by \( D \) a general member of \( K_t \). For \( \Gamma \) fixed the family \( D \) of all \( D \)'s satisfying (10) is a Dynkin class. If \( \mathcal{M} \) is the family of all \( M \)'s in the case **[I.]** then, this family is multiplicative and \( \mathcal{M} \subset D \). In this way it follows

\[ D(\mathcal{M}) \subset D = K(\mathcal{M}) = K_t \]

and one can conclude that, for \( \Gamma \) in the case **[I.]** and for \( D \) general in \( K_t \), the equality (10) holds.

**[III.]** (General case.) This case can be obtained in a same manner from **[II.]** by fixing an arbitrary \( D \in K_t \).

It will be obtained that \( P_a(\Gamma) \) is Borel measurable in \( a \) for any \( \Gamma \in K \).

**Corollaire 1.**

\[ E_a(G \circ \theta_t, D) = E_a(E_{X_t}(G), D) \quad \text{for} \quad G \in B(K), D \in K_t, \]

\[ E_a(F \circ (G \circ \theta_t)) = E_a(F \circ E_{X_t}(G)) \quad \text{for} \quad G \in B(K), F \in B(K_t), \]

\[ E_a(G \circ \theta_t | K_t) = E_{X_t}(G) \quad (a.s.) (P_a) \quad \text{for} \quad G \in B(K). \]

[For more details see [4], [5], [6], [1]].
The probability of extinction of a genotype

We now consider an application to genetics discussed by W. Feller [2]. As it is known the heritable characters depend on special carries, called genes, which appear in pairs. Each gene of a particular pair can suggest two forms A and a which determine a genetic type in a population. Therefore, three different pairs can be formed AA, Aa, aa, such that the organism belongs to one of these three genotypes. On the other hand, the reproductive cells, called gametes, are assumed to have one gene. In this way, the gametes of an organism of genotype AA or aa have the gene A or respective the gene a, whereas the gametes of an organism of genotype Aa may have the gene A or the gene a with equal probability. We can view such a problem in the context of the binomial distribution

$$P_n(k) = \binom{n}{k} p^k q^{n-k}.$$  

We can imagine an urn with 2N elements (that is the genes of types A and a). Then, the genotype structure of N offsprings will be the result of 2N independent drawings from the urn. Furthermore, it is considered that a population consists of N individuals in each generation. Now, if i of the genes are of the type A (0 \leq i \leq 2N) in some generation, then it is said that the generation is in the state \(E_i\). In this way a Markov chain is connected to such a genetic process. It has 2N + 1 states \(E_0, E_1, E_2, \ldots, E_{2N}\). Calculating the transition probability from the state \(E_i\) to \(E_j\), in one generation, one gets

$$P_{ij}^N = C_{2N}^j \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}. \quad (11)$$

Now, one considers a population of N individuals consisting of \(X_N(n) = i\) individuals of type A in the n-th generation. Then, the next generation consists of N individuals randomly selected from a practically infinite offspring of the previous generation. Obviously the selection process is binomial with probability \(x = \frac{i}{N}\) for type A, the proportion \(\frac{i}{N}\) of A types being equal to the probability in a large offspring population. Therefore, the transition probability is given by the following equality

$$P_{ij}^N = P(X_N(n + 1) = j \mid X_N(n) = i) = C_N^j x^i (1 - x)^{N-j}.$$ 

Let now denote by s the fitness of A relative to a when selection forces act on the population

$$s = \frac{x - z}{x(1 - x)}$$

where

$$\bar{x} = \frac{x(1 + s)}{1 + sx}.$$ 

Therefore, one obtains

$$P_{ij}^N = C_N^j \bar{x}^i (1 - \bar{x})^{N-j}.$$ 

Now, if \(s = s_N(n)\) is a random variable then, the probability of extinction of a genotype, or the time until extinction, or the total A population, or other characteristics of interest, become very hard to calculate. For this reason the Markov chains \(X_N(n)\) can be approximated by a diffusion process, or more exactly, by a solution of a stochastic differential equation.

To this end the following form is considered for the process

$$X_N([Nt]) = N x_N(t)$$
where \( [Nt] \) is the greatest integer not exceeding \( Nt \) and \( t \) is any positive number. Thus, \( x_N(t) \) represents the proportion of \( A \) types in population. Now it is supposed that
\[
N E x_N(n) \to \sigma(t)
\]
as \( N \to \infty \), \( n = [Nt] \), and \( N E x_N^k(n) \to \nu(t) \) and \( N E x_N^k(n) \to 0 \) for all \( k > 2 \).

Let now be \( \Delta t = \frac{1}{N} \). Then, one gets
\[
a_N \equiv \frac{1}{\Delta t} E_{x,t}[x_N(t + \Delta t) - x_N(t)] = E[x_N([Nt] + 1) - x_N([Nt])] \quad \text{for } x_N([Nt]) = x_N.
\]
But \( X_N(n) \) is a binomial variable \( B(N, \pi) \) such that it results
\[
a_N = NE(\pi - x) = NE \left( \frac{(1 + s_N(n))x}{1 + s_N(n)x} - x \right) = \left[ \sigma(t) - \nu(t)x(1 - x) \right] \equiv a(x, t)
\]
and respective
\[
\frac{1}{\Delta t} E_{x,t}[x_N(t + \Delta t) - x_N(t)]^2 \to x(1 - x)[1 + \nu(t)x(1 - x)] = b(x, t). \quad (13)
\]
The moments of higher order tend to zero as \( N \to \infty \). Now it can be shown that the convergence is sufficiently rapid as to satisfy the imposed conditions. Let now consider again the specified model. Thus, we have
\[
E[x_N(t + \Delta t) - x_N(t) | x_N(t)] = 0
\]
and
\[
E[|x_N(t + \Delta t) - x_N(t)|^2 | x_N(t)] = \frac{x_N(t)[1 - x_N(t)]}{N}
\]
\[
E[|x_N(t + \Delta t) - x_N(t)|^4 | x_N(t)] \leq \frac{K}{N^2}
\]
for a constant \( K \).

Now, it can be seen that the conditions for convergence hold, provided that the stochastic differential equation
\[
dx(t) = a(x, t)dt + \sqrt{b(x, t)}
\]
\[
t(0) = x
\]
(14)
has a unique solution (with absorption at \( x = 0 \) and \( x = 1 \)). But the conditions of the existence and uniqueness theorem are not satisfied by the coefficient \( \sqrt{b(x, t)} \) in (14).

To show the existence and uniqueness it is necessary to consider (14) in the interval \( I_{c,d} \] with absorption at the boundary \( \delta I \). The conditions of the existence and uniqueness theorem are satisfied in \( I_{c,d} \). Thus, it is a unique solution \( x_s(t) \) in \( I_{c,d} \) up to the time
\[
\tau_x = \inf\{t | x_s(t) \in \delta I \}
\]
If \( \varepsilon_1 < \varepsilon_2 \) then, \( \tau_{\varepsilon_1} \geq \tau_{\varepsilon_2} \) and
\[
x_{\varepsilon_1}(t) = x_{\varepsilon_2}(t) \quad \text{where } 0 \leq t \leq \tau_{\varepsilon_2},
\]
by the so-called the localization principle.\(^2\) Now, as \( \varepsilon \to 0 \), one gets that \( x_{\varepsilon}(t) \) converges

\(^2\) The Localization principle. Let us consider \( a_1(x, t) = a_2(x, t) \) and \( b_1(x, t) = b_2(x, t) \) for \( c \leq x \leq d \) and \( t \geq 0 \). Suppose that \( a_i(x, t) \) and \( b_i(x, t) \), with \( i = 1, 2 \), satisfy the conditions of the existence and uniqueness theorem in \( R \) for the stochastic differential equations
\[
dx_i(t) = a_i(x, t)dt + b_i(x, t)dw(t), \quad i = 1, 2
\]
\[
x_i(0) = x_0 \in [c, d].
\]
Let us denote \( \tau_i = \inf\{t | x_i(t) \notin (c, d) \} \) with \( i = 1, 2 \). Then, \( \tau_1 = \tau_2 \) a.s. and \( x_1(t) = x_2(t) \) for all \( t \leq \tau_1 \) a.s.
to a limit $x(t)$ and $\tau_e \to \tau$ (here $\tau$ is the absorption of $x(t)$).

In this way it results that $x_N(t) \to x(t)$, with $x(t)$ the solution of the stochastic differential equation

$$dx(t) = a(x,t)dt + \sqrt{b(x,t)}dw(t)$$

with absorbing boundaries at $x = 0$ and $x = 1$.

Thus, once a genotype is extinct, it will stay extinct for all future generations unless mutation occurs. Therefore, the probability of extinction is the probability of exit of $x(t)$ from the interval $(0, 1)$.

**Conclusion.** Obviously, various situations may exist when the survival of a particular genotype can be very dynamic.

But, as we have already emphasized, for a random variable $f = f_N(n)$ the **probability of extinction** of a genotype, or the **time until extinction**, or the **total A population**, or other characteristics of interest, become very hard to calculate. And this is the reason for which a Markov chain $\{X_N(n)\}$ is useful to be approximated by a solution of a stochastic differential equation.

**References**


