

---

## An application of probabilistic calculus to a problem of genetics

**MIOARA VARGA**

*Department of Biotechnology, Faculty of Biotechnologies, University of Agriculture and Veterinary Medicine Bucharest, Bd. Marasti 59, Sector 1, Bucharest, Romania;*

### Abstract

*In the current paper, we discuss an application of genetics, discussed by Feller, ever since the 60's, and also recently presented by other researchers. We know that inherited features depend on special carriers, called genes, which appear in pairs and each gene of a particular pair produces two forms  $A$  and  $a$ ; these determine a genetic type in a population. Thus, three different pairs can be formed:  $AA$ ,  $Aa$ ,  $aa$  and the organism belongs to one of these three genotypes. On the other hand, it is also known the gametes (the reproductive cells) are assumed to have one gene. The gametes of an organism of genotype  $AA$  or  $aa$  have the gene  $A$  or respectively 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 present in this paper, this kind of problem, in the context of the binomial distribution (the probability of that to happen is of  $k$  independent events from  $n$  possible cases, when  $p$  is the probability with it appears one of these  $k$  events), i.e.*

$$P(n, k) = \binom{n}{k} p^k q^{n-k},$$

where  $q = 1 - p$ .

Keywords: Markov chain, transition probabilities, random variable, stochastic differential equation.

### Introduction

S. Wright, and R.A. Fisher discussed firstly on this problem; lately, G. Malecot talks in: *Sur un problem de probabilite en chaine que pose la genetique*, Comptes Rendus de l'Academie des Sciences, 1944, about its Markovian nature and it was also presented at length in G.V. Orman: *Chapters Applied Mathematics*, Ed. Blue, Cluj-Napoca, 1999.

In this paper, we used some results obtained by D. Bocu [1,2] and G.V. Orman [5] in their search for genetic problems.

In what follows, we present some of base notions about the Markov chain, because we use these notions in this paper.

In many applications of the Probability Theory, we meet some random variables depending by one or many parameter. Thus, in the Feasibility Theory (with the large implications in economy, transport, etc.) the time is in this case such of parameter.

The theory of stochastic processes presents the study of the families defined on a probability space.

Let  $\{\Omega, \mathcal{K}, P\}$  be a probability space, and  $T$  an arbitrary set.

**Definition.**

A stochastic process is a parameterized collection of random variables  $\{X_t\}_{t \in T}$  defined on the probability space  $\{\Omega, \mathcal{K}, P\}$  assuming values in  $R^n$ .

We can remark that:

$$X_t : \Omega \rightarrow R^n, (\forall) t \in T .$$

The parameter space  $T$  may be:

- a)  $Z$  or  $N$  (in this case, one say that the stochastic processes depend by a discreet parameter, and their name is chain.
- b)  $R$  or the half line  $[0, +\infty)$ , or an interval  $[a, b]$  -in this case as one say that *the stochastic processes depend of a continuous parameter or the processes with continuous time.*
- c)  $A$ , where  $A \subseteq R^n$ , for  $n \geq 1$ .

The random variables  $\{X_t(\omega) / t \in T\}$  can take as values complex numbers, points in Euclidian  $k$ -dimensional space, and we call it, *the state space of the process.*

**Markov processes depending on a discreet parameter.**

Let  $X(t, \omega)$  be a stochastic process, defined on a probability space  $\{\Omega, \mathcal{K}, P\}$ , where  $T=N$ .

In this particular case, let us denote the process,  $X_n(\omega), n \in N$ . We assume that even random variable is discreet, and let us denote  $S$ , the following set:

$$S = \bigcup_i \{X_n(\omega) = i \mid n \in N\}$$

I.e.  $S$  is the state space of the process.

**Definition 1.**

The family of random variables  $\{X_n\}_{n \in N}$  is called *Markov chain* if satisfies the Markov relation:

$$P[X_{t_n}(\omega) = i_n / X_{t_{n-1}}(\omega) = i_{n-1}, X_{t_{n-2}} = i_{n-2}, \dots, X_{t_1}(\omega) = i_1] = P[X_{t_n}(\omega) = i_n / X_{t_{n-1}}(\omega) = i_{n-1}] \tag{1}$$

$(\forall) n \geq 2, 0 \leq t_1 < \dots < t_n$  and  $i_1, \dots, i_n \in S$ .

Let one define now, the transition probability function. If  $\{\Omega, \mathcal{K}, P\}$  is the probability space,  $\mathcal{K}_0$  is the  $\sigma$ -field of measurable subsets of  $\Omega$ ; if all the random variables are discreet, and the state space is  $S$ , the probabilities on all the sets of  $\mathcal{K}_0$ , are determined by the finite-dimensional probabilities:

$$P[X_n(\omega) = i_n, \dots, X_0(\omega) = i_0], \tag{2}$$

$(\forall) n \in N$  and  $i_0, \dots, i_n \in I$ .

The (4) probability may be defined as a product of conditional probabilities:

$$P[X_0(\omega) = i_0] \cdot \prod_{t=1}^n P[X_t(\omega) = i_t / X_h = i_h, 0 \leq h < t] \tag{3}$$

For a Markov chain, transition probabilities have the following expression:

$$P[X_t(\omega) = i_t / X_{t-1}(\omega) = i_{t-1}] = p(t; i_{t-1}, i_t), 1 \leq t \leq n, \tag{4}$$

i.e. the functions that depend by  $t, i_{t-1}$  and  $i_t$ .

Let us denote

$$P[X_0(\omega) = i_0] = p_{i_0},$$

then the relation (5) became:

$$P[X_t(\omega) = i_t, 0 \leq t \leq n] = p_{i_0} \prod_{t=1}^n p(t; i_{t-1}, i_t) \tag{5}$$

Let  $\Pi = (p_{ij})_{i,j \in S}$  the matrices of transition probabilities. Thus, the initial probabilities, and the transition probabilities satisfy the following conditions:

$$\begin{cases} p_i \geq 0, \sum_{i \in S} p_i = 1 \\ p_{ij} \geq 0, \sum_{j \in S} p_{ij} = 1 \end{cases} \tag{6}$$

and the transition matrices is a stochastic matrices.

Furthermore, let us define the transition probabilities after  $n$  steps i.e.  $p_{ij}^{(n)}$ . Let us convent to initial step one have:

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 0, i \neq j \\ 1, i = j. \end{cases} \tag{7}$$

To the first-step,

$$p_{ij}^{(1)} = p_{ij} \tag{8}$$

and the numbers  $p_{ij}^{(n)}$  let one define successively, by relation:

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(1)}, \tag{9}$$

where  $n \geq 1$ .

The formula may be generalized as follows:

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}, \tag{10}$$

The last relation and the (11) relation are called *the Chapman-Kolmogorov relation*; this resumes the probabilistic calculus to matrices calculus.

It is interesting, in the case of discrete parameter processes, the asymptotical behavior after  $n$  steps.

### Markov processes depending on a continuous parameter

Often, in nature and in the social or economical life, we meet processes for that we know a state of the system to the arbitrary time  $t$ . It not determines uniquely their states in the

following moments of time, but it determines the probability as the system transits in one of the state of the system.

Further, let us consider the parameter space  $T$ , an interval of non – negative number with the signification of time, and the state space  $S$ , that may be a set of complex numbers, a set of points in Euclidian  $\mathcal{K}$ – dimensional space, or even points in an abstract space. Here, we suppose that

$$S \subseteq R .$$

The most important criterion in classification of stochastic processes is referring to the connection of the random variables  $\{X_t\}_{t \in T}$ . This may be defined specifying the transition probability for any finite number of such random variables, and the definition is presented in the consistent waym [4].

**Definition 2.**

Let be the parameter space the interval  $[0, T]$

The stochastic process  $\{X_t\}_{t \in T}$  is called a Markov process, if for  $n \in N^*$  and any sequences  $0 \leq t_0 < t_1 < \dots < t_n \leq T$  and  $x_0, x_1, \dots, x_n \in S$  (the space state), following relation is verified:

$$P(X(t_n) < x_n | X(t_{n-1}) = x_{n-1}, \dots, X(t_0) = x_0) = P(X(t_n) < x_n | X(t_{n-1}) < x_{n-1}) \tag{11}$$

The relation (14) means the process forget the past, provided that  $t_{n-1}$  is locked as the present; i.e. the notion of Markov process may be associated with a dynamic system without post action, i.e. a system, with an evolution in future ( to the  $t_n$  ) depend only if present state ( to the  $t_{n-1}$  ) and not depend on the state of system in past.

Before the Markov processes represent the mathematical models of the many applications from various domains of activity, the theoretical knowledge of these is very important; in this article, it was given a simplified presentation of some notion and results of them.

Let us consider an urn with  $2N$  elements (we can suppose that is the genes of types  $A$  and  $a$ ). Then, the genotype structure of  $N$  offsprings will be result of  $2N$  independent drawings from the urn.

We want to know the probability of extinction of a genotype or the time until extinction, or the total  $A$  population, or other characteristics of interest, but some time, these are very hard to calculated.

**Method**

For the reasons previously presented, the Markov chains  $\{X_N(n)\}$ , which represent the number of individuals by type  $A$  in the  $n$ -th generation from a population with  $N$  individuals, can be approximated by a stochastic differential equation. In the following, it was presented a calculus of the probability for the genotype extinction.

## Result

In the following, we suppose that a population consists of  $N$  individuals in each generation. If that  $k$  of the genes are of the type  $A$  ( $0 \leq k \leq 2N$ ) in some generation, we say that the generation is in the state  $E_k$ . Thus, a Markov chain is connected to such a genetic process.

It exists  $2N+1$  states:  $E_0, E_1, \dots, E_{2N}$ . We are interested now, by the transition probability from the state  $E_k$  to  $E_j$  in one generation. This is calculated after the following formula:

$$P_{kj}^N = \binom{2N}{j} \left(\frac{k}{2N}\right)^j \left(1 - \frac{k}{2N}\right)^{2N-j}, \tag{12}$$

because  $p$  - the probability that to appear a gene of type  $A$  is the rapport from the number of favorable cases for this events (i.e.  $k$ ) and the number of possible cases (i.e.  $2N$ ), and  $q = 1 - p$  (the probability that do not appear  $k$  genes of type  $A$ ).

Now, we consider a population of  $X_N(n) = k$  individuals of type  $A$  in the  $n$ -th generation. It remarks that the selection process is binomial, with the probability  $x = \frac{k}{n}$  for type  $A$ , the proportion  $\frac{k}{n}$  of type  $A$  is equal to the probability in a large offspring population.

Then, the transition probability is given by the following equality:

$$P_{kj}^N = P(X_N(n+1) = j / X_N(n) = k) = \binom{N}{j} x^j (1-x)^{N-j}. \tag{13}$$

If we denote “ $u$ ”, the fitness of  $A$  relative to  $a$  when the selection forces act on the population, one have:

$$u = \frac{\bar{x} - x}{x(1-\bar{x})}. \tag{14}$$

By (14) it results:

$$\bar{x} = \frac{x(1+u)}{1+ux}. \tag{15}$$

Thus, the transition probability is calculated by the following formula:

$$P_{kj}^N = \binom{N}{j} \bar{x}^j (1-\bar{x})^{N-j}. \tag{16}$$

We can observe that if  $s = s_N(n)$  is a random variable, then the calculus of the extinction probability is very hard to obtain; in this case, the Markov chains  $\{X_N(n)\}$  can be approximated by a solution of a stochastic differential equation. If one denote  $[Nt]$  the greatest integer not exceeding  $Nt$ , and  $t$  is a positive number, we consider the following form for the process  $\{X_N(n)\}$ :

$$X_N([Nt]) = Nx_N(t), \tag{17}$$

where  $x_N(t)$  is the proportion of types  $A$  in population.

Now, as  $N \rightarrow \infty$  and  $n = [Nt]$ , we suppose that:

$$\begin{cases} NEs_N(n) \rightarrow \sigma(t) \\ NEs_N^2(n) \rightarrow \nu(t) \end{cases} \quad (18)$$

where  $Es_N(n)$  represents the expectation of random variables  $s_N(n)$ , and  $Es_N^2(n)$  is the expectation of random variables  $s_N^2(n)$ , while  $NEs_N^k(n) \rightarrow 0$ , for all  $k > 2$ .

Now one take  $\Delta t = \frac{1}{N}$ , then:

$$a_{N\equiv} \frac{1}{\Delta t} E_{x,t} [x_N(t + \Delta t) - x_N(t)] = E[X_N([Nt] + 1) - X_N([Nt]) / X_N([Nt]) = x_N]. \quad (19)$$

Because  $X_N(n)$  is a binomial variable (in  $N$  and  $\bar{x}$ ), it results:

$$a_N = NE(\bar{x} - x) = NE\left(\frac{(1 + s_N(n))x}{1 + s_N(n)x} - x\right) \rightarrow [\sigma(t) - \nu(t)x]x(1 - x) \equiv a(x, t) \quad (20)$$

and

$$\frac{1}{\Delta t} E_{x,t} [x_N(t + \Delta t) - x_N(t)]^2 \rightarrow x(1 - x)[1 + \nu(t)x(1 - x)] \equiv b(x, t) \quad (21)$$

Because the conditions for the convergence hold, then, the stochastic differential equation:

$$\begin{cases} dx(t) = a(x, t)dt + \sqrt{b(x, t)}dw(t) \\ x(0) = x \end{cases} \quad (22)$$

has a unique solution (with absorption at  $x = 0$  and  $x = 1$ ).

The relations (22) have a unique solution [3].

The requirements of existence and uniqueness solution for (22) are met, supposing that we consider this system on an interval  $I_\varepsilon$ , where  $I_\varepsilon = [\varepsilon, 1 - \varepsilon]$ , with absorption at the boundary  $\delta I$  of  $I_\varepsilon$ . Thus, (22) has the unique solution  $x_\varepsilon(t)$  in  $I_\varepsilon$  up to the time

$$\tau_\varepsilon = \inf\{t / x_\varepsilon(t) \in \delta I_\varepsilon\}.$$

If it uses the *localization principle* (a result for the stochastic differential equations), we obtain:

for  $\varepsilon_1 < \varepsilon_2$ , it results  $\tau_{\varepsilon_1} \geq \tau_{\varepsilon_2}$  and

$$x_{\varepsilon_1}(t) = x_{\varepsilon_2}(t) \quad (23)$$

where  $0 \leq t \leq \tau_{\varepsilon_2}$ .

If  $\varepsilon \rightarrow 0$ , then one admit that  $x_\varepsilon(t)$  converges to a limit  $x(t)$  and  $\tau_\varepsilon \rightarrow \tau$  ( $\tau$  is the absorption of  $x(t)$ )

Then it results that

$$x_N(t) \rightarrow x(t),$$

where  $x(t)$  is 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  $x=1$ .

## Conclusions

In conclusion, we can say that once the 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  $(0,1)$ .

## References

1. BHARUHA-REID, A.T., *Elements of the theory of Markov processes and their applications*, Dover Publications Inc., Mineola, New-York, 1997.
2. BOCU, D., ORMAN, G.V., *On the Convergence of Standard Problems of Genetic Search ((II))*, Proc. International Ics Symposium on Engineering of Intelligent System EIS'98, Tenerife, Spain, 1998.
3. IKEDA, N., WATANABE, S., *Stochastic Differential Equations and Diffusion Processes*, 2<sup>nd</sup> edition, North-Holland, 1989.
4. KARATZAS, I., SHREVE, S., *Brownian Motion and Stochastic Calculus*, Springer-Verlag, 1998.
5. ORMAN V.G., *Handbook of Limit Theorems and Stochastic Approximation*, Ed. "Transilvania", Brasov, 2003.