*2012 4th International Conference on Computer Modeling and Simulation (ICCMS 2012)*
***IPCSIT vol.22 (2012) © (2012) IACSIT Press, Singapore***

# Random Number Generator by Ehrenfest Chain: Data-Oriented Approach

Ahmad Habibizad Navin[1+], Rana Fathalipoor[2], Raheleh Olfatkhah[1], Mir Kamal Mirnia[1]

[1]Computer Department, Science and Research Branch, Islamic Azad University, Tabriz, East Azerbaijan, Iran.

[2]Computer Department, Islamic Azad University Tabriz Branch, Tabriz, Iran.

**Abstract.** A new Normal random number generator is presented based on data-oriented theory in this paper. By this approach digits are generated and then the random number is achieved by appending the generated digits. To produce Normal random number by this approach, first digit is generated by Ehrenfest chain and the remaining digits are generated by Uniform distribution. Mathematical discussions are provided for validity and simulation results show the suitable accuracy of this approach.

**Keywords***:* Ehrenfest chain, Normal random number, data oriented

## 1. Introduction

Random number generation is an important topic in computer science and engineering that is used in the following areas: computer games, encryption, genetic algorithms, modeling and simulation of complicated systems, sensor networks, and mathematical calculation by using Monte Carlo method and etc. The main contribution of this paper is to present a new method for generating normal random numbers by using Ehrenfest chain based on dataoriented theory. A part of data-oriented theory discusses about the digit probability of random numbers providing a new theory named "Numerical Probability". Digit probabilities of random numbers can be calculated by this theory [1]. To achieve the proposed random number generator, the paper is organized as follows: formal definitions to outline the new model are defined in the next section. Mathematical discussions are presented in section 3. New model is explained in section 4. Finally, conclusion provided.

## 2. Formal Definitions

In this section, the following formal definitions are provided to discuss the proposed method formally.

### 2.1. Random variable

A random experiment with an outcome space S, a function X that assigns to each element s in S one and only one real number X(s)=x is called a random variable [4]. The space of X is the set of real numbers $\{x: X(s) = x, s \in S\}$, where $s \in S$ means the element s belongs to the set S.

### 2.2. Discrete random variable

Let X denotes a random variable in one dimensional space S, which is a subset of the real numbers. Suppose that the space S contains a countable number of points; that is, S contains either a finite number of points or the points of S can be put into a one-to-one correspondence with the positive integers. Such a set S

+ Corresponding author.
*E-mail address*: ah_ habibi@iaut.ac.ir.

is called a set of discrete points or simply a discrete outcome space. Furthermore, the random variable X is called a random variable of the discrete type, and X is said to have a distribution of the discrete type [4]. For a random variable X of the discrete type, the probability P(X=x) is frequently denoted by f(x), and this function is called the probability mass function. The probability mass function of a discrete random variable X is a function that satisfies the properties in (1).

(a) $f(x) > 0, \quad x \in S;$

(b) $\sum_{x \in s} f(x) = 1;$ (1)

(c) $P(X \in A) = \sum_{x \in s} f(x), \quad where\ A \subset S.$

### 2.3. Continuous random variable

There are some random variables whose set of possible values is uncountable. Let X be such a random variable. We say that X is a continuous random variable if probability density function of the random variable X with space S is an integrable function f(x) satisfying conditions in (2) [4].

(a) $f(x) > 0, \quad x \in s;$

(b) $\int_s f(x)dx = 1;$ (2)

(c) $The\ probability\ of\ the\ event\ X \in A$

$$is\ P(X \in A) = \int_A f(x)dx.$$

### 2.4. Binomial distribution

The binomial distribution is used when there are exactly two mutually exclusive outcomes of a trial [4]. These outcomes are appropriately labeled "success" and "failure". The binomial distribution is used to obtain the probability of observing x successes in N trials, with the probability of success on a single trial denoted by p. The binomial distribution assumes that p is fixed for all trials. Equation (3) is the formula for the binomial probability mass function.

$$P(z, p, n) = \binom{n}{x}(p)^x(1-p)^{(n-x)}, \text{ for } x = 0,1,2,\ldots,n \quad (3)$$

where

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

### 2.5. Normal random variable

A random variable is said to be uniformly distributed over the interval (0,1) if its probability density function is, as in (4) [8]

$$Ux(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & otherwise. \end{cases} \quad (4)$$

### 2.6. Normal random variable

X is a normal random variable with mean λ and variance $\delta^2$ if and only if its probability density function satisfies (5) [6].

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}. \quad (5)$$

The normal distribution is probably the most frequently used distribution. Its graph looks like a bell-shaped function, which is why it is often called bell distribution. The normal distribution is important in probability theory and statistics. Empirically, many observed distributions, such as people's heights, test scores, experimental errors are normal and theoretically, the normal distribution arises as a limiting distribution of averages of large numbers of samples, justified by the central limit theorem. Different kind of normal distribution is used one of which is standard normal distribution. In this paper special normal distribution is defined below.

### 2.7. Standard normal random variable

Let Z be a normal random variable with μ = 0 and δ = 1. Then Z is the standard normal random variable, with probability density function as in, (6) [5]

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \tag{6}$$

The Fig. 1 shows the graph of Φ (z) and the probabilities for the values of Z in the intervals of width δ about the mean μ. The normal random variable arises when observations of a quantity are disturbed by a large number of random fluctuations which are independent of each other.

### 2.8. Special normal random variable

Let S be a normal random variable with $\mu = \frac{1}{2}$ and $\delta = \frac{1}{6}$. Then S is the *special normal random variable*, with the probability density function, as in (7)

$$f_s(s) = \frac{3}{\sqrt{2\pi}} e^{-\frac{9}{2}s^2} \tag{7}$$

Most values of S are distributed between 0 and 1. Because of importance of S in our discussion it is named Special normal random variable. The graph of f(s) and the probabilities for the values of s are shown in Fig. 2.

### 2.9. Stochastic process

*Stochastic process* (SP) is a sequence of random variables {X(t)| t ∈T} defined on a given probability space indexed by the time variable t, where t varies over an index set T[2]. Stochastic process is a statistical process involving a number of random variables depending on a variable parameter (which is usually time).

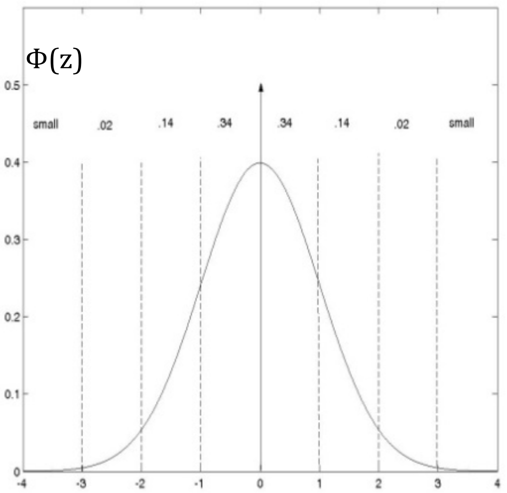


Fig.1 The standard normal probability density function; μ = 0, and δ= 1.

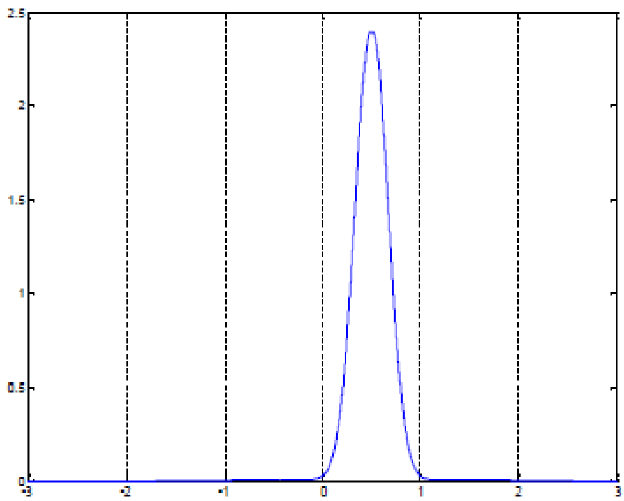

Fig.2 The normal probability density function; μ =1/2 and δ= 1/6.

### 2.10. Markov process

*Markov process* is a SP whose dynamic behavior is such that probability distribution for its further development depends only on its present state and not the process leading present state. If the present state space is discrete, then it is named a discrete Markov chain. Let S be a countable set. We consider random variables Xn with values in S [2]. The official definition of Markov chain is by the Markov property, as in (8)

$$P[X_{n+1} = b \,|X_0 = a_0, X_1 = a_1, \dots, X_n = a_n] = P[X_{n+1} = b \,|X_n = a_n] \tag{8}$$

In other words, the next state depends only on the present state. We shall also assume that the process is time homogeneous, so that the transition probabilities are always the same as in (9)

$$P[X_{n+1} = b \,|X_n = a] = P[X_1 = b \,|X_0 = a] \tag{9}$$

### 2.11. Transition probabilities

$Let\ P[X_{n+1} = y|X_n = x] = P(x, y)$ be the *transition probabilities.* In fact, the chain is associated just with the transition probabilities. The transition probabilities determine the Markov chain.

### 2.12. Stationary distribution

For each y a stationary distribution is a probability mass function π(y) as in (10) [4].

$$\sum_x \pi(x)P(x, y) = \pi(y) \tag{10}$$

### 2.13. Ehrenfest chain

The *Ehrenfest chain* is a Markov chain which consisting a process on the d + 1 points 0; 1; 2; 3; : : : ; d. This model consists of 2 boxes, labeled A and B, containing d balls labeled {1,2,..d}. An integer is randomly selected from {1,2,...d} and the ball corresponding to this integer is removed from its box and placed in the other box. This process is repeated indefinitely, by assuming independence between trials. The state of the chain is the number of particles in box A. In Ehrenfest chain transition probabilities are as in (11) [3]

$$P_{xy} = \begin{cases} \frac{d-x}{d}, y = x + 1 \\ \frac{x}{d}\ ,\ y = x - 1 \\ 0 \qquad\quad else \end{cases} \tag{11}$$

## 4. Mathematical Discussions

Equation (12) is the stationary distribution of Ehrenfest chain.

$$\pi(x) = \binom{d}{x}\frac{1}{2}x\frac{1}{2}d - x = P\left(x, \frac{1}{2}, d\right) \tag{12}$$

With the binomial distribution [3]. Many interesting problems can be addressed via the binomial distribution. However, for large Ns, the binomial distribution can become quite awkward to work with. Fortunately, as N becomes large, the binomial distribution becomes more and more symmetric, and begins to converge to a normal distribution. That is, for an appropriate N, a binomial variable X is approximately ≅N(Np, Npq). Hence, the normal distribution can be used to approximate the binomial distribution. Therefore the stationary distribution of Ehrenfest chain is approximated to a normal distribution [6]. Therefore the stationary distribution of Ehrenfest chain can be approximated by normal distribution as in (13)

$$\pi(x) \approx N(\frac{d}{2}, \frac{d}{4}) \tag{13}$$

Let d=9. Therefore (14) is achieved

$$\pi(x) \approx N(4.5, 2.25\,) \tag{14}$$

This is the stationary distribution for generating number between 0 and 9. By scaling these numbers we obtain (15)

$$\pi(x) \approx N(0{,}45, 0.0225\,) \approx N(\frac{1}{2}, \frac{1}{36}) \tag{15}$$

# 5. New Modeling

To generate random numbers with new method in Ehrenfest chain we set d=9. New random number generation procedure is shown in Fig. 3.To produce Normal random number in this method, using random walk, the first digit is generated by Ehrenfest chain and the remaining digits are generated by Uniform distribution. Average and variance of generated numbers are in (16)

$$\bar{x} \approx \frac{1}{2}, S^2 \approx \frac{1}{36} \tag{16}$$

This is an approximation for N(1/2,1/36) which is obtained in mathematical discussions. Generated numbers density function is approximated by the probability function in (17)

$$f_s(s) = \frac{6}{2\pi}e^{-18(s-\frac{1}{2})^2} \tag{17}$$

This is the special normal function defined before. The simulation result in Fig.4 shows that the distribution of random numbers is the same as normal distribution.

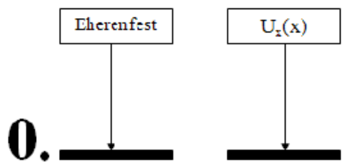


Fig.3 New random number generation method

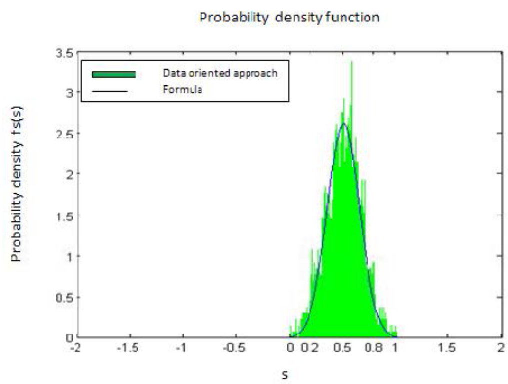


Fig.4 Distribution of random numbers with new method

Simulation is performed by Matlab software. To study the similarity between generated numbers with normal distribution and also the generated numbers quality, mean square error (MSE) is achieved from (18)

$$MSE = \frac{1}{10}\sum_{i=0}^{9}(rf_i - \int_{0.i}^{0.(i+1)} f_s\,(s)ds)^2 \tag{18}$$

Where $\mathrm{rf_i}$ is the relative frequency of generated numbers in the interval [0.i, 0.(i+1)] and $\int_{0.i}^{0.(i+1)} f_s\, ds$ is the area of the region bounded by function f(s) in interval [0.i, 0.(i+1)]. Resulting numbers with Total relative frequency are [0.0014 0.0152 0.0679 0.1636 0.2473 0.2494 0.1689 0.07390.0197 0.0024]. With the numbers for the corresponding area [0.0035 0.0188 0.0672 0.01576 0.2432 0.2468 0.1647 0.0723 0.0208 0.0039]. Now the mean square error is MSE=0.000010109. Desirable similarity of proposed generator and quality of generated numbers are concluded from MSE calculation.

# 6. Conclusion

Data-oriented theory provides some methods to model the concepts with data structure. By using efficient data in modeling, required results will be obtained by less calculation. A new random number generator by Ehrenfest Chain based on data-oriented theory is presented in this paper. By this approach digits are generated and then the random number is achieved by appending the generated digits. To produce Normal random number by this approach, first digit is generated by Ehrenfest chain and the remaining digits are generated by Uniform distribution. Obtained result shows the suitable quality of generated numbers.

# 7. References


[1] Ahmad Habibizad-Navin; Mehdi Naghian Fesharaki; Mohammad Teshnelab, Uniform distribution simulation by using digit probability, In proceeding of The 5th Seminar on Probability and Stochastic processes, Birjand, Iran, 2005, pp: 365-368.

[2] K. S. Trivedi, Probability, Statistics with Reliability, Queuing and Computer Science Applications, 2nd ed., Wiley, 2002, ISBN: 0-471-33341-7. pp.337-338.

[3] E. A. Pasha. Stochastic Process, 13th ed., Payame Noor University Press, 2002, ISBN: 378/1750955, pp.79-83

[4] R.V. Hogg; E.A. Tanis. Probability and Statistical Inference, 3rd ed., Pearson, 2004, ISBN: 81-7808-631-X,pp.108-166.

[5] Raheleh Olfatkhah; Ahmad Habibizad Navin; Mirkamal Mirnia. A Novel Model of Normal Random Variable Based on Data-Oriented Theory, 2010.

[6] Prem S. Mann; Christopher Jay Lacke. Introductory Statistics, John Wiley and Sons, 2010, pp. 283-285.

[7] Sheldon Ross. A First Course in Probability, 9th ed., 1976, ISBN: 0-02-979620-2, pp.124-125.