

## Mathematical Discussions About Data Oriented Modeling of Uniform Random Variable

<sup>1,2</sup>Ahmad Habibizad Navin, <sup>3</sup>Mehdi Naghian Fesharaki, <sup>1</sup>Mirkamal Mirnia and <sup>4</sup>Mohammad Teshnehlab

<sup>1</sup>Islamic Azad University, Tabriz Branch, Tabriz, Iran, IAUT-CRL

<sup>2</sup>Islamic Azad University, Science and Research Branch, Tehran, Iran

<sup>3</sup>Malek-Ashtar University of Technology, Tehran, Iran

<sup>4</sup>Khaje Nasir Toosi University of Technology, Tehran, Iran

---

**Abstract:** In this study, to show that UDPD is an acceptable and efficient model of uniform random variable, we introduce mathematical and numerical discussions for identification and characterization of it and also we show how to calculate some probabilistic parameters as mean, variance and mathematical expectation by using UDPD. Thus we provide a way to discuss random variables with n digits and make a new theory named numerical probability. This theory is enables us to discuss about random variables with predefined number of digits as required in applied and engineering systems.

**Key words:** Uniform random variables, numerical probability, data oriented modeling, prodigraph, uniform digital probability digraph

---

### INTRODUCTION

Recalling Hogg's saying (Hogg and Tanis, 2004) "Those who can use computer and statistics together will become stars in their field of study", the importance of joining the computer and statistical science is clear. Hogg's saying can be understood more when we notice today's computer developments and their successful performance in all computing fields. Computers have accelerated advancements of sciences and technology.

Since the memory is cheap and easy accessible nowadays thus we can use more memory locations or large set of data for modeling concepts freely. Data oriented modeling is a technique to model the concepts by data structures, sizeable amount of data, such as vectors, matrices, trees, graphs and etc., which notifies the title Data Oriented Modeling. Data oriented modeling can be used in applied systems and sciences. We have used it for modeling fuzzy controller of Anti-lock Brake System, ABS, for the first time, so we could reduce time complexity of fuzzy controller to obtain the near real time fuzzy based controller (Habibizad *et al.*, 2007d). Then, we have introduced data oriented models of Uniform Random Variable to create a new field of study which connects computer science, statistical and probability together (Habibizad *et al.*, 2007a). Subsequently we present a data oriented model of image and characterize a framework for image processing tasks such as segmentation, clustering and measuring images similarity with desired precision

and speed (Habibizad *et al.*, 2007c). Finally, a new method introduced for generating uniform random numbers based on data oriented modeling (Habibizad *et al.*, 2007b). This method is a simulation technique of UDPD model. Generated numbers by this method have better uniformity than conventional methods and this is why, is called Uniformity Improving Method, UIM.

The main idea of this study is using mathematical and numerical discussions to identify and characterize the Data Oriented Model of Uniform Random Variable (Habibizad *et al.*, 2007a) and to show that it is an acceptable and efficient model of uniform random variable. Data Oriented modeling takes care the concepts with data structure. Up to now, statisticians have used mathematical functions to model random variables. But this approach conforms to human brain. In other words, for human, it is easier to use mathematical functions for modeling. In this approach a random variable, X, is modeled by its distribution function,  $f(x)$ . Specification, characteristic and frequency of X values are inferred from  $f(x)$  by mathematical calculations. In our approach we will propose new data oriented model for random variables, such that models X based on sizeable amount of data (data structures). These models are developed upon some data structures such as vectors, matrices, trees, graphs and Digital Probability Digraphs. Specification, characteristic and frequency of data represented by X, can be inferred by data processing. These models of random variables are well-matched with the digital

computing systems, so can be used in probability and statistical inference by computers efficiently. Also Data oriented modeling of random variables led us to present a theory named numerical probability and provide the framework and area for designing special computer to do probabilistic and statistical calculations very fast. Our research group is going to design such computers; some of them will appear soon.

**FUNDAMENTAL CONCEPTS**

Statistics and Probability, as one of the most applied and essential sciences have become strongly dependent on computers. Our contribution (Habibizad *et al.*, 2007a) is to introduce a new model of random variable in a certain way which is consistent to today’s computers structure. According to the Hogg saying (Hogg and Tanis, 2004), “we look forward to that these models become stars in the Statistics, Probability and Computer science”.

In our approach random variable X is modeled with data structure. UDPD is a model of uniform random variable based on Data-oriented modeling. In other words it is a weighted digraph which is explained further.

To characterize the data oriented model of uniform random variable, UDPD and analyze it numerically following Definitions and theorems are given. It is worth to mention that some of the following definitions are quoted from Habibizad *et al.* (2007a) to clear some likely ambiguities and prepare a good platform for mathematical discussions and numerical analyses of UDPD and uniform random variable.

**Definition 1:** Weighted directed graph G is called probability digraph or prodigraph in short if and only if for any vertex  $a \in v$  we have:  $\sum_{b \in v} P_{ab} = 1$ , where  $p_{ab}$  is the weight of edge  $a \rightarrow b$ .

Note that  $P_{ij} = 0$  if and only if there is not any edge from  $i$  to  $j$ . Therefore like the adjacency matrix we can represent prodigraph by vector of vertices  $V = [v_i]$  and probability matrix  $P = [P_{ij}]_{|V| \times |V|}$ . Here after we can denote it by  $G = ([V], [P])$ . For example prodigraph of Fig. 1 can be denoted by:

$$G_p = \left( [a, b, c], \begin{bmatrix} 0.2 & 0.8 & 0.0 \\ 0.6 & 0.0 & 0.4 \\ 0.3 & 0.0 & 0.7 \end{bmatrix} \right)$$

To get involved in the world of numbers, we introduce a special case of prodigraph. This prodigraph is produced with digital vertices and is called Digital Prodigraph which is defined as follows:

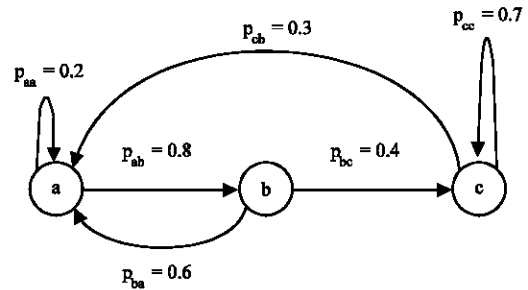


Fig. 1: A prodigraph

**Definition 2:** Let  $G = ([V], [P])$  be a prodigraph. Then G is a Digital Prodigraph if and only if  $V = [0, 0, 1, 2, \dots, 9]$ .

**Definition 3:** Let  $G = ([V], [P])$  be a Digital Prodigraph and  $w = 0, a_1, a_2, \dots, a_n$  be a walk on this graph, where  $a_i$  is a vertex of G. Value Of Walk of this walk is denoted by  $VOW_w$  and is defined as follows:

$$VOW_w = a_1 \times 10^{-1} + a_2 \times 10^{-2} + \dots + a_n \times 10^{-n} = \sum_{i=1}^n a_i \times 10^{-i} = 0.a_1a_2\dots a_n.$$

In other words, VOW of each walk is obtained by appending each vertex of walk as we traverse digital prodigraph from 0. to  $a_n$ .

For each VOW which is obtained from digital prodigraph we assign a probability value which is defined as follows:

**Definition 4:** Let  $G = ([V], [P])$  be a Digital Prodigraph and  $w = 0, a_1, a_2, \dots, a_n$  be a walk on this graph. Let  $VOW_w = y = 0.a_1a_2\dots a_n$ . Then  $f(y)$  is the Probability of  $VOW_w$  if and only if  $f(y) = P_{0a_1} \times P_{a_1a_2} \times \dots \times P_{a_{n-1}a_n} = P_{0a_1} \prod_{i=2}^n P_{a_{i-1}a_i}$ .

**Theorem 1:** Let  $(VOW_{G_p})_0^n$  be the set of VOW of all walks with length n starting from 0. On the Digital Prodiagraph  $G_p$ . Then  $\sum_{y \in (VOW_{G_p})_0^n} f(y) = 1$ , where  $f(y)$  is the probability of y.

**Proof:** See Habibizad *et al.* (2005a).

**Definition 5:** Let  $G = (V, E)$  be a prodigraph and  $a, b \in V$ . a is adjacent to b, if  $P_{ab} > 0$ . And a is adjacent from b, if  $P_{ba} > 0$ .

**Theorem 2:** Let U be a random variable which is distributed uniformly in  $[0, 1]$  and  $V_u = \{0, 1, 2, \dots, 9\}$  be the set of digits. The probability that  $v \in V_u$  is the nth digit of U equals 0.1.

**Proof:** See Habibizad *et al.* (2007b).

Theorem 2 is known as digits uniformity Theorem.

**Data oriented model of random variable:** Let X be a random variable and f(x) is the probability distribution function of it. The main advantage of Data Oriented Modeling of random variable is determining Digital Probability Digraph in a way that their VOWs matches the values of random variable X and P<sub>VOW</sub> matches the probability inferred from f(x).

To be able to model random variable with Digital Probability Digraph we need to have Digit Probabilities in hand so that it can be used as the edge weights in the above structure. We have presented the calculations of these Digit Probabilities in Habibizad *et al.* (2005c) and we have proved the uniform distribution, Digit Probabilities are equal to 0.1 for all edges by the Digits Uniformity Theorem. The uniform random variable is modeled by using the data oriented modeling explained further.

**Uniform digital probability digraph:** Similar to random variable U, which is distributed uniformly in [0, 1], we define G<sub>u</sub> = (V<sub>u</sub>, P<sub>u</sub>), Uniform Digital Probability Digraph, UDPD in short, as a data oriented model of U, based on digits uniformity theorem as follows:

$$V_u = [0., 0.1, 2., \dots, 8, 9]$$

$$P_u = \begin{bmatrix} 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

Based on digits uniformity theorem, if we let w = 0., a<sub>1</sub>, a<sub>2</sub>, ..., a<sub>n</sub> be a walk on the G<sub>u</sub> = (V<sub>u</sub>, P<sub>u</sub>) then the VOW<sub>w</sub> will be a random number which is distributed uniformly in [0, 1]. In other words, each digit of VOW<sub>w</sub> is generated with the probability of 0.1 and by putting all these digits together we get an n-digit uniform random number.

**Lemma 1:** Let y be the VOW with length n, obtained from G<sub>u</sub>. Then the probability of y is equal to 0.1<sup>n</sup>.

**Proof:** It is trivial by definition 4.

**Lemma 2:** Let y be the VOW with length n obtained from G<sub>u</sub>. Then we have:

$$\lim_{n \rightarrow \infty} f(y) = 0$$

**Proof:** Let y be the VOW with length n obtained from G<sub>u</sub>. Then by lemma 1 we have:

$$\lim_{n \rightarrow \infty} f(y) = \lim_{n \rightarrow \infty} 0.1^n = 0.$$

### DISCUSSION

We claim that G<sub>u</sub> is a uniform random variable which is distributed uniformly in distance [0,1] and thus can be use in probabilistic problems and statistical inference instead of U. The basic relation of Calculations of the probabilistic parameters is presented in this section to clarify this claim. The main contribution of this paper is to identify and characterize Data Oriented Models of Uniform Random Variable, G<sub>u</sub> by mathematical and numerical discussions. In other words, in this section, we present basic relations for calculating the Probability mass function, Expectation value, Mean, Variance and Moment-generating function of G<sub>u</sub>.

**Probability mass function of G<sub>u</sub>:** The probability mass function (p.m.f) f(x) of a discrete random variable X is a function that satisfies the following properties (Hogg and Tanis, 2004):

- (a) f(x) > 0, x ∈ S;
- (b) ∑<sub>x ∈ S</sub> f(x) = 1;
- (c) P(X ∈ A) = ∑<sub>x ∈ A</sub> f(x); where A ⊂ S.

The probability mass function of G<sub>u</sub> with n digits is shown by f(G<sub>u</sub><sup>n</sup>) and satisfies the following properties:

- (a) f(y) > 0, y ∈ (VOW<sub>G<sub>u</sub></sub>)<sub>0</sub><sup>n</sup>;
- (b) ∑<sub>y ∈ (VOW<sub>G<sub>u</sub></sub>)<sub>0</sub><sup>n</sup></sub> f(y) = 1;
- (c) P(y ∈ A) = ∑<sub>y ∈ A</sub> f(y), where A ⊂ (VOW<sub>G<sub>u</sub></sub>)<sub>0</sub><sup>n</sup>,

where, f(y) is given by definition 4 and n is the number of digits in y.

The property (a) is satisfied because f(y) is the product of n positive numbers, digit probabilities and the property (b) is satisfied by theorem 1.

The satisfaction of property (c) is evident.

The function f(y) given by definition 4 is the probability mass function of G<sub>u</sub>. Hereafter we call f(y) the probability mass function of G<sub>u</sub> with n digits and denoted by f(G<sub>u</sub><sup>n</sup>).

**Mathematical expectation of  $u(G_u)$  with  $n$  digits:** If  $f(x)$  is the probability mass function of the random variable  $X$  of the discrete type with space  $S$  and if the summation  $\sum_{x \in S} u(x)f(x)$ , which is sometimes written as  $\sum_S u(x)f(x)$  exists, then the sum is called the mathematical expectation or the expected value of the function  $u(X)$  and it is denoted by  $E[u(X)]$ . That is  $E[u(X)] = \sum_{x \in S} u(x)f(x)$  (Hogg and Tanis, 2004).

Similarly if  $f(G_u^n)$  is the probability mass function of  $G_u$  with  $n$  digits and if the summation  $\sum_{y \in (V \circ W_{G_u})^n} u(y)f(y)$ , sometimes written as  $\sum_{(V \circ W_{G_u})^n} u(y)f(y)$ , exists, then the sum is called the mathematical expectation or the expected value of the function  $u(G_u)$  with  $n$  digits and is denoted by  $E[u(G_u^n)]$ . That is  $E[u(G_u^n)] = \sum_{(V \circ W_{G_u})^n} u(y)f(y)$ .

**Mean of  $G_u$  with  $n$  digits:** The mean value of  $G_u$  with  $n$  digits is equal to Expected value of  $G_u$  with  $n$  digits and is denoted by  $E[G_u^n]$  where  $E[G_u^n] = \sum_{y \in (V \circ W_{G_u})^n} y f(y)$ .

**Variance of  $G_u$  with  $n$  digits:** The variance of random variable  $X$  is given by  $\sigma_x^2 = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$ . Similarly the variance of  $G_u$  with  $n$  digits is given by  $\sigma_{G_u^n}^2 = E[(G_u^n - E(G_u^n))^2] = E(G_u^{n2}) - [E(G_u^n)]^2$ .

**Moment-generating function of  $G_u$  with  $n$  digits:** Let  $X$  be a random variable of the discrete type with p.m.f.  $f(x)$  and space  $S$ . If there is a positive number  $h$  such that  $E(e^{tx}) = \sum_{x \in S} e^{tx} f(x)$

exists and is finite for  $-h < t < h$ , then the function defined by  $M(t) = E(e^{tx})$  is called the moment-generating function of  $X$  (or of the distribution of  $X$ ). This is often abbreviated m.g.f. (Hogg and Tanis, 2004).

Similarly the moment-generating function of  $G_u$  with  $n$  digits is defined by  $M(t) = E(e^{tG_u^n})$ .

All above relations give probabilistic parameters by using  $G_u$  with  $n$  digits instead of random variable. These relations provide formal arguments to discuss random variable with  $n$  digits and make a new theory named numerical probability which is explained further.

**Numerical probability:** We know that there are two basic types of random variables continuous and discrete. Let  $X$  be a continuous type random variable. Then the probability of  $X$  to be exactly a specific value is zero. However in the engineering world and applied systems, each continuous random variable is presented by a predefined precision and finite number of digits. This means that continuous type random

variables are converted to discrete type random variables when they are used in applied and engineering systems. Since the data oriented modeling models the random variable with  $n$  digits, hence, data oriented approach can be used to model both continuous and discrete random variables. For example from above discussions specially lemmas 1, 2 and theorem 1 we have

$$\lim_{n \rightarrow \infty} (V \circ W_{G_u})_0^n = [0,1] \text{ also } \lim_{n \rightarrow \infty} f(G_u^n) = f(U)$$

and vice versa. For a given number of digits of  $U$ ,  $n$  we can model it by  $G_u$  with  $n$  digits.

Numerical probability theory discusses mathematical inference and analyses the methods which provides numerical answer of probabilistic problems. This theory is based on data oriented modeling of random variable. This models the random variable by a data structure as vector, matrix, graph, tree and etc. These structures were made up by digit probability regularly. In contrast to conventional methods, the new theory provides the answer of probabilistic and statistical problems in terms of digits probability. For example an algorithm has been given to estimate population distribution (Habibzad *et al.*, 2005b) and we introduced a novel simulation method of uniform random variable (Habibzad *et al.*, 2007b) and also an algorithm has been presented to test random number generators based on digit probability by our group which is going to appear soon.

## CONCLUSION

A short view of Uniform Digital Probability Digraph, UDPD, the data oriented model of uniform random variable are presented and to prove that it is an acceptable and efficient model, mathematical and numerical discussions are presented to identify and characterize it. These arguments are enable us to discuss about random variable with predefined number of digits as we need in applied and engineering systems and are very important because:

- It is the base of numerical probability as a new and efficient theory.
- It joins the computer science with the probability and statistical science strongly. Then we are able to use the computer in statistical inference potently.
- It provides the framework and area for designing special computer to do probabilistic and statistical calculations very fast.

By considering the above key notes we hope the data oriented models of random variables will become stars of computer science, statistics and probability.

**REFERENCES**

- Habibizad, A., M. Naghian and M. Teshnelab, 2005a. Probability digraph and some of its important structures. 5th Seminar on Probability and Stochastic Processes. Birjand-Iran, pp: 155-160.
- Habibizad, A., M. Naghian and A. Alayi, 2005b. Presenting an algorithm to estimate distribution of a statistical population. 35th Annual Iranian Mathematics Conference. Ahwaz-Iran, pp: 80-82.
- Habibizad, A., M. Naghian and M. Lotfi, 2005c. Presenting a probabilistic problem and solution of it. 35th Annual Iranian Mathematics Conference. Ahwaz-Iran, pp: 100-105.
- Habibizad, A., M. Naghian, M.K. Mirnia, M. Teshnelab and E. Shahamatnia, 2007a. Data oriented modeling of uniform random Variable: Applied approach. In: Proceedings of World Academy of Science, Engineering and Technology. Vienna, Austria, 21: 382-385.
- Habibizad, A., M. Naghian, M.K. Mirnia and M. Teshnelab, 2007b. A novel method for improving the uniformity of random number generator based on data oriented modeling. *Int. J. Comput. Sci. Network Security*, 7 (7): 269-273.
- Habibizad, A., A. Sadighi, M. Naghian, M.K. Mirnia, M. Teshnelab and R. Keshmiri, 2007c. Data oriented modeling of image: As a framework for image processing. In: Proceeding of World Academy of Science, Engineering and Technology. Berlin, Germany, 21: 390-393.
- Habibizad, A., M. Naghian, M. Teshnelab and E. Shahamatnia, 2007d. Fuzzy based problem-solution data structure as a data oriented model for ABS controlling. In: Proceedings of World Academy of Science, Engineering and Technology. Barcelona, Spain, 20: 364-369.
- Hogg, R.V. and E.A. Tanis, 2004. Probability and Statistical Inference. Pearson Education.