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

A Suggestion about Optimal Size of Digit Bank in Data Oriented Random Number Generator

Ahmad Habibizad Navin[1], Elnaz Safar Ali Khani[2], Seyyed Yaser Torabi[3]

[1]Computer Research Lab, Science and Research, Branch, Islamic Azad University, Tabriz, Iran

[2]Computer Research Lab, Islamic Azad University, Tabriz Branch, Tabriz, Iran

[3]Computer Research Lab, Amirkabir University of Technology, Tehran, Iran.

Abstract. Recently data oriented theory to model random variables has been presented by our research group. Data oriented theory uses data structures to model concepts instead of functions. By using these data structures, a Data oriented Random Number Generator (DRNG) has been made for any random variable. For making DRNG digit banks are used. In this paper the size of digit banks in DRNG is discussed and an optimal size of it is suggested.

Keywords :Random number generator, probability, digit bank, random variable, data-oriented modeling

1. Introduction

Data-oriented is a new theory that presents methods by which any concepts can be modeled in terms of data structures so that much more data about problems is saved.

Due to the use of advanced memory technology in modern computing systems, concepts modeled with data-oriented theory can be processed more quickly. The questions can be answered by data processing or by fewer amounts of mathematical operations using these models. The methods to find the solution by using large amount of data in data structures are called data-oriented methods. Tree, graph, matrix, vector or some other data structures can be used for this purpose.

The basic structure of data-oriented modeling has been presented in [1] and requirement tools, definition and important mathematical theorems for these models were presented in [2] .Recently some methods have been presented to generate data oriented models of some continues random variables, For example data oriented models of uniform [3], normal [4] and chi Square random variable [5] have been presented. Based on models of chi square random variables we have developed chi square random variable generator (chi square DRNG). For making DRNG we use some digit banks along with some operations are done on it. Selecting appropriate size for digit bank in DRNG is one of the important challenges. By increasing the size of digit bank, cost increases, speed reduces and also similarity of the relative frequency of generated numbers to the probability density function increase.

The main contribution of this paper is to suggest a suitable size for digit banks that could generate numbers with acceptable precision and optimize it's size with appropriate speed.

2. Basic Concept

To outline the models of random variables and DRNG construction, some introductory definitions are required. Probability diagraph, Digital Prodigraph, Value Of Walk (VOWw) and Probability of VOW (P_{VOW}) are defined in [6,7]. In this paper we use some other definitions as follow:

Definition: To construct new model of random variables with probability tree**,** digit probabilities should be determined. Ahmad fact is used to compute these probabilities [4,5] which gives probability of a random variable's digits by using its density function. Suppose that the random variable X has probability density function f(x) thus based on Ahmad fact, the probability of the first digit of X after the point to be i, is denoted by P_i and computed by:

$$p_i = p_{(0.i<x<i+1)} = \int_{0.i}^{0.i+1} f(x)dx.$$

For example: $p_1 = \int_{0.1}^{0.2} f(x)dx.$

Second digit probability can also be calculated by using Ahmad fact if the first digit is given. Let i be the first digit of *X*, the probability of second digit of random variable *X*, to be *j*, is denoted with P_{ij} and computed by

$$p_{ij} = p_{(0.ij<x<ij+1)} = \int_{0.ij}^{0.ij+1} f(x)dx.$$

For example, let the first digit of *X* be 2. The probability of second digit of random variable *X*, to be 9, is denoted with P_{29} and computed by

$$P_{29} = P_{(0.28<x<0.29)} = \int_{0.28}^{0.29} f(x)dx.$$

If the prior digits are given to be $i_1 i_2 .. i_{k-1}$, the probability of *kth* digit of *X* to be *j*, is computed by

$$p_{i_1 i_2 \dots i_{k-1} j} = p_{(i_1 i_2 \dots i_{k-1} j<x<i_1 i_2 \dots i_{k-1} j+1)} = \int_{0.i_1 i_2 \dots i_{k-1} j}^{0.i_1 i_2 \dots i_{k-1} j+1} f(x)dx.$$

3. Data Oriented Models of Random Variables

A various types of random variables including normal random variables, uniform and chi square variables have been modeled by data oriented method [3,4,5] which Normal and Chi square random variable are given here briefly

3.1. Normal Random variable

A new hierarchical discrete model of special normal random variable as a data-oriented model has been introduced [4]. Special normal random variable is a normal random variable with μ=0 and σ=1/3. Digital 2-complete probability tree is used to model special normal random variable. To make this structure, digit's probabilities of normal random variables are required. Ahmad fact is used to obtain them by using probability density function. That tree is shown in Fig.1.

3.2.Chi square random variable

In probability theory and statistics, the chi square distribution (or χ^2-distribution) with k degrees of freedom is the distribution of a sum of the squares of k independent standard random variables. Degree of Freedom (DF) is the only parameter of this random variable [7].

Chi square random variable has been modeled with data structure in [5]. The probability complete tree is used for modeling this random variable. Fig.2 shows the data oriented models of chi square random variable with 3 degree of freedom. Weights of edges show the digits with corresponding probabilities that have been calculated with AF [5]. Based on this model we designed a chi square DRNG that is explained as follows.

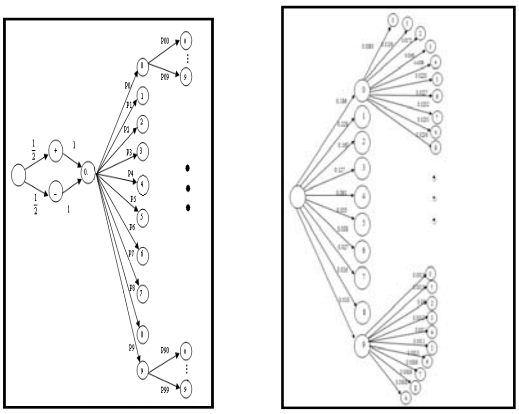

Fig.1.(a) digital probability compelet tree of spacial normal random variable; (b) probability compelet tree of chi square random variable (DF=3)

4. Drng With Data Oriented Approach

Suppose X denotes the attribute of the result, in a random test. To identify this non-deterministic phenomenon, the probability distribution function f(x) must be estimated. To simulate random variable X by data oriented method, the numbers should be generated so that its relative frequency be almost close to f(x). To produce such random numbers DRNG is used. Construction method of DRNG is as follows:

In order to create DRNG, a digit bank named main bank is considered so that some operations should be done on it.

The digits in this bank are extracted from the data structure that is made for random variable. In this section we explain DRNG obtained from first levels of tree.

To make digit bank, weights of edges are used. So that from the values obtained for P_i , the first two digits after decimal point is selected as the number of the iteration number for i. For example if P_1=0.228 then the number 22 is selected which is considered to be the number of times that 1 are repeated and 1 is inserted in the bank 22 times.

After placing these numbers in the digit bank, one number is selected from this bank randomly (it is the DRNG out puts). To generate another random number, by repeated shuffling, the digits placed in the bank will be exchanged and then a number is chosen from the bank randomly again. By repeating this operation for many times (about 1000 times), the relative frequency diagram can be obtained from the generated numbers. This diagram is similar to f(x). Simulation result for chi square random variables is shown in Fig. 2.

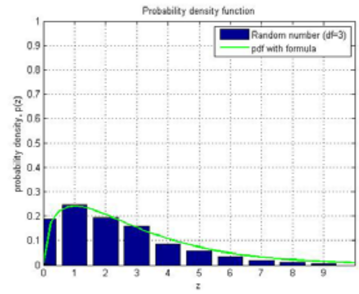

Fig.2.DRNG of chi square random variable (DF=3)

5. Suggestion about Optimal Size of Digit Banks

As discussed above, the relative frequency diagram of DRNG is almost similar to probability density functions of random variables. The similarity depends on two parameters: size of digit bank and the used depth of probability tree. In this paper we discuss on selecting suitable size of digit bank, from cost, precision and speed point of view.

To increase the similarity of relative frequency of generated numbers to f(x), the size of digit bank and the depth of tree used in modeling should be increased.

In this paper we used complete probability tree with depth 2 and the size of digit bank will be discussed.

To model random variable to the second level of tree, ten new banks are considered which is labeled from 0 to 9.

From the values obtained for P_{ij}, the first two digits after decimal point is selected as the number of the iteration number for j in bank i. For example if P_{25}=0.0493 so that the number 04 is selected which is considered to be the number of times that 5 are repeated and 5 inserted in the bank with label 2 four times. By calculating P_{ij} for all i, j and inserting selection results in digit banks, DRNG can start**.** One digit is selected from the main bank randomly First (as explained before) and then depending on the selected number, one of the other ten bank is selected. For example if the selected number from main bank is 8 the bank with 8 label is selected. Then from this newly selected bank, one element is chosen randomly too.

Simulation result is shown in Fig.3.a for this case. As it is clear the simulation results in this case are not good and similarity does not increase as expected. To achieve better results, three and four digits after decimal point from P_{ij} are selected. In this way from P_{25}=0.0493, number 049 is selected and 5 inserted in bank2, 49 times. Simulation result is shown in Fig.3.b

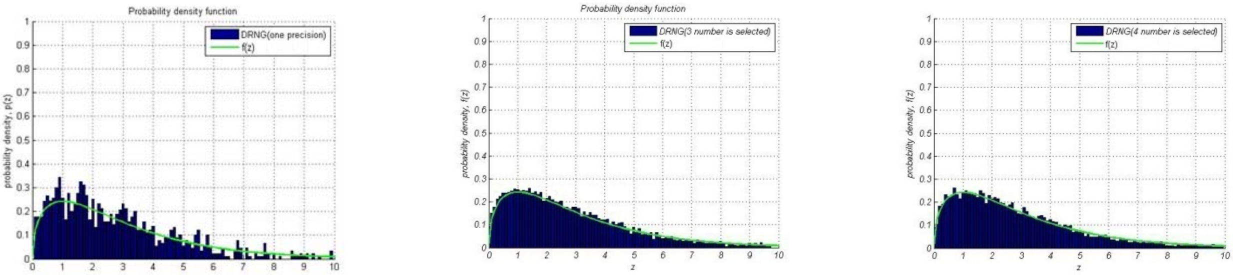

Fig.3. (a) DRNG -for digit banks, 2 number after decimal point is selected; (b) DRNG-for digit banks, 3 number after decimal point is selected; (c) DRNG-for digit banks, 4 number after decimal point is selected.

In this case result is good and reasonable and frequency diagram is almost similar to f(x).

To improve results, four digit after decimal point are selected (493 from 0.0493), which increases the similarity (Fig.3.c).

These three cases are compared from memory usage and MSE[1] of results point of view in table1. When three digits are selected, similarity increases, MSE decreases and memory usage increases. Increasing the costs is suitable because results get better. When four digits after decimal point are selected, similarity increases, MSE decreases and memory usage increases too and increasing size of memory causes to reduce speed of memory access. In this case memory increasing is so much and MSE decreasing is very low which it is not reasonable to increase cost to this extent. Therefore we suggest three digits after decimal point to be selected to memory usage be efficient and calculation speed be appropriate.

Finally we achieve these results: For modeling random variable to one digit, we use two digits after decimal point from P_i values and one bank is needed. For modeling to two digits, three digits after decimal point is selected from P_{ij} and 11 banks are needed and so on For modeling random variable to k digits, k+1 digits should be selected from $P_{i1i1..ik-1\ ij}$ values to place in the bank and 10^{k-1} +1 bank are needed.

Table1.Comparing some parameters of digit bank

Number of selected digits after the decimal point from P_{ij}	2	3	4
Used memory cell	50	1000	100000
MSE	0.064	0.024	0.025

6. Conclusion

In [5] we presented Data oriented Random Number Generator (DRNG) for chi square random variable. For making DRNG some digit banks are considered. The size of digit banks is one of the important challenges in DRNG. Whatever the size increases, the cost increases more, speed decreases and also gains more accuracy vice versa. In this paper we discuss about optimal size of digit banks so that efficiently uses from memory. An optimal size of digit banks is suggested. Simulation results show that presented suggestion for the size of digit banks is suitable from precision, speed and cost point of view.

7. References

[1] Habibizad-Navin A, Mirnia M. Alternative view of the shortest path problem. In proceeding of the 30th International Mathematical Conference, Iran; 1999, pp. 122-124.

[2] Habibizad-Navin A, Naghian Fesharaki M, Teshnelab M. Probability graph and some of its important structure. In proceeding of The 5th Seminar on probability and Stochastic Processes. Birjand, Iran, 2005, pp. 155-160.

[3] Habibizad-Navin A, Naghian Fesharaki M, Mirnia M, Teshnelab M. Mathematical discussions about data-Oriented modeling of uniform random variable. Journal of applied sciences 8(6), 2007, pp. 1113-1117.

[4] Olfatkhah R, Habibizad-Navin A, Mirnia M. A Novel Model of Normal Random Variable Based on Data-Oriented Theory. ICCEA. InternationalConference on Computer Engineering and Applications, 2010

[1] Mean Square Error

[5] Habibizad-Navin A, Alikhani S.E ,Mirnia M, Torabi S.Y, Chi Squre Random Variable Generator, Computational Intelligence and Communication Networks (CICN),India, 2010

[6] Habibizad-Navin A, Naghian Fesharaki M, Mirnia M ,Kargar M, Modeling of random variable with digital probability hyper digraph: data-oriented approach. In proceeding of World, Academy of Science, Engineering and Technology. Venice,Italy, 2007, pp. 365-367.http://www.waset.org/pwaset/v25/v2566.pdf