ISSN:1306-3111
e-Journal of New World Sciences Academy
2007, Volume: 2, Number: 2
Article Number: A0026

NATURAL AND APPLIED SCIENCES
ENGLISH (Abstract: TURKISH)
Received: January 2007
Accepted: April 2007

Erkan Tanyıldızı
Asaf Varol
Mehmet Güngör
University of Firat
etanyildizi@firat.edu.tr
Elazig-Türkiye

STATISTICAL ANALYSIS OF THE DATA IN SEQUENTIAL PROCESSES BY COMPUTER AIDED SOFTWARE

ABSTRACT

In this study, the abilities and usability of the Arena program in simulation are examined. The Arena simulation program was investigated and an application example in statistics was given. The results of simulations in our study from Arena and conventional methods were compared.

Keywords: Probability Distribution, Truncated Distribution, Arena Simulation Program

BİLGİSAYAR DESTEKLİ YAZILIM İLE ARDIŞIK SÜREÇLERDEKİ VERİLERİN İSTATİSTİK ANALİZİ

ÖZET

Bu çalışmada; bir sistemde yer alan ardışık süreçlerdeki verilerin, en uygun olasılık dağılımlarını belirleyen, en kısa ve en doğru şekilde istatistik olarak analizinde başvurulan Arena simülasyon programı incelenmiş ve bununla ilgili bir uygulama verilmiştir.

Anahtar Kelimeler: Olasılık Dağılımı, Kesilmiş Dağılım, Arena Simülasyon Programı

1. INTRODUCTION (GİRİŞ)

Simulation deals with a wide collection of the methods including models and techniques formed for a real system. It is a model of systems or processes including the relations defined among the system objects. In other words, it is the analogy of the running of a real system during a period. Modeling with simulation observes and defines the behavior of system, develops theories and hypothesis and uses this theory to predict the behavior in the future [1].

A Model is the representation of plan or an object. The aim of the model is to help better understanding and explaining or improving of the system. Simulation works with real or planned models. Simulation done in computer can both processes the information more rapidly and allows obtaining exact results as in the real systems.

2. RESEARCH SIGNIFICATION (ARAŞTIRMANIN ÖNEMİ)

The determination of most suitable probability distribution of sequential processes in a system takes very long time. These requires cumbersome mathematical analysis. Hence there is a need for a tool to overcome these difficulties. The Arena Software is developed for computer aided probability and statistical analysis which can easily be used for determining most suitable distributions.

3. THE ARENA SIMULATION PROGRAM (ARENA BENZETİM PROGRAMI)

There are a lot of simulation software's. Among these software's, new and preferable Arena simulation program is largely practical use and establishes the base of our study. In order to see the advantage of Arena we give and example which is manipulated both in classical methods and Arena. Now we give an application with classical method.

3.1. Application (Uygulama)

Let us consider the simulation program which simulates the frequency, the arrival time, the service time and leaving time of people coming to the ATM of bank for 30 minutes of period.

Table 1. The arrival times, service times and leaving times for people in an ATM
(Tablo 1. ATM'ye gelen kişilerin geliş, hizmet ve çıkış süreleri)

Customer Number	Arrival Time	Service Time	Leaving Time	Customer Number	Arrival Time	Service Time	Leaving Time
1	0.00	2.00	2.00	20	9.80	4.15	57.63
2	1.04	2.04	4.04	21	11.00	3.42	61.05
3	1.54	1.70	5.74	22	12.15	2.80	63.85
4	1.58	1.00	6.74	23	12.90	3.25	67.10
5	2.36	2.50	9.24	24	13.20	4.45	71.55
6	2.46	2.10	11.34	25	13.55	2.30	73.85
7	2.55	1.50	12.84	26	15.00	4.40	78.25
8	2.80	3.16	16.00	27	15.52	5.20	83.45
9	2.90	4.20	20.20	28	17.00	4.70	88.15
10	3.01	3.20	23.40	29	19.30	5.10	93.25
11	3.11	2.18	25.58	30	19.74	1.50	94.75
12	4.20	1.90	27.48	31	20.01	2.60	97.35
13	4.90	4.50	31.98	32	20.50	3.25	100.60
14	6.00	2.70	34.68	33	20.85	6.15	106.75
15	6.70	1.80	36.48	34	21.43	2.78	109.53
16	6.90	3.15	39.63	35	25.60	3.98	113.51
17	8.00	7.00	46.63	36	27.80	4.60	118.11
18	8.50	5.50	52.13	37	29.67	2.30	120.41
19	9.04	1.35	53.48				

37 people came to the ATM in 30 minutes. It is necessary to know sample mean and standard error to determine the distribution of arrival intervals of people [2].

Table 2. Calculation of sample mean and standard deviation
(Tablo 2. Ortalama ve standart hata hesabı)

X	f_i frequency	m_i	$m_i.f_i$	$f_i\ (m_i-\overline{X})^2$
0-5	13	2.5	32.5	912.56
5-10	7	7.5	52.5	79.89
10-15	5	12.5	62.5	13.14
15-20	6	17.5	105.0	263.07
20-25	3	22.5	67.5	405.18
25-30	3	27.5	82.5	828.83
	37	90.0	402.5	2502.70

k = Class or interval number,
m_i = Mean value of i *th* class,
f_i = The absolute frequency of i *th* class.

$$\overline{X}=\frac{\sum_{i=1}^{k} f_i.m_i}{\sum_{i=1}^{k} f_i}=10.9 \tag{1}$$

Sample standard deviation:

$$s=\sqrt{\frac{\sum_{i=1}^{k} f_i(m_i-\overline{X})^2}{\sum_{i=1}^{k} f_i-1}}=8.3 \tag{2}$$

We must test consistency of exponential distribution which is frequently used for the time of between and exposition process models in random occurrence. Since it is used in the truncated distribution of our application, we first consider the structure of the truncated distribution.

It is well know that there is the following relation between (*pdf*) *f* and (*cdf*) *F* of the continuous random variable *X*:

$$F(x)=\int_{-\infty}^{x} f(t)dt \tag{3}$$

Also, for *F* we have

$$0\le F(x)\le 1, \tag{4}$$

and for $0<p<1$,

$$F^{-1}(p)=inf\{x: F(x)\ge p\}=sup\{x: F(x)<p\} \tag{5}$$

is considered as the general inverse function of *F* [3].

3.2. Truncated Distribution (Kesilmiş Dağılım)

The domain of a continuous random variable with any distribution when it is bounded from both sides or any side called the truncated distribution of main distribution.

Truncated examples are consisting of main population which is formed by removing of the small and/or big observations [4].

Now let us define two concepts which we shall use later. Let $\alpha(F)$ and $w(F)$ denote the end points of *F* at the left and right respectively. Hence we have,

$$\alpha(F)=inf\{x: F(x)>0\}=F^{-1}(0) \tag{6}$$

and

$$w(F)=sup\{x: F(x)<1\}=F^{-1}(1) \tag{7}$$

where (6) is $-\infty$ or finite; (7) is $+\infty$ or finite [5]. ***pdf*** of ***f'*** which is truncated from the both sides can be expressed as,

$$f_{uv}(x) = kf(x) \tag{8}$$

where

$$\alpha(F_{uv}) = u \quad \text{and} \quad w(F_{uv}) = v\,. \tag{9}$$

Since $f_{uv}(x)$ is a ***pdf***,

$$k = \frac{1}{F(v) - F(u)}\,. \tag{10}$$

Hence ***pdf*** of ***f*** which is truncated from both sides is given by,

$$f_{uv}(x) = \frac{f(x)}{F(v) - F(u)}\,. \tag{11}$$

From (3), ***cdf*** of ***f*** which is truncated from both sides is given as,

$$F_{uv}(x) = \frac{F(x) - F(u)}{F(v) - F(u)} = 1 - \frac{F(v) - F(x)}{F(v) - F(u)}\,. \tag{12}$$

From (12), ***pdf*** and ***cdf*** of ***f*** which are truncated from the right side are,

$$f_v(x) = \frac{f(x)}{F(v)} \tag{13}$$

and

$$F_v(x) = \frac{F(x)}{F(v)} = 1 - \frac{F(v) - F(x)}{F(v)} \tag{14}$$

where

$$\alpha(F_v) = \alpha(F) \text{ and } w(F_v) = v\,. \tag{15}$$

And ***pdf*** and ***cdf*** of ***f*** which are truncated from the left side are,

$$f_u(x) = \frac{f(x)}{1 - F(u)} \tag{16}$$

and

$$F_u(x) = \frac{F(x) - F(u)}{1 - F(u)} = 1 - \frac{1 - F(x)}{1 - F(u)} \tag{17}$$

where

$$\alpha(F_u) = u \text{ and } w(F_u) = w(F)\,. \tag{18}$$

For example, from (11) and (12) the ***pdf*** and ***cdf*** of standard exponential distribution which is truncated from the both sides are given by the following expressions:

$$f_{uv}(x) = \frac{e^{-x}}{e^{-u} - e^{-v}} \tag{19}$$

and

$$F_{uv}(x) = \frac{e^{-u} - e^{-x}}{e^{-u} - e^{-v}}\,. \tag{20}$$

As a special case, if we set $u = 0$ ($\alpha(F) = 0$) in (19) and (20) then the ***pdf*** and the ***cdf*** of standard exponential distribution which is truncated from the right side is given with the following expression:

$$f_v(x) = \frac{e^{-x}}{1 - e^{-v}} \tag{21}$$

and

$$F_v(x) = \frac{1 - e^{-x}}{1 - e^{-v}} \tag{22}$$

As a special case, if we set $v = \infty$ ($w(F)=\infty$) in (19) and (20) then the *pdf* and the *cdf* of standard exponential distribution which is truncated from the left side is given with the following expression:

$$f_u(x) = \frac{e^{-x}}{e^{-u}} = e^{-(x-u)} \tag{23}$$

and

$$F_u(x) = \frac{e^{-u} - e^{-x}}{e^{-u}} = 1 - e^{-(x-u)}\,. \tag{24}$$

Equations (23) and (24) are *pdf* and *cdf* of exponential distribution with the origin u.

It is well known that, mean and standard deviations of distributions in standard form are different from that of distributions which are truncated from any side.
The distributions by truncation way, brought a convenient form to observed data. For instance, the *pdf* and *cdf* of exponential distribution which is truncated from the left side are an admissible model of the *pdf* and *cdf* of exponential distribution with the origin u as (23) and (24) [6].

In the light of (3)-(9) as (13)-(15) for our application in the study, we give the *cdf* of the distribution which is truncated from right as follows,

$$F_v(x) = \frac{1 - e^{-x/\beta}}{1 - e^{-v/\beta}} \tag{25}$$

where

$$f(x) = \begin{cases} \dfrac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0 & , \text{otherwise.} \end{cases} \tag{26}$$

After determining the distribution function we must test its consistency. In order to make this we can use Chi-Square and Kolmogorov-Smirnov test.

3.3. Chi-Square Test (Ki-Kare Testi)

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{27}$$

Degrees of freedom;

$$v = k - 1 - m \tag{28}$$

o_i = Observed frequency,
e_i = Theoretical frequency of each groups,
k = Number of interval,
v = Degrees of freedom,
m= For calculation of theoretical frequency, sample data parameters or experiment number.

From (28) degrees of freedom in application is v=4-1-1=2.

Calculation of Chi-Square test is given in Table 3.

Table 3. Calculation of Chi-Square Test
(Tablo 3. Ki-Kare testinin hesaplanması)

$F_V(x)$	e	o	$\frac{(o-e)^2}{e}$
0.394380	14.5572	13	0.1666
0.248464	9.1932	7	0.5232
0.156910	5.8057	5	0.1118
0.099091	7.4440	12	2.7885
0.062578			
0.039519			
1	37	37	3.5901

Since χ^2=3.59 is smaller than the value in table of χ^2, the data is convenient for exponential distribution.

3.4. Kolmogorov-Smirnov Test (Kolmogorov-Smirnov Testi)

$D = max|F_o - F_n|$ (29)

F_o, Theoretical *cdf*,
F_n, Observed *cdf* ,
D, Absolute deviation of difference of two distributions.
Calculation of Chi-Square test is given in Table 4.

Table 4. Calculation of Kolmogorov-Smirnov Test
(Tablo 4. Kolmogorov-Smirnov Testinin Hesaplanması)

Observed frequency	Observed probability (F_n)	Theoretic probability (F_o)	Absolute deviation (D)
13	0.351351	0.393438	0.017130
7	0.189189	0.248464	0.043510
5	0.135135	0.156910	0.011820
6	0.162162	0.099091	0.069357
3	0.081081	0.062578	0.022473
3	0.081081	0.039519	0.044069

From (29) the biggest absolute deviation is found D= 0.069.
The value in the table of Kolmogorov-Smirnov is,

$$D = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{37}} = 0.22 \,. \quad (30)$$

Since 0.069 is smaller than (30), the data is convenient for exponential distribution.

Now, let us make same application by Arena. This program performs the examination of data and statistical analysis methods [1, 2, and 7]. Results of analysis are summarized below.

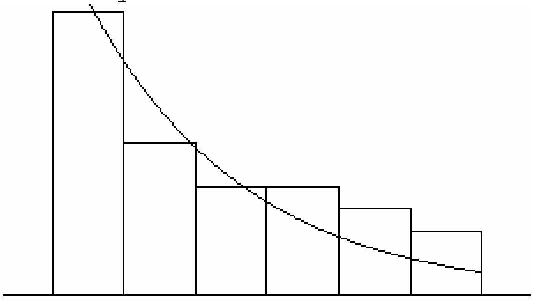

Figure 1. The exponential distribution of data (-0.001+EXPO (10.9))
(Şekil 1. Verilerin üstel dağılımı)

Distribution Summary:

Distribution	:Exponential
Expression	:-0.001 + EXPO(10.9)
Square Error	:0.008511

Data Summary:

Number of Data Points	= 37
Min Data Value	= 0
Max Data Value	= 29.7
Sample Mean	= 10.9
Sample Std Dev	= 8.32

Chi Square Test:

Number of intervals	= 4
Degrees of freedom	= 2
Test Statistic	= 6.05
Corresponding p-value	= 0.0489

Kolmogorov-Smirnov Test:

Test Statistic	= 0.105
Corresponding p-value	> 0.15

Summary of the convenient distributions:

Function	Sq Error
Beta	0.000614
Gamma	0.00822
Erlang	0.00851
Exponential	0.00851
Weibull	0.00981
Lognormal	0.0261
Triangular	0.0279
Uniform	0.0474
Normal	0.0579

The same idea is applied to establish the convenient distribution related to service time. Hence, distribution of service time is TRIANGULAR (0.999, 1.76, 7).

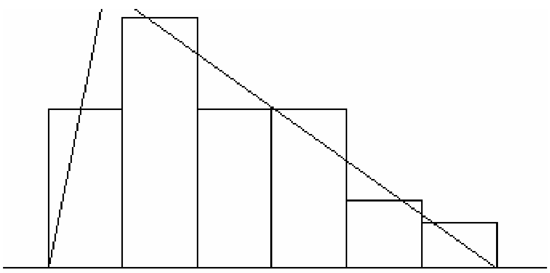

Figure 2. The triangular distribution of data (TRIA (0.999, 1.76,7))
(Şekil 2. Verilerin üçgen dağılımı)

4. DISCUSSION AND RESULT (TARTIŞMA VE SONUÇ)

Among the softwares the Arena simulation program is the one which is used most widely. Statistical analysis, queue and usage rates, process time and transmitting intervals can be arranged according to the requirements of the work. The effects of the alternations done in the system or the system itself can be observed by animation after modeling. In addition, the detailed results of the simulation can be seen in the program.

REFERENCES (KAYNAKLAR)

1. Kelton, W.D., Sadowski, R.P., and Sadowski, D.A., (1998). Simulation with arena, McGraw-Hill, Inc., New York.
2. Tanyıldızı, E., (2002). Bilgisayar destekli yazılım için uygun bir yazılım geliştirilmesi, Yüksek Lisans Tezi, F.Ü. Fen Bilimleri Enstitüsü, Elazığ.
3. Balakrishnan, N., and Cohen, A.C., (1991). Order statistics and inference, Academic Press, Inc., London.
4. Bayazıt, M. and Oğuz, B., (1991). Probability and statistic for engineers, Birsen Yayınevi Ltd. Şti., İstanbul.
5. Galambos, J., (1987). The asymptotic theory of extreme order statistics, Robert E.Krieger Publishing.
6. Balakrishnan, N., and Basu, A.P., (1995). The exponential distribution, Gordon and Breach Publishers, The Netherlands.
7. Varol, A., and Tanyıldızı, E., (2002). Benzetim, Otomasyon, August 2002, pp:134-139.