SIMULATION OF DISCRETE STOCHASTIC PROCESSES

S. Georgievska, M.Kon-Popovska

Institute of Informatics, Faculty of Natural Sciences and Mathematics
Sts. Cyril and Methodius University
Arhimedova bb, P.O.Box 162, Skopje, Macedonia
{sonja, margita}@ii.edu.mk

Abstract: Simulation is the imitation of the operation of a real-world process or system over time. As systems become more and more complicated, computer simulation tends to be the most appropriate technique for predicting the behavior of a system or finding its optimal design.

Analysis of discrete stochastic processes is one field where this method is applied. As these processes include a lot of random variables, a great attention must be paid to the analysis of the output of a simulation program. Here we take a look at some of the techniques developed for this purpose.

Keywords: distribution of a random variable, discrete stochastic processes, queuing systems, terminating systems, nonterminating systems.

1. Introduction

Simulation is the process of designing a mathematical or logical model of a real system and then conducting computer-based experiments with the model to describe, explain, and predict the behavior of the real system.

In this text the simulation of discrete stochastic processes will be a basic concern. A process is discrete if changes in state occur only at discrete points in time, and it is stochastic if any random variables are present. For example, a queuing system, where clients wait in queues to be served, is a discrete stochastic process. Changes occur when a client enters or leaves the system. The distribution of the inter-arrival time of the entrance of the clients is usually exponential, and the distribution of the serving time might be exponential, uniform, normal, etc. We might be interested in the parameters of the system, such as: the average time clients wait in queue, the fraction of the customers that wait, the average length of the queue, the probability that a client will be served (in case of finite queue), average idle time for a server etc. There are analytic models for most of the simple systems, but very often simulation is a necessary technique.

2. Output analysis

Discrete-event simulation models are different from most other types of models. Because a discrete-event simulation model brings together the confluence of many random variables, the output of the model is, itself, a random variable. The output of a simulation model can easily be misinterpreted, resulting in false conclusions about the system it represents.

$$\frac{dP_s(t)}{dt} = 0 \tag{1}$$

Discrete-event systems (in fact all dynamic systems) can be categorized as being either terminating or nonterminating. A system is classified as terminating if the events that drive the system cease occurring at some point in time, while in nonterminating systems the discrete events reoccur indefinitely. When analyzing the output of a simulation model, it is necessary to differentiate between data gathered when the system was in a transient phase and when it was in steady state. The system is in steady state relative to state variable s when where s(t) is the state of the system at time t and $P_s(t)$ is the probability that the system is in state s at time t. Otherwise, the system has not achieved steady state and is said to exhibit transient behavior.

3. Output analysis for terminating systems

For the sample (x_i, i=1,2,...,n)

$$\overline{X} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{2}$$

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{X})^2}{n-1}} \tag{3}$$

It can be easily shown that if $E(x_i) = m$ for all i then $E(\overline{X}X) = m$. In order that $E(s^2) = \sigma^2$ where $\sigma^2 = D(x_i)$ (the variance of x_i) it is necessary that each x_i have a common expected value and all x_i are independent.

For terminating simulations, the most commonly used method of insuring that the observations x_i are independent and have a common expected value is replication. During the course of each simulation, observations are made at designated points in time or upon the occurrence of designated events.

For a simulation replicated R times, with K intermediate observations in each simulation let x_{ij} = be the j-th observation of the i-th replication, where i=1,2,...R and j=1,2,...K and let y_i= be some overall performance measure during i-th replication. Then

$$\overline{X}_j = \sum_{i=1}^{R} \frac{x_{ij}}{R} \qquad j = 1,2,...K \tag{4}$$

$$\overline{Y} = \sum_{i=1}^{R} \frac{y_i}{R} \tag{5}$$

$$s_j^{\,2} = \sum_{i=1}^{R} \frac{(x_{ij} - \overline{X}_j)^2}{R-1} \tag{6}$$

$$s_y^{\,2} = \sum_{i=1}^{R} \frac{(y_i - \overline{Y})^2}{R-1} \tag{7}$$

For example, in the i-th simulation of a queuing system, we would measure the waiting time of an arriving customer at each of K different time points, x_{ij}, and Y_i, the overall waiting time of all customers arriving during the simulation. We now have independent and unbiased estimates of the expected value and variance of the system's performance at K different points in time, as well as an unbiased estimate of the mean and variance of the overall performance measure. Once the mean and point estimators have been established using equations (4) through (7), we can set approximate confidence intervals for $E(x_{ij})$ and $E(y_i)$ using

$$P(\mu_j = \overline{X}_j \pm t_{\alpha/2,(R-1)} \frac{s_j}{\sqrt{R}}) = 1 - \alpha \tag{8}$$

$$P(\mu_Y = \overline{Y} \pm t_{\alpha/2,(R-1)} \frac{s_y}{\sqrt{R}}) = 1 - \alpha \tag{9}$$

where $t_{\alpha/2,(R-1)}$ is the number satisfying that for a random variable T with t-distribution and (R-1) degrees of freedom it holds

$P(-t_{\alpha/2,(R-1)} < T < t_{\alpha/2,(R-1)}) = 1-\alpha.$

The half width of the confidence interval is

$$I = t_{\alpha/2,(R-1)} \frac{s}{\sqrt{R}} \tag{10}$$

and indicated how accurately we have estimated the performance measure in question.

Illustrative problem 1. Let us consider a simulation of the classic M|M|1|∝ queuing system (one server, infinite queue, exponential distribution of the serving time and the inter-arriving time). In this simulation we assume that the arrival rate is 60 arrivals per hour and the mean service time is 48 seconds, resulting in a traffic intensity of 0.8.

Table 1 shows x_{ij}, the sample waiting times for the customers, arriving immediately after minute 10, 20, ..., 80, the average waiting times for the customers arriving at these points of time for 25 replications and the appropriate standard deviations of the sample (it is the output of the appropriate program).

In practice the system should be simulated hundreds of times before drawing any firm conclusions about the mean and the variance of the waiting times, or even the number of simulations needed to set confidence intervals with the intended half width. As a first guess of the number of simulations required to establish reasonable confidence intervals on mean waiting times, let us estimate the number of observations required to set a 95 percent confidence interval, with a half width of 1/6 minutes, for a customer arriving immediately after the 10-th minute. Letting the sample variance act as an estimate for the true variance of waiting times, then from equation (10) the half width of the confidence interval is $I = 1.96s_1/\sqrt{R}$ and solving for R we have $R=(1.96s/I)^2=830$.

Upon completing these 830 simulations we would most likely have a more accurate estimate of the true standard deviation, leading to more replications of the simulation. In fact, since the true standard deviation of the waiting time is approximately 130, to set the 95 confidence interval on the mean time a customer waits, at a half width of 1/6 minutes, we would in actuality need to simulate the system approximately 600 times. It should be pointed out that we generally do not have this kind of information on the variance of the property being estimated, since in most simulation projects good analytic models are not available.

4. Output analysis for nonterminating systems

When analyzing the output of simulation models of nonterminating systems, we must deal with several problems:

Initial Condition Bias. The data collected during the early part of the simulation may be biased by the initial state of the system. The behavior of the system during this early phase of the simulation may be irrelevant to the questions we expect the model to answer.

Simulation of M|M|1 system

Tarrivals.: 60 Tserving.: 48 Speed: 0

Rpl	1	2	3	4	5	6	7	8
0	1	0	16	152	0	0	0	27
1	166	502	498	335	308	163	0	91
2	0	0	0	0	0	0	0	0
3	227	545	370	358	398	156	82	0
4	250	0	81	0	0	0	116	124
5	36	194	169	452	244	37	47	0
6	382	480	324	283	69	116	0	0
7	173	0	0	132	0	0	46	51
8	129	186	169	67	0	0	327	111
9	0	401	184	304	446	439	220	90
10	35	0	0	152	47	10	0	0
11	29	0	135	91	0	58	0	27
12	583	0	449	67	107	216	359	0
13	181	0	39	0	0	102	0	33
14	0	0	0	0	17	0	195	101
15	0	84	10	0	0	0	0	0
16	0	0	0	62	59	12	0	0
17	0	0	0	82	78	37	299	0
18	8	0	187	0	0	0	0	0
19	233	0	256	504	454	188	55	0
20	13	18	156	113	441	473	536	0
21	0	307	145	170	81	0	376	111
22	0	0	345	316	3	375	314	0
23	0	0	120	209	224	18	431	0
24	0	192	0	0	42	97	31	0
Average:	97.84	116.36	146.12	153.96	120.72	99.88	137.36	30.64
St. dev:	147.0	183.0	152.0	152.0	162.0	141.0	168.0	44.0

Start Stop Pause Continue

Table 1. Simulation of M/M/1/∝ system

Covariance Between Samples. Groups or sets of data gathered during the simulation are generally not independent of one another. Therefore the variance estimates will be biased.

Run length. Although the system itself may be nonterminating, the simulation of the system must eventually? be terminated. If we terminate the system too early, we may not have a representative simulation.

We will present a method for analyzing the output of a simulation model of a nonterminating system. We will demonstrate the application of this method, using the output of a simulation model of a blood bank inventory system.

Illustrative problem 2. A blood bank maintains an inventory of whole blood of a specific type. The blood is collected from voluntary donors, who come to the blood bank randomly, at the mean rate of two donors per day. Each donor contributes exactly one unit of blood (1 pint). The mean number of blood recipients is one per day and is also randomly distributed. Each recipient will request at least one unit of blood. The number of units requested after the first unit, is Poisson distributed with a mean of 1.0. Blood has a limited storage life and any blood that is older than 21 days cannot be used. The blood bank does not place an upper limit on the number of units kept in inventory, but lets the supply rise and fall by the random occurrence of donor and recipients. If the supply falls to zero, blood is obtained from a second source but at a considerable expense. The simulation model would estimate the average blood inventory, the mean number of units that become outdated each year, and the mean number of blood units that must be obtained from the second source each year.

5. The method of replication

For nonterminating simulations we must avoid (or at least minimize) the effect of the initial conditions on the output of the model. The most common way of doing it is to discard the observations gathered during the early phase of the simulation model and use only data gathered when the system has reached a steady-state condition. Here we must decide when steady-state conditions begin.

Illustrative problem 3. We run the blood bank simulation model for 1000 days (approximately three years) recording the average inventory every 100 days. From 25 replications it appears that after 500 days the average inventory of blood becomes approximately constant, that is, the system should be simulated for about 500 days before collecting statistics to estimate the steady-state properties of the blood bank. Based on these exploratory simulations, the simulation is now replicated 20 times, (run length of 2000 days) discarding statistics gathered during the first 500 days, and then collecting statistics during the next 1500 days.

Table 2 shows the results of these replications.

Using the results of the replications we can set a 95 percent confidence interval on the average number of blood units on hand as well as the expected number of units short and outdated each year. For illustration purposes we will set a 95 percent interval on the average number of units short per year. Based on a 1500-day simulation an unbiased estimate of the mean number of units short per year is Average short per year == (365/1500) Average short per 1500 days, and since Variance (Shortage/Year) = $(365/1500)^2$ Variance (Shortage/1500 days) we will estimate the variance of the annual number of units short as $s^2_{365} = (365/1500)^2 s^2_{1500}$.

If we assume that the distribution of the number of units short per year is normal, we can set a 95 percent confidence interval on units short per year using

$$\overline{X}_{shortage/year} = 31$$

$$S_{shortage/year} = 17.5$$

$$\text{or } P[E(\text{annual units short}) = 31 \pm 2.093\sqrt{\frac{17.5^2}{20}}] = 0.95.$$

$$P[E(\text{annual units short}) = \overline{X} \pm t_{\alpha/2,(n-1)}\frac{s}{\sqrt{n}}] = 0.95$$

In addition, we will mention three more techniques for analyzing the output of nonterminating systems: the method of batch means, autocorrelation methods and the regenerative method.

Tdonors: 12 Trecipients: 24 NPoisson: 1 Days:

Rpl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
0	31	12	43	26	36	24	31	37	9	31	39	13	36	21	31	28
1	37	22	38	23	26	24	28	12	2	11	14	35	25	39	26	24
2	16	28	15	33	15	35	18	20	31	0	9	1	0	43	51	21
3	17	17	36	43	28	26	17	20	32	43	0	33	43	8	25	25
4	38	28	14	1	17	36	29	7	9	6	1	4	32	9	23	16
5	31	37	43	34	28	27	25	10	26	20	25	4	7	30	15	24
6	38	8	40	33	25	37	44	15	17	13	11	41	18	23	21	25
7	0	1	11	8	25	12	5	37	21	26	40	1	5	2	2	13
8	39	20	9	7	3	38	34	34	7	28	36	17	6	26	3	20
9	7	5	20	30	33	30	3	29	39	27	29	49	25	36	38	26
10	2	37	37	34	51	48	19	9	5	49	30	29	36	1	25	27
11	47	22	33	8	43	55	19	23	4	8	15	20	8	5	7	21
12	16	37	50	35	12	7	30	30	1	3	7	15	7	13	8	18
13	36	41	18	10	45	42	24	35	46	38	36	16	20	20	13	29
14	35	27	59	8	0	6	22	0	36	35	26	24	50	18	39	25
15	3	6	5	6	23	0	1	6	2	54	37	17	37	24	15	15
16	16	34	34	23	14	6	13	45	8	2	27	31	39	44	24	24
17	23	4	17	36	15	27	5	16	35	31	32	17	28	29	41	23
18	0	6	7	7	19	42	16	25	20	36	4	36	32	12	36	19
19	3	4	23	49	31	1	7	33	9	10	39	17	1	7	28	17

Av.units outdated:32.24167
St.dev.:15.294383
Av.units short:31.000668
St.dev.:17.537138

Table 2. Simulation of the blood bank system

6. Conclusion

The output of a simulation model requires careful analysis. Classical statistical techniques seldom directly apply. The observations collected during the simulation are often not independent nor time invariant. When analyzing the output of simulation models we must distinguish between terminating and nonterminating simulations, and between simulations that reach a steady-state condition versus simulation in which a steady-state condition cannot be achieved.

For terminating simulations, the model can be run a number of times, with each run being an independent observation. For nonterminating simulation, we must let the model warm up and pass through any transient conditions. We then collect the data in a manner that will allow us to establish point and interval estimates on the measures of performance needed to answer our questions concerning the system being simulated.

7. References

1. Stewart V. Hoover, Ronald F. Perry (1990), *Simulation: a problem solving approach,* Addison-Wesley Publishing Company, Inc.