Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

SENSITIVITY OF OUTPUT PERFORMANCE MEASURES TO INPUT DISTRIBUTIONS IN QUEUEING SIMULATION MODELING

Donald Gross

Department of Operations Research
and Engineering
George Mason University
Fairfax, VA 22030, U.S.A.

Man Juttijudata

Investment and Securities Public Company Limited
138 Silom Road
Bangkok 10500 THAILAND

ABSTRACT

With modern simulation packages, the modeler can choose almost any standard statistical distribution for generating input random variables. The question arises as to how sensitive are output performance measures to the particular type of distributions selected for modeling the input. We investigate the sensitivity of two output performance measures (average queue wait and the 95 percentile value of the queue wait distribution) for a G/G/1 queue, simulated using GPSS/H.

1 INTRODUCTION

The major question we attempt to answer in this study is just how important is it to expend resources (time and money) in determining the **input** distributions (interarrival and service times, for example) for a queueing simulation model? That is, how **sensitive** are the **output** measures of performance such as queue sizes and waiting times to the particular choice of input distributions? Note that we are not concerned here with how well the input distributions chosen fit, in a statistical sense, the data we may be using to determine distribution choice, but how sensitive the output performance measures are to the distribution choice. In other words, we want to find out if it suffices to choose distributions with roughly the same density shape and first two moments matched to the data at hand. Or is it necessary to collect large amounts of data and spend great effort in precisely choosing input distributions, and, if so, which procedures are best?

There is some evidence that, **perhaps**, matching only the first two moments might suffice. For M/G/1, the Pollaczek-Khintchine (P-K) formula tells us we need only the mean ($1/\lambda$) of the interarrival time distribution (this is also the standard deviation) and the mean ($1/\mu$) and standard deviation (σ_s) of the service time distribution to calculate expected value measures of performance (MOP). The expected queue size, for example, is given in Equation (1) as

$$L = \rho + \frac{\rho^2 + \lambda^2 \sigma_s^2}{2(1-\rho)} \quad (\rho = \lambda/\mu).$$

Also, the Kingman/Marshall upper bound for G/G/1 queues

$$Wq = \frac{\lambda(\sigma_A^2 + \sigma_B^2)}{2(1-\rho)},$$

is given by Equation (2),
which is also the heavy traffic approximation, and depends only on the first two moments of the interarrival and service time distributions $(1/\lambda, \sigma_A^2, 1/\mu, \sigma_B^2)$.

There is a great deal of material in the statistical literature concerning fitting known statistical distributions to data. However, we are not particularly interested in how well a distribution fits the data with respect to classical statistical measures, but how sensitive the MOPs are to distribution choice.

In this study, we conduct a series of experiments using a G/G/1 simulation model written in GPSS/H. The two MOPs we consider are the mean queue wait, W_q, and the 95th percentile of the queue wait distribution which we denote as $W_q(.95)$; that is, letting T_q represent the random variable, wait in queue, then $\Pr\{T_q > W_q(.95)\} = .05$. Twenty replications of 20,000 customers with a warm-up period of 2,000 were used in all cases considered and confidence interval estimates were produced for the MOPs. The simulation model was validated for cases where known theoretical steady-state results existed (e.g., M/M/1 for both MOPs and M/G/1 for W_q), and simulation error for mean values was generally within 5%, with the confidence bound range, for most cases, being ± 5%, with a worst case of ±13%.

2 FIRST EXPERIMENT

This experiment was performed to determine whether the first two moments of interarrival and service time distributions are sufficient in estimating the MOPs. The Γ/Γ/1 (Γ = Gamma distribution) was selected as a base case model and compared to LN/LN/1 (LN = lognormal), W/W/1 (W = Weibull) and P5/P5/1 (P5 = Pearson, type 5) models. Traffic intensities (ρ) of 0.2, 0.5, and 0.8 were considered along with coefficients of variation of interarrival and service time distributions (CV_{at} and CV_{st} respectively) of 0.5, 1.0 and 2.0, making a total of 108 cases in all (each with 20 replications of 20,000 customers and a 2,000 customer warm-up period).

Table 1 presents some sample results showing that there can be significant sensitivity in the output MOPs, and that, in general, matching only the first two moments is not sufficient. We see that for higher ρ values (.8), the percentage differences are, for the most part, modest. But for ρ = .5, percentage differences are sizable, and for ρ = .2, some turned out to be very large, although the absolute values of the MOPs were near zero, which partially accounts for large percentage differences. Even so, there is ample evidence that it is not always sufficient to match only the first two moments. Twenty-eight of the 108 cases run had percent differences for W_q greater than 20%, twelve of these being cases with ρ values of .5 or .8. These were either P5 models, or W and LN models with CV_{at} =2. Further, some of the P5 cases of the original 108 cases had to be eliminated from consideration due to slow convergence of the second moment in the generated variates. One curious result was that in many cases, the percentage differences seemed to be somewhat less for $W_q(.95)$ than for W_q - there were only 21 cases where percentage differences for $W_q(.95)$ were greater than 20%, nine of these for ρ values of .5 or .8, with conditions similar to those 12 for W_q.

Table 1: Some % Difference Examples

Case	CV_{at}	CV_{st}	Model	$\%\Delta W_q$	$\%\Delta W_q(.95)$
ρ = .8	0.5	2	LN	2.05	0.52
			WB	3.55	1.28
			P5	slow convrg of 2nd mom	
ρ = .8	0.5	1	LN	0.07	6.87
			WB	3.88	2.90
			P5	3.35	17.14
ρ = .5	2	.5	LN	54.62	50.41
			WB	21.18	18.69
			P5	slow convrg of 2nd mom	

Max %Δs ≈ 100 for some ρ = .2 cases (MOPs ≈ 0)

For thirty six of the 108 cases run, analytical solutions exist for W_q, i.e., those cases for which CV_{at} = 1, the Γ/ Γ /1 and W/W/1 reduce to M/G/1 models. Further, when CV_{st} is also 1, the M/G/1 models further reduce to M/M/1 models and analytical results are available for both W_q and $W_q(.95)$. Tables 2 and 3 show these results. Table 2 shows the M/M/1 results and we see both the simulated Γ (=GM) and W (=WB) models are very close to the theoretical values, again validating the simulation. We can also see the "heavy traffic" effect in that the large percent differences of the LN and PT models diminish as ρ increases. But there **is** sizable differences in these models, again indicating that matching the first two moments only is, in general, not sufficient. Table 3 shows similar effects for the M/G/1 cases.

3 SECOND EXPERIMENT

This experiment investigated selecting two-parameter distributions by matching the first and third quartiles of the base case Γ/ Γ /1 to W/W/1, LN/LN/1 and P5/P5/1. The same 108 cases were attempted, although many of these resulted in infeasible fits, or feasible fits with $\rho \geq 1$. For the feasible cases, the percent deviations from the base case were much larger, and we concluded from this experiment that matching the first and third quartile is considerably poorer than matching the first two moments.

4 THIRD EXPERIMENT

In this experiment, we compared Maximum Likelihood estimation (MLE) to MOM for fitting two-parameter distributions. The experimental design had ρ =.5, CV_{at} = 1, CV_{st} = .5,1,2. Data were generated for sample sizes of 20, 50, 100, 200, 500, 1000 for the above three cases for Γ / Γ /1 and LN/LN/1 models. Γ and LN parameters were estimated for each of the sample sizes by both MOM and MLE. The empirical distribution, MOM and MLE cases were simulated for the various sample sizes and compared to the actual Γ / Γ /1 and LN/LN/1 base case simulations. The conclusions drawn from this experiment were (i) as sample size increases, the empirical model converges to actual, (ii) MOM appears somewhat "better" than MLE for small sample sizes, but both yield similar results for large sample sizes, and (iii) percent differences from actual are still quite large in many cases, especially for small sample sizes.

The general conclusion from experiments 1, 2 and 3 is that two parameter distributions are not sufficient in general. It appears that, although the first two moments are important, that it is necessary, in many cases, to capture higher-order moments. To attempt to determine just how many may be necessary, we did a final experiment using a discrete point distribution. While there are families of distributions that have more than two parameters (e.g., Johnson translation distributions, phase-type distributions, Coxian distributions, generalized hyper exponential distributions), fitting can be a formidable task (see, for example, Johnson and Taffe, 1991, Johnson, 1993, and Harris and Marchal, 1997). We chose discrete point distributions, which were very easy to fit by MOM and allowed us to see what happened as we increased the number of moments matched.

Table 2: M/M/1 Theoretical Cases

	M/M/1 (CV[AT]=1, CV[ST]=1)				
	Model	Wq	%Diff	Wq(.95)	%Diff
Rho =. 2	Theo.	0.05	0	0.3466	0
	Model	Wq	%Diff	Wq(.95)	%Diff
Rho = .2	GM/GM/1	0.0502	0.4	0.3467	0.028852
	WB/WB/1	0.0501	0.2	0.3486	0.577034
	LN/LN1	0.0258	-48.4	0.1527	-55.94345
	PT/PT/1	0.0189	-62.2	0.0001	-99.97115
Rho = .5	Theo.	0.5	0	2.3026	0
	GM/GM/1	0.4959	-0.82	2.2869	-0.681838
	WB/WB/1	0.5042	0.84	2.315	0.5385217
	LN/LN1	0.3934	-21.32	2.0149	-12.49457
	PT/PT/1	0.3255	-34.9	1.6339	-29.04108
Rho = .8	Theo.	3.2	0	11.0804	0
	GM/GM/1	3.1328	-2.1	10.9238	-1.413306
	WB/WB/1	3.2602	1.88125	11.1705	0.8131475
	LN/LN1	2.9477	-7.884375	10.7616	-2.877152
	PT/PT/1	2.621	-18.09375	10.332	-6.754269

Table 3: M/G/1 Theoretical Cases

		M/G/1: (CV[AT]=1, CV[ST]=.5)		(CV[AT=1, CV[ST]=2)	
	Model	Wq	%Diff	Wq	%Diff
Rho =. 2	Theo.	0.0313	0	0.125	0
	GM/GM/1	0.0313	0	0.1271	1.68
	WB/WB/1	0.0314	0.319488 8	0.1243	-0.56
	LN/LN1	0.0089	-71.5655	0.0931	-25.52
	PT/PT/1	0.0025	-92.0127 8	0.0694	-44.48
Rho = .5	Theo.	0.3125	0	1.25	0
	GM/GM/1	0.3141	0.512	1.2885	3.08
	WB/WB/1	0.3137	0.384	1.2607	0.856
	LN/LN1	0.1928	-38.304	1.0837	-13.304
	PT/PT/1	0.1135	-63.68	0.8318	-33.456
Rho = .8	Theo.	2	0	8	0
	GM/GM/1	1.9959	-0.205	8.4837	6.04625
	WB/WB/1	2.0337	1.685	8.2794	3.4925
	LN/LN1	1.6943	-15.285	7.4834	-6.4575
	PT/PT/1	1.2943	-35.285	5.9483	-25.64625

5 FOURTH EXPERIMENT

The experimental design for this final experiment had $\rho = .5$, $CV_{at} = 1$, and $CV_{st} = .5, 1, 2$. Data were generated for sample sizes (n) of 20, 50, 100, 200, 500, 1000 for above three cases for $\Gamma / \Gamma / 1$ and LN/LN/1 models (same as the third experiment). For each sample size and each G/G/1 model, a discrete k-point distribution was fit by nonlinear programming, matching two up to five moments respectively. The nonlinear programming problem to be solved is given by (3) below.

$$\min \; g(p_1, \ldots, p_k, t_1, \ldots, t_k) = [(\sum_{j=1}^{kp_j} t_j^{n+1} - m_{n+1}]^2$$

$$s.t. \quad m_i = \sum_{j=1}^{k} p_j t_j^i \qquad (i = 1, 2, \ldots, n)$$

$$\sum_{j=1}^{k} p_j = 1$$

$$p_j \geq 0 \qquad (j = 1, 2, \ldots, k)$$

The actual, empirical and the four discrete k-point fits were simulated, and W_q and $W_q(.95)$ compared. Figure 1 shows, for the $CV_{st} = 1$ case, the percent differences in W_q and W_q (.95) from the gamma and lognormal simulated cases with simulations using the empirical distribution, and discrete distributions with two, three and four moments matched for sample data generated from gamma and lognormal distributions for sample sizes of 50, 200 and 1000 observations. Results are similar for the other two cases ($CV_{st} = .5$ and 2). It is interesting to note that using the discrete distribution with a five moment match is almost as good as using the entire empirical distribution and for large n, they both are fairly close to the actual simulated MOPs.

6 CONCLUSIONS

Output measures of performance can, indeed, be sensitive to the particular "shape" of the input distribution. Using standard two-parameter distributions for which only the first two moments are captured in many cases is not sufficient, unless the system is in heavy traffic (probably $\rho > .9$). Results here indicate that it may be necessary to capture at least five moments, even though the lower order moments dominate in importance. Sample sizes when determining input distributions should be at least 200 and preferably closer to 1000 which should provide good results when using empirical distributions from which to generate inputs. Using a discrete k-point distribution for which the first five moments are matched to the data is almost as good. These are easy to determine by non-linear programming and easily generate sample data for simulation.

This research shows that attention must be paid to proper input modeling, and that it is not a trivial task. Some general papers in this area include Fox (1981), Kelton (1984), Cheng (1993), Cheng, et al (1996), Leemis (1995, 1996) and Nelson et al (1995). Three rather "non-classical" classes of distributions which appear to have potential in input modeling are the Johnson translation distributions (see Storer et al, 1988), Phase-type distributions (see Johnson and Taaffe, 1991), and Bezier distributions (see Wagner and Wilson, 1996).

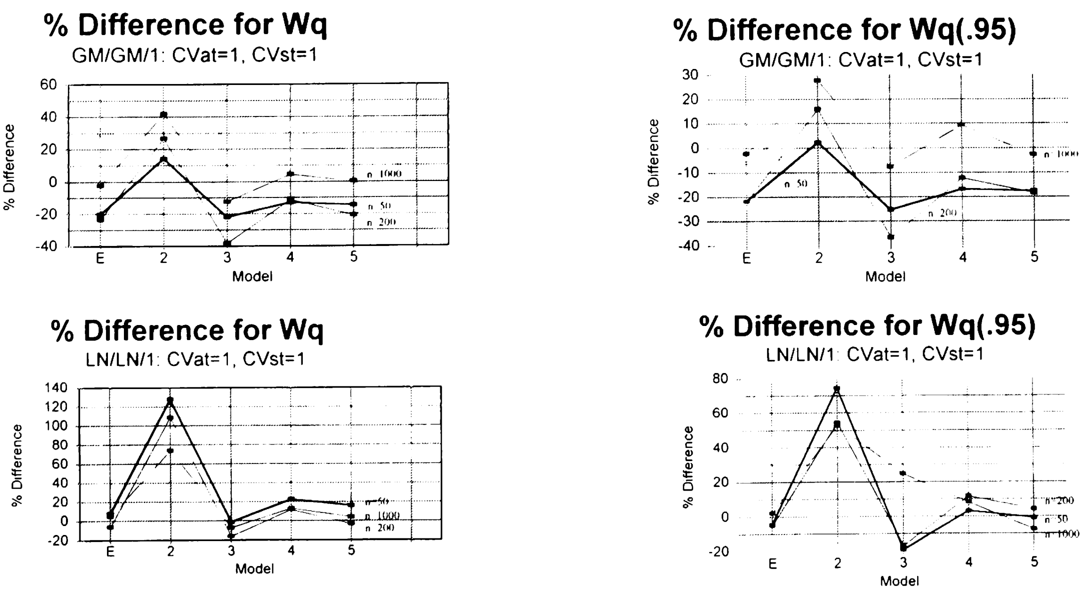

Figure 1: $CV_{at} = 1$, $CV_{st} = 1$.

APPENDIX - SOME REAL DATA

In a master level discrete-event simulation course at The George Washington University, a student team did a project involving a local bank. A G/G/c queue was simulated. The interarrival time and teller service time data were run through the UNIFIT II data fitting package and Table A.1 shows the moment comparison given by UNIFIT II for the actual data and the best five theoretical distributions it suggests. For c = 7 (giving a = .63), two sets of simulations were run. Using the empirical distribution (EM) for service times, G/EM/7 runs were made for G = EM , the first five theoretical distributions given by UNIFIT and the four discrete k-point distributions matching two, three, four and five moments to the data. A similar set of runs were made using EM for interarrival times (EM/G/7) and G= EM, the first five distributions given by UNIFIT and the four k-point distributions. Percent differences of the G/EM/7 and EM/G/7 models from the EM/EM/7 model are given in Table A.2. Note the anomaly when G is Pearson 6. The moments vary considerably from the data, yet the percent differences in output measures are modest. Except for this, "generally" the percent differences are smallest when the percent differences in moments are smallest.

REFERENCES

Cheng, R.C.H. 1993. Selecting Input Models and Random Variate Generation. In *Proceedings of the 1993 Winter Simulation Conference,* ed. G.W. Evans, M. Mollaghasemi, E. C. Russell, W.E. Biles, 34--40.

Cheng,R.C.H., Holland, W. and Hughes, N.A. Selection of Input Models Using Bootstrap Goodness-Of-Fit. In *Proceedings of the 1996 Winter Simulation Conference,* ed. J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain, 199-206.

Fox, B.L. 1981. Fitting "Standard" Distributions to Data Is Necessarily Good: Dogma or Myth? In *Proceedings of the 1981 Winter Simulation Conference*, ed. T. I. Oren, C.M. Delfosse, C.M. Shub, 305--307.

Harris, C.M. and Marchal, W.G. 1997. Distribution Estimation Using Laplace Transforms. Department of Operations Research and Engineering Technical Report, George Mason University, Fairfax, Virginia.

Johnson, M.A. 1993. An Empirical Study of Queueing Approximations Based on Phase-Type Distributions. *Commun. Statist.-Stochastic Models*, 9:531--561.

Johnson, M. A. and M. R. Taaffe. 1991. An Investigation of Phase-Distribution Moment-Matching Algorithms for use in Queueing Models. *Queueing Systems*, 8:129--147.

Kelton, W. D. 1984. Input Data Collection and Analysis. In *Proceedings of the 1984 Winter Simulation Conference*, ed. S. Sheppard, U. Pooch, D. Pegden, 305--307.

Leemis, L.M. 1996. Discrete-Event Simulation Input Process Modeling. In *Proceedings of the 1996 Winter Simulation Conference,* ed. J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain, 39-46.

Leemis, L. M. 1995. Input Modeling for Discrete-Event Simulation. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulos, K. Kang, W.R. Lilegdon,D. Goldsman, 16--23.

Nelson, B. L, Ware, P. and Cario, Harris, Jamison, Miller, Steinbugl, Yang. 1995. Input Models When Simple Models Fail. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulos, K.Kang, W.R. Lilegdon, D. Goldsman, 93--100.

Storer, R.H., Swain, J. J., Venkatraman, S. and J. R. Wilson. 1988. Comparison of Methods For Fitting Data Using Johnson Translation Distributions. In *Proceedings of the 1988 Winter Simulation Conference*, ed. M. Abrams, P. Hughes, J. Comfort, 476-481.

Wagner, M.A.F., and J. R. Wilson. 1996. Using Univariate Bezier Distributions to Model Simulation Input Processes. *IIE Transactions*, 28:699--711.

Wagner, M.A.F., and J.R. Wilson. 1996. Recent Developments in Input Modeling With Bezier Distributions. In *Proceedings of the 1996 Winter Simulation Conference,* ed. J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain, 1448-1456.

AUTHOR BIOGRAPHIES

DONALD GROSS is a research professor in the Department of Operations Research and Engineering at George Mason University and Professor Emeritus, Department of Operations Research, at The George Washington University. His research interests are queueing theory and simulation input modeling.

MAN JUTTIJUDATA received his D.Sc. degree from The George Washington University and is a specialist in economics and financial modeling with the Thai Investment and Securities Public Company Limited. His dissertation research was on input modeling.

Table A.1 - Model Moment Comparisons

Interarrival-Time Data

Model	Mean	Variance	Skewness	Kurtosis
Arrival Data n = 168				
Sample Values	.51458	.36501	2.41691	11.9585
1-Rand. Walk	.51458	.40588	2.63788	13.5547
2-Weibull	.51360	.35416	2.63788	13.5547
3-Pearson 6	.54844	.94431	Does Not Exist	Does Not Exist
4-Gamma	.51458	.32572	2.21817	10.3804
5-Gamma (E)	.51458	.43244	2.60641	13.1900

Service-Time Data

Model	Mean	Variance	Skewness	Kurtosis
Service Data n = 132				
Sample Values	2.26887	5.71426	3.53355	19.2950
1-Log logistic	2.23157	41.6280	Does Not Exist	Does Not Exist
2-Pearson 6	2.25786	6.30176	26.2730	Does Not Exist
3-Lognormal	2.22235	4.11634	3.49973	30.7117
4-Inv. Gaussian	2.26887	4.40487	2.77510	15.8353
5-Random Walk	2.26887	3.46828	2.03991	9.52663

Table A.2 - Simulation Results

G/EM/7

Mod-el (G)	Mean	% Δ	Std. Dev.	% Δ	Skew	% Δ	Kurt	% Δ	% Δ Wq	% Δ Wq (.95)
EM	.52	0	.60	0	2.44	0	12.09	0	0	0
RW	.51	-.01	.64	5.58	2.62	7.64	13.38	10.67	13.17	10.93
WB	.51	-.29	.59	-1.28	2.48	1.76	12.59	4.16	-2.19	-1.43
P6	.55	6.89	.98	62.73	28.89	>999	>999	>999	10.61	8.23
GM	.51	-.04	.57	-5.12	2.22	-8.75	10.40	-14.0	-10.9	-8.80
GME	.51	-.05	.66	9.28	2.61	7.17	13.21	9.26	25.22	19.98
KP2	.51	-.09	.61	.51	.33	-87	1.11	-91	39.56	38.37
KP3	.52	.01	.61	.61	2.44	.09	7.00	-42	-11.5	-11.3
KP4	.51	.06	.61	.54	2.44	.26	12.12	.28	2.65	4.55
KP5	.51	-.08	.61	.58	2.45	.41	12.14	.46	2.47	1.65

EM/G/7

Model (G)	Mean	% Δ	Std. Dev.	% Δ	Skew	% Δ	Kurt	% Δ	% Δ Wq	% Δ Wq (.95)
EM	2.27	0	2.38	0	3.58	0	19.65	0	0	0
LL	2.23	-1.56	3.19	33.96	147	>999	>999	>999	-9.44	-8.31
P6	2.26	-.46	2.53	6.34	19.73	451	3714	>999	-3.86	-2.48
LN	2.22	-2.02	2.03	-14.9	3.46	-3.39	29.09	48.05	-17.0	-12.9
IG	2.27	-.06	2.10	-11.9	2.78	-22.3	15.93	-18.9	-4.79	-2.33
RW	2.27	.06	1.86	-21.9	2.03	-43.4	9.38	-52.3	-8.24	-5.29
KP2	2.27	-.01	2.40	.72	.11	-96.9	1.01	-94.8	21.27	33.77
KP3	2.27	-.09	2.40	.58	3.57	-.20	13.78	-29.9	-2.31	-7.61
KP4	2.27	.04	2.40	.72	3.57	-.37	19.60	-.24	2.27	4.44
KP5	2.27	-.02	2.40	.67	3.57	-.38	19.61	-.21	-.13	.73