

1992-8
ET

Faculteit der Economische Wetenschappen en Econometrie

05348

Serie Research Memoranda

Comparison of two approximations for the loss
probability in finite-buffer queues

M. Miyazawa
H.C. Tijms

Research Memorandum 1992-8
maart 1992



**COMPARISON OF TWO APPROXIMATIONS FOR THE LOSS PROBABILITY
IN FINITE-BUFFER QUEUES**

Masakiyo Miyazawa, Science University of Tokyo
Dept. of Information Sciences
Noda City, Chiba 278, Japan

Henk Tijms, Vrije University
Dept. of Econometrics
1081 HV Amsterdam, The Netherlands



Abstract This note deals with two related approximations that were recently proposed for the loss probability in finite-buffer queues. The purpose of the paper is two-fold. First, to provide better insight and more theoretical support for both approximations. Second, to show by an experimental study how well both approximations perform. An interesting empirical finding is that in many cases of practical interest the two approximations provide upper and lower bounds on the exact value of the loss probability.

1 INTRODUCTION

Consider the finite-capacity GI/GI/c/N+c queue, where any arriving customer finding all c servers busy and all N other waiting places occupied is lost. It is assumed that the traffic intensity $\rho = \lambda E(S)/c$ is smaller than 1, where λ is the arrival rate of customers and $E(S)$ is the mean service time of a customer. A problem of considerable practical interest is to find the loss probability P_{loss} being defined as the long-run fraction of customers that are lost. The recent papers Sakasegawa et al (1990) and Tijms (1991) address this problem. The first paper proposes the approximation

$$(1) \quad P_{\text{app}}(\text{time}) = \frac{(1-\rho) \left(1 - \sum_{i=0}^{N+c-1} p_i^{(\infty)} \right)}{1 - \rho + \rho \sum_{i=0}^{N+c-1} p_i^{(\infty)}} ,$$

and the second paper gives the approximation

$$(2) \quad P_{\text{app}}(\text{cus}) = \frac{(1-\rho) \left(1 - \sum_{i=0}^{N+c-1} \pi_i^{(\infty)} \right)}{1 - \rho + \rho \sum_{i=0}^{N+c-1} \pi_i^{(\infty)}} .$$

Here $\{p_i^{(\infty)}\}$ and $\{\pi_i^{(\infty)}\}$ denote for the corresponding infinite-capacity queue the equilibrium distributions of the number of customers present at an arbitrary point in time and just prior to an arrival epoch, respectively.

In Tijms (1991) the approximation for the loss probability is extended to the batch-arrival $GI^X/GI/c/N+c$ queue with partial overflow:

$$(3) \quad P_{\text{app}}(\text{cus}) = \frac{(1-\rho) \left(1 - \sum_{i=0}^{N+c-1} q_i^{(\infty)} \right)}{1 - \rho + \rho \sum_{i=0}^{N+c-1} q_i^{(\infty)}},$$

where

$$(4) \quad q_i^{(\infty)} = \frac{1}{E(X)} \sum_{j=0}^i \pi_j^{(\infty)} P(X > i-j).$$

In the batch-arrival queue with partial overflow only those arriving customers finding no free waiting place are lost. For the batch-arrival case the traffic intensity ρ is defined as $\rho = \lambda E(X)E(S)/c$, where λ denotes the arrival rate of batches and the random variable X denotes the batch size. The distribution $\{\pi_j^{(\infty)}\}$ has the same meaning as before. It is noted that $\{q_i^{(\infty)}\}$ represents the stationary distribution of the number of customers left behind at a service-completion epoch in the infinite-capacity queue. In view of the approximations (1)-(3), it is natural to consider for the batch-arrival case the alternative approximation

$$(5) \quad P_{\text{app}}(\text{time}) = \frac{(1-\rho) \left(1 - \sum_{i=0}^{N+c-1} z_i^{(\infty)} \right)}{1 - \rho + \rho \sum_{i=0}^{N+c-1} z_i^{(\infty)}},$$

where

$$(6) \quad z_i^{(\infty)} = \frac{1}{E(X)} \sum_{j=0}^i p_j^{(\infty)} P(X > i-j).$$

Note that the approximations (3) and (5) contain the approximations (1) and (2) as special cases. Also, note that by the PASTA property the approximations (3) and (5) are identical when the arrival process of batches is a Poisson process. The approximation (3) is exact for the single-server $M^X/GI/1/N+1$ queue with Poisson arrivals and the multi-server $M^X/M/c/N+c$ queue with exponential services, see e.g. Tijms (1991).

The remainder of the paper is organized as follows. In the sections 2 and 3 we discuss some theoretical aspects of the two approximations. Under certain conditions on the arrival process of batches it will be shown that either $P_{app}(time) \geq P_{app}(cus)$ or $P_{app}(time) \leq P_{app}(cus)$ is always true regardless of the service-time distribution and the batch-size distribution. Section 4 deals with an experimental study of the two approximations. An interesting empirical finding is that in many cases the two approximations provide sharp upper and lower bounds on the exact value of the loss probability P_{loss} .

2 SUFFICIENT CONDITIONS

The analysis in this section is given under a relaxation of the assumption of independently and identically distributed interarrival times of the batches. We are concerned with the $G^X/GI/c/N+c$ queue in which the arrival process of batches is a stationary point process. Denoting by X_n the number of customers in the n^{th} batch and by S_k the service time of the k^{th} customer, it is assumed that $\{X_n\}$ and $\{S_k\}$ are i.d.d. sequences, where $\{X_n\}$ and $\{S_k\}$ are mutually independent and are independent of the arrival process. In the following the generic variables X and S denote the batch size and the service time of a customer. The assumption of partial overflow is made, that is, only those arriving customers finding no free waiting place are lost. There are $N+c$ waiting places for the customers including any customer in service. The long-run fraction of arriving customers that are lost is denoted by $P_{loss}^{(N)}$. The stationary distribution of the number of customers in the system at an arbitrary point in time and just prior to an arrival epoch are denoted by $\{p_i^{(N)}\}$ and $\{\pi_i^{(N)}\}$ respectively. It is assumed that the traffic intensity $\rho = \lambda E(X)E(S)/c$ is less than 1, where λ is the mean arrival rate of batches. This assumption guarantees the existence of the stationary queue-length distributions $\{p_i^{(\infty)}\}$ and $\{\pi_i^{(\infty)}\}$ for the corresponding infinite-capacity queue.

In Sakasegawa et al. (1990) three heuristic assumptions, called Assumptions 1, 2 and 3, are made to get the approximation (1). We modify those assumptions into suitable forms. Assumption 1 automatically holds for our model. We first replace Assumption 2 by the stronger condition

$$(A.1) \quad \text{There exists a constant } \gamma \text{ satisfying } p_i^{(N)} = \gamma p_i^{(\infty)} \text{ for } i=0,1,\dots,N+c-1.$$

Remark 2.1 Since $\sum_{i=0}^{N+c} p_i^{(N)} = 1$, (A.1) implies $\gamma = (1 - p_{N+c}^{(N)}) / \sum_{i=0}^{N+c-1} p_i^{(\infty)}$. Hence, γ is in general different from the normalizing constant for the truncation $(p_i^{(\infty)})_{i=0}^{N+c}$.

Assumption A.1 holds for both the $M^X/GI/1/N+1$ queue and the $M^X/M/c/N+c$ queue. A direct proof of this result can be found in Tijms (1991). Alternatively, this result can be deduced from results in Miyazawa (1989) and Miyazawa and Shanthikumar (1991). The assumption (3) in Sakasegawa et al. (1990) could be expressed by $p_{N+c}^{(N)} = \pi_{N+c}^{(N)}$. For the batch-arrival case, we extend this assumption to the following form:

$$(A.2) \quad \lambda \left(\sum_{i=0}^{N+c-1} p_i^{(N)} E[(X+i-N-c)^+] + E(X) p_{N+c}^{(N)} \right) = \lambda E(X) P_{\text{loss}}^{(N)},$$

where $a^+ = \max(0, a)$. The assumption (A.2) is motivated by considering the flow-balance equation for the overflow customers. The assumption is exact if $p_i^{(N)} = \pi_i^{(N)}$ for $i=0, 1, \dots, N+c$. Clearly, the latter equality holds when the arrival process is Poisson.

Theorem 2.1 If the assumptions (A.1) and (A.2) hold, then

$$(7) \quad P_{\text{loss}}^{(N)} = \frac{(1-\rho) \left[E(X) - \sum_{i=0}^{N+c-1} p_i^{(\infty)} \sum_{j=0}^{N+c-1-i} P(X > j) \right]}{(1-\rho) E(X) + \rho \sum_{i=0}^{N+c-1} p_i^{(\infty)} \sum_{j=0}^{N+c-1-i} P(X > j)}.$$

Proof. From Little's formula for the average number of busy servers, we have

$$c - \sum_{i=0}^{c-1} (c-i) p_i^{(N)} = \lambda E(X) E(S) (1 - P_{\text{loss}}^{(N)}),$$

cf. also Sakasegawa et al (1990). Dividing both sides of this equation by c yields

$$(8) \quad 1 - \sum_{i=0}^{c-1} \left(1 - \frac{i}{c}\right) p_i^{(N)} = \rho (1 - P_{\text{loss}}^{(N)}).$$

In particular, noting that $P_{\text{loss}}^{(N)} = 0$ for $N = \infty$,

$$(9) \quad 1 - \sum_{i=0}^{c-1} \left(1 - \frac{i}{c}\right) p_i^{(\infty)} = \rho .$$

We next substitute $p_i^{(N)} = \gamma p_i^{(\infty)}$ of (A.1) into (8). Then, by using (9), we have

$$(10) \quad 1 - (1-\rho)\gamma = \rho(1-P_{\text{loss}}^{(N)}) .$$

Similarly, (A.2) implies

$$\gamma \sum_{i=0}^{N+c-1} p_i^{(\infty)} E[(X+i-N-c)^+] + E(X) p_{N+c}^{(N)} = E(X) P_{\text{loss}}^{(N)} .$$

By using this and (A.1), we get

$$(11) \quad 1 = \sum_{i=0}^{N+c-1} P_i^{(N)} + P_{N+c}^{(N)} \\ = \frac{\gamma}{E(X)} \sum_{i=0}^{N+c-1} p_i^{(\infty)} \sum_{j=0}^{N+c-1-i} P(X>j) + P_{\text{loss}}^{(N)} .$$

Thus we get (7) from (10) and (11).

Theorem 2.1 suggests to use the right-hand side of (7) as an approximation for the loss probability in the general $G^X/GI/c/N+c$ queue. We denote this approximation by $P_{\text{app}}^{(\text{time})}$. It is a matter of simple algebra to verify that the right-hand sides of (5) and (7) are the same. The approximation $P_{\text{app}}^{(\text{time})}$ is exact for the $M^X/GI/1/N+1$ and $M^X/M/c/N+c$ queues, since (A.1) and (A.2) hold for these models. The approximation $P_{\text{app}}^{(\text{cus})}$ can be obtained by using similar heuristic assumptions. The assumptions

$$(B.1) \quad \pi_i^{(N)} = \gamma \pi_i^{(\infty)} \quad \text{for } i=0,1,\dots,N+c-1$$

$$(B.2) \quad \pi_1^{(N)} = p_1^{(N)} \quad \text{and} \quad \pi_1^{(\infty)} = p_1^{(\infty)} \quad \text{for } i=0,1,\dots,c-1$$

lead to the approximation $P_{\text{app}}^{(\text{cus})}$, cf. also Tijms (1991) for an alternative derivation. By the PASTA property, the two approximations $P_{\text{app}}^{(\text{time})}$ and $P_{\text{app}}^{(\text{cus})}$ are identical when the arrival process of batches is Poisson. How-

ever, it seems very hard to see which of the two approximations is in general better because the assumptions of their heuristic derivations are difficult to compare. Nevertheless, the comparison of the two approximations is an interesting issue. In the next section, we consider this issue for a restricted class of arrival processes.

3 STOCHASTIC ORDERING

In the following, we use the notion of stochastic ordering. For two distributions μ and ν on the real line, we call μ to be stochastically less than ν and denote it by $\mu \leq_{st} \nu$ if

$$1 - F(x) \leq 1 - G(x) \quad \text{for all real } x,$$

where F and G are the cumulative distribution functions of μ and ν , respectively. It is well known (e.g. see Stoyan (1983)) that $\mu \leq_{st} \nu$ is equivalent to

$$\int_{-\infty}^{+\infty} \phi(x) \mu(dx) \leq \int_{-\infty}^{+\infty} \phi(x) \nu(dx) \quad \text{for any nondecreasing function } \phi.$$

From the relations (3)-(7), we obtain the following result.

Proposition 3.1 The approximations $P_{app}(cus)$ and $P_{app}(time)$ increase if the corresponding distributions $\{\pi_i^{(\infty)}\}$ and $\{p_i^{app(\infty)}\}$ are stochastically increased.

This monotonicity is very natural and should be satisfied by any approximation for the loss probability that uses the equilibrium probabilities of the corresponding infinite-capacity queue.

From now on, we consider subclasses of the $GI^X/GI/c/N+c$ queue. We first assume that the interarrival-time distribution F is NBUE (New Better than Used in Expectation), see Stoyan (1983) for the definition of NBUE. Then the same argument as used in Miyazawa (1989) to get relation (4.8) of his paper yields

$$\{\pi_i^{(\infty)}\} \leq_{st} \{p_i^{(\infty)}\}.$$

Hence Proposition 3.1 implies that $P_{app}(cus) \leq P_{app}(time)$. If F is NWUE, then all inequalities are reversed. Thus we get the following result.

Proposition 3.2 For the $GI^X/GI/c/N+c$ queue with an NBUE (NWUE) interarrival-time distribution we have

$$P_{app}(cus) \leq (\geq) P_{app}(time).$$

Finally, it is a practically important question whether $P_{app}(cus)$ and $P_{app}(time)$ provide bounds on the exact value of the loss probability P_{loss} . Of course, this is a very hard problem to answer. We could only deal with this problem in an experimental way. Our numerical investigations to be discussed in the next section lead to the following conjecture.

Conjecture. For both the $GI^X/GI/1/N+1$ queue and the $GI^X/M/c/N+c$ queue, it holds that

$$P_{app}(cus) \leq P_{loss} \leq P_{app}(time) \quad (\text{NBUE interarrival time})$$

and

$$P_{loss} \leq P_{app}(cus) \quad (\text{NWUE interarrival time}).$$

4. NUMERICAL DISCUSSION

Let us first give a number of numerical results. Table 1 deals with the single-server $D^X/E_k/1/N+1$ queue with batch arrivals. For several constant batch sizes and several Erlangian distributions, the approximate values $P_{app}(cus)$ and $P_{app}(time)$ are given together with the exact value of the loss probability P_{loss} . In the table these values are respectively denoted by $appc$, $appt$ and $exact$. Table 2 deals with the multi-server $C_2/M/c/N+c$ queue with single arrivals, where the interarrival time has a Coxian-2 distribution. Note that a Coxian-2 distributed interarrival time A can be represented as $A=A_1$ with probability $1-b$ and $A=A_1+A_2$ with probability b , where A_1 and A_2 are independently, exponentially distributed random variables with respective means $1/\lambda_1$ and $1/\lambda_2$. A Coxian-2 distribution is not uniquely determined by its mean $E(A)$ and its coefficient of variation c_A (= the ratio of standard deviation and mean). To fix uniquely the three parameters, we consider the following normalizations:

Table 1 The loss probability for the $D^X/E_k/1/N+1$ queue

		$\rho=0.8$	$N=10$	$\rho=0.9$	$N=25$	$\rho=0.95$	$N=50$
		E_4	E_2	E_4	E_2	E_4	E_2
X=1	appc	8.45×10^{-10}	1.06×10^{-5}	3.45×10^{-11}	1.69×10^{-6}	4.38×10^{-11}	1.41×10^{-6}
	appt	2.46×10^{-9}	1.74×10^{-5}	5.47×10^{-11}	2.11×10^{-6}	5.43×10^{-11}	1.57×10^{-6}
	exact	1.24×10^{-9}	1.26×10^{-5}	4.26×10^{-11}	1.85×10^{-6}	4.89×10^{-11}	1.48×10^{-6}
X=3	appc	9.36×10^{-9}	2.68×10^{-5}	8.60×10^{-11}	2.34×10^{-6}	6.43×10^{-11}	1.66×10^{-6}
	appt	4.40×10^{-7}	1.47×10^{-4}	4.05×10^{-10}	4.96×10^{-6}	1.28×10^{-10}	2.30×10^{-6}
	exact	2.40×10^{-8}	4.74×10^{-5}	1.66×10^{-10}	3.40×10^{-6}	9.41×10^{-11}	1.98×10^{-6}
X=5	appc	1.79×10^{-7}	9.06×10^{-5}	2.71×10^{-10}	3.85×10^{-6}	1.02×10^{-10}	2.01×10^{-6}
	appt	2.08×10^{-4}	2.02×10^{-3}	4.55×10^{-9}	1.36×10^{-5}	3.42×10^{-10}	3.53×10^{-6}
	exact	5.92×10^{-7}	2.10×10^{-4}	7.23×10^{-10}	6.74×10^{-6}	1.91×10^{-10}	2.76×10^{-6}

Table 2 The loss probability for the $C_2/M/c/N+c$ queue with $c=10$

		$\rho=0.5$	$N=25$	$\rho=0.8$	$N=100$	$\rho=0.9$	$N=200$
		B	G	B	G	B	G
$c_A^2=5$	appc	1.84×10^{-5}	3.93×10^{-4}	3.71×10^{-5}	9.00×10^{-5}	5.58×10^{-5}	7.95×10^{-5}
	appt	1.29×10^{-5}	2.48×10^{-4}	3.22×10^{-5}	7.75×10^{-5}	5.21×10^{-5}	7.41×10^{-5}
	exact	1.35×10^{-5}	3.25×10^{-4}	3.20×10^{-5}	8.53×10^{-5}	5.20×10^{-5}	7.76×10^{-5}
$c_A^2=0.7$	appc	5.78×10^{-12}	1.04×10^{-11}	2.05×10^{-13}	2.46×10^{-13}	9.38×10^{-13}	1.01×10^{-13}
	appt	6.77×10^{-12}	1.20×10^{-11}	2.14×10^{-13}	2.56×10^{-13}	9.56×10^{-13}	1.03×10^{-13}
	exact	5.90×10^{-12}	1.08×10^{-11}	2.06×10^{-13}	2.49×10^{-13}	9.41×10^{-13}	1.02×10^{-12}

- (i) *Gamma normalization (G)*. The parameters b , λ_1 and λ_2 are chosen such that the third moment of the Coxian-2 distribution is the same as the third moment of the gamma distribution that is uniquely determined by $E(A)$ and c_A .
- (ii) *Balanced means (B)*. For $0.5 \leq c_A^2 < 1$, the parameter b is taken equal to 1, i.e. the interarrival time is the sum of two exponentials. For $c_A^2 \geq 1$ the parameters of the Coxian-2 density $p_1 \lambda_1 e^{-\lambda_1 t} + p_2 \lambda_2 e^{-\lambda_2 t}$ are chosen such that $p_1/\lambda_1 = p_2/\lambda_2$.

The following conclusions can be drawn from our numerical investigations:

- The two approximations are of a comparable quality with a few exceptions in the batch-arrival queue. Each of the two approximations is accurately enough to be used for dimensioning the buffer size when a (very) small loss probability is required (as is the case in many telecommunication problems).
- In many cases of practical interest the two approximations provide sharp upper and lower bounds on the exact value of the loss probability.

It is striking how remarkably accurate the approximations are for extremely small loss probabilities. This finding was also seen for highly variable interarrival times. For example, for the $C_2/M/c/N+c$ queue with $c_A^2=50$ (gamma), $\rho=0.9$, $c=10$ and $N=3000$, we have the approximate values $P_{app}(cus)=1.32 \times 10^{-12}$ and $P_{app}(time)=1.10 \times 10^{-12}$ and the exact value $P_{loss}=1.21 \times 10^{-12}$. Also, it is remarkable how well the two approximations match the dependency of the loss probability on the shape of the arrival process.

To conclude this section, the following recommendations are made for the use of the approximations for engineering purposes:

- (1) In case both approximations are computable, use the approximation

$$\frac{1}{2} \left\{ P_{app}(cus) + P_{app}(time) \right\}.$$

- (2) For both the $GI^X/GI/1/N+1$ queue and the $GI^X/M/c/N+c$ queue, use the bounds

$$P_{app}(cus) \leq P_{loss} \leq P_{app}(time) \quad \text{if } c_A^2 \leq 1$$

and

$$P_{app}(cus) \geq P_{loss} \quad \text{if } c_A^2 \geq 1.$$

For the $GI^X/M/c/N+c$ queue with $c_A^2 > 1$ the approximation $P_{app}(time)$ does not provide consistently an upper or lower bound on P_{loss} as can be seen from Table 2. Also, for the $GI^X/GI/c/N+c$ with $c > 1$ and nonexponential services, it is not to be expected that $P_{app}(cus) \leq (\geq) P_{loss}$ generally holds if $c_A^2 \leq (\geq) 1$; otherwise, it would be true that $P_{app}(cus) = P_{loss}$ for the $M/GI/c/N+c$ queue, but this equality does not hold for this multi-server case when the service times are non-exponential.

Acknowledgement

We are indebted to Professor H. Sakasegawa for valuable discussions.

References

1. Miyazawa, M. (1989), Comparison of the loss probability of the $GI^X/GI/1/k$ queues with a common traffic intensity, *Journal of the Operations Research Society of Japan* 32: 505-516.
2. Miyazawa, M. and Shanthikumar, J.G. (1991), Monotonicity of the loss probabilities of single server finite queues with respect to convex order of arrival or service processes, *Probability in the Engineering and Informational Sciences*, 5: 43-52.
3. Sakasegawa, H., Miyazawa, M. and Yamazaki, G. (1990), Evaluating the overflow probability using the infinite queue (submitted for publication to *Management Science*).
4. Stoyan, D. (1983), *Comparison methods for queues and other stochastic models*. Edited with revision by D.J. Daley, John Wiley & Sons, New York.
5. Tijms, H.C. (1991), Heuristics for finite-buffer queues (to appear in *Probability in the Engineering and Informational Sciences* (1992)).