Probability in the Engineering and Informational Sciences, **21**, 2007, 551–556. Printed in the U.S.A.
DOI: 10.1017/S0269964807000319

A BAYESIAN APPROACH TO FIND RANDOM-TIME PROBABILITIES FROM EMBEDDED MARKOV CHAIN PROBABILITIES

WINFRIED K. GRASSMANN
Department of Computer Science
University of Saskatchewan
Saskatoon, S7N 5C9 Canada
E-mail: grassman@cs.usask.ca

JAVAD TAVAKOLI
Department of Mathematics, Statistics and Physics
University of British Columbia Okanagan
Kelowna, V1V 1V7 Canada
E-mail: javad.tavakoli@ubc.ca

The embedded Markov chain approach is widely used in queuing theory, in particular in $M/G/1$ and $GI/M/c$ queues. In these cases, one has to relate the embedded equilibrium probablities to the corresponding random-time probabilities. The classical method to do this is based on Markov renewal theory, a rather complex approach, especially if the population is finite or if there is balking. In this article we present a much simpler method to derive the random-time probabilities from the embedded Markov chain probabilities. The method is based on conditional probability. Our approach might also be applicable in such situations.

1. INTRODUCTION

In the $M/G/1$ and the $GI/M/1$ models in statistical equilibrium, one typically finds first the probabilities at the times after a departure and the probabilities before an arrival, respectively, and one then uses these probabilities to find random-time probabilities

© 2007 Cambridge University Press 0269-9648/07 $25.00

https://doi.org/10.1017/S0269964807000319 Published online by Cambridge University Press

(see, e.g., [7, 9, 10]). For doing this, one normally uses Markov renewal theory (see, e.g., [11]). However, in [4], we suggested using Bayes' law instead. This gives much shorter derivations. In this article, this method will be explored further and generalized. In particular, we are looking at the $M/G/1$ queue and the $GI/M/c$ queue, and in both cases, we allow arrival rates to depend on the number in the system. We also consider a tandem queue with the first line never starved, and the first server having a general service time. However, queuing networks are also amenable to embedded Markov chain analysis if only one service time has a general distribution and all other service and interarrival times are Markovian.

2. THE $M/G/1$ QUEUE

Consider the queue with Poisson arrivals, general service times, and one server, where the arrival rates λ_i depend on i, the number in the system. The service rate is denoted by μ. It is well known that π_i, the probability of having i elements in system after a departure, can be found by using an embedded Markov chain [4]. In many practical applications, such as finding costs, one needs the random-time probabilities P_i rather than the π_i. However, as shown in [4], we have the following relation between P_i and π_i:

$$\pi_i = \frac{P_i \lambda_i}{\sum_{j=0}^{\infty} P_j \lambda_j}. \tag{1}$$

Since similar ideas are used later, we repeat the proof of (1). If, at time t, the system is in a statistical equilibrium, one has, following the arguments of Bayes' theorem,

$$P\{i \text{ at } t, \text{arrival in } (t, t+h)\} = P_i \lambda_i h + o(h),$$

$$P\{\text{arrival in } (t, t+h)\} = \sum_{j=0}^{\infty} P_j \lambda_j h + o(h),$$

$$P\{i \text{ at } t \mid \text{arrival in } (t, t+h)\} = \frac{P_i \lambda_i}{\sum_{j=0}^{\infty} P_j \lambda_j} + \frac{o(h)}{h}.$$

As $h \to 0$, $P\{i \text{ at } t \mid \text{arrival in } (t, t+h)\}$ converges to Q_i, the probability of having i elements in the system before an arrival. As is shown for instance by [6], $Q_i = \pi_i$. This completes the proof of (1).

For all i with $\lambda_i > 0$, one finds from (1)

$$P_i = \frac{\pi_i U}{\lambda_i}, \tag{2}$$

https://doi.org/10.1017/S0269964807000319 Published online by Cambridge University Press

where

$$U = \sum_{j=0}^{\infty} P_j \lambda_j. \tag{3}$$

However, since we do not know the P_j, we cannot use (3) to find U. However, one can interpret U as the factor that ensures that $\sum_{i=0}^{\infty} P_i = 1$. This requires that we know P_i for all i. If we have a buffer of size N, such that $\lambda_i = 0$ for $i \geq N$, then $\lambda_N = \pi_N = 0$, and, unfortunately, (2) cannot be used to find P_N, and neither can we find U by (3). However, we have an alternative formula for U, namely

$$U = \frac{1}{\frac{1}{\mu} + \frac{\pi_0}{\lambda_0}}. \tag{4}$$

To prove (4), note that U is really the throughput. This follows because $P\{\text{arrival in } (t, t+h)\} = Uh + o(h)$ and because the probability of more than one arrival is $o(h)$. Alternatively, we can invoke the PASTA principle. The throughput is both the unconditional arrival rate and the unconditional departure rate, which, in turn, implies that the expected time between departures is $1/U$. On the other hand, the expected time between departures is $1/\mu$ if, after the previous departure, there were elements in the system and $1/\mu + 1/\lambda_0$ if the previous departure left the system empty. This completes the proof of (4).

Once U is obtained from (4), (2) can be used, except that for P_N, one has

$$P_N = 1 - \sum_{i=0}^{N-1} P_i = 1 - U \sum_{i=0}^{N-1} \frac{\pi_i}{\lambda_i}.$$

Hence, the problem of relating the departure time probabilities to the random-time probabilities is solved.

If the arrival rate is independent of i, $\lambda_i = \lambda$, then $U = \lambda$, and by (2), $P_i = \pi_i$. Also, in this case, (4) works out correctly.

In [7, p. 232], the following formula is given when $\lambda_i = \lambda$, $i < N$, and $\lambda_i = 0$, $i \geq N$:

$$P_i = \frac{\pi_i}{\pi_0 + \frac{\lambda}{\mu}}, \qquad i < N,$$

in agreement with (2) and (4). This shows that our result generalizes the one given in [7].

3. THE *GI/M/c* QUEUE

In this section we discuss the $GI/M/c$ queue with balking; that is, we assume that arrivals join with probability q_i, where i is the number in the system. An embedded Markov chain can be used to find the probabilities π_i of having i elements in the system before an arrival. As earlier, the π_i are also equal to the probabilities that a departure leaves i elements behind. To find the transition probabilities of the embedded Markov chain, one can apply randomization, similar to the methods used in [2–4].

https://doi.org/10.1017/S0269964807000319 Published online by Cambridge University Press

In order to obtain a relation between π_i and the random-time probabilities P_i, we use a similar argument as earlier. For this, we again need the throughput U. Since the total probability that an arrival joins is $\sum_{j=0}^{\infty} q_j\pi_j$, $U = \lambda\sum_{j=0}^{\infty} q_j\pi_j$. On the other hand, if $\mu_j = j\mu$ for $j < c$ and $\mu_j = c\mu$ otherwise, then $U = \sum_{j=1}^{\infty} P_j\mu_j$. This relation, which can be obtained by the same arguments as earlier means that "departures see time averages." Applying Bayes' theorem now yields

$$\pi_i = \frac{P_{i+1}\mu_{i+1}}{U},$$

and, consequently,

$$P_i = \frac{\pi_{i-1}U}{\mu_i}, \qquad i = 1, 2, \ldots; \tag{5}$$

$$P_0 = 1 - \sum_{i=1}^{\infty} P_i.$$

Using $U = \lambda\sum_{i=0}^{\infty} q_i\pi_i$, (5) allows us to find all P_i once the π_i are known.

As a check, consider the $GI/M/1$ queue with $\lambda_i = \lambda$. In this case, $U = \lambda$, and (5) yields $P_i = \pi_{i-1}\lambda/\mu$, which matches with the formulas give in the literature (see, e.g., [8, p. 196]). In [5], an algorithm is given for finding both π_i and P_i, given constant interarrival times, using classical Markov renewal-type arguments. We used this algorithm to calculate π_i and P_i for some cases with $c = 2$. For instance, for $\lambda = 1$ and $\mu = 1$, one finds from (5) that $P_1 = \pi_0$, whereas $P_2 = \pi_1/2$. Our calculations showed that $P_1 = \pi_0 = .490871$, whereas $P_2 = .202840$, and $\pi_1 = 0.405680$, which works out exactly. Similar results were obtained for other values of λ and μ as well.

As an example of a finite-size buffer, consider the $E_2/M/3$ queue with a buffer size $N = 6$, $\lambda = 2$, and $\mu = 1$. In this case, $q_i = 1$ for $i < N$, and $q_i = 0$ otherwise. We formulated this problem as a continuous-time Markov chain and solved it as such. This provided us with the joint distribution of the arrival phase and the number in the system and, from this, we obtained the P_i and π_i as follows. Notice that the third row is obtained by (5), which matches exactly row 1.

i	1	2	3	4	5	6
P_i	0.24921	0.28211	0.18179	0.10443	0.05974	0.03338
π_{i-1}	0.12770	0.28913	0.27947	0.16055	0.09183	0.05131
$\frac{\pi_{i-1}U}{\mu_i}$	0.24921	0.28211	0.18179	0.10443	0.05974	0.03338

4. A TANDEM QUEUE

We are now looking at a tandem queue in which the first line is never starved. The first server has a general service time with rate μ_1, whereas the second server has an exponential service time with rate μ_2. There is a buffer of size N, but when the buffer is

https://doi.org/10.1017/S0269964807000319 Published online by Cambridge University Press

full, the first server is blocked. We want to find a relation between π_i, the probability of having i elements in the second line immediately before an arrival from line 1, and P_i, the random time probability of having i elements in line 2. In this case, the expected time between arrivals to line 2 becomes

$$\frac{1}{\mu_1} + \frac{\pi_{N-1}}{\mu_2}.$$

This is because a service completion with the first server at a point in time when there are $N - 1$ elements in the second line causes the first server to block for a time $1/\mu_2$. After that, it takes an expected time $1/\mu_1$ for server 1 to complete service. The throughput U is the reciprocal of expected time between arrivals to line 2; that is,

$$U = \frac{1}{\frac{1}{\mu_1} + \frac{\pi_{N-1}}{\mu_2}}.$$

The relation between P_i and π_i is similar to the one given in (5):

$$P_i = \frac{\pi_{i-1} U}{\mu_2}, \qquad i = 1, 2, \ldots, N;$$

$$P_0 = 1 - \sum_{i=1}^{N} P_i.$$

As an example, consider a tandem queue with $\mu_1 = 0.75$, $\mu_2 = 1$, and buffer size $N = 6$. The service times of the first server are Erlang-2. In this case, we found the values of P_i and π_i by a continuous-time Markov chain as follows. Again, the third row exactly matches the first row.

i	1	2	3	4	5	6
P_i	.26386	.17862	.12078	.08123	.05317	.02991
π_{i-1}	0.36266	0.24550	0.16601	0.11165	0.07308	0.04111
$\frac{\pi_{i-1}U}{\mu_2}$	0.26386	0.17862	0.12078	0.08123	0.05317	0.02991

5. CONCLUSION

The approach presented here to link embedded Markov chain probabilities and random-time probabilities is simpler and numerically more efficient than the classical Markov renewal approach. To see this, we refer the reader to [1, 7, 8]. There are many systems where an embedded Markov chain analysis is feasible. For instance, networks with all but one server having an exponential service time can be so analyzed, and in all these cases, the Baysian approach outlined in this article will work. Our method will also work in the case of the $GI/M/c$ queue with heterogeneous server, a model analyzed in [5] by Markov renewal methods. In some problems, it might be difficult to find the throughput, in which case one probability cannot be determined, and the

https://doi.org/10.1017/S0269964807000319 Published online by Cambridge University Press

others can only be obtained up to a factor. However, even in this case, our approach can be used as an accuracy check, and it provides useful insights.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada. We also thank the editor for his constructive remarks.

References

1. Chen, X. (1993). The use of derivatives for optimizing steady state queues. Master's thesis, University of Saskatchewan.
2. Ferreira, F. & Pacheco, A. (2006). Analysis of GI/M/s/c queues using uniformization. *Computers and Mathematics with Applications* 51: 291–304.
3. Grassmann, W.K. (1982). The GI/PH/1 queue: A method to find the transition matrix. *INFOR* 20: 144–156.
4. Grassmann, W.K., Chen, X., & Kashyap, B.B.K. (2001). Optimal service rates of the state-dependent M/G/1 queues in steady state. *Operations Research Letters* 29: 57–63.
5. Grassmann, W.K. & Zhao, Y.Q. (1996). Heterogeneous multiserver queues with general input. *INFOR* 35: 208–224.
6. Gross, D. & Harris, C.M. (1971). On one-for-one ordering inventory policies with state-dependent lead times. *Operations Research* 19: 735–760.
7. Gross, D. & Harris, C.M. (1998). *Fundamentals of queueing theory*, 2nd ed. New York: Wiley.
8. Jain, J.L., Mohanty, S.G., & Böhm, W. (2007). *A course on queueing models*. London: Chapman & Hall/CRC.
9. Kleinrock, L. (1975). *Queueing systems*, Vol. 1. New York: Wiley.
10. Prabhu, N. (1965). *Queues and inventories: A study of their basic stochastic properties*. New York: Wiley.
11. Ross, S.M. (1970). *Applied probability models with optimization applications*. San Francisco: Holden Day.

https://doi.org/10.1017/S0269964807000319 Published online by Cambridge University Press