CONVERGENCE OF GENERALIZED TRUNCATION METHOD IN RETRIAL QUEUES

Y. Shin

Department of Statistics, Changwon National University
Changwon, Korea
ywshin@changwon.ac.kr

In many cases, the queue length processes in retrial queues are described by the spatially inhomogeneous Markov chain caused by transitions due to repeated attempts. This lack of homogeneity is one of the causes of analytical complexity of retrial queues and leads to an approximation method. Many authors approximate the original inhomogeneous system by the so-called generalized truncated system in which the retrial rates are restricted to be constant over a given level and the level is enlarged until the satisfactory solution is obtained. However, the rigorous mathematical proofs for the convergence of generalized truncation method are a few and are treated case by case. In this paper, we provide a proof of convergence of the approximation by using the tightness and the stochastic comparison between retrial queues. Some examples are presented to show the usefulness of our approach.

Keywords: retrial queue, tightness, stochastic comparison, generalized truncated system.

1. INTRODUCTION

The queueing systems with repeated attempts (called retrial queues) are characterized by the following feature. When an arriving customer finds that all servers are busy and no waiting position is available, the customer joins a virtual pool of blocked customers called 'orbit' and repeat its request after a random time until the customer gets into the service area. The queue length processes in retrial queues are described by the Markov chain with spatially inhomogeneous infinitesimal generator (or transition probability matrix for discrete time case) caused by transitions due to repeated attempts. This spatial inhomogeneity often leads the analytical complexity and numerical approximations are needed. The detailed overviews of the related references with retrial queues can be found in Falin and Templeton [7] and Artalejo [3].

Falin and Templeton [7] listed several truncation methods to compute the stationary distribution using the other calculable system which is given by varying the retrial rates. There have been drawing the researchers' attention for augmented truncation methods where one truncates the chain to the first N states, makes the resulting matrix stochastic and irreducible in some convenient way, and then solves the finite system (eg. see Seneta [13], Heyman [8] and the references therein). Seneta [12] presented a necessary and sufficient condition for the augmented truncation method to converge to the original stationary distribution is that the sequence of approximating distribution is tight. Zhao and Liu [17] showed that the censored Markov chain provides the best approximation among finite truncation methods in the sense of minimal l_1-sum of errors between the exact distribution and approximation. However, the truncation method uses a Markov chain with a finite state space. So, if the stationary distribution of the infinite-state Markov chain has a long tail, and averages and variances are heavily affected by truncation, the truncation level may have to be very large to get a good approximation. This drawback can be improved by using another calculable system with infinite state space. This method is said to be generalized truncation method [7]. Shin [16] and Shin and Pearce [15] showed that a generalized truncation method can provide the better results than those of censored chain for the Markov chain with transition matrix of upper Hessenberg form.

A variety of generalized truncation models have been considered to approximate the original intricate retrial queues e. g., [11, 7, 6, 1]. However, rigorous mathematical proofs for the convergence of generalized truncation methods are a few and are treated case by case.

In this paper we present a criterion for the convergence of the generalized truncation model to the original system in terms of tightness of probability distributions. Comparison methods are used for the proof of tightness in some examples. In section 2 we provide some preliminary results on the convergence of probability distributions. In section 3 we apply the results in section 2 to some retrial queues from the literature.

2. PRELIMINARIES

We start with recalling the notion of tightness of probability measures. A sequence $\{\boldsymbol{\pi}_n\}$ of probability distributions on $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ is said to be *tight* if for each $\epsilon > 0$ there exists a finite set $E \subset \mathbb{N}_0$ such that $\boldsymbol{\pi}_n(E) > 1 - \epsilon$ for all n. The definition of tightness and its properties including the theorem 2 below can be found in many probability text books, e.g. Bilingsley [5, pp. 336]. Assume that $U_n(t) \leq U(t)$ for all $t \geq 0$, $n = 1, 2, \dots$ and $W(t)$ and $W_n(t)$ converge weakly to the proper random variables $W(\infty)$ and $W_n(\infty)$, $n = 1, 2, \dots$ as $t \to \infty$, respectively. Then $U_n(\infty) \leq_{st} U(\infty)$ for all $n = 1, 2, \dots$ and the distributions of $\{W_n(\infty), n = 1, 2, \dots\}$ is tight, where $X \leq_{st} Y$ means that X is stochastically less than Y, that is, $P(X > x) \leq P(Y > x)$ for all x.

Theorem 2.1 *If $\{\boldsymbol{\pi}_n, n \geq 1\}$ is a tight sequence of probability distributions on $\mathbb{N}_0$, and if each subsequence that converges weakly at all converges weakly to the probability measure $\boldsymbol{\pi}$, then $\{\boldsymbol{\pi}_n\}$ converges weakly to $\boldsymbol{\pi}$.*

Now consider a stochastic process $\boldsymbol{W} = \{(U(t), V(t)), t \geq 0\}$ and a sequence $\{\boldsymbol{W}_n, n = 1, 2, \dots\}$ of stochastic processes $\boldsymbol{W}_n = \{(U_n(t), V_n(t)), t \geq 0\}$ on the same probability space and with the same state space $\mathbb{N}_0 \times S$, where S is a finite set. The following is immediate from stochastic comparison result of Kamae et al. [9, Proposition 3] and the definition of tightness.

Theorem 2.2 *Assume that $U_n(t) \leq U(t)$ for all $t \geq 0$, $n = 1, 2, \dots$ and $W(t)$ and $W_n(t)$ converge weakly to the proper random variables $W(\infty)$ and $W_n(\infty)$, $n = 1, 2, \dots$ as $t \to \infty$, respectively. Then $U_n(\infty) \leq_{st} U(\infty)$ for all $n = 1, 2, \dots$ and the distributions of*

$\{W_n(\infty), n = 1, 2, \ldots\}$ *is tight, where* $X \le_{st} Y$ *means that* X *is stochastically less than* Y*, that is,* $P(X > x) \le P(Y > x)$ *for all* x.

Let $\boldsymbol{p} = (p_j, j \in \mathbb{N}_0)$ and $\boldsymbol{p}^n = (p_j^n, j \in \mathbb{N}_0)$, $n = 1, 2, \ldots$ be a sequence of probability distributions on $\mathbb{N}_0$. Note that the weak convergence of $\boldsymbol{p}^n$ to $\boldsymbol{p}$ is equivalent to $\lim_{n\to\infty} p_j^n = p_j$ for all $j \in \mathbb{N}_0$ and is also equivalent to $\lim_{n\to\infty} \|\boldsymbol{p}^n - \boldsymbol{p}\| \equiv \lim_{n\to\infty} \sum_{j\in\mathbb{N}_0} |p_j^n - p_j| = 0$.

Proposition 2.1 *Let* $Q = (q_{ij})$ *be an infinitesimal generator of an irreducible and positive recurrent Markov chain on* $\mathbb{N}_0$ *and* $\boldsymbol{\pi} = (\pi_j, j \in \mathbb{N}_0)$ *be the stationary probability distribution of* Q*. Let* $Q^n = (q_{ij}^n)$, $n = 1, 2, \ldots$ *be a sequence of infinitesimal generators such that* $\lim_{n\to\infty} q_{ij}^n = q_{ij}$ *for all* $i, j \in \mathbb{N}_0$*. Assume that each of* Q^n *has stationary probability distribution* $\boldsymbol{\pi}^n = (\pi_j^n, j \in \mathbb{N}_0)$, $n = 1, 2, \ldots$*. If* $\{\boldsymbol{\pi}^n\}$ *is tight, then* $\lim_{n\to\infty} \|\boldsymbol{\pi}^n - \boldsymbol{\pi}\| = 0$.

Let $\mathbf{N}$ be any infinite subset of $\{1, 2, \ldots\}$ such that $\{\boldsymbol{\pi}^n, n \in \mathbf{N}\}$ converges weakly to a probability distribution, say $\tilde{\boldsymbol{\pi}}$. Then for all $j \in \mathbb{N}_0$,

$$\tilde{\pi}_j(-q_{jj}) = \lim_{n\in\mathbf{N}} \sum_{i\in\mathbb{N}_0-\{j\}} \pi_i^n q_{ij}^n \ge \sum_{i\in\mathbb{N}_0-\{j\}} \lim_{n\in\mathbf{N}} \pi_i^n q_{ij} = \sum_{i\in\mathbb{N}_0-\{j\}} \tilde{\pi}_i q_{ij}.$$

Thus $\tilde{\boldsymbol{\pi}} Q \le 0$. Since $\tilde{\boldsymbol{\pi}}$ is a probability measure, $\tilde{\boldsymbol{\pi}} = \boldsymbol{\pi}$. It follows from theorem 2 that the assertion is proved.

3. EXAMPLES

In this section we apply the results in the previous section to several examples drawn from the literature of retrial queues. It is easily seen that the infinitesimal generators of the generalized truncation models presented below converge to that of original model. So we focus on the stochastic dominance between the original models and truncation models. Following the same procedures as Bhaskaran [4] or Shin and Kim [14], the comparison results for the examples below are obtained and we omit the proof.

3.1. $BMAP/M/s/K$ **retrial queue.** We consider the queueing system in which there are s identical servers and $K - s$ waiting positions in the service facility. Service times of customers are independent of each other and have a common exponential distribution with parameter μ. When a customer arrives and finds an idle server, the customer receives service immediately. When a customer finds that all the servers are busy and a waiting position is available upon its arrival, the customer occupies the waiting position. Otherwise, the customer who finds all the waiting position is full joins orbit and retries its luck after random time. The retrial is independent and identically repeated until a server or a waiting position is seized. We assume that the time intervals between the successive attempts of the customers in orbit to get service are exponentially distributed with parameters γ_k when there are $k \ge 1$ customers in orbit.

Customers arrive at the system according to a batch Markovian arrival process ($BMAP$) with representation representation $\{D_n, n = 0, 1, 2, \ldots\}$, where D_n, $n = 1, 2, \ldots$ are nonnegative $m \times m$ matrices and the matrix D_0 of size m has strictly negative diagonal elements and nonnegative off-diagonal elements and $D = \sum_{n=0}^{\infty} D_n$ is an infinitesimal generator of irreducible Markov chain. For the more details about $BMAP$, see Lucantoni [10].

Let $\eta(t)$ and $\xi(t)$ be the numbers of customers in orbit and service facility, respectively at time t and let $J(t)$ be the phase of arrival process at time t. Then the stochastic process

$\boldsymbol{X} = \{X(t), t \geq 0\}$ with $X(t) = (\eta(t), \xi(t), J(t))$ is a continuous time Markov chain on the state space $\mathcal{S} = \{(k, i, j),\ k \geq 0, 1 \leq i \leq K, 1 \leq j \leq m\}$.

We consider two retrial queueing systems $\Sigma^{(i)}$ with retrial rate $\gamma_n^{(i)}$, $i = 1, 2$, when there are n customers in the orbit and the other features are the same as those of the model described above. Let $\eta^{(i)}(t)$ and $\xi^{(i)}(t)$ be the numbers of customers in the orbit and the service facility in the system $\Sigma^{(i)}$, $i = 1, 2$ at time t, respectively and denote $\boldsymbol{X}^{(i)} = \{X^{(i)}(t) = (\eta^{(i)}(t), \xi^{(i)}(t), J(t)),\ t \geq 0\}$ on the state space $\mathcal{S}^*$. We assume that $\boldsymbol{X}^{(i)}$, $i = 1, 2$ are regular Markov chains. Define a relation $\prec$ on $\mathcal{S}$ by $(k, l, i) \prec (k', l', i')$ if and only if $k \leq k'$, $k + l \leq k' + l'$ and $i' = i$. Then it is easily seen that $\prec$ is a partial order on $\mathcal{S}$.

Proposition 3.1 *Assume that* $\gamma_n^{(1)} \geq \gamma_n^{(2)}$ *and* $\gamma_n^{(1)} \leq \gamma_{n+1}^{(1)}$, $\gamma_n^{(2)} \leq \gamma_{n+1}^{(2)}$ *for all* $n = 0, 1, 2, \ldots$. *Then we can construct two stochastic processes* $\boldsymbol{X}^{(i)}$, $i = 1, 2$ *on the same probability space whose sample paths satisfy the relation* $X^{(1)}(t) \prec X^{(2)}(t)$ *for all* $t \geq 0$, *that is,* $\eta^{(1)}(t) \leq \eta^{(2)}(t)$ *and* $\eta^{(1)}(t) + \xi^{(1)}(t) \leq \eta^{(2)}(t) + \xi^{(2)}(t)$.

For each positive integer N, define a retrial queueing system Σ_N and $\hat{\Sigma}_N$ which are the same as the system described above except that when there are $k \geq 1$ customers in the orbit, the retrial rate $\gamma_{k,N} = \alpha + \min(k, N)\beta$ in Σ_N and the retrial rate in $\hat{\Sigma}_N$ is

$$\hat{\gamma}_{k,N} = \begin{cases} \alpha + k\beta, & 1 \leq k \leq N \\ \infty, & k \geq N + 1. \end{cases}$$

Let $\boldsymbol{X}^N = \{X^N(t),\ t \geq 0\}$ and $\hat{\boldsymbol{X}}^N = \{\hat{X}^N(t),\ t \geq 0\}$ denote the Markov chains that describe the system states of Σ_N and $\hat{\Sigma}_N$, repectively on the state space $\mathcal{S}$.

Corollary 3.1 *For all positive integers M and N, there exist the versions of* $\boldsymbol{X}^N$ *and* $\hat{\boldsymbol{X}}^N$ *on the common probability space satisfying the following relation with probability 1,*

$$\hat{X}^N(t) \prec \hat{X}^{N+1}(t) \prec X(t) \equiv X^{\infty}(t) \prec X^{M+1}(t) \prec X^M(t),\ t \geq 0.$$

Corollary 3.2 *If* $\boldsymbol{X}^N$ *is ergodic for some positive integer N, then* $\boldsymbol{X}^M$, $M \geq N$ *is ergodic and the set of stationary distributions* $\{\boldsymbol{\pi}_n, n \geq N\}$ *of* $\{\boldsymbol{X}^n, n \geq N\}$ *is tight. Furthermore,* $\lim_{n\to\infty} \|\boldsymbol{\pi}_n - \boldsymbol{\pi}\| = 0$, *where* $\boldsymbol{\pi}$ *is the stationary distribution of* $\boldsymbol{X}$.

3.2. $MAP_1, MAP_2/M/c$ **retrial queue with guard channel [6].** Let $\tilde{\Sigma}$ be the MAP_1, $MAP_2/M/c$ retrial queue with two types of customers, say type 1 and type 2, in which there are c identical servers and $b - c$ waiting positions in the service facility. For the type 1 customers, $c_h \geq 0$ servers are reserved. The service times of each type of customers are independent of each other and have a common exponential distribution with parameter μ.

Type 1 customer who finds an idle server upon its arrival receives service immediately. When a type 1 customer finds that all the servers are busy and a waiting position is available upon its arrival, the customer occupies the waiting position. If all waiting positions are occupied upon arrival of the type 1 customer, then it is lost. If the sojourn of type 1 customer in the waiting position is greater than the exponential time with parameter ν_1, then it leave the system without being served.

On the arrival of type 2 customer, if the number of servers occupied is more than $c - c_h$, then it either leaves the system forever with probability $1 - p$ or leaves the system temporarily with probability p to retry for service after random amount of time. The retrial is independent and identically repeated and the retrial times are exponentially distributed with parameter γ. On the retrial, if retrial customer finds that the number of servers more than $c - c_h$ are still occupied, then it returns to the retrial group as a retrial customers with probability q or it leaves the system forever with probability $1 - q$. The capacity of retrial group is infinite.

The arrival process of type i customers is assumed to be a Markovian arrival process (MAP) with representation (C_i, D_i), $i = 1, 2$. Let $J_i = \{J_i(t), t \geq 0\}$ be the underlying Markov chain of MAP on the state space $\{1, 2, \ldots, m_i\}$ with generator $C_i + D_i$, $i = 1, 2$ and $J(t) = (J_1(t), J_2(t))$. Let $\mathbf{X} = \{X(t), t \geq 0\}$ with $X(t) = (\eta(t), \xi(t), J(t))$, where $\eta(t)$ and $\xi(t)$ denote the numbers of customers in orbit and service facility, respectively at time t. Let $\tilde{\Sigma}_N$ be the retrial queue described above except that the retrial rate is $\gamma_k = \min(N, k)\gamma$ when there are k customers in the orbit. Let $\mathbf{X}^N = \{X^N(t), t \geq 0\}$ with $X^N(t) = (\eta^N(t), \xi^N(t), J^N(t))$ be the corresponding process to the system $\tilde{\Sigma}_N$.

Proposition 3.2 *For each pair of integers (N, M) with $M > N$, we can construct two stochastic processes $\mathbf{X}^M$ and $\mathbf{X}^N$ on the same probability space whose sample paths satisfy the relations $X^M(t) \prec X^N(t)$ for all $t \geq 0$.*

3.3. $M/M/c/c$ **retrial queue with negative arrivals [1].** Let Σ be the $M/M/c/c$ retrial queue with positive and negative customers which arrive to the system according to independent Poisson processes with rates λ and δ, respectively. Let μ be the service rate of each server, respectively. The service facility consists of c identical servers without any waiting places, so an arriving customer who finds all servers busy is blocked and joins the retrial group, called orbit. The time intervals describing the repeated attempts are assumed to be independent and exponentially distributed with rate $\gamma_n = \alpha + n\beta$, when there are $n \geq 1$ customers in the orbit and $\gamma_0 = 0$. A negative arrival has the effect of removing a random batch of customers from the retrial group. Let p_k be the probability of deleting k customers when a negative arrival occurs. Negative customer only act when all servers are busy. The input flows of positive and negative arrivals, intervals between repeated attempts and service times are mutually independent. Let $\eta(t)$ and $\xi(t)$ be the number of customers in orbit and service facility, respectively at time t. Then $\mathbf{X} = \{X(t), t \geq 0\}$ with $X(t) = (\eta(t), \xi(t))$ is a continuous time Markov chain on the state space $\mathcal{S} = \{(k, i),\ k \geq 0,\ 0 \leq i \leq c\}$.

Let Σ_N be the $M/M/c/c$ retrial queue with positive and negative customers described above except that the retrial rate is $\gamma_{N,n} = \alpha + \min(N, n)\beta$, when there are $n \geq 1$ customers in the orbit and $\gamma_{N,0} = 0$. Let $\mathbf{X}^N = \{X^N(t), t \geq 0\}$ with $X^N(t) = (\eta^N(t), \xi^N(t))$ be the corresponding process to the system Σ_N.

Proposition 3.3 *Let N be a fixed positive integer. For each $M > N$, we can construct two stochastic processes $\mathbf{X}^M$ and $\mathbf{X}^N$ on the same probability space whose sample paths satisfy the relations: for all $t \geq 0$, (1) $\eta^M(t) \leq \eta^N(t) + N$, (2) $\eta^M(t) + \xi^M(t) \leq \eta^N(t) + \xi^N(t) + N$, and (3) if $\eta^M(t) \geq N$, then $\xi^M(t) \geq \xi^N(t)$.*

Acknowledgements: This research was supported by University IT Research Center Project.

REFERENCES

1. *Anisimov V. V., Artalejo J. R.* Analysis of Markov multiserver retrial queues with negative arrivals // Queueing Systems. 2001. V. 39. P. 157–182.
2. *Anisimov V. V., Artalejo J. R.* Approximation of multiserver retrial queues by means of generalized truncation models // Top. 2002. V. 10. P. 51–66.
3. *Artalejo J. R.* A classified bibliography of research on retrial queues: Progress in 1990–1999 // Top. 1999. V. 7. № 2. P. 187–211.
4. *Bhaskaran B. G.* Almost sure comparison of birth and death processes with application to $M/M/s$ queueing systems // Queueing Systems. 1986. V. 1. P. 103–127.
5. *Billingsley P.* Probability and Measure. New York: John Wiley & Sons, 1995.
6. *Choi B. D., Chang Y., Kim H.* $MAP_1, MAP_2/M/c$ retrial queue with guard channels and its applications to cellular networks // Top. 1999. V. 7. P. 321–248.
7. *Falin G. I., Templeton J. G. C.* Retrial Queues. London: Chapman & Hall, 1997.
8. *D. P. Heyman.* Approximating the stationary distribution of an infinite stochastic matrix // J. of Applied Probability. 1991. V. 28. P. 96–103.
9. *Kamae T., Krengel U., O'Brien G. L.* Stochastic inequalities on partially ordered spaces // The Annals of Probability. 1977. V. 5. № 6. P. 899–912.
10. *Lucantoni D. M.* New results on the single server queue with a batch Markovian arrival process // Stochastic Models. 1991. V. 9. P. 1–46.
11. *Neuts M. F., Rao B. M.* Numerical investigation of a multiserver retrial model // Queueing Systems. 1990. V. 7. P. 169–190.
12. *Seneta E.* Computing the stationary distribution for infinite Markov chains // Linear Algebra and Its Applications. 1980. V. 34. P. 259–267.
13. *Seneta E.* Non-negative Matrices and Markov Chains. New York: Springer-Verlag, 1980.
14. *Shin Y. W., Kim Y. C.* Stochastic comparisons of Markovian retrial queues // J. Korean Statistical Society. 2000. V. 29. P. 473–488.
15. *Shin Y. W., Pearce C. E. M.* An algorithmic approach to the Markov chain with transition probability matrix of upper block-Hessenberg form // Korean J. of Computational and Applied Mathematics. 1998. V. 5. № 2. P. 403–426.
16. *Shin Y. W.* Approximations for Markov chains with upper Hesenberg transition matrices // J. of the Operations Research Society of Japan. 2001. V. 44. № 1. P. 90–98.
17. *Zhao Y. Q., Liu D.* The censored Markov chain and the best augmentation // J. of Applied Probability. 1996. V. 33. P. 623–629.