The Censored Markov Chain and the Best Augmentation

Y. Quennel ZHAO

Department of Mathematics and Statistics

University of Winnipeg

Winnipeg, Manitoba

CANADA R3B 2E9

Danielle Liu[*]

Department of Systems Engineering

Case Western Reserve University

Cleveland, Ohio 44106

U.S.A.

September 2, 2004

Abstract: Computationally, when we solve for the stationary probabilities for a countable-state Markov chain, the transition probability matrix of the Markov chain has to be truncated, in some way, into a finite matrix. Different augmentation methods might be valid such that the stationary probability distribution for the truncated Markov chain approaches that for the countable Markov chain as the truncation size gets large. In this paper, we prove that the censored (watched) Markov chain provides the best approximation in the sense that for a given truncation size the sum of errors is the minimum and show, by examples, that the method of augmenting the last column only is not always the best.

[*]Current address: AT & T Bell Laboratories, Holmdel, NJ 07733-3030, U.S.A.

1 Introduction

Approximating a countable-state Markov chain by using finite-state Markov chains is an interesting and often a challenging topic, which has attracted many researchers' attention. Computationally, when we solve for the stationary distribution, when it exists, of a countable-state Markov chain, the transition probability matrix of the Markov chain has to be truncated in some way into a finite matrix as a first step. We then compute the stationary distribution of this finite-state Markov chain as an approximation to that of the countable-state one. We expect that as the truncation level (or size) increases to infinity, the solution for the finite Markov chain would converge to that of the countable-state Markov chain. While for many application problems the justification of the convergence could be made by the physical meanings of the finite- and the countable-state Markov chains, it is not always easy to formally justify this claim.

The study of approximating the stationary probabilities of an infinite Markov chain by using finite Markov chains was initiated by Seneta [12] in 1967. Many up-to-date results were obtained by him and several collaborators. Most of their results are included in a paper by Gibson and Seneta [3]. Other references could be found therein and/or in another paper [4] published in the same year by the same authors. Other researchers, including Wolf [14], used some different approaches from that of Seneta *et al.*. For instance, Heyman provided a probabilistic treatment of the problem in [7]. Later on, Grassmann and Heyman [6] justified the convergence for infinite-state Markov chains with repeating rows. All the above results are for approximating stationary distributions. Regarding more general issues of approximating a countable-state Markov chain, see the book by Freedman [2].

In this paper we are interested in only approximating stationary distributions of infinite Markov chains. Let P be the transition probability matrix of an infinite Markov chain which has a unique stationary distribution, and let P_{nw} be the northwest corner of P. Notice that P_{nw} is a sub-stochastic matrix. The procedure to make P_{nw} stochastic by adding appropriate values to its entries is called augmentation. Whether or not the stationary probabilities of the augmented Markov chain converge to that of the original Markov chain as the size of the northwest corner matrix goes to infinity is not the central topic in this paper. We are interested in determining which augmentation method provides the best approximation when the convergence has been established. Numerical

evidence often shows that the method to augment the last column only often provides the best approximation. Gibson and Seneta seem to believe such a conclusion after numerically comparing five different augmentation methods for the imbedded Markov chain at the service completion epochs of the $M/M/1$ queue. Based on available analytic results, it is unknown which augmentation method is the best in the sense that the sum of absolute errors is the minimum. Also, no comparison has been made between the method by watching the Markov chain only when it is in a state belonging to the northwest corner and other augmentation methods. Our main contributions here are: a) to analytically show that for any given truncation level the censored Markov chain gives the best approximation in the sense that the sum of the absolute errors is the minimum; and b) to show, by examples, that the method of augmenting the last column only is not always the best.

2   The censored Markov chain

Censored Markov chains, also called watched Markov chains, were first considered by Lévy [9, 10, 11]. Since then, censored Markov chains have been found very useful in different aspects of study of Markov chains. The censoring operation or technique was used by Kemeny, Snell and Knapp [8] to prove that every recurrent Markov chain has a positive-valued regular measure unique up to multiplication by a scalar. It was also introduced as a new technique to do approximations by Williams [13]. Recently, Grassmann and Heyman [5, 6] used this technique to deal with the block elimination for transition probability matrices of infinite Markov chains. Freedman used this operation to approximate countable Markov chains for the limiting behavior and also for more general issues. Results were presented in his book [2]. In this section, after giving the definition, we provide some basic properties of censored Markov chains. They are used to show that the censoring is the best method to approximate an infinite Markov chain in the sense that the error sum of the stationary probabilities is the minimum for any given truncation level. In general, there is no easy way to compute the transition probability matrix of the censored Markov chain from that of the original Markov chain, as commented by Freedman (page 20 in [2]).

Since we are only interested in approximating the stationary probabilities of Markov chains, we will not distinguish between the Markov chain itself and its transition prob-

ability matrix. Let P be a countable-state Markov chain with the state space $S = \{0, 1, 2, \ldots\ldots\}$. Let E be a subset of S. Let P^E be the stochastic process whose nth transition is the nth time of the Markov chain P being in the set E. In other words, the sample paths of process P^E are obtained from the sample paths of P by omitting all parts in E^c, where E^c is the complement of E. Therefore, P^E is the process obtained by watching P only when in E, or by censoring P from E^c. A rigorous definition of the censored process can be found on page 14 of [2]. The following lemma is essentially Lemma 6-6 in [8]. The result in Lemma 2 is also stated in Grassmann and Heyman [5].

Lemma 1 *Let P be the transition probability matrix of an arbitrary Markov chain partitioned according to subsets E and E^c:*

$$P = \begin{array}{c} \\ E \\ E^c \end{array} \begin{array}{c} E \quad\quad E^c \\ \left[\begin{array}{cc} T & U \\ D & Q \end{array} \right] \end{array}. \tag{1}$$

Then, the censored process is a Markov chain and its transition probability matrix is given by

$$P^E = T + U\hat{Q}D \tag{2}$$

with $\hat{Q} = \sum_{k=0}^{\infty} Q^k$.

Lemma 2 *If the Markov chain P is irreducible, then so is the censored Markov chain. If P is ergodic with stationary probabilities $\{\pi_k\}$, then the stationary probabilities $\{\pi_k^E\}$ of the censored Markov chain are given by*

$$\pi_k^E = \frac{\pi_k}{\sum_{i \in E} \pi_i}, \quad k \in E. \tag{3}$$

Equation (2) usually gives us little help in computing P^E from P. Only for some special cases can P^E be explicitly determined. This means that only for some simple cases can $U\hat{Q}D$ be explicitly determined. One of such cases is given in Example 1. The entries of $U\hat{Q}D$ are taboo probabilities with the taboo set E (see I.9 of Chung [1]), i.e., the (i,j)th entry of $U\hat{Q}D$ is the probability, starting from the state i, of ever going to state j under the restriction that none of the states in the taboo set E is visited in-between.

Example 1: Let $E = \{0, 1, \ldots, K\}$. If $P = (p_{ij})_{i,j \in S}$ is an upper Hessenberg matrix ($p_{ij} = 0$ whenever $i > j + 1$), then the censoring operation is the same as that of the last

4

column augmented only. This case includes the imbedded Markov chain of the $M/G/1$ queue as a special case. This result is also asserted by Gibson and Seneta (the last sentence of section 5) [3].

3 The best augmentation

While augmenting the last column might be the best operation for many cases, it is not difficult to construct examples where the augmentation of the last column only is <u>not</u> the best. By using the results in Section 2, we show that the censored Markov chain does best in the sense that it gives the smallest truncation error.

In the rest of the paper, for the censored Markov chain we always use $E = \{0, 1, 2, \ldots, K\}$. Let us consider an ergodic countable-state Markov chain P and let $P^{(K)}$ be the finite Markov chain with the states in $\{0, 1, 2, \ldots K\}$, which is obtained by some method of augmenting the northwest corner P_{nw} of the transition probability matrix P. Denote the stationary probabilities for P and $P^{(K)}$ by π_k and $\pi_k^{(K)}$ respectively. Define the l_1 norm of the errors of stationary probabilities between the countable Markov chain and the finite Markov chain by

$$l_1(K, \infty) = \sum_{k=0}^{K} |\pi_k^{(K)} - \pi_k| + \sum_{k=K+1}^{\infty} \pi_k. \tag{4}$$

The following result will be used often in the comparison of different augmentation methods later in the paper.

Theorem 1 *The error sum of the stationary probabilities between the countable Markov chain and the augmented Markov chain is given by*

$$l_1(K, \infty) = 2\left(1 - \sum_{k=0}^{K} \pi_k\right) + 2 \sum_{\substack{k=0 \\ \pi_k > \pi_k^{(K)}}}^{K} \left(\pi_k - \pi_k^{(K)}\right) \tag{5}$$

$$= 2 \sum_{\substack{k=0 \\ \pi_k^{(K)} > \pi_k}}^{K} \left(\pi_k^{(K)} - \pi_k\right). \tag{6}$$

The proof is straight forward. Now we immediately have some corollaries.

Corollary 1 *For a given truncation level K, the l_1 norm $l_1(K, \infty)$ of the errors is the same and the minimum for all augmentation methods such that $\pi_k^{(K)} \geq \pi_k$ for all $k =$*

$0, 1, \ldots, K$. *In this case, the minimum error sum is*

$$l_1(K, \infty) = 2 \left(1 - \sum_{k=0}^{K} \pi_k \right). \tag{7}$$

The proof is directly from (5) or (6).

Corollary 2 *The censoring is an augmentation method such that the error sum $l_1(K, \infty)$ is the minimum.*

The proof is from Lemma 2 and Corollary 2.

Corollary 3 *For an ergodic Markov chain whose transition probability matrix is upper Hessenberg, the method of only augmenting the last column kept in the northwest corner is such that the error sum $l_1(K, \infty)$ is the minimum.*

In fact, in this case, the censoring operation and augmenting the last column only result the same transition probability matrix $P^{(K)}$ as we have already remarked in Example 1 in Section 2.

In Section 5 of [3], the imbedded Markov chain at the service completion epochs for the $M/M/1$ queue was used for studying which augmentation method is the best. Based on numerical comparisons among the error sums for five different augmentation methods, Gibson and Seneta concluded that for a fixed transition probability matrix P and a fixed truncation level K, the method of augmenting the last column only provides the best approximation to the stationary probabilities π_k of the countable Markov chain in the sense that the l_1 norm is the minimum. Even though the conclusion is true for this specific example (a special case of Corollary 3), it is not the case in general. In fact, it is not so difficult to provide an explicit expression for estimating the error sum for this special example. Gibson and Seneta seemed to believe that this conclusion should also be true in more general cases and remarked that further plausible arguments as to why the method to augment the last column only should be the "best" augmentation may be provided on account of P being stochastically monotone. Some arguments were provided in another paper [4] of theirs, but it does not lead to the conclusion that the last column augmented is the best approximation in the sense of l_1 norm. What they proved is that

if P is stochastically monotone, then for any $j = 0, 1, \ldots K$,

$$0 \leq \sum_{k=0}^{j} \left(\nu_k^{(K)} - \pi_k \right) \leq \sum_{k=0}^{j} \left(\pi_k^{(K)} - \pi_k \right),$$

where π_k, $\nu_k^{(K)}$ and $\pi_k^{(K)}$ are stationary probabilities of the original Markov chain, the finite Markov chain by augmenting the last column only, and the finite Markov chain obtained by any augmentation method, respectively. This result only tells us that the stationary probability vector of the finite Markov chain by augmenting the last column only is the minimum in the stochastic monotone sense, or that any partial sum of the errors of the stationary probabilities is the minimum for the last column augmentation method. We provide an example (Example 4) in the following to show that the method of augmenting the last column only is not the best in the sense of l_1 norm even when stochastic monotonicity is satisfied.

Example 2: In this example, we show that a) augmenting the last column is not always the best and b) augmenting the first column is not always the worst either. Consider the imbedded Markov chain of the $D/M/1$ queue with unit interarrival times and service rate μ. In this case, the traffic intensity is $\rho = 1/\mu$. The stationary probabilities π_k for the infinite-state Markov chain depend on only σ, the unique solution of $x = e^{-\mu(1-x)}$ inside the unit circle. There is no explicit expression for the stationary probabilities $\pi_k^{(K)}$ for the finite Markov chain by augmenting the last column only. In order to provide correct numerical values, we used different ways to compute both π_k and $\pi_k^{(K)}$. When $K = 2$, for various values of μ tested: $1.25 \leq \mu \leq 2.32$, we always have $\pi_1^{(2)} < \pi_1$, which means that the method of augmenting the last column only is not the best method in the sense of the l_1 norm. In fact, when $K \geq 3$, we have not found any value of μ such that the l_1 norm of the errors is the minimum for the method of augmenting the last column only. Notice that the transition probability matrix of the $D/M/1$ queue is stochastically monotone, therefore the method of augmenting the last column is the best method in the sense of stochastic monotonicity. The $D/M/1$ queue can also be served as an example to show that the method of augmenting the first column only is not always the worst. For example, when μ is not so large, say $\mu \leq 2$, the method to augmenting the first column only is better than that of uniformly augmenting the last row of P_{nw}.

Another reasonable criterion for comparing different augmentation methods is the l_∞

norm defined by

$$l_\infty(K, \infty) = \max \left\{ \max_{0 \le k \le K} |\pi_k^{(K)} - \pi_k|, \max_{k \ge K+1} \pi_k \right\}. \tag{8}$$

According to the definition, the l_∞ norm only compares the maximal error of the stationary probabilities. When the l_1 norm is the same for two different augmenting methods it makes much sense to further compare their l_∞ norms in order to decide which one is better. We give the following example to show that the method of augmenting the last column only is not always the best either.

Example 3: Consider the imbedded Markov chain at the arrival epochs for the $M/M/1$ queue with the arrival rate λ and the service rate μ. Without loss of generality, assume $\lambda + \mu = 1$. It is not difficult to see that the stationary probabilities for the censored Markov chain are given by

$$\pi_k^{(K)} = \frac{(1 - \rho)\rho^k}{\sum_{i=0}^{K} \pi_i} = \frac{\rho^k(1 - \rho)}{1 - \rho^{K+1}}, \quad k = 0, 1, \dots K \tag{9}$$

with $\rho = \lambda/\mu$. Also, it is not difficult to solve the stationary equations directly to give the following expressions of the stationary probabilities $\nu_k^{(K)}$ for the augmented Markov chain by augmenting the last column only.

$$\begin{aligned}
\nu_0^{(K)} &= \frac{1 - \rho}{1 - \rho^{K+2}} \\
\nu_k^{(K)} &= \nu_0^{(K)} \rho^k, \quad k = 1, 2, \dots, K - 1 \\
\nu_K^{(K)} &= \frac{1}{\mu} \nu_0^{(K)} \rho^K.
\end{aligned}$$

Notice that $\nu_k^{(K)} > \pi_k$ for all $k = 0, 1, \dots, K$. So, the l_1 norms for the censored Markov chain and the Markov chain by augmenting the last column only are the same and given by

$$l_1(K, \infty) = 2 \left(1 - \sum_{k=0}^{K} \pi_k \right) = 2\rho^{K+1}. \tag{10}$$

For the censored Markov chain,

$$l_\infty(K, \infty) = \pi_0^{(K)} - \pi_0 = \frac{\rho^{K+1}(1 - \rho)}{1 - \rho^{K+1}}$$

and for the Markov chain by augmenting the last column only,

$$l_\infty(K, \infty) = \nu_K^{(K)} - \pi_K = \frac{\rho^{K+1}(1 - \rho)(1 + \rho^{K+1})}{1 - \rho^{K+2}}.$$

The comparison of these two maximal error tells us that for this example the maximal error of the stationary probabilities for the censored Markov chain is smaller than that for the Markov chain by augmenting the last column only. Specifically,

$$\frac{l_\infty \text{ for censored}}{l_\infty \text{ for last column augmented}} = \frac{1 - \rho^{K+2}}{1 - \rho^{2K+2}} < 1.$$

As the final remark, we emphasize that even though the censoring operation provides the minimum l_1 truncation error, the censoring augmentation or the transition matrix of the censored Markov chain is usually difficult to compute. Augmenting the last column only is always applicable.

References

[1] Chung, K.L. (1967) *Markov Chains with Stationary Transition Probabilities*, 2nd edn, Springer-Verlag, Berlin.

[2] Freedman, David (1983) *Approximating Countable Markov Chains*, 2nd edn, Springer-Verlag, New York.

[3] Gibson, D. and Seneta, E. (1987a) Augmented truncations of infinite stochastic matrices. *J. Appl. Prob.* **24**, 600–608.

[4] Gibson, D. and Seneta, E. (1987b) Monotone infinite stochastic matrices and their augmented truncations. *Stoch. Proc. Appl.* **24**, 287–292.

[5] Grassmann, W.K. and Heyman, D.P. (1990) Equilibrium distribution of block-structured Markov chains with repeating rows. *J. Appl. Prob.* **27**, 557–576.

[6] Grassmann, W.K. and Heyman, D.P. (1993) Computation of steady-state probabilities for infinite-state Markov chains with repeating rows. *ORSA J. on Computing* **5**, 292–303.

[7] Heyman, D.P. (1991) Approximating the stationary distribution of an infinite stochastic matrix. *J. Appl. Prob.* **28**, 96–103.

[8] Kemeny, J.G., Snell, J.L. and Knapp, A.W. (1976) *Denumerable Markov Chains*, 2nd edn, Springer-Verlag, New York.

[9] Lévy, Paul (1951) Systèmes markoviens et stationnaires. Cas dénombrable. *Ann. Sci. École Norm. Sup.* **68** (3), 327–381.

[10] Lévy, Paul (1952) Complément à l'étude des processus de Markoff. *Ann. Sci. École Norm. Sup.* **69** (3), 203–212.

[11] Lévy, Paul (1958) Processus markoviens et stationnaires. Cas dénombrable. *Ann. Inst. H. Poincaré* **18**, 7–25.

[12] Seneta, E. (1967) Finite approximation to infinite non-negative matrices. *Proc. Camg. Phil. Soc.* **63**, 983–992.

[13] Williams, D. (1966) A new method of approximation in Markov chain theory and its application to some problems in the theory of random time substitution. *Proc. Lond. math. Soc.* **16** (3), 213–240.

[14] Wolf, D. (1980) Approximation of the invariant probability measure of an infinite stochastic matrix. *Adv. Appl. Prob.* **12**, 710–726.