On convergence and error bounds for augmented truncation approximations of Markov chains

Yuanyuan Liu (刘源远)

Central South University, Changsha

Based on joint research with Peter Glynn, Alex Infanger and Na Lin

The 17th Workshop on Markov Processes and Related Topics, November 26, 2022

Outline

Consider a discrete-time Markov chain $X = \{X_n : n \geq 0\}$ on a complete separable metric space E with transition kernel $P = (P(x, dy) : x, y \in E)$. We assume that X has a unique stationary distribution $\pi = (\pi(dx) : x \in E)$.

Consider a sequence of truncated sets $A_n \subset E$, such that A_n tends to E. Define P_n to be the truncated and augmented chain, i.e. $P_n \geq P_{A_n}$ and P_n is stochastic. Let Π_n be the set of all the invariant distributions ${}_{(n)}\pi$ of P_n. Suppose that Π_n is not empty.

Question: ${}_{(n)}\pi \to \pi$?

We now provide an example showing that even when X is very well-behaved, ${}_{(n)}\pi$ may fail to converge to π as $n \to \infty$.

Example 1. Suppose that $E = \mathbb{Z}_+$ with $P(2i, 2i+1) = 1/2 = P(2i, 0)$ for $i \in \mathbb{Z}_+$, and $P(2i+1, 2i+2) = 1$ for $i \in \mathbb{Z}_+$. Then,

$$P^2(2i, 0) \geq P(2i, 0)P(0, 0) = \frac{1}{4}$$

for $i \in \mathbb{Z}_+$ whereas,

$$P^2(2i+1, 0) \geq P(2i+1, 2i+2)P(2i+2, 0) = \frac{1}{2},$$

for $i \in \mathbb{Z}_+$. Hence $P^2(x, 0) \geq 1/4$ for $x \in \mathbb{Z}_+$, so that the two-step transition matrix is a Markov matrix, and X is uniformly ergodic.

Suppose we use last state augmentation. Then, when $A_n = \{0, 1, \ldots, n\}$ with n odd, state n is absorbing, and the single closed communicating class corresponding to ${}_{(n)}P$ is just $\{n\}$. It follows that

$$ {}_{(n)}\pi(x) = \begin{cases} 1, & \text{if } x = n, \\ 0, & \text{otherwise}, \end{cases} \qquad \text{for } n \text{ odd}, $$

so that ${}_{(n)}\pi$ fails to converge to the stationary distribution π of X, despite the fact that X is uniformly ergodic.

For augmented truncation approximations, one often wants to know

▷ whether the equilibrium behavior of the truncated and augmented chain converges to that of the untruncated system;

▷ the quantitative bounds on the difference between them when the stability is robust.

For augmented truncation approximations, one often wants to know

▷ whether the equilibrium behavior of the truncated and augmented chain converges to that of the untruncated system;

▷ the quantitative bounds on the difference between them when the stability is robust.

Let f be a finite-valued positive function on E. For a finite measure μ, define its f-norm by

$$\|\mu\|_f := \sup_{|g|\le f} |\mu(g)|.$$

For $f \equiv 1$, the f-norm coincides with the total variation norm which will be denoted by $\|\cdot\|_1$.

Related literature

For convergence:

▷ Seneta (1980): convergence holds iff $\{_{(n)}\pi, n \geq 1\}$ is tight;

▷ Liu and Zhao (1995): the censored MC is the best.

▷ Tweedie (1998): addressed the two issues well for the first and the last column augmentation when P is a Markov matrix or P is geometrically ergodic and monotone.

▷ Liu (2010): investigated an arbitrary augmentation and truncation bounds for polynomially ergodic and monotone MCs.

▷ Hart and Tweedie (2012): studied the convergence for CT Markov processes.

▷ $\cdots \cdots$

For error bounds:

▷ Tweedie (1998): f-norm bounds for monotone and geometrically ergodic MCs by Ergodicity Method.

▷ Herve L. and Ledoux J. (2014, 2022): general Mcs EM.

▷ Masuyama (2016): f-norm bounds for the LCBA truncation under drift conditions for matrix-analytic models Perturbation Method.

▷ Liu and Li (2018): computable f-norm bounds via the Poisson equation, the residual matrix, and the norm ergodicity coefficient. PM

▷ Kuntz et al. (2021) (Siam Review): convergence and error bounds for stochastic reaction networks.

▷ $\cdots$ $\cdots$

In this talk, we study general truncation-augmentation schemes of Markov chains on "general" state spaces.

▷ We give some conditions under which one can be assured that arbitrarily augmented truncation approximation is convergent.

▷ We derive the upper and lower bounds of a solution to the Poisson equation, based on which the error bounds of the truncation schemes is further obtained.

Outline

A function $V : E \to \mathbb{R}_+$ is said to be coercive if the sublevel set

$$B_n := \{x \in E : V(x) \leq n\}$$

is either empty or compact in E for each $n \geq 1$. Let P_n be defined on the truncated subset B_n.

Consider the following assumptions.

A1. There exists a coercive function f, a non-negative function $V : E \to \mathbb{R}_+$, and $b < \infty$ for which for $x \in E$

$$\int_E P(x, dy)V(y) \leq V(x) - f(x) + b$$

A2. P is weakly continuous, i.e., P maps a bounded continuous function into a bounded continuous function.

Theorem 1. (IGL 2022) Suppose A1 and A2 hold with V continuous. If ${}_{(n)}\pi \in {}_{(n)}\Pi$, then

$$ {}_{(n)}\pi \Rightarrow \pi $$

as $n \to \infty$, where $\Rightarrow$ denotes weak convergence in E.

Sketch of proof:

▷ Show the tightness of the sequence $({}_{(n)}\pi, n \geq 1)$ (A1). Then there exists a probability π' on E such that

$$ {}_{(n'_k)}\pi \Rightarrow \pi', \quad \text{as } k \to \infty. $$

▷ Prove that $\pi'(dy) = \int_E \pi'(dx) P(x, dy)$ (A2).

X is called strongly uniformly recurrent if there exists $\lambda > 0$ and a probability ϕ such that for $x, y \in E$

$$P(x, dy) \geq \lambda \phi(dy).$$

Theorem 2. (IGL 2022) Suppose that X is strongly uniformly recurrent under P with a unique stationary distribution π. Then, X is strongly uniformly recurrent under ${}_{(n)}P$ with a unique stationary distribution ${}_{(n)}\pi$, and

$$\|{}_{(n)}\pi - \pi\|_1 \to 0, \quad \text{as } n \to \infty.$$

Outline

Perturbation formula

Let $\psi_{x,n}(\cdot)$ be a probability distribution on A_n, which depends on n and x. The generally augmented truncation of P to the set A_n, denoted by ${}_{(n)}P$, is defined as

$$
{}_{(n)}P(x, dy) := \begin{cases} P(x, dy) + \psi_{x,n}(dy)P(x, A_n^c), & \text{if } x \in E,\ y \in A_n, \\ 0, & \text{if } x \in E,\ y \in A_n^c\ . \end{cases}
$$

In other words, whenever P would make a transition from x to A_n^c, ${}_{(n)}P$ transitions to $y \in A_n$ with probability $\psi_{x,n}(dy)$; otherwise, ${}_{(n)}P$ is the same as P.

Perturbation formula: $(\pi - \hat{\pi})g = \hat{\pi}(P - \hat{P})\tilde{g}$.

Upper bounds via the Poisson equation

The Poisson equation is known as

$$\tilde{g} - P\tilde{g} = \bar{g} \tag{1}$$

where $\bar{g} = g - (\pi g)\mathbf{e}$.

The function g is called the forcing function and is assumed to satisfy $\pi g < \infty$. The function $\tilde{g}$ satisfying (1) is called the solution to the Poisson equation, which is not unique (not even unique up to additive constants).

A3. There exists a positive constant λ, probability distributions φ and $a = (a(n), n \in \mathbb{Z}_+)$ such that

$$\sum_{n=0}^{\infty} a(n) P^n(x, \cdot) \geq \lambda \varphi(\cdot), \quad x \in C.$$

A4. There exists a positive constant $b < \infty$, a set C and finite functions $V \geq 1$ and $f \geq 1$ such that

$$PV(x) \leq V(x) - f(x) + b \cdot \mathbf{1}_{\{C\}}(x), \quad x \in E,$$

where $\mathbf{1}_{\{C\}}(\cdot)$ is the indictor function in the set C.

Lemma 1. (GLL 2022) Suppose that A3 and A4 hold (for the same C). Then for any measurable function g satisfying $0 \le g \le f$, there exists a solution $\tilde{g}$ to Poisson equation (1) such that

$$-b\,(V(x)+bd) \le \tilde{g}(x) \le V(x)+bd, \quad \text{for } x \in E,$$

where $d = \frac{1}{\lambda}\sum_{n=0}^{\infty} na(n)$.

Sketch of proof:

▷ Construct a split chain and obtain the solution of Poisson equation

$$\tilde{g}(x) = E_x \sum_{j=0}^{\tau-1} \bar{g}(X_j).$$

▷ Apply the comparison theorem and properties of the split chain to bound $\tilde{g}(x)$.

Theorem 3. (GLL 2022) Let $N(C) = \min\{n : C \subseteq A_n\}$ and V be a coercive function. Suppose that $\{{}_{(n)}P, n \geq N(C)\}$ uniformly satisfies A3 and A4. Then, for ${}_{(n)}\pi \in {}_{(n)}\Pi$, we have

$$\begin{aligned} \left\| \pi - {}_{(n)}\pi \right\|_f &\leq b[n + d(b+1)]\pi(A_n^c) + \int_{y \in A_n^c} \pi(dy)V(y) \\ &= H_1(n, b, d, V), \end{aligned}$$

where $d = \frac{1}{\lambda}\sum_{n=0}^{\infty} na(n)$. Moreover, if $\pi V < \infty$, then $H_1(n, b, d, V) \to 0$ as $n \to \infty$ and

$$\left\| \pi - {}_{(n)}\pi \right\|_f \to 0, \quad n \to \infty.$$

Upper bounds via the norm ergodicity coefficient

The V-norm ergodicity coefficient is defined by $\Lambda_V(B)$ of a kernel B by

$$\Lambda_V(B) = \sup\left\{\|\mu B\|_V : \|\mu\|_V \leq 1, \mu \boldsymbol{e} = 0\right\}.$$

Theorem 4. (GLL 2022) Let V be a coercive function. Suppose that $\pi V < \infty$, $\beta = \|_{(n)}P\|_V < \infty$ and there exists a positive integer m and a positive constant $\rho < 1$ such that $\Lambda_V\left({}_{(n)}P^m\right) \leq \rho$ for any n. Then there exists a unique ${}_{(n)}\pi$ such that

$$\left\|\pi - {}_{(n)}\pi\right\|_V \leq H_2(n, m, \rho, V), \tag{2}$$

where

$$H_2(n, m, \rho, V) = \begin{cases} \frac{1-\beta^m}{(1-\beta)(1-\rho)} \left(n\pi(A_n^c) + \int_{y \in A_n^c} \pi(dy)V(y) \right), & \text{if } \beta \neq 1, \\ \frac{m}{1-\rho} \left(n\pi(A_n^c) + \int_{y \in A_n^c} \pi(dy)V(y) \right), & \text{if } \beta = 1, \end{cases}$$

and $H_2(n, m, \rho, V) \to 0$ as $n \to \infty$.

Remark. Since $\|_{(n)}P\|_V \leq \|P\|_V$ and $\Lambda_V({}_{(n)}P) \leq \Lambda_V(P)$, $\|P\|_V < \infty$ and $\Lambda_V(P) \leq \rho$ are verifiable sufficient conditions for $\|_{(n)}P\|_V < \infty$ and $\Lambda_V({}_{(n)}P) \leq \rho$ uniformly.

However, it is worth noting that for $m > 1$, we cannot make the condition $\Lambda_V({}_{(n)}P^m) \leq \rho$ satisfied by requiring $\Lambda_V(P^m) \leq \rho$. Considering Example 1, one can easily calculate that $\Lambda_1(P^2) = 3/4 < 1$ while $\Lambda_1({}_{(n)}P^2) = 1$ for $n > 1$ odd.

For $V \equiv 1$ (although not coercive), applying the similar arguments yields the following results.

Corollary 1. Suppose that there exists a positive integer m and a positive constant $\rho < 1$ such that $\Lambda_1\left({}_{(n)}P^m\right) \le \rho$ for any n. Then there exists a unique ${}_{(n)}\pi$ such that

$$\left\|\pi - {}_{(n)}\pi\right\|_1 \le \frac{2m\pi(A_n^c)}{1-\rho}.$$

In particular, if $\Lambda_1\left(P\right) \le \rho$, then the above result holds for $m = 1$.

Considering the uniformly recurrent Markov chains, we derive the following corollary.

Corollary 2. Suppose that there exists a positive integer m, a positive constant $\lambda < 1$ and a probability measure ϕ_m such that

$$ {}_{(n)}P^m(x, A) \geq \lambda \phi_m(A) \tag{3} $$

for any n and for all $x \in E$ and $A \subseteq E$. Then, for ${}_{(n)}\pi \in {}_{(n)}\Pi$, we have

$$ \|\pi - {}_{(n)}\pi\|_1 \leq \frac{2m\pi(A_n^c)}{\lambda}. $$

Note when $m = 1$, X is strongly uniformly recurrent.

Lower bounds in the total variation norm

Proposition 1 (GLL 2022) The total variation error of stationary probabilities between the original Markov chain and the augmented Markov chain is given by

$$\|\pi - {}_{(n)}\pi\|_1 \geq 2\pi(A_n^c).$$

Remark. For uniformly recurrent Markov chains satisfying (3), we have

$$2\pi(A_n^c) \leq \|\pi - {}_{(n)}\pi\|_1 \leq \frac{2m\pi(A_n^c)}{\lambda}.$$

Outline

Example 1. Consider a DTMC $X = \{X_n, n \geq 0\}$ with transition probabilities

$$P(i,0) = \frac{1-2a}{1-a}, \quad P(i,i+k) = a^k, \; k = 1,2,\ldots, \quad \text{for any } i \geq 0,$$

where $0 < a < \frac{1}{2}$. It is obvious that the chain X is strongly uniformly recurrent. By Corollary 2, we immediately have

$$\|\pi - {}_{(n)}\pi\|_1 \leq \frac{2(1-a)}{1-2a} \sum_{i=n+1}^{\infty} \pi(i).$$

For $1 < s < \frac{1}{2a}$, let $V(i) = s^i$, $i \geq 0$. One can easily verify that

$$PV(i) \leq \beta V(i) + b, \quad \text{for } i \geq 0,$$

where $\beta = \frac{as}{1-as} < 1$ and $b = 1 - \frac{a}{1-a}$.

Multiplying both sides of the above drift condition by π yields $\pi V \le \frac{b}{1-\beta}$, from which it follows that

$$\sum_{i=n+1}^{\infty} \pi(i) \le \frac{1}{s^{n+1}} \sum_{i=n+1}^{\infty} \pi(i) s^i \le \frac{b s^{-(n+1)}}{1-\beta}$$

and

$$\|\pi - {}_{(n)}\pi\|_1 \le \frac{2a}{1-2as} s^{-n}.$$

Then for any fixed $\varepsilon > 0$, we can set $N_\varepsilon = -\log_s \frac{(1-2as)\varepsilon}{2a}$ and guarantee that

$$\|\pi - {}_{(n)}\pi\|_1 \le \varepsilon, \quad \text{for } n \ge N_\varepsilon.$$

Observe that the results of Liu and Li (2018) are also valid for this example. However, they rely on the quantity

$$\sum_{i=0}^{n} {}_{(n)}\pi(i) \sum_{j=n+1}^{\infty} P(i,j)(V(n)+V(j)).$$

In order to find the minimum truncation size that controls the error $\|\pi - {}_{(n)}\pi\|_V$ within some fixed precision, applying their results requires calculating the above quantity in terms of every possible n in advance, which is computationally expensive and cumbersome.

Our results are more straightforward and effective in determining the threshold N_ε of the truncation size.

Outline

Discussion for CTMCs

Parallel to the results of DTMCs, we can obtain the similar upper bounds on augmented truncation approximations for CTMCs via the Poisson equation.

For CTMCs with bounded generators, this result works well. However, it may be invalid for those with unbounded generators.

Thank you for your attention!