Bounding large Markov chains using stochastic comparison and censoring techniques

Sana Younès

PRiSM Laboratory, Versailles university
LACL laboratory, Paris-Est university

9^{ième} Atelier d'Évaluation de Performances

Outline

1. Stochastic comparison
2. Censoring techniques
3. Bounding performability measures by censoring techniques
4. Algorithms for bounding censored Markov chains

Stochastic comparison (1)

Let $S = \{1, 2, \cdots, n\}$ be a finite state space.

Definition ($\leq_{st}$ order)

Let p and q two probability distributions
$p \leq_{st} q \text{ iff } \sum_{j=k}^{n} p_j \leq \sum_{j=k}^{n} q_j \quad \forall k = 1, 2, \ldots, n$

Definition ($\leq_{st}$ comparison of two discrete-time Markov chain)

Let $\{X(t),\ t > 0\}$ and $\{Y(t),\ t \geq 0\}$ be two DTMC taking values in S. $\{X(t),\ t \geq 0\}$ is said to be less than $\{Y(t),\ t \geq 0\}$ in the strong stochastic sense, that is,

$$\{X(t),\ t \geq 0\} \leq_{st} \{Y(t),\ t \geq 0\} \text{ iff } X(t) \leq_{st} Y(t) \quad \forall t.$$

Stochastic comparison (1)

Let $S = \{1, 2, \cdots, n\}$ be a finite state space.

Definition ($\leq_{st}$ order)

Let p and q two probability distributions

$p \leq_{st} q \;$ iff $\; \sum_{j=k}^{n} p_j \leq \sum_{j=k}^{n} q_j \;\; \forall k = 1, 2, \ldots, n$

Definition ($\leq_{st}$ comparison of two discrete-time Markov chain)

Let $\{X(t),\ t > 0\}$ and $\{Y(t),\ t \geq 0\}$ be two DTMC taking values in S. $\{X(t),\ t \geq 0\}$ is said to be less than $\{Y(t),\ t \geq 0\}$ in the strong stochastic sense, that is,

$$\{X(t),\ t \geq 0\} \leq_{st} \{Y(t),\ t \geq 0\} \text{ iff } X(t) \leq_{st} Y(t) \;\; \forall t.$$

Stochastic comparison (2)

Definition ($\leq_{st}$ monotonicity)

Let P be a stochastic matrix. P is said to be stochastically st-monotone if for any probability vectors p and q:

$$p \leq_{st} q \implies p\,P \leq_{st} q\,P$$

.

- Let $P[i,*]$ be the row i of the matrix P.
- P is $\leq_{st}$ monotone iff $P[i,*] \leq_{st} P[i+1,*], \forall i \in S$.
- P is not monotone, Q is monotone.

$$P = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.0 & 0.6 & 0.4 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} \quad Q = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$$

Stochastic comparison (3)

$P \leq_{st} Q$ iff $P[i,*] \leq_{st} Q[i,*], \forall i \in S$.

Theorem (Sufficient conditions for DTMC comparison)

Let $\{X(t), t \geq 0\}$ and $\{Y(t), t \geq 0\}$ be two time-homogeneous DTMC and P and Q be their respective probability transition matrices. Then:

$$\{X(t), t > 0\} \leq_{st} \{Y(t), t > 0\}$$

if:

- $X(0) \leq_{st} Y(0)$,
- *st-monotonicity of P or Q*
- *st-comparability of the matrices holds, that is,* $P[i,*] \leq_{st} Q[i,*]\ \forall i$.

Construction of $\leq_{st}$ monotone upper bound

- For a matrix P Vincent's algorithm construct a matrix Q such that
  - $P \leq_{st} Q$
  - Q is $\leq_{st}$ monotone.
- Inequalities denoting the two sufficient conditions are replaced by equalities to construct optimal bounds.
- $$\begin{cases} \sum_{k=j}^{n} Q[1,k] & = & \sum_{k=j}^{n} P[1,k] \\ \sum_{k=j}^{n} Q[i+1,k] & = & max(\sum_{k=j}^{n} Q[i,k], \sum_{k=j}^{n} P[i+1,k]) \end{cases}$$
- Bounds obtained by this algorithm are optimal.

steady-state distribution

Let Q be a monotone, upper bounding matrix for P for the st-ordering. If the steady-state distributions (π_P and π_Q) exist, then:

$$\pi_P \leq_{st} \pi_Q$$

- Suppose that $\mathcal{Y}$ is an absorbing DTMC, k is an absorbing state.
- Let $\mathcal{Z}$ be an $\leq_{st}$ monotone upper bound : $\mathcal{Y} \leq_{st} \mathcal{Z}$.
- Assume that k is placed at the end of S.

Absorption probability

Let $\pi_{\mathcal{Y}}[i,k]$ (resp. $\pi_{\mathcal{Z}}[i,k]$) the absorption probability in k for chain $\mathcal{Y}$ (resp. $\mathcal{Z}$) when initial state is i:

$$\pi_{\mathcal{Y}}[i,k] \leq \pi_{\mathcal{Z}}[i,k]$$

- Transition matrices of $\mathcal{Y}$ and $\mathcal{Z}$ can be written respectively:

$$\begin{bmatrix} I & 0 \\ R & Y \end{bmatrix} \qquad \begin{bmatrix} I & 0 \\ R' & Z \end{bmatrix}$$

- Fundamental matrix of $\mathcal{Y}$ and $\mathcal{Z}$ (states of Y and Z are transient):

$$M_{\mathcal{Y}} = (I - Y)^{-1} \qquad M_{\mathcal{Z}} = (I - Z)^{-1}$$

Mean first passage time

Let $T_{\mathcal{Y}}[i]$ (resp. $T_{\mathcal{Z}}[i]$) be the random variable denoting the absorption time in chain $\mathcal{Y}$ (resp. $\mathcal{Z}$) where i is the initial state:

- $T_{\mathcal{Z}}[i] \leq_{st} T_{\mathcal{Y}}[i]$
- $\mathbf{E}(T_{\mathcal{Z}}[i]) = \sum_j M_{\mathcal{Z}}[i,j] \leq \mathbf{E}(T_{\mathcal{Y}}[i]) = \sum_j M_{\mathcal{Y}}[i,j]$

Censoring techniques (1)

- Consider a DTMC with transition matrix Q.
- Consider a partition of the state space (E, E^c), Q is written:

$$Q = \begin{pmatrix} Q_E & Q_{EE^c} \\ Q_{E^cE} & Q_{E^c} \end{pmatrix} \begin{matrix} E \\ E^c \end{matrix}$$

- The censored Markov chain introduced by Levy 57 (called watched Markov chain).
- The CMC only watches the chain when it is in E.
- Transition matrix of CMC is defined as:

$$S_E = Q_E + Q_{EE^c} \left(\sum_{i=0}^{\infty} (Q_{E^c})^i \right) Q_{E^cE}$$

Censoring techniques (2)

- Computing S_E is not easy if Q is large: If (Q_{E^c}) does not contain any recurrent class, the fundamental matrix is:

$$\sum_{i=0}^{\infty}(Q_{E^c})^i = (I - Q_{E^c})^{-1}$$

- If the chain is finite but not ergodic, all states of E^c must be transient (no reccurent class or absorbing states)
- When Q is very large: difficult to analyse Q
- It is difficult also to compute $(I - Q_{E^c})^{-1}$
- **Proposed approach:** we derive stochastic bounds to S_E (without knowing all informations about Q neither S_E).

What can $\leq_{st}$ Bounds provide?

- $\mathcal{X}$ the exact chain (state space S).
- $\mathcal{Y}$ censored chain (state space E).
- $\mathcal{Z}$ upper bound to $\mathcal{Y}$, $\leq_{st}$ monotone (state space E).
- What can we deduce for performability measures of $\mathcal{X}$ to $\mathcal{Z}$.
  1. **Upper bounds** to exact steady-state probabilities.
  2. **Upper bounds** to exact steady-state rewards.
  3. **Upper and lower bounds** to exact absorption probabilities.
  4. **Lower bound** to exact mean first passage time.

Bounds to steady state measures

Sum of steady-state probabilities

Assuming that $E = S' \cup S''$ is the censored subset and that states of S'' are placed at the end of E, then:

$$\sum_{i \in S''} \pi_E(i) \leq \sum_{i \in S''} \pi_{S_E}(i) \leq \sum_{i \in S''} \pi_{S_E^{sup}}(i)$$

Steady-state rewards

Let $\rho : S \rightarrow \mathbb{R}$ be the reward function that assigns to each state $i \in S$ a reward value $\rho(i) \geq 0$ for all i. Let E be the set of states which has non zero rewards. Assuming that we sort the states in E such that function ρ is non decreasing, then:

$$\sum_{i \in E} \rho(i)\pi_E(i) \leq \sum_{i \in E} \rho(i)\pi_{S_E}(i) \leq \sum_{i \in E} \rho(i)\pi_{S_E^{sup}}(i)$$

Bounds to steady state measures

Sum of steady-state probabilities

Assuming that $E = S' \cup S''$ is the censored subset and that states of S'' are placed at the end of E, then:

$$\sum_{i \in S''} \pi_E(i) \leq \sum_{i \in S''} \pi_{S_E}(i) \leq \sum_{i \in S''} \pi_{S_E^{sup}}(i)$$

Steady-state rewards

Let $\rho : S \rightarrow \mathbb{R}$ be the reward function that assigns to each state $i \in S$ a reward value $\rho(i) \geq 0$ for all i. Let E be the set of states which has non zero rewards. Assuming that we sort the states in E such that function ρ is non decreasing, then:

$$\sum_{i \in E} \rho(i)\pi_E(i) \leq \sum_{i \in E} \rho(i)\pi_{S_E}(i) \leq \sum_{i \in E} \rho(i)\pi_{S_E^{sup}}(i)$$

Bounds to absorption probability

- $\mathcal{X}$ contains a finite number of absorbing states.
- E contains all absorbing states and the states which immediately precede absorbing states and the initial state i.

Absorption probabilities

The absorbing probabilities in each absorbing state are the same in both chains (the exact $\mathcal{X}$ and the censored $\mathcal{Y}$).

Mean number of passages

Let $M_{\mathcal{X}}[i,j]$ (resp. $M_{\mathcal{Y}}[i,j]$) be the mean number of passages in j before absorption knowing that the initial state is i for chain $\mathcal{X}$ (resp. $\mathcal{Y}$), then:

$$M_{\mathcal{X}}[i,j] = M_{\mathcal{Y}}[i,j] \text{ if } j \in E$$

Bounds to absorption probability

- $\mathcal{X}$ contains a finite number of absorbing states.
- E contains all absorbing states and the states which immediately precede absorbing states and the initial state i.

Absorption probabilities

The absorbing probabilities in each absorbing state are the same in both chains (the exact $\mathcal{X}$ and the censored $\mathcal{Y}$).

Mean number of passages

Let $M_{\mathcal{X}}[i,j]$ (resp. $M_{\mathcal{Y}}[i,j]$) be the mean number of passages in j before absorption knowing that the initial state is i for chain $\mathcal{X}$ (resp. $\mathcal{Y}$), then:

$$M_{\mathcal{X}}[i,j] = M_{\mathcal{Y}}[i,j] \text{ if } j \in E$$

Bounds to absorption time

- $T_{\mathcal{X}}[i]$ be the random variable denoting the absorption time in chain $\mathcal{X}$ (resp. $\mathcal{Y}$), i is the initial state.
- $T_{\mathcal{Y}}[i]$ be the random variable denoting the absorption time $\mathcal{Y}$.
- $\leq_{st}$ comparison of $T_{\mathcal{X}}[i]$ and $T_{\mathcal{Y}}[i]$ is defined on dates that $\in \mathbb{N}$ and not on states.

Mean first passage time

The mean absorption time (first passage time) in chain $\mathcal{Y}$ is less or equal than the mean absorption time in chain $\mathcal{X}$:

$$\mathbf{E}(T_{\mathcal{Y}}[i]) \leq \mathbf{E}(T_{\mathcal{X}}[i])$$

Algorithms for bounding CMC

1. Truffet's approach: Based on Q_E published in Applied probability journal 1997 by Truffet.
2. DPY (Dayar Pekergin Younes) approach: based on Q_E and Q_{E^cE}

Tugrul Dayar, Nihal Pekergin and Sana Younes

Conditional steady-state bounds for a subset of states in Markov chain, SMCtools 2006

3. FPY (Fourneau Pekergin Younes) approach: based on Q_E and some information about E^c.

Jean Michel Fourneau, Nihal Pekergin and Sana Younes

Censoring Markov Chains and Stochastic Bounds, EPEW 2007

Truffet's approach

- Use only Q_E.
- Two steps:
  1. First add the slack probability in the last column of Q_E to make it stochastic
  2. Make it monotone by apply Vincent algorithm
- Simple, optimal if we know only Q_E but needs to obtain something more accurate
- A lower bound is obtained by adding slack probabily to the first column of Q_E.

Truffet's approach

$$Q = \left[\begin{array}{ccc|cc} 0.2 & 0.3 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0 & 0.2 \\ 0.2 & 0.3 & 0.3 & 0.1 & 0.1 \\ \hline 0.1 & 0.2 & 0.2 & 0.3 & 0.2 \\ 0 & 0.3 & 0.3 & 0.3 & 0.1 \end{array}\right] \quad \text{slack probability} = \left[\begin{array}{c} 0.3 \\ 0.2 \\ 0.2 \end{array}\right]$$

- Add slack probability in the last column

$$\left[\begin{array}{ccc} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.2 & 0.4 \\ 0.2 & 0.3 & 0.5 \end{array}\right]$$

- Make it monotone

$$T(Q_E) = \left[\begin{array}{ccc} 0.2 & 0.3 & 0.5 \\ 0.2 & 0.3 & 0.5 \\ 0.2 & 0.3 & 0.5 \end{array}\right] \geq_{st} S_E = \left[\begin{array}{ccc} 0.23 & 0.43 & 0.33 \\ 0.41 & 0.29 & 0.29 \\ 0.22 & 0.38 & 0.38 \end{array}\right]$$

DPY

- Use Q_E and Q_{EE^c}.
- Gives a better bound than Truffet's bound.
- If Q_{EE^c} is rank-1, DPY gives the exact censored matrix.
- For simplicity we illustrate the algorithm by the following example.

DPY:Example (1)

$$Q = \left[\begin{array}{ccc|cc} 0.2 & 0.3 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.2 & 0 & 0.2 \\ 0.2 & 0.3 & 0.3 & 0.1 & 0.1 \\ \hline 0.1 & 0.2 & 0.2 & 0.3 & 0.2 \\ 0 & 0.3 & 0.3 & 0.3 & 0.1 \end{array}\right] \quad \text{slack probability} = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.2 \end{bmatrix}$$

- Compute G such that:

$$G = \begin{bmatrix} 1 & 0.4/0.5 & 0.2/0.5 \\ 1 & 1 & 0.3/0.6 \end{bmatrix}$$

- Determine $Max(G) = [1\ 1\ 0.5]$
- To obtain what we will add to Q_E to obtain an upper bound to S_E, we compute:

$$[0\ (1-0.5)\ 0.5] * \begin{bmatrix} 0.3 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0 & 0.15 & 0.15 \\ 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 \end{bmatrix}$$

DPY:Example (2)

- Add to Q_E to obtain:
$$\begin{bmatrix} 0.2 & 0.45 & 0.35 \\ 0.4 & 0.3 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$$
- Make it monotone

$$S_E \leq_{st} DPY(Q_E) = \begin{bmatrix} 0.2 & 0.45 & 0.35 \\ 0.2 & 0.45 & 0.35 \\ 0.2 & 0.4 & 0.4 \end{bmatrix} \leq_{st} T(Q_E)$$

$$\begin{bmatrix} 0.23 & 0.43 & 0.33 \\ 0.41 & 0.29 & 0.29 \\ 0.22 & 0.38 & 0.38 \end{bmatrix} \leq_{st} \begin{bmatrix} 0.2 & 0.45 & 0.35 \\ 0.2 & 0.45 & 0.35 \\ 0.2 & 0.4 & 0.4 \end{bmatrix} \leq_{st} \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.2 & 0.3 & 0.5 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

FPY:Approach based on paths and graph algorithm

Theorem

Let L_E be an element-wise lower bound to S_E, $Q_E \leq L_E \leq S_E$. Then

$$S_E \leq_{st} T(L_E) \leq_{st} T(Q_E)$$

Main idea to compute L_E

- $(\sum_{i=0}^{\infty}(Q_{E^c})^i)[j,k]$ is the sum of all probability of paths entering in E^c from j and leaving it after an arbitrary number of visits inside E^c from k.
- We select some paths instead of generating all of them
- We adapt several well-known graph algorithms, shortest path , Breadth First search, to select some paths and compute their probability.

FPY:Approach based on paths and graph algorithm

Theorem

Let L_E be an element-wise lower bound to S_E, $Q_E \leq L_E \leq S_E$. Then

$$S_E \leq_{st} T(L_E) \leq_{st} T(Q_E)$$

Main idea to compute L_E

- $(\sum_{i=0}^{\infty}(Q_{E^c})^i)[j,k]$ is the sum of all probability of paths entering in E^c from j and leaving it after an arbitrary number of visits inside E^c from k.
- We select some paths instead of generating all of them
- We adapt several well-known graph algorithms, shortest path , Breadth First search, to select some paths and compute their probability.

Paths selection

BFS

- We start from an initial state belonging to E.
- The probability of a path is the product of the probability of the arcs.
- We fix the depth for the tree selected.

Shortest Path

- We adapt Dijkstra algorithme for our use.
- The weight in the path is $-log(Q(i,j))$.
- The shortest path according to this weight is the path with the highest probability.

Improve SP by taking self loops

- Let $\mathcal{P}$ be a path selected with probability p and x a node of $\mathcal{P}$.
- If there is a self loop on x that has probability q, the probability of $\mathcal{P}_x$ the path obtained by considering the self loop is pq.
- By considering i passage times in x the obtained probability is $pq^i > p$.
- If we consider all i times the probability is $p/1 - q > pq^i > p$.
- So if we take under consideration a self loop we obtain a better probability that is good for the accuracy of the bound.

Conclusion and perspective

- Applying censored techniques and stochastic comparison in DTMCs model checking (submitted).
- Extend to infinite Markov chains .
- Study transient time between exact and the censored Markov chain.
- Some remarks for DPY:
  - We think that DPY is optimal if we know only Q_E and Q_{E^cE} (need a proof).
  - Implementation: Is it easy to generate only Q_E and Q_{E^cE} without generating the remaining blocks?