Model Reduction for DTMCs

Annalisa Paladino, Emanuele Ruoppolo

1 Introduction

The verification and analysis of probabilistic models such as *Discrete-Time Markov Chains* (DTMCs) are foundational in formal methods, probabilistic programming, systems biology, and performance evaluation. However, as model complexity grows, the state space becomes prohibitively large, a phenomenon known as the *state-space explosion problem*. In order to study models with a very large state space without resorting to approximation or simulation techniques we can attempt to *reduce* the state space of the underlying Markov chain by aggregating states with equivalent behaviours. An interesting class of these aggregation methods that can be decided by the structural analysis of the original Markov chain is known as **lumping**. In the literature, several notions of lumping have been introduced: strong and weak lumping, exact lumping, and strict lumping. This idea traces back to Kemeny and Snell (1960) [1] on Markov chains, and has since been expanded in contexts such as probabilistic bisimulation and Markov decision processes.

This project investigates the application of **strong lumpability** to a DTMC featuring five states, reducing it to a three-state model while preserving its essential probabilistic behavior.

2 Preliminaries and Specific Objectives

2.1 Background

In this section we rapidly review the fundamentals of Discrete-Time Markov Chains and the concept of lumpability.

Discrete-Time Markov Chain. A Discrete-Time Markov Chain (DTMC) is a time-homogeneous Markov process with discrete parameter T and discrete state space S.

A DTMC $\mathcal{D}$ is a tuple $(S, \mathbf{P}, \iota_{init}, AP, L)$ with

- S is a countable nonempty set of states
- $\mathbf{P} : S \times S \to [0, 1]$ is the transition probability function, such that $\sum_{s'} \mathbf{P}(s, s') = 1$ for all $s \in S$.
- $\iota_{init} : S \to [0, 1]$ is the initial distribution with $\sum_{s \in S} \iota_{init}(s) = 1$
- AP is a set of atomic propositions
- $L : S \to 2^{AP}$ is the labeling function, assigning to state s, the set $L(s)$ of atomic propositions that are valid in s

Lumpability. The notion of lumpability provides a model aggregation technique that can be used for generating a Markov chain that is smaller than the original one but allows one to determine exact results for the original process. The concept of lumpability [2] [3] can be

formalized in terms of equivalence relations over the state space of the Markov chain. Any such equivalence induces a partition on the state space of the Markov chain and aggregation is achieved by clustering equivalent states into macro-states, thus reducing the overall state space.

If the partition can be shown to satisfy the so-called *strong* lumpability condition, then the equilibrium solution of the aggregated process may be used to derive an exact solution of the original one.

Now let:

- $\mathcal{X} = \{1, 2, \ldots, n\}$ be the finite state space of a discrete-time Markov chain (DTMC),
- $P \in \mathbb{R}^{n \times n}$ be the transition matrix of the chain,
- $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_k\}$ be a partition of $\mathcal{X}$ into $k < n$ disjoint blocks, i.e.,

$$\ell_i \subset \mathcal{X}, \quad \ell_i \cap \ell_j = \emptyset \text{ for } i \neq j, \quad \text{and} \quad \bigcup_{i=1}^{k} \ell_i = \mathcal{X}.$$

Strong Lumpability. The DTMC is said to be strongly lumpable with respect to the partition $\mathcal{L}$ if and only if:

$$\sum_{y \in \ell_j} P(x, y) = \sum_{y \in \ell_j} P(x', y)$$

for all $\ell_j \in \mathcal{L}$ and all $x, x' \in \ell_i$ for some i.

This means that if you pick any two states x, x' from the same block ℓ_i, then the probability of transitioning to any other block ℓ_j is the same. Moreover the transition probability between blocks is well-defined and independent of the representative state chosen within a block.

This condition allows the definition of a **reduced DTMC** over the aggregated states $\{\ell_1, \ldots, \ell_k\}$, with transition probabilities defined as:

$$\bar{P}(\ell_i, \ell_j) = \sum_{y \in \ell_j} P(x, y) \quad \text{for any } x \in \ell_i.$$

Consequences of strong lumpability are:

- It can be proven that the process over the aggregated blocks is itself a Markov chain.
- It preserves stationary distributions: if the original DTMC has stationary distribution π, then the induced (lumped) chain has stationary distribution

$$\bar{\pi}(\ell_i) = \sum_{y \in \ell_i} \pi(y).$$

- Strong lumpability ensures *exact aggregation*, not just approximate behavior.

2.2 Objectives

The given DTMC comprises five states $\{s_0, s_1, s_2, s_3, s_4\}$. Assuming you can observe only state labels (depicted in different colors), and traces starting at s_0, states s_1 and s_2 are indistinguishable from each other, and the same holds for states s_3 and s_4.

Our specific goals were:

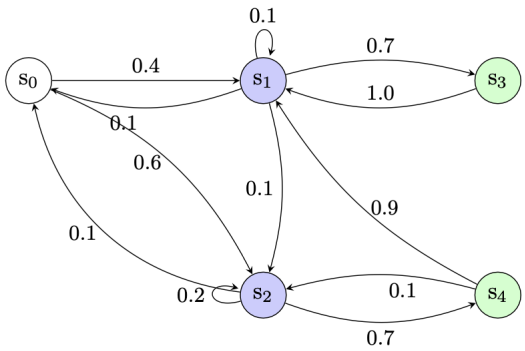

Figure 1: Initial DTMC.

- To verify whether the model is strongly lumpable with respect to the partitions $\{s_0\}, \{s_1, s_2\}, \{s_3, s_4\}$;
- To construct the reduced DTMC over these blocks;
- To demonstrate preservation of reachability and long-run distributions in the lumped system;
- To discuss general applicability and scalability of lumpability.

3 Methods

The project followed the steps below:

3.1 Model Extraction

The state transition structure was inferred from the provided diagram, converting edge weights into transition probabilities. This is the resulting transition matrix:

$$P = \begin{bmatrix} 0.0 & 0.4 & 0.6 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.1 & 0.7 & 0.0 \\ 0.1 & 0.0 & 0.2 & 0.0 & 0.7 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.9 & 0.1 & 0.0 & 0.0 \end{bmatrix}$$

3.2 State Partitioning

Based on indistinguishability of states under observation, the partition was defined as:

$$\ell_0 = \{s_0\}, \quad \ell_1 = \{s_1, s_2\}, \quad \ell_2 = \{s_3, s_4\}$$

3.3 Lumped Model Construction

We now verify and compute the lumped transitions.

1. Transitions from $\ell_0 = \{s_0\}$

Since ℓ_0 contains only a single state, the lumpability condition is trivially satisfied. The lumped transition probabilities are calculated directly from the transitions of s_0:

$$\begin{aligned} P(\ell_0 \to \ell_0) &= 0 \\ P(\ell_0 \to \ell_1) &= P(s_0 \to s_1) + P(s_0 \to s_2) && = 0.4 + 0.6 = 1.0 \\ P(\ell_0 \to \ell_2) &= 0 \end{aligned}$$

2. Transitions from $\ell_1 = \{s_1, s_2\}$

To find the transition probabilities from block ℓ_1, we compute the transition probabilities from each of its constituent states (s_1 and s_2) to the lumped blocks ℓ_0, ℓ_1, ℓ_2.

For state s_1:

$$\begin{aligned} P(s_1 \to \ell_0) &= P(s_1 \to s_0) && = 0.1 \\ P(s_1 \to \ell_1) &= P(s_1 \to s_1) + P(s_1 \to s_2) && = 0.1 + 0.1 = 0.2 \\ P(s_1 \to \ell_2) &= P(s_1 \to s_3) && = 0.7 \end{aligned}$$

For state s_2:

$$\begin{aligned} P(s_2 \to \ell_0) &= P(s_2 \to s_0) && = 0.1 \\ P(s_2 \to \ell_1) &= P(s_2 \to s_2) && = 0.2 \\ P(s_2 \to \ell_2) &= P(s_2 \to s_4) && = 0.7 \end{aligned}$$

Since the transition probabilities from s_1 and s_2 to each target block are identical, the lumpability condition is satisfied. The lumped transition probabilities are:

$$\begin{aligned} P(\ell_1 \to \ell_0) &= 0.1 \\ P(\ell_1 \to \ell_1) &= 0.2 \\ P(\ell_1 \to \ell_2) &= 0.7 \end{aligned}$$

3. Transitions from $\ell_2 = \{s_3, s_4\}$

Finally, we check the states within ℓ_2.

For state s_3:

$$\begin{aligned} P(s_3 \to \ell_0) &= 0 \\ P(s_3 \to \ell_1) &= P(s_3 \to s_1) && = 1.0 \\ P(s_3 \to \ell_2) &= 0 \end{aligned}$$

For state s_4:

$$\begin{aligned} P(s_4 \to \ell_0) &= 0 \\ P(s_4 \to \ell_1) &= P(s_4 \to s_1) + P(s_4 \to s_2) && = 0.9 + 0.1 = 1.0 \\ P(s_4 \to \ell_2) &= 0 \end{aligned}$$

Again, the transition probabilities from s_3 and s_4 to each block are identical, so the condition holds. The lumped transition probabilities are:

$$P(\ell_2 \to \ell_0) = 0$$
$$P(\ell_2 \to \ell_1) = 1.0$$
$$P(\ell_2 \to \ell_2) = 0$$

When the condition above was satisfied, transitions between blocks were aggregated to form the reduced transition matrix (Table 1).

4 Results

The lumped transition matrix over states ℓ_0, ℓ_1, ℓ_2 is:

From \ To	ℓ_0	ℓ_1	ℓ_2
ℓ_0	0.0	1.0	0.0
ℓ_1	0.1	0.2	0.7
ℓ_2	0.0	1.0	0.0

Table 1: Transition matrix of the lumped DTMC.

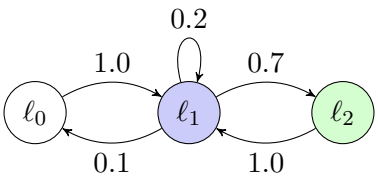

Figure 2: The lumped Markov chain.

Although it can be formally proved that strongly lumped systems are still DTMC, we can verify that the lumped system we obtained respects the properties of a DTMC, indeed:

- It has a **finite** set of states $S = \{\ell_0, \ell_1, \ell_3\}$.
- It respects the **Markov property**, from each state the probability of transitioning to another state does not depend on the previous steps, and it is described by the above mentioned transition matrix $\tilde{P}$.
- It is **time-homogeneous**, the values of $\tilde{P}$ are constant and do not change over time.
- The sum of outgoing transition probabilities for every state is 1, we can easily verify that each row sums to 1.

4.1 Property Evaluation

We checked that both reachability properties and long-run behavior were preserved in the aggregated model.

Reachability properties

We can observe from the original transitions:

- $s_0 \to s_1, s_2 \in \ell_1 \Rightarrow \ell_0 \to \ell_1$
- $s_1 \to s_3 \in \ell_2, s_2 \to s_4 \in \ell_2 \Rightarrow \ell_1 \to \ell_2$
- $s_3 \to s_1 \in \ell_1, s_4 \to s_1, s_2 \in \ell_1 \Rightarrow \ell_2 \to \ell_1$

These paths exist in both the original and the lumped DTMC. Thus every path of blocks in the lumped DTMC corresponds to at least one path of states in the original DTMC, and vice versa. This ensures that if a block l_j is reachable from l_i in the lumped DTMC, then some state in l_j is reachable from some state in l_i in the original DTMC. The structure of the state transitions is preserved at the block level.

Long-run behavior

Steady-state probabilities over aggregates are preserved. To prove it we must compute the stationary distribution $\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4)$ relative to the original system with transition matrix P, such that:

$$\pi = \pi P$$

and show that the stationary distribution of the lumped system $\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_1, \tilde{\pi}_2)$ is such that:

$$\tilde{\pi}_0 = \pi_0, \quad \tilde{\pi}_1 = \pi_1 + \pi_2, \quad \tilde{\pi}_2 = \pi_3 + \pi_4$$

Avoiding to show all the computation of the two distributions, which consists in solving two linear systems we have:

$$\pi = \left(\frac{1}{18}, \frac{335}{747}, \frac{80}{747}, \frac{469}{1494}, \frac{56}{747}\right)$$

and

$$\tilde{\pi} = \left(\frac{1}{18}, \frac{5}{9}, \frac{7}{18}\right)$$

We then have:

$$\begin{aligned} \pi_0 &= \frac{1}{18} = \tilde{\pi}_0 \\ \pi_1 + \pi_2 &= \frac{335}{747} + \frac{80}{747} = \frac{415}{747} = \frac{5}{9} = \tilde{\pi}_1 \\ \pi_3 + \pi_4 &= \frac{469}{1494} + \frac{56}{747} = \frac{581}{1494} = \frac{7}{18} = \tilde{\pi}_2 \end{aligned}$$

We have then proved that on the long run, according to its stationary distribution, the lumped system keeps the behavior of the original system.

4.2 Conditions for safe application of lumpability

It's possible to apply safely lumpability under the following conditions:

1. States in the same partition class must exhibit identical transition behavior to other classes

2. The partition must respect the observable distinctions in the system (e.g., different colors/labels). Each class should consist of states with similar dynamics, to avoid information loss in the aggregated Markov chain

3. For exact preservation, the initial distribution should be constant within each partition class

4.3 Scalability

A practical verification algorithm is:

Algorithm 1 Check Strong Lumpability

```
for each partition class C such that |C| > 1 do
    for each target partition class D do
        Pick reference state i ∈ C
        Compute p_ref = Σ_{j∈D} P(i, j)
        for each other state k ∈ C do
            if Σ_{j∈D} P(k, j) ≠ p_ref then
                return FALSE
            end if
        end for
    end for
end for
return TRUE
```

Time Complexity: $\mathcal{O}(n^2)$, where n is the number of states.

Space Complexity: $\mathcal{O}(1)$ additional space (in-place computation).

In our example: $n = 5$, $k = 3$ (number of partition classes).

To ensure the method is scalable to larger systems:

- *Detection*: Finding lumpable partitions is generally hard, but domain knowledge often suggests natural partitions
- *Sparse matrices*: Utilizing sparse (rather than dense) transition matrices allows scaling to systems with tens of thousands of states
- *Hierarchical lumping*: Multiple levels of lumping can be applied successively for very large models
- *Approximate lumping*: When exact lumping fails, near-lumpability can still provide useful approximations. For large-scale models, approximate reachability analysis techniques (e.g., Monte Carlo methods) can be employed effectively

5 Discussion and Outlook

Our analysis confirmed:

- Strong lumpability conditions are satisfied for ℓ_1 and ℓ_2;
- The reduced model is a valid DTMC;

- Reachability from s_0 to terminal blocks ℓ_2 via ℓ_1 is preserved;
- The stationary distribution over blocks is consistent with the aggregate behavior of the original model.

This study demonstrates that lumpability is a powerful model reduction technique when the Markov chain exhibits structural symmetry or observational equivalence. The reduced model preserves key behavioral characteristics of the original chain, such as reachability and long-run statistics.

In conclusion, lumpability offers a formal and computationally sound method for simplifying the analysis of complex probabilistic systems, preserving both correctness and efficiency.

References

[1] John Kemeny & James L. Snell (1960): Finite Markov Chains. Van Nostrand.

[2] Sumita, U., & Rieders, M. (1989). Lumpability and time reversibility in the aggregation-disaggregation method for large markov chains. Communications in Statistics. Stochastic Models, 5(1), 63–81. https://doi.org/10.1080/15326348908807099

[3] Buchholz, P. (1994). Exact and Ordinary Lumpability. J. Applied Probability.