Policy Iteration for Sum-Product Inferences

Pieter-Jan Vandaele PIETERJAN.VANDAELE@UGENT.BE
Jasper De Bock JASPER.DEBOCK@UGENT.BE
Foundations Lab for imprecise probabilities, Ghent University, Belgium

Markov chains (MCs) are a popular probalistic model for describing the uncertain evolution of the state of a process in a state space $\mathcal{X}$, which we will take to be finite. Compared to general stochastic processes, these MCs additionally satisfy the Markov property $P(X_{n+1}|X_{1:n}) = P(X_{n+1}|X_n)$, and often also a time-homogeneity condition $P(X_{n+1}|X_n) = P(X_2|X_1)$. Imprecise Markovs chains (IMCs) are a generalisation of MCs that allow for imprecision in the transition probabilities between the states, typically in the form of bounds. An IMC can be seen as a set of stochastic processes $\mathcal{P}$, of which three types are of particular interest. An IMC under repetition independence ($\mathcal{P}_{RI}$) contains all time-homogeneous MCs that are compatible with the imprecise transition probabilities. An IMC under complete independence ($\mathcal{P}_{CI}$) contains all compatible (but not neccessarily time-homogeneous) MCs. Finally, IMCs under epistemic irrelevance ($\mathcal{P}_{EI}$) contain all compatible (but not necessarily Markovian or time-homogeneous) stochastic processes. For an IMC $\mathcal{P}$ and a gamble $f(X_{1:n})$ that takes a sequence of n states, the lower expectation conditional on its initial state x_1 is defined as $\underline{\mathrm{E}}_{\mathcal{P}}(f(X_{1:n})|x_1) = \inf_{p\in\mathcal{P}} E(f(X_{1:n})|x_1)$, and similarly for the conditional upper expectation. For many specific choices of the gambles f, these correspond to inferences that are of interest.

IMCs share a great deal of similarity with Markov Decision Processes (MDPs) [3], where one typically has a choice between actions that influence the transition probabilities of an MC. A *policy* chooses which action to perform at each time instant in order to minimise or maximise some reward accumulated during the stochastic process. These policies can choose the same action at each time instant, allow the actions to vary throughout the process or even let them depend on the history of the stochastic process. These three types of policies are reminiscent of the IMCs under repectively RI, CI and EI. It is well known that, for many of the inferences in which the MDP community is interested, the result will be the same regardless of what type of policy is considered. Furthermore the MDP community has a vast amount of efficient algorithms [3] for computing specific such inferences, including total sums, time averages and discounted rewards. Our poster aims to explore to what extent these results and algorithms can be adapted and generalised to IMCs.

In comparison to MPDs, research on IMCs focuses on more general inferences. For instance, De Bock et al. consider so-called sum-product inferences, which correspond to gambles of the form $f(X_{1:n}) := \sum_{k=1}^{n} g(X_k) \prod_{l=1}^{k-1} h(X_l)$, where g can be any gamble and h must be non-negative. For this family of functions, it was shown that $\underline{\mathrm{E}}_{\mathcal{P}_{CI}}(f(X_{1:n})|x_1)$ and $\underline{\mathrm{E}}_{\mathcal{P}_{EI}}(f(X_{1:n})|x_1)$ coincide, and can be efficiently computed using a recursive scheme, and similarly for upper expectations [1]. Interestingly, for the specific case of hitting times, Krak [2] also describes a non-recursive algorithm that resembles the *policy iteration* algorithm that is standard in MPD literature, and notices that for this inference, at least in the limit for large n, not only EI and CI coincide, but RI as well.

Inspired by this connection, we show in our poster that for sum-product gambles f with $(\forall x \in \mathcal{X})\ 0 \leq h(x) < 1$, at least for n going to infinity, EI, CI and RI coincide. Furthermore, the common value $\lim_{n\to\infty} \underline{\mathrm{E}}(f(X_{1:n})|x_1)$ is finite regardless of the values of $g(x)$. It is this observation that allows us to employ an algorithm that resembles policy iteration. At each step of the iteration, an MC is chosen with a limit expectation that, for each initial state, is lower than in the previous step. This expectation can be found by solving a set of linear equations. This is repeated until no more improvement is possible, and consequently $\lim_{n\to\infty} \underline{\mathrm{E}}_{\mathcal{P}_{RI}}(f(X_{1:n})|x_1)$ has been found, which is then also the limit lower expectation for $\mathcal{P}_{CI}$ and $\mathcal{P}_{EI}$. An advantage of our method over the recursive method is that, under conditions satisfied in most practical applications, it finds the exact solution for the lower expectation in a finite number of steps, whereas the recursive scheme approaches the solution in norm. Similar results and algorithms apply to the upper expectations as well.

References

[1] Jasper De Bock, Alexander Erreygers, and Thomas Krak. Sum-product laws and efficient algorithms for imprecise Markov chains. In *Proceedings of UAI 2021*, volume 161 of *PMLR*, pages 1476–1485. PMLR, 2021.

[2] Thomas Krak. Computing expected hitting times for imprecise Markov chains. In *Advances in Uncertainty Quantification and Optimization Under Uncertainty with Aerospace Applications*, pages 185–205. Springer, 2021.

[3] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 2005.