Errata for Abstract Reward Processes: Leveraging State Abstraction for Consistent Off-Policy Evaluation, NeurIPS 2024

Shreyas Chaudhari
University of Massachusetts Amherst
schaudhari@cs.umass.edu

April 2026

This note highlights that in its current form Theorem 4.3 as stated in in our paper [Chaudhari et al., 2024] is false. Specifically, one step in its proof does not hold, invalidating the proof. We state a stricter updated condition under which the theorem holds, while noting that weaker conditions may also exist that are sufficient.

Brief description of the original mistake

In Appendix B.5, the proof of Theorem 4.3 invokes Equation (63), which asserts a conditional independence between the "past" and "future" importance-weight products, conditioned on a length-c abstract-state window:

$$\rho_{0:(t-c)^+} \perp \rho_{(t-c+1)^+:t} \mid (Z_i)_{i=(t-c+1)^+}^{t}. \tag{63}$$

However, the stated assumption that the abstract process defined over Z is c-th order Markov—as defined in Definition 4.2—does *not* by itself imply (63). The reason is that $\rho_{0:(t-c)^+}$ and $\rho_{(t-c+1)^+:t}$ are functions of the underlying *state–action* trajectory, and conditioning on $(Z_i)_{i=(t-c+1)^+}^{t}$ continues to leave ambiguity about the distribution of the underlying state $S_{(t-c+1)^+}$ at the clipping boundary.

In general, it is possible for multiple underlying states to produce the same abstract-state window $(Z_i)_{i=(t-c+1)^+}^{t}$. Consequently, the past and future importance weight products can remain dependent through the (unobserved) boundary state if the state, or its distribution, cannot be inferred from the abstract states.

Updated sufficient condition (strong)

We believe that Definition 4.2 may be replaced by the following condition:

Definition 1 (c-step decodability)**.** There *exists* a (arbitrary, latent) deterministic function f such that

$$S_{t-c+1} = f\big((Z_i)_{i=(t-c+1)^+}^{t}; \pi\big) \quad \text{for } \pi \in \{\pi_b, \pi_e\}. \tag{1}$$

Equivalently, the boundary state S_{t-c+1} is determined by the length-c abstract-state window $(Z_i)_{i=(t-c+1)^+}^{t}$. We defer an updated proof for later work, however, it can be seen that under this condition, Equation (63) and the proof follows without further modification. The rest of the paper and its theoretical and empirical results remain unchanged.

Acknowledgements: We thank Nikos Vlassis for finding and alerting us to this error.

References

Shreyas Chaudhari, Ameet Deshpande, Bruno C da Silva, and Philip S Thomas. Abstract reward processes: Leveraging state abstraction for consistent off-policy evaluation. *Advances in Neural Information Processing Systems*, 37:17069–17105, 2024.