Missingness-MDPs: Bridging the Theory of Missing Data and POMDPs

Joshua Wendland[1] , **Markel Zubia**[1] , **Roman Andriushchenko**[2] , **Maris F. L. Galesloot**[3] , **Milan Češka**[2] , **Henrik von Kleist**[4] , **Thiago D. Simão**[5] , **Maximilian Weininger**[1] , **Nils Jansen**[1,3]
[1]Ruhr University Bochum
[2]Brno University of Technology
[3]Radboud University Nijmegen
[4]Harvard University
[5]Eindhoven University of Technology
{joshua.wendland, markel.zubia}@ruhr-uni-bochum.de

Abstract

We introduce *missingness-MDPs* (miss-MDPs), a novel subclass of partially observable Markov decision processes (POMDPs) that incorporates the theory of missing data. A miss-MDP is a POMDP whose observation function is a *missingness function*, specifying the probability that individual state features are *missing* (i.e., unobserved) at a time step. The literature distinguishes three canonical missingness types: (1) missing completely at random (MCAR), (2) missing at random (MAR), and (3) missing not at random (MNAR). The problem is to compute near-optimal policies for a miss-MDP with *unknown* missingness function, given a dataset of action-observation histories. Achieving such optimality guarantees for policies requires learning the missingness function from data, which is infeasible for general POMDPs. To overcome this challenge, we exploit structural properties of different missingness types to derive probably approximately correct (PAC) algorithms that learn the missingness function. These algorithms yield an approximate, fully specified, miss-MDP that we solve using off-the-shelf planning methods. We prove that, with high probability, the resulting policies are ε-optimal in the true miss-MDP. Empirical results confirm the theory and demonstrate the superior performance of our approach over two model-free methods.

1 Introduction

Markov decision processes (MDPs) [Puterman, 1994] capture sequential decision-making under uncertainty. A typical assumption is that all *state features* are fully observable at all times. In practice, however, features may be *missing*, e.g. due to sensor failures, so decisions must be made from incomplete information. For instance, a medical doctor may diagnose a patient based on recorded measurements such as heart rate and temperature that are potentially incomplete.

Our problem is to compute optimal policies for MDPs with dynamically missing state features. While the MDP structure and historical data are given, the process governing missingness is unknown. Partially observable Markov decision processes (POMDPs) [Åström, 1965] naturally capture this setting, with an observation function specifying the probability of receiving partial state information, of which missing features are a special case. However, the observation function is unknown and must, in principle, be learned from action–observation histories. Yet, learning the observation function of a POMDP is notoriously challenging: In general, it is unidentifiable from histories alone [Allman *et al.*, 2009; Finesso, 1990; Gilbert, 1959; Tune *et al.*, 2013], and even without requiring identifiability (i.e. merely seeking a model consistent with the data), learning remains both statistically [Krishnamurthy *et al.*, 2016; Xiong *et al.*, 2022; Golowich *et al.*, 2022] and computationally intractable [Terwijn, 2002; Mossel and Roch, 2005].

Our core contribution is exploiting the theory of *missing data* [Schafer and Graham, 2002; Buuren, 2018; Little and Rubin, 2019] to define suitable identifiability assumptions for POMDPs. We introduce *missingness-MDPs* (miss-MDPs) as a proper subclass of POMDPs. In miss-MDPs, the observation function is a *missingness function* that generates either perfect or missing information for each state feature. We classify this function as *missing completely at random* (MCAR), *missing at random* (MAR), or *missing not at random* (MNAR), following Rubin [1976], detailed in Section 4.

We refine our problem statement: Given a miss-MDP with an *unknown* missingness function and a dataset of action-observation histories, the goal is to compute a policy that maximizes the expected discounted reward. We assume that the underlying MDP and, thus, the transition function are known. In the doctor-patient example, the transition function captures changes in the patient's health status [Komorowski *et al.*, 2018]; see Figure 1. Similarly, in a robot navigation task, the probability of transitioning to a subsequent state may be known, while sensor failures can still introduce missingness and, therefore, partial observability. To obtain guarantees on the result, we approximate the unknown missingness function from the dataset, and thereby the original miss-MDP. For this approximate miss-MDP, we compute a near-optimal policy through off-the-shelf POMDP solvers such as SARSOP [Kurniawati *et al.*, 2008]. Missingness functions are *not learnable*

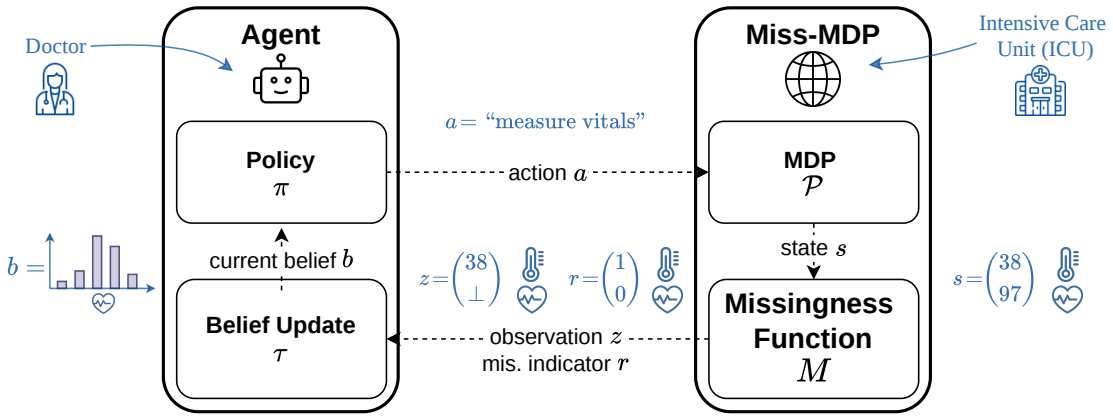

Figure 1: A doctor-treating-patient example (blue annotations) of an agent interacting with a miss-MDP. The missingness function causes the heart rate feature of the state to go missing, indicated as $\perp$. The missingness indicator evaluates to 0 for missing features and to 1 otherwise.

in general [Bhattacharya *et al.*, 2019], yet, we identify and focus on missingness functions that are tractable to learn.

In summary, our contributions are:

1. We introduce miss-MDPs, which integrate and define the semantics of missing data in a proper subclass of the more general POMDP framework (Section 4).
2. We prove that state beliefs do not always depend on the missingness function's probabilities (Remark 2), similar to *ignorability* of missing data [Little and Rubin, 2019].
3. We present *probably approximately correct* (PAC) learning algorithms for tractable subclasses of the three missingness types (Sections 5.1 and 5.2).
4. We prove that by using these algorithms, we approximate the ε-optimal policy for the miss-MDP under the correct assumption on the missingness function (Section 5.3).

Our empirical evaluation (Section 6) confirms our theoretical results and highlights the practical advantages of our approach: (1) For datasets of moderate size to learn the missingness function, the performance of the resulting policies converges to that of the optimal policy; and (2) our policies show superior performance in comparison to POMCP [Silver and Veness, 2010] and PPO [Schulman *et al.*, 2017], which use the direct POMDP formulation of the missingness problem. We provide all proofs and reproducibility details of our experiments in an extended version of this paper [Wendland *et al.*, 2026].

2 Related Work

We discuss literature on missing data, sequential decision-making, and reinforcement learning (RL).

We build on a rich body of works in missing data analysis, see e.g. [Tsiatis, 2006; Little and Rubin, 2019] where classical assumptions such as MCAR, MAR, and MNAR provide high-level categories. More refined tools, such as missingness graphs, allow one to encode assumptions about the missingness in a structured way [Mohan *et al.*, 2013; Shpitser *et al.*, 2015], leading to highly specific learnability results [Bhattacharya *et al.*, 2019; Nabi *et al.*, 2020]. Our setting departs from the standard missing-data paradigm in several important respects. In particular, the concept of missingness is embedded in the setting of solving POMDPs, enabling a more principled understanding of missingness in the context of sequential decision-making under uncertainty. Classical missing data approaches [Dempster *et al.*, 1977; Van den Broeck *et al.*, 2015] are interested in the parameters of the state distribution, as if there were no missingness (referred to as "full-data distribution"). In contrast, we are interested in finding the optimal policy under missingness. Existing methods that estimate the states given the observations – such as imputation – do not estimate the target distribution necessary for planning under uncertainty (see Remark 1 for details).

Prior work on MDPs with missing observations has focused mainly on RL, where missing data is treated as incidental rather than explicitly modeled. Both, full observations [Chen *et al.*, 2023] and individual features may be missing [Shim *et al.*, 2018; Yoon *et al.*, 2019; Böck *et al.*, 2022]. Some approaches pre-process missingness in observations for RL agents [Wang *et al.*, 2019], while others adopt model-based methods, often restricted to simpler settings such as MCAR [Futoma *et al.*, 2020]. Planning approaches typically neglect distinctions between MCAR, MAR, and MNAR [Liu *et al.*, 2022; Yamaguchi *et al.*, 2020; Futoma *et al.*, 2020]. Other works combine deep learning with POMDP solvers, but without explicitly modeling the missingness process [Liu *et al.*, 2022]. Some approaches include imputation strategies such as Bayesian multiple imputation [Lizotte *et al.*, 2008] and expectation-maximization [Yamaguchi *et al.*, 2020].

Complementary research establishes PAC guarantees for RL in POMDPs with unknown observation functions [Krishnamurthy *et al.*, 2016; Guo *et al.*, 2016; Uehara *et al.*, 2023]. These methods impose structural conditions on the underlying POMDP, such as full-column-rank transition or observation matrices [Guo *et al.*, 2016; Uehara *et al.*, 2023], deterministic transitions [Krishnamurthy *et al.*, 2016; Uehara *et al.*, 2023], optimality of memoryless policies [Krishnamurthy *et al.*, 2016], or reachability of every state within a few steps [Guo *et al.*, 2016]. We make no such assumptions

on the transition dynamics: Miss-MDPs violate the full-rank condition for observations, since missing features can cause distinct states to yield identical observation distributions.

To summarize: Our approach directly learns the missingness function and provides PAC guarantees on the resulting policy. To our knowledge, no existing work bridges missingness and POMDPs to (1) learn the missingness function with statistical guarantees, and (2) leverage it to guarantee the optimality of the resulting policies.

3 Preliminaries

A *probability distribution over a finite set* X is a function $\mu\colon X \to [0,1]$ with $\sum_{x\in X}\mu(x) = 1$. The set of such distributions is $\Delta(X)$. Writing $\mu = \{x_1 \mapsto p_1, \dots, x_k \mapsto p_k\}$ indicates that $\mu(x_1) = p_1$ and so on. The *support* of distribution $\mu \in \Delta(X)$ is $\text{supp}(\mu) = \{x \in X \mid \mu(x) \neq 0\}$. Sampling a random variable $\boldsymbol{x}$ from μ is denoted by $\boldsymbol{x} \sim \mu$. Given $\sigma\colon X \to \Delta(Y)$, we let $\sigma(y \mid x) := \sigma(x)(y)$. The indicator function $\mathbf{1}_\varphi$ returns 1 if predicate φ holds and 0 otherwise.

Definition 1 (POMDPs)**.** A *partially observable Markov decision process* is a tuple $\mathcal{P} = (S, A, T, b_0, \varrho, Z, O, \gamma)$ with finite factored *state space* $S = \times_{i=1,\dots,n} S_i$ and the set of *feature indices* $I = \{1, \dots, n\}$, finite *action space* A, *transition function* $T\colon S \times A \to \Delta(S)$, *initial state distribution* $b_0 \in \Delta(S)$, *reward function* $\varrho\colon S \times A \to \mathbb{R}$, finite factored *observation space* $Z = \times_{i=1,\dots,m} Z_i$, *observation function* $O\colon S \to \Delta(Z)$, and *discount factor* $\gamma \in [0,1)$.

The observation function is action-independent without loss of generality, as we may augment the state space to incorporate the last performed action [Chatterjee *et al.*, 2016].

A *trajectory* in a POMDP $\mathcal{P}$ is a sequence of states, observations, and actions. A *history* $h = \left(z^{(0)}, a^{(0)}, z^{(1)}, a^{(1)}, \dots\right) \in \mathcal{H} \subseteq (Z \times A)^*$ is the observable fragment of a trajectory, i.e., a sequence of observations and actions. A history can be summarized by a *sufficient statistic* known as a *belief* $b \in \mathcal{B} \subseteq \Delta(S)$, a probability distribution over underlying states induced by a history $h \in \mathcal{H}$. The *belief update* $\tau\colon \mathcal{B} \times A \times Z \to \mathcal{B}$ computes a *successor belief* b' via Bayes' rule [Spaan, 2012].

A *policy* $\pi\colon \mathcal{B} \to \Delta(A) \in \Pi$ maps beliefs to probability distributions over actions. The *objective* is to find a policy $\pi \in \Pi$ that maximizes its *value*, representing the infinite-horizon expected cumulative discounted reward: $V_\mathcal{P}(\pi) = \mathbb{E}^\pi\left[\sum_{t=0}^\infty \gamma^t \varrho(s^{(t)}, a^{(t)})\right]$. As the decision problem of finding the optimal policy is undecidable in general [Madani *et al.*, 2003], we focus on ε-optimal policies [Hauskrecht, 2000].

4 Missingness in MDPs

We introduce *missingness-MDPs* (miss-MDPs), the different types of missingness functions and ways to visualize them.

Definition 2 (Miss-MDP)**.** A missingness-MDP is a tuple $\mathcal{P} = (S, A, T, b_0, \varrho, Z, M, \gamma)$, where S, A, T, b_0, ϱ, and γ are as in a POMDP, the finite *observation space* is $Z = \times_{i\in I}(S_i \cup \{\perp\})$, with $\perp$ denoting *missing information*, and function $M\colon S \to \Delta(Z)$ is the *missingness function* such that $\forall s \in S, \forall z \in \text{supp}(M(s)), \forall i \in I$ either $z_i = s_i$ or $z_i = \perp$.

The state space S and observation space Z share the feature indices I, and we have that $Z \supsetneq S$ as some features can go *missing* in Z, being replaced by the symbol $\perp$. This process of "poking holes" is governed by the stochastic missingness function M. While M may take actions into account, we use an action-independent M w.l.o.g (see Section 3).

Missingness indicators. Missingness functions can equivalently be described as a map to vectors of *missingness indicators* [Mohan *et al.*, 2013], i.e. $M\colon S \to \Delta(R)$, where $R = \{0,1\}^n$. A vector $r \in R$ has $r_i = 0$ if feature i is missing ($z_i = \perp$), and otherwise $r_i = 1$. The function $f_R\colon Z \to R$ maps observations to their missingness indicators.

Example 1. Let $\mathcal{P}$ be a miss-MDP with $S = \{x, y\}^2$, $Z = \{x, y, \perp\}^2$, and missingness function M defined as: $M((s_1, s_2)) = \{(s_1, s_2) \mapsto 0.5, (s_1, \perp) \mapsto 0.5\}$. We have $f_R((y, x)) = (1, 1)$ and $f_R((y, \perp)) = (1, 0)$. Visiting state (y, x) yields either (y, x) or $(y, \perp)$, each with probability 0.5.

Problem statement. We are given a miss-MDP $\mathcal{P}$ with an *unknown* missingness function M, a required precision $\varepsilon > 0$ and confidence threshold $\delta > 0$, and a dataset $\mathcal{D} = (h_1, \dots, h_k)$ of k histories $h_i \in \mathcal{H}$ collected from $\mathcal{P}$ under an unknown but fair policy π_b (i.e. it has positive probability to visit all reachable states). The goal is to (1) approximate the missingness function $\widehat{M} \approx M$ for all reachable states and (2) use it to compute a policy $\pi^* \in \Pi$ such that with probability at least δ, we have $(\sup_\pi V_\mathcal{P}(\pi)) - V_\mathcal{P}(\pi^*) \leq \varepsilon$.

The resulting policy π^* comes with a PAC guarantee [Valiant, 1984; Ashok *et al.*, 2019]: In *finite time* and with probability at least δ, we return a policy that is ε-optimal.

Advantages of miss-MDPs over POMDPs. For computing ε-optimal policies in POMDPs, knowledge of the observation function is required. However, for arbitrary POMDPs, learning the observation function from a dataset of histories is impossible [Gilbert, 1959; Tune *et al.*, 2013]. In contrast, miss-MDPs exhibit structure in the missingness function M that, for certain types of missingness function, allows for learning M from histories (see also the paragraph "Missingness types in focus" in Section 5). Consequently, miss-MDPs are a *proper and more feasible subclass* of POMDPs.

4.1 Types of Missingness Functions

There are three major types of missingness functions [Rubin, 1976]: *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR). We define these in the context of miss-MDPs. This way, we connect the theory of missing data to POMDPs, which allows us to identify cases in which M can be learned from histories.

The simplest type is MCAR, where the probability of features going missing is independent of the *feature values* of the state. In the doctor-diagnosing example (Figure 1), the temperature feature may be missing due to a loosely attached thermometer. The miss-MDP in Example 1 is MCAR.

Definition 3 (**MCAR**)**.** The missingness function $M\colon S \to \Delta(Z)$ of a miss-MDP $\mathcal{P}$ is MCAR if and only if $\forall r \in R$, $\exists p_r \in [0,1], \forall s \in S, \mathbb{P}\left(f_R(\boldsymbol{z}) = r \mid \boldsymbol{z} \sim M(s)\right) = p_r$.

Admittability and I_{always}. We introduce a notion of admittability that indicates whether an observation z could originate from a state s. We say that z is *admittable* by s, denoted $z \preceq s$, if and only if $\forall i \in I$, $z_i = \perp$ or $z_i = s_i$. In Example 1, we have $(y, \perp) \preceq (y, x)$ and $(y, x) \preceq (y, x)$ but $(x, \perp) \not\preceq (y, x)$. Further, $I_{\text{always}} = \{i \in I \mid \forall s' \in S\colon \mathbb{P}(\boldsymbol{z}_i = \perp \mid \boldsymbol{z} \sim M(s')) = 0\} \subseteq I$ is the set of indices of always observed features, and $I_{\text{mis}} = I \setminus I_{\text{always}}$ is its complement.

Missingness probabilities in MAR depend only on observed state features – in the example of Figure 1, the observed temperature influences the missingness of the heart rate feature. We distinguish two MAR variants: a restricted one we call *simple MAR* [Mohan and Pearl, 2021], and the general one [Rubin, 1976]. For simple MAR, a missingness probability is only influenced by the observable features that *never go missing*, i.e., by z_i for $i \in I_{\text{always}}$. For MAR, a missingness probability is only influenced by the non-missing features of a given *observation*, including features that may go missing. Any MCAR missingness function is also (simple) MAR.

Definition 4 ((Simple) MAR). The missingness function $M\colon S \to \Delta(Z)$ of a miss-MDP $\mathcal{P}$ is:

- **Simple MAR** if and only if for all $s, s' \in S$ that agree on always-observed features (i.e. $\forall i \in I_{\text{always}}$, $s_i = s'_i$), the missingness probability is the same for all missingness indicators $r \in R$, formally: $\mathbb{P}(f_R(\boldsymbol{z}) = r \mid \boldsymbol{z} \sim M(s)) = \mathbb{P}(f_R(\boldsymbol{z}') = r \mid \boldsymbol{z}' \sim M(s'))$.
- **MAR** if and only if for all $s, s' \in S$ and $z \in Z$, if $z \preceq s, s'$, the probability of its indicator $r := f_R(z)$ is equal for states s and s': $\mathbb{P}(f_R(\boldsymbol{z}') = r \mid \boldsymbol{z}' \sim M(s)) = \mathbb{P}(f_R(\boldsymbol{z}'') = r \mid \boldsymbol{z}'' \sim M(s'))$.

Example 2. We redefine M in the miss-MDP from Example 1 to be simple MAR: $M((s_1, x)) = \{(s_1, x) \mapsto 1\}$, and $M((s_1, y)) = \{(s_1, y) \mapsto 0.5, (\perp, y) \mapsto 0.5\}$. Here, the missingness probability of feature 1 depends on the *always* observed value of feature 2. As an example of MAR, which is not simple MAR, consider: $M((s_1, x)) = \{(s_1, x) \mapsto 0.5, (\perp, \perp) \mapsto 0.5\}$, and $M((s_1, y)) = \{(s_1, y) \mapsto 0.25, (\perp, y) \mapsto 0.25, (\perp, \perp) \mapsto 0.5\}$. Here, feature 2 may go missing as well. The missingness probability of feature 1 depends on the value of feature 2, only when it is *observed*!

Definition 5 (MNAR). The missingness function M of a miss-MDP $\mathcal{P}$ is MNAR if and only if it is not MAR.

For MNAR, missingness probabilities may depend on the values of missing features – e.g., in Figure 1 the temperature feature influences its own missingness. In particular, in *self-censoring* missingness functions, a feature's missingness probability depends on its own value.

Example 3. We adapt Example 1 to make M MNAR and self-censoring for feature 2: $M((s_1, x)) = \{(s_1, x) \mapsto 0.5, (s_1, \perp) \mapsto 0.5\}$ and $M((s_1, y)) = \{(s_1, y) \mapsto 0.1, (s_1, \perp) \mapsto 0.9\}$.

4.2 Missingness Graphs

Missingness graphs (m-graphs) visualize the dependencies of missingness functions. We adopt the definition from Mohan and Pearl [2021] for our miss-MDP framework. An m-graph is a causal diagram [Pearl, 1995] in the form of a directed acyclic graph. Its vertices correspond to variables, and the edges correspond to the relationships between the variables.

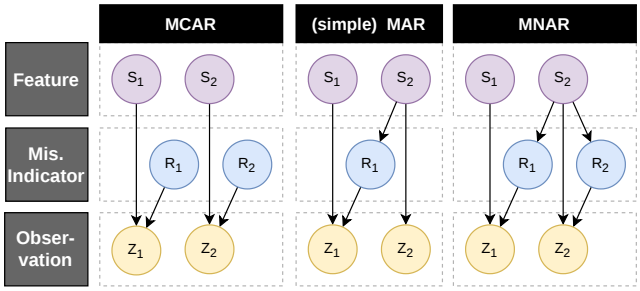

Figure 2: Example missingness graphs visualizing relations between miss-MDP elements for the three types of missingness functions.

Vertices are grouped into three categories: Ⓢ–nodes represent features of the state space, Ⓩ–nodes represent the observation features, and Ⓡ–nodes represent the missingness indicators.[1] We omit the Ⓡ–nodes for always observed features. Arrows between nodes represent a causal relationship: The parent node is a direct cause of the child node. The absence of an edge intuitively indicates that two variables do not directly influence each other; formally, it means they are conditionally independent, given other variables in the graph, according to the d-separation criteria [Pearl, 2009].

Visualizing types of missingness. Figure 2 uses m-graphs to illustrate the conditional independence assumptions of different types of missingness functions. For MCAR, both Ⓡ–nodes are purely stochastic, having no incoming arrows and thus not depending on any feature value. For (simple) MAR, there are two changes: Feature S_2 is always observable (R_2 is absent), and it affects missingness indicator R_1 (red arrow). For MNAR, S_2 can go missing, so R_1 depends on information that can go missing. Note that m-graphs cannot represent *context-specific* independence assumptions, which are needed to, e.g., represent non-simple MAR functions such as the one in Example 2; but the missingness functions we focus on may all be represented by m-graphs. We provide the corresponding m-graphs for all experiments in [Wendland *et al.*, 2026, Appendix D].

5 Approximating Missingness-MDPs

Our goal is to compute ε-optimal policies for miss-MDPs with an unknown M. For this, we first compute an approximation $\widehat{M} \approx M$ from the given dataset $\mathcal{D}$ of histories. This yields an approximated, but fully specified miss-MDP $\widehat{\mathcal{P}}$, which can be solved using any off-the-shelf POMDP solver. Recall that in general, it's impossible to approximate unknown observation functions in POMDPs. Thus, we identify miss-MDPs where approximating M with guarantees is possible.

Missingness types in focus. A necessary condition is that the missingness function can be approximated solely from observations, a property that missing data literature calls *identifi-*

[1] We exclude the set of unobserved features U used in Mohan and Pearl [2021], as in our setting $U = \emptyset$, since M depends on states.

ability [Bhattacharya *et al.*, 2019]. Establishing new identifiability results is not the focus of this paper. Instead, we provide PAC guarantees for types that are known to be identifiable. Thus, we focus on missingness functions of type (1) MCAR, (2) simple MAR and (3) non-self-censoring MNAR with independent missingness indicators. Additionally, in Section 6, we experiment on non-identifiable MNAR.

Outline. Remark 1 explains the necessity of approximating the missingness function for planning in miss-MDPs and why we do not use imputation methods. Remark 2 presents an interesting insight orthogonal to our problem: For maintaining a belief during policy execution, certain types of missingness can in fact be *ignored*. Sections 5.1 and 5.2 describe our algorithms for approximating missingness functions. Both are structured as follows: First state assumptions, define how to compute $\widehat{M}$, prove that the approximation is probably approximately correct, and explain how to utilize additional knowledge on the missingness function to reduce sample complexity. Section 5.3 uses these algorithms to compute near-optimal policies by means of the approximated missingness functions.

Remark 1 (Why estimate the missingness function?)**.** In the missing data literature, imputation methods [Dempster *et al.*, 1977; Van den Broeck *et al.*, 2015] recover $\mathbb{P}(s \mid z)$, the conditional distribution of the state given the observation. Using this distribution, one can infer a distribution over the states to mitigate missingness. This quantity does not suffice for planning under partial observability. We require the reverse: $\mathbb{P}(z \mid s)$, the missingness function M, which specifies the likelihood of each observation given the underlying state. While the two conditional distributions are related by Bayes' rule, leveraging imputation methods would require estimating both marginal distributions $\mathbb{P}(s)$ and $\mathbb{P}(z)$. However, our methods require to compute belief updates which in turn requires M: the belief transition $\mathbb{P}(b' \mid b, a)$ for a belief $b(s)$ over states depends on $\mathbb{P}(z \mid s)$. Further, belief-based planning methods such as SARSOP [Kurniawati *et al.*, 2008] also require the missingness function to maintain and update beliefs.

Remark 2 (Ignorability)**.** Missing data literature defines *ignorability* as cases where any quantity of interest can be consistently estimated from observations alone, and it is not necessary to model the missingness process [Little and Rubin, 2019]. This holds under MCAR, and also under MAR, whenever the quantity depends only on the observed features. We identify a similar notion of ignorability for miss-MDPs: If the missingness function M is MAR (and therefore including MCAR), then belief updates can be computed without knowledge of the precise probabilities of M, since these cancel out in Bayes' rule; see [Wendland *et al.*, 2026, Appendix B] for a formal proof. Thus, MAR missingness is ignorable for maintaining a belief when executing a policy in a miss-MDP. However, we stress that *the missingness function remains a requirement for computing belief-based policies*, see Remark 1.

Occurrence counts. Both algorithms extract the number of occurrences of every observation using the dataset $\mathcal{D} = (h_1, \dots, h_k)$ of k histories $h_i \in \mathcal{H}$. For each $h_i = \big(z^{(0)}, a^{(0)}, \dots, z^{(l)}, a^{(l)}\big)$, we denote the j-th observation $z^{(j)}$ by $h_i^{(j)}$. The number of occurrences of an observation $z \in Z$ is: $\#_{\mathcal{D}}(z) = \sum_{i=1}^{k} \sum_{j=0}^{|h_i|} \mathbf{1}_{h_i^{(j)}=z}$. To lift occurrences for a set $Z' \subseteq Z$, we define $\#_{\mathcal{D}}\big(Z'\big) = \sum_{z \in Z'} \#_{\mathcal{D}}(z)$.

5.1 Approximating M for MCAR and Simple MAR

If a missingness function is of type simple MAR, we can approximate it using the *approximation for simple MAR* algorithm, `AsMAR`. The modifications to obtain the algorithm for the more restricted MCAR-type functions, `AMCAR`, are described at the end of the section.

Always-observed features. Based on the dataset $\mathcal{D}$, we partition the feature indices I into those that are always observed and those that can go missing: $\hat{I}_{\text{always}} = \{i \in I \mid \#_{\mathcal{D}}(\{z \in Z \mid z_i = \perp\}) = 0\}$ and $\hat{I}_{\text{mis}} = I \setminus \hat{I}_{\text{always}}$, respectively. Note, this partitioning is based on empirical data ($\hat{I}_{\text{always}} \supseteq I_{\text{always}}$) and we might misclassify a feature index to be in $\hat{I}_{\text{always}}$ even though it can go missing.

Computing $\widehat{M}$. We use the fact that M can be seen as a mapping $S \to \Delta(R)$ (see paragraph "Missingness indicators", Section 4). Consequently, for every state, we want to approximate the probability of a certain vector of missingness indicators. The simple MAR assumption tells us that the probabilities may only depend on the features in $\hat{I}_{\text{always}}$, meaning that $M(z \mid s) = M(z \mid s')$ when $s_i = s'_i$ for the always observed features $i \in I_{\text{always}}$. Thus, for all possible values of the always-observable features of a state $s \in S$ and missingness indicator vector $r \in R$, we can compute the occurrence count $\#_{\mathcal{D}}(s, r) = \#_{\mathcal{D}}\left(Z_s^r\right)$, where

$$Z_s^r = \left\{ z \in Z \;\middle|\; \forall i \in I\colon \begin{array}{l} (i \in \hat{I}_{\text{always}} \implies z_i = s_i) \\ \text{and} (r_i = 0 \implies z_i = \perp) \end{array} \right\}.$$

Using $\#_{\mathcal{D}}(s, r)$, we obtain $\widehat{M}(z \mid s)$ as the fraction of observing $(s, f_R(z))$ and the sum of counts for s and all possible missingness indicators values:

$$\widehat{M}(z \mid s) = \frac{\#_{\mathcal{D}}(s, f_R(z))}{\sum_{r \in R} \#_{\mathcal{D}}(s, r)}. \tag{1}$$

Probably approximately correct. With enough data, our approach yields an arbitrarily precise approximation of the true missingness function. We formalize this in Theorem 1 as a PAC guarantee, not only proving that it becomes ε-precise for every $\varepsilon > 0$, but that we can also bound the probability of an error (through unlucky sampling). Additionally, we can adapt the claim to bound the imprecision of the resulting $\widehat{M}$ for a given dataset. The proof is provided in [Wendland *et al.*, 2026, Appendix C.2].

Theorem 1 (PAC guarantee for `AsMAR`)**.** Let $\mathcal{P}$ be a miss-MDP with a simple MAR missingness function. For every precision ε and confidence threshold δ, there exists a number n^*, such that every dataset $\mathcal{D}$ of n^* histories satisfies the following: With probability at least δ, for $\widehat{M}$ computed on $\mathcal{D}$ according to Equation (1), for all reachable states $s \in S$ and observations $z \in Z$, we have $|\widehat{M}(z \mid s) - M(z \mid s)| \leq \varepsilon$. The number n^* is polynomial in $1/\varepsilon$, $1/\delta$, and the size of the miss-MDP (in particular the smallest probability m to visit a state under the sampling policy). Dually, for a dataset $\mathcal{D}$ and confidence threshold δ, there exists an ε such that

with probability at least δ, for all reachable states $s \in S$ and observations $z \in Z$, we have the same inequality, i.e. $|\widehat{M}(z \mid s) - M(z \mid s)| \leq \varepsilon$.

Using additional assumptions on the missingness function. Beyond the necessary simple MAR assumption, we can exploit additional assumptions to improve the approximation of M for the same $\mathcal{D}$. Consider a feature i that is always observable, but does not affect the missingness probability of other features. We can exclude such i from $\hat{I}_{\text{always}}$, combining the occurrence counts of states that differ only in this feature. Hence, if we assume M to be MCAR, $\hat{I}_{\text{always}}$ reduces to an empty set. Consequently, we get that $\#_{\mathcal{D}}(s, r)$ is independent of s, and we then only count occurrences of missingness indicators, resulting in the algorithm `AMCAR`. We prove the correctness of these improvements in [Wendland *et al.*, 2026, Appendix C.2]. In Section 6, we empirically show the improved precision of $\widehat{M}$ estimated from the same $\mathcal{D}$ with this approach.

5.2 Approximating M With Independent Missingness Indicators

This section presents the *approximation for independent missingness indicators* algorithm, `AIMI`. It may be applied to M of types MCAR, simple MAR as well as a subset of identifiable MNAR which satisfy the following requirements:

1. **Independence of missingness indicators:** The fact that one feature is missing must not influence the missingness-probability of any other feature. Formally, for $s \in S$ and $z \in Z$, $\mathbb{P}(\boldsymbol{z} \mid \boldsymbol{z} \sim M(s)) = \Pi_{i \in I} \mathbb{P}(\boldsymbol{z}_i \mid \boldsymbol{z} \sim M(s))$.
2. **No self-censoring:** Intuitively, a feature may not influence its own missingness probabilities. Formally, for all $i \in I$ and every pair of states $s, s' \in S$ that differ only in the i-th feature ($s_i \neq s'_i$, but for all $j \neq i$ we have $s_j = s'_j$) we have $\mathbb{P}(\boldsymbol{z}_i = \perp \mid \boldsymbol{z} \sim M(s)) = \mathbb{P}(\boldsymbol{z}_i = \perp \mid \boldsymbol{z} \sim M(s'))$.
3. **Positivity:** Intuitively, if a feature affects the missingness probabilities of other features, we need to observe its value to learn the missingness probabilities. However, this is impossible if it always misses. Therefore, we require a *positivity assumption* [Hernán and Robins, 2020]: For all $i \in I$ and $s \in S$, we have $\mathbb{P}(\boldsymbol{z}_i \neq \perp \mid \boldsymbol{z} \sim M(s)) > 0$.

Computing $\widehat{M}$. We compute the occurrence count for every state $s \in S$, feature $i \in I$ and value of a corresponding i-th missingness indicator $r_i \in \{0, 1\}$ as $\#_{\mathcal{D}}(s, i, r_i) = \#_{\mathcal{D}}(Z_s^{i, r_i})$, where

$$Z_s^{i,r_i} = \left\{ z \in Z \;\middle|\; \forall j \in I \setminus \{i\} \colon \begin{array}{l} (z_j = s_j) \text{ and} \\ (r_i = 0 \iff z_i = \perp) \end{array} \right\}.$$

By positivity, a large enough dataset almost surely contains observations to make the counters non-zero (i.e. for all s and i, we have $\#_{\mathcal{D}}(s, i, 0) + \#_{\mathcal{D}}(s, i, 1) > 0$). The probability of a non self-censoring feature i depends only on the other features $j \in I \setminus \{i\}$. Finally, using the independence assumption, we can infer $\widehat{M}$ by taking the product of the individual missingness probabilities of all features.

$$\widehat{M}(z \mid s) = \prod_{i \in I} \frac{\#_{\mathcal{D}}(s, i, f_R(z)_i)}{\#_{\mathcal{D}}(s, i, 0) + \#_{\mathcal{D}}(s, i, 1)}. \tag{2}$$

Probably approximately correct. In [Wendland *et al.*, 2026, Appendix C.3], we prove Theorem 2 that provides the same kind of guarantee as in Theorem 1; the only differences are the assumptions on the missingness function and the approach for calculating $\widehat{M}$.

Theorem 2 (PAC guarantee for `AIMI`)**.** Let $\mathcal{P}$ be a miss-MDP where the missingness function satisfies independence, non-self-censoring, and positivity. Then, $\widehat{M}$ computed using Equation (2) offers the same PAC guarantees as in Theorem 1.

Using additional assumptions on the missingness function. In general, `AIMI` maintains a counter for every combination of the feature valuations of other features $j \in I \setminus \{i\}$. If we know that a certain feature j does not affect the missingness probability of i – there is no edge between the j-th Ⓢ-node and the i-th Ⓡ-node – we merge the counters for all values of the j-th feature. This knowledge comes from **(a)** an m-graph, **(b)** assuming simple MAR while observing feature j goes missing in $\mathcal{D}$, or **(c)** assuming MCAR, in which case we drop the dependency on s in the counters. We prove in [Wendland *et al.*, 2026, Appendix C.3] that all these modifications retain the PAC guarantees.

5.3 Computing a Policy With the Approximations

Here, we provide the final step of our problem statement: computing a PAC policy from the approximated miss-MDP. We show in [Wendland *et al.*, 2026, Appendix C.4] that after finitely many samples, $\widehat{M}$ is accurate enough to yield an ε-optimal policy. We highlight that learning $\widehat{M}$ to precision ε is insufficient, as the errors in $\widehat{M}$ may aggregate when solving the miss-MDP.

Theorem 3 (Computing ε-optimal Policies)**.** Let $\mathcal{P}$ be a miss-MDP with a missingness function that is simple MAR or that satisfies independence, no self-censoring, and positivity. Assume we can sample histories collected under a fair policy, and we know a lower bound on the smallest missingness probability $p \leq \min_{s \in S, z \in Z} M(z \mid s)$. Then, for every given precision ε and confidence threshold δ, by sampling a number of histories polynomial in ε, δ, and the size of the miss-MDP, we can compute a policy π^* such that with probability at least δ it is ε-optimal, i.e. $(\sup_\pi V_{\mathcal{P}}(\pi)) - V_{\mathcal{P}}(\pi^*) \leq \varepsilon$.

Practical considerations. Theorem 3 concerns asymptotic convergence to an ε-optimal policy, providing the theoretical foundation of our approach. In practice, the required number of samples is very large, and we work with datasets that are not necessarily sufficient to provide the ε-optimality guarantees. Datasets of limited size may cause a practical problem: For an observation z with $\#_{\mathcal{D}}(s, f_R(z)) = 0$, for any $s \in S$ we obtain $\widehat{M}(z \mid s) = 0$, leading to a division by zero for s when performing the belief update τ. We circumvent this case by setting $\#_{\mathcal{D},\kappa}(s, r) = \#_{\mathcal{D}}(s, r) + \kappa$, i.e. we add a small $\kappa > 0$ to every count. The influence of κ diminishes with an increasing dataset size $|\mathcal{D}|$.

6 Experiments

We conduct an experimental evaluation to test our theoretical results and to compare our methods against several baselines. We aim to answer the following questions:

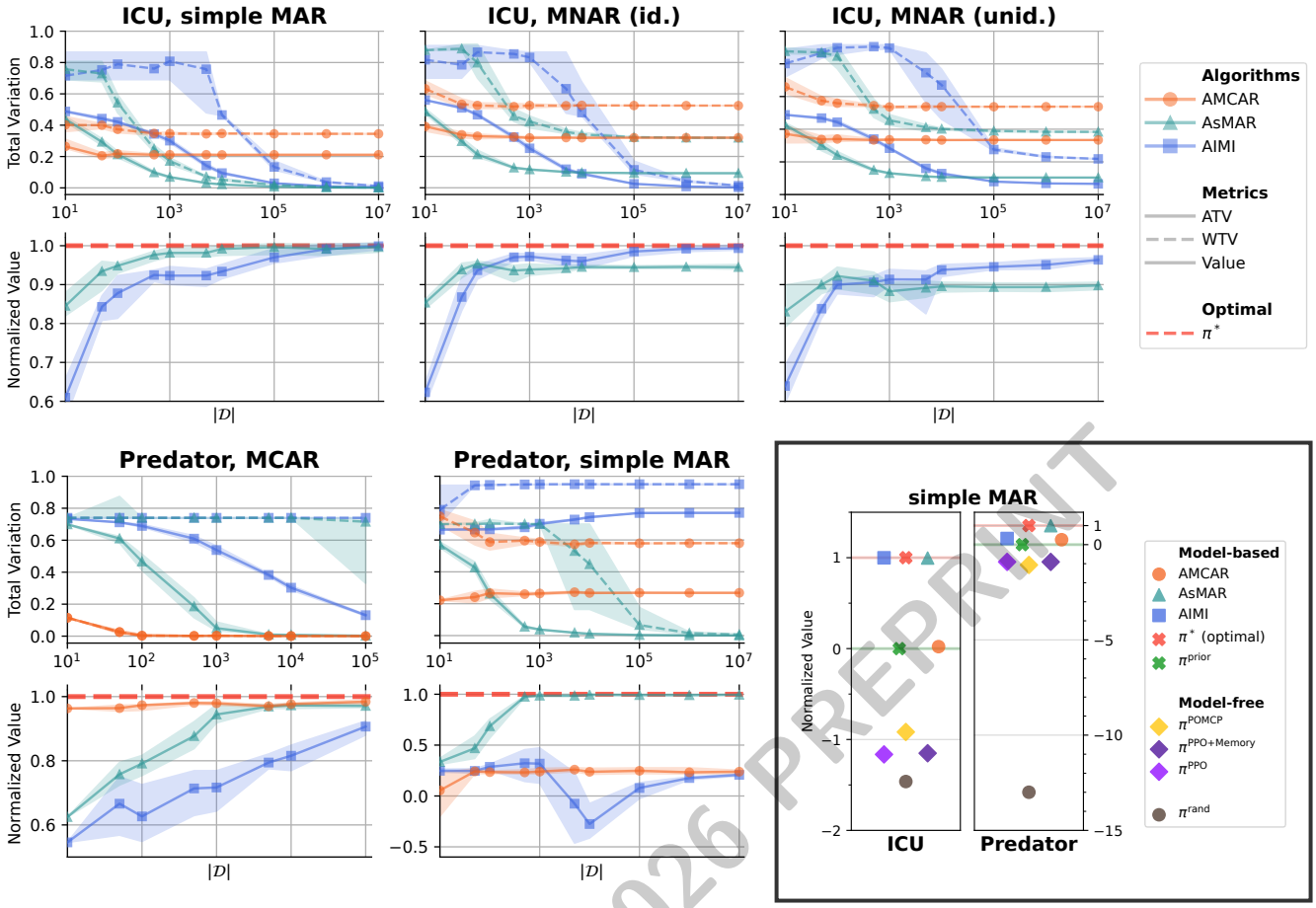

Figure 3: **Left** (and top)**:** Empirical results for the *ICU* and *Predator* benchmarks, with average/worst TV (ATV/WTV) of $\widehat{M}$ and resulting policy values $V_{\mathcal{P}}(\pi)$. Values are normalized so that 1 corresponds to the *optimal* policy π^* using the true M and 0 to the *prior* policy π^{prior}. **Bottom right:** Comparison of normalized values on ICU_{sMAR} and $Predator_{sMAR}$ between model-based and the model-free baseline methods.

Q1. Do our proposed learning algorithms adequately approximate the missingness function?

Q2. How do (in)correct assumptions on the missingness function affect the approximation?

Q3. With increasing dataset size, does the value of the policy computed on the approximated miss-MDP converge to that of the optimal policy on the true miss-MDP?

Q4. How does the value computed from the approximated miss-MDP compare against model-free baselines?

Below, we first describe the experimental setup. Then, we analyze the results in the order of the above questions.

Benchmarks. We consider two environments with varying missingness types: *ICU* models a doctor treating a patient, whose vital measurements are not always available [Johnson *et al.*, 2022], and in *Predator*, a variant of the Tag environment [Pineau *et al.*, 2003], a predator chases a partially hidden prey. For our benchmarks, we consider the following four missingness types: (1) *MCAR*, (2) *sMAR*, a simple MAR function, (3) *MNAR (id.)*, an identifiable MNAR function (fulfilling the required assumptions of `AIMI`), and (4) *MNAR (unid.)*, an unidentifiable MNAR function with self-censoring. For *Predator* and all missingness functions, the (x, y)-coordinates of the prey always go missing jointly, i.e. the missingness indicators are dependent; for *ICU*, the missingness indicators are independent. *ICU* has 800 reachable states and *Predator* up to 1311 [Wendland *et al.*, 2026, Appendix D].

Setup. For a range of dataset sizes $|\mathcal{D}|$, we collect data using the uniform random policy π^{rand} with $\forall a \in A, \pi^{\text{rand}}(a \mid \cdot) = 1/|A|$, and compute the estimate $\widehat{M} \approx M$ using our proposed algorithms: `AMCAR`, `AsMAR`, and `AIMI` (Section 5). Each $\widehat{M}$ yields an approximated miss-MDP $\widehat{\mathcal{P}}$, for which we compute a policy $\hat{\pi}$ using the POMDP solver SARSOP [Kurniawati *et al.*, 2008]. We consider the following baselines: **(1)** *optimal*: the SARSOP policy π^* computed for the true M (the upper bound); **(2)** *prior*: the SARSOP policy π^{prior} with an uninformed prior of M, where each feature independently goes missing with probability 0.5; **(3)** *POMCP* [Silver and Veness, 2010]: generating an online POMDP planning policy π^{POMCP}; **(4)** *PPO* [Schulman *et al.*, 2017]: the poli-

cies $\pi^{\text{PPO+Memory}}$ and π^{PPO}, with and without memory. We group π^{PPO} and π^{POMCP} as *model-free* baselines as they rely solely on simulation access to the miss-MDP. We provide more extensive details of the setup in [Wendland *et al.*, 2026, Appendix D]. Moreover, we publicly provide the code of our implementation[2]. For further details on reproducibility, see [Wendland *et al.*, 2026, Appendix A].

Metrics. For every dataset size and method, we perform 20 independent runs and report the average together with the interquartile range (shaded area) of the following metrics:

1. To empirically assess the quality of the approximation $\widehat{M}$ compared to the true M, we compute the **total variation** (TV) of the distributions at a state $s \in S$ as $TV(s) = \frac{1}{2}\sum_{z \preceq s} \left|\widehat{M}(z \mid s) - M(z \mid s)\right|$. We aggregate the TV across states by the average TV (ATV): $1/|S| \sum_s TV(s)$, and the worst TV (WTV): $\max_s TV(s)$.
2. We assess how the various $\hat{\pi}$ produced by the algorithms perform on the true miss-MDP $\mathcal{P}$ by comparing their **value** $V_{\mathcal{P}}(\hat{\pi})$ to $V_{\mathcal{P}}(\pi^{\text{prior}})$ and the optimum $V_{\mathcal{P}}(\pi^*)$. All policy values are normalized s.t. 1 and 0 correspond to the values of the *optimum* and *prior* baselines, respectively.

Results. The experimental results in Figure 3 (left, top) present how the TV of the approximated missingness function $\widehat{M}$ and the value of the associated policy $\hat{\pi}$ evolve with dataset size $|\mathcal{D}|$. Figure 3 (right) compares our algorithms to several baselines for selected benchmarks. We report the best baseline performance across multiple seeds. We provide baseline results for all benchmarks as well as results for an additional $\textit{Predator}_{\textit{MNAR (unid.)}}$ benchmark in [Wendland *et al.*, 2026, Appendix D].

Q1: With sufficient data, our algorithms adequately approximate the missingness function. Under the appropriate missingness assumptions, each algorithm learns the missingness function (bringing the TV near zero): `AMCAR` learns the exact missingness function in $\textit{Predator}_{\textit{MCAR}}$ within 100 observations. We observe similar results for `AsMAR` in $\textit{ICU}_{\textit{sMAR}}$ and $\textit{Predator}_{\textit{sMAR}}$, and for `AIMI` in $\textit{ICU}_{\textit{MNAR (id.)}}$.

Q2: The assumptions on the missingness function significantly affect the quality of the approximation. Under correct assumptions on the missingness function, the TVs of the approximated $\widehat{M}$ approach zero (e.g. `AIMI` on $\textit{ICU}_{\textit{sMAR}}$ and $\textit{ICU}_{\textit{MNAR (id.)}}$). Further, relaxations on the missingness (e.g. assuming MAR despite MCAR missingness) allows the approximation of M, but may lead to significant data inefficiency (see `AMCAR` vs. `AsMAR` on $\textit{Predator}_{\textit{MCAR}}$). Breaking the missingness assumptions results in converging to a poor WTV (see `AIMI` on all Predator benchmarks).

Q3: Convergence to the optimal policy follows the quality of the approximation. With a sufficiently accurate approximation of the missingness function, the values of our method's policies converge to the optimal policy's π^* values (e.g. `AIMI` on $\textit{ICU}_{\textit{sMAR}}$ and $\textit{ICU}_{\textit{MNAR (id.)}}$, `AMCAR` and `AsMAR` on $\textit{Predator}_{\textit{MCAR}}$, `AsMAR` on $\textit{Predator}_{\textit{sMAR}}$). Despite the missingness of $\textit{ICU}_{\textit{MNAR (unid.)}}$ not adhering to the necessary assumptions, `AIMI` approximates the relevant parts of M well enough to yield a policy that is considerably close to π^*.

Q4: Baseline methods, including PPO and POMCP, are not competitive. Figure 3 (right) compares our algorithms to several baselines. Our algorithms, under correct missingness assumptions (see `AsMAR` or `AIMI` for *ICU*, or `AsMAR` for *Predator*), consistently outperform policies derived from an uninformed prior over the missingness function (π^{prior}). Meanwhile, all model-free approaches—π^{PPO}, $\pi^{\text{PPO+Memory}}$, and π^{POMCP}—significantly underperformed on both benchmarks. We attribute this gap to two factors. First, model-free methods solely access the transition function through simulation. Second, in a separate experiment without missingness, these methods successfully learned the optimal policy for the fully observable MDP. This indicates that the poor performance stems from these methods' inability to effectively handle the specific type of partial observability of miss-MDPs. We attribute the better performance (compared to the random policy π^{rand}) of the model-free approaches in *Predator*, in contrast to *ICU*, to the lower stochasticity of the underlying transition function of *Predator*.

7 Conclusion

Miss-MDPs integrate the theory of missing data with decision-making under uncertainty. From data generated by a miss-MDP with known underlying MDP, we approximate the unknown missingness function, which – under assumptions about its type – enables the computation of ε-optimal policies. We show that incorrect assumptions about the missingness type may lead to misspecified models and suboptimal policies. Our experiments support the theory and demonstrate the practical benefits of our approach, highlighting the superiority over model-free baselines in our problem setting. Future work may lift the assumption of a known transition function and extend miss-MDPs to the more general setting of miss-POMDPs.

Acknowledgements

This work was supported by the European Research Council (ERC) Starting Grant 101077178 (DEUCE), BUT Internal Grant Agency (FIT-S-26-9011), Czech Science Foundation (GA25-18318S), and by the project VASSAL: "Verification and Analysis for Safety and Security of Applications in Life" which is funded by the European Union under Horizon Europe WIDERA Coordination and Support Action/Grant Agreement No. 10116002.

References

[Allman *et al.*, 2009] Elizabeth S Allman, Catherine Matias, and John A Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, 37(6A):3099–3132, 2009.

[Ashok *et al.*, 2019] Pranav Ashok, Jan Kretínský, and Maximilian Weininger. PAC statistical model checking for Markov decision processes and stochastic games. In *CAV (1)*, volume 11561 of *Lecture Notes in Computer Science*, pages 497–519. Springer, 2019.

[2] https://github.com/ai-fm/missingness-mdps

[Åström, 1965] Karl Johan Åström. Optimal control of Markov processes with incomplete state information. *J. Math. Anal. Appl.*, 10(1):174–205, 1965.

[Bhattacharya *et al.*, 2019] Rohit Bhattacharya, Razieh Nabi, Ilya Shpitser, and James M. Robins. Identification in missing data models represented by directed acyclic graphs. In *UAI*, Proceedings of Machine Learning Research, pages 1149–1158. AUAI Press, 2019.

[Buuren, 2018] Stef van Buuren. *Flexible imputation of missing data*. CRC Press, Taylor and Francis Group, 2018.

[Böck *et al.*, 2022] Markus Böck, Julien Malle, Daniel Pasterk, Hrvoje Kukina, Ramin Hasani, and Clemens Heitzinger. Superhuman performance on sepsis MIMIC-III data by distributional reinforcement learning. *PLOS ONE*, 17(11):e0275358, November 2022.

[Chatterjee *et al.*, 2016] Krishnendu Chatterjee, Martin Chmelík, Raghav Gupta, and Ayush Kanodia. Optimal cost almost-sure reachability in POMDPs. *Artificial Intelligence*, 234:26–48, 2016.

[Chen *et al.*, 2023] Minshuo Chen, Yu Bai, H. Vincent Poor, and Mengdi Wang. Efficient RL with impaired observability: Learning to act with delayed and missing state observations. In *NeurIPS*, 2023.

[Dempster *et al.*, 1977] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[Finesso, 1990] Lorenzo Finesso. *Consistent estimation of the order for Markov and hidden Markov chains*. University of Maryland, College Park, 1990.

[Futoma *et al.*, 2020] Joseph Futoma, Michael C. Hughes, and Finale Doshi-Velez. POPCORN: partially observed prediction constrained reinforcement learning. In *AISTATS*, Proceedings of Machine Learning Research, pages 3578–3588. PMLR, 2020.

[Gilbert, 1959] Edgar J Gilbert. *On the identifiability problem for functions of finite Markov chains*. Sandia Corporation, 1959.

[Golowich *et al.*, 2022] Noah Golowich, Ankur Moitra, and Dhruv Rohatgi. Learning in observable POMDPs, without computationally intractable oracles. In *NeurIPS*, pages 1458–1473, 2022.

[Guo *et al.*, 2016] Zhaohan Daniel Guo, Shayan Doroudi, and Emma Brunskill. A PAC RL algorithm for episodic POMDPs. In *AISTATS*, JMLR Workshop and Conference Proceedings, pages 510–518, 2016.

[Hauskrecht, 2000] Milos Hauskrecht. Value-function approximations for partially observable Markov decision processes. *JAIR*, 13:33–94, 2000.

[Hernán and Robins, 2020] Miguel A Hernán and James M Robins. *Causal Inference: What If*. CRC Press, 2020.

[Johnson *et al.*, 2022] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV, 2022.

[Komorowski *et al.*, 2018] Matthieu Komorowski, Leo A. Celi, Omar Badawi, Anthony C. Gordon, and A. Aldo Faisal. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24:1716–1720, 2018.

[Krishnamurthy *et al.*, 2016] Akshay Krishnamurthy, Alekh Agarwal, and John Langford. PAC reinforcement learning with rich observations. In *NIPS*, pages 1840–1848, 2016.

[Kurniawati *et al.*, 2008] Hanna Kurniawati, David Hsu, and Wee Sun Lee. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *RSS*. MIT Press, 2008.

[Little and Rubin, 2019] Roderick Little and Donald Rubin. *Statistical Analysis with Missing Data, Third Edition*. Wiley Series in Probability and Statistics. Wiley, 2019.

[Liu *et al.*, 2022] Zeyu Liu, Anahita Khojandi, Xueping Li, Akram Mohammed, Robert L Davis, and Rishikesan Kamaleswaran. A machine learning-enabled partially observable Markov decision process framework for early sepsis prediction. *INFORMS Journal on Computing*, 34(4):2039–2057, July 2022.

[Lizotte *et al.*, 2008] Daniel J Lizotte, Lacey Gunter, Eric Laber, and Susan A Murphy. Missing data and uncertainty in batch reinforcement learning. In *NIPS*, 2008.

[Madani *et al.*, 2003] Omid Madani, Steve Hanks, and Anne Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artif. Intell.*, 147(1-2):5–34, 2003.

[Mohan and Pearl, 2021] Karthika Mohan and Judea Pearl. Graphical models for processing missing data. *J. Am. Stat. Assoc.*, 116(534):1023–1037, 2021.

[Mohan *et al.*, 2013] Karthika Mohan, Judea Pearl, and Jin Tian. Graphical models for inference with missing data. In *NIPS*, pages 1277–1285, 2013.

[Mossel and Roch, 2005] Elchanan Mossel and Sébastien Roch. Learning nonsingular phylogenies and hidden Markov models. In *STOC*, pages 366–375. ACM, 2005.

[Nabi *et al.*, 2020] Razieh Nabi, Rohit Bhattacharya, and Ilya Shpitser. Full law identification in graphical models of missing data: Completeness results. In *ICML*, Proceedings of Machine Learning Research, pages 7153–7163. PMLR, 2020.

[Pearl, 1995] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

[Pearl, 2009] Judea Pearl. *Causality*. Cambridge University Press, 2nd edition, 2009.

[Pineau *et al.*, 2003] Joelle Pineau, Geoffrey J. Gordon, and Sebastian Thrun. Point-based value iteration: An anytime algorithm for POMDPs. In *IJCAI*, pages 1025–1032. Morgan Kaufmann, 2003.

[Puterman, 1994] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics. Wiley, 1994.

[Rubin, 1976] Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[Schafer and Graham, 2002] Joseph L. Schafer and John W. Graham. Missing data: our view of the state of the art. *Psychological Methods*, 7(2):147–177, June 2002.

[Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017.

[Shim *et al.*, 2018] Hajin Shim, Sung Ju Hwang, and Eunho Yang. Joint active feature acquisition and classification with variable-size set encoding. In *NeurIPS*, pages 1375–1385, 2018.

[Shpitser *et al.*, 2015] Ilya Shpitser, Karthika Mohan, and Judea Pearl. Missing data as a causal andprobabilistic problem. In *UAI*, pages 802–811, 2015.

[Silver and Veness, 2010] David Silver and Joel Veness. Monte-Carlo planning in large POMDPs. In *NIPS*, pages 2164–2172. Curran Associates, Inc., 2010.

[Spaan, 2012] Matthijs T. J. Spaan. Partially observable markov decision processes. In *Reinforcement Learning*, Adaptation, Learning, and Optimization, pages 387–414. Springer, 2012.

[Terwijn, 2002] Sebastiaan Terwijn. On the learnability of hidden Markov models. In *ICGI*, volume 2484 of *Lecture Notes in Computer Science*, pages 261–268. Springer, 2002.

[Tsiatis, 2006] Anastasios A Tsiatis. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer, 2006.

[Tune *et al.*, 2013] Paul Tune, Hung X Nguyen, and Matthew Roughan. Hidden Markov model identifiability via tensors. In *2013 IEEE International Symposium on Information Theory*, pages 2299–2303. IEEE, 2013.

[Uehara *et al.*, 2023] Masatoshi Uehara, Ayush Sekhari, Jason D. Lee, Nathan Kallus, and Wen Sun. Computationally efficient PAC RL in POMDPs with latent determinism and conditional embeddings. In *ICML*, Proceedings of Machine Learning Research, pages 34615–34641. PMLR, 2023.

[Valiant, 1984] Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[Van den Broeck *et al.*, 2015] Guy Van den Broeck, Karthika Mohan, Arthur Choi, Adnan Darwiche, and Judea Pearl. Efficient algorithms for Bayesian network parameter learning from incomplete data. In *UAI*, pages 161–170. AUAI Press, 2015.

[Wang *et al.*, 2019] Yuhui Wang, Hao He, and Xiaoyang Tan. Robust reinforcement learning in POMDPs with incomplete and noisy observations. *CoRR*, abs/1902.05795, 2019.

[Wendland *et al.*, 2026] Joshua Wendland, Markel Zubia, Roman Andriushchenko, Maris F. L. Galesloot, Milan Češka, Henrik von Kleist, Thiago D. Simão, Maximilian Weininger, and Nils Jansen. Missingness-MDPs: Bridging the theory of missing data and POMDPs. *CoRR*, abs/2605.12262, 2026.

[Xiong *et al.*, 2022] Yi Xiong, Ningyuan Chen, Xuefeng Gao, and Xiang Zhou. Sublinear regret for learning POMDPs. *Production and Operations Management*, 31(9):3491–3504, 2022.

[Yamaguchi *et al.*, 2020] Nobuhiko Yamaguchi, Osamu Fukuda, and Hiroshi Okumura. Model-based reinforcement learning with missing data. In *CANDAR (Workshops)*, pages 168–171, 2020.

[Yoon *et al.*, 2019] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. ASAC: active sensing using actor-critic models. In *MLHC*, volume 106 of *Proceedings of Machine Learning Research*, pages 451–473. PMLR, 2019.