EFFECTIVE SPARSITY IN DISTRIBUTION AND ERROR ESTIMATION AND OFF-POLICY EVALUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating a discrete probability distribution from a finite sample is a fundamental problem in machine learning and statistics. Classical guarantees typically depend on the alphabet size K or the observed support $\mathbb{T}$, although in many applications the quantity of interest depends only on a structured, data-dependent combination of distributional errors. We identify *effective sparsity* as a useful notion of complexity for such quantities and develop a concentration bound that adapts to the residual structure of data-dependent coefficients while separately accounting for the probability mass of unseen categories. *Our bound removes the direct dependence on $\sqrt{K}$ and $\sqrt{|\mathbb{T}|}$*, which can be crucial in high-dimensional problems with approximately sparse structure. We further demonstrate substantial improvements in two important problems. First, in generalization analysis via Algorithmic Robustness, our bound replaces the $\sqrt{K}$ or $\sqrt{|\mathbb{T}|}$ dependence in existing guarantees by the effective residual mass of local model errors. Experiments on 22 pretrained ImageNet models show that our uncertainty term is consistently substantially smaller than the recent robustness-based bound. Second, in off-policy evaluation for finite-horizon confounded partially observable decision processes, our analysis replaces the exponential dependence on the observation-action history space in existing first-order error bounds by polynomial dependence on the horizon under realistic effective-sparsity conditions, *yielding an exponential improvement in the dependence on history size*. These results establish effective sparsity as a general mechanism for obtaining substantially sharper distribution and error estimation guarantees in high-dimensional problems.

1 INTRODUCTION

Estimating a discrete probability distribution from a finite sample is a fundamental problem in machine learning and statistics. Let $\boldsymbol{p} = (p_1, \ldots, p_K)$ be an unknown distribution over a finite alphabet and $\hat{\boldsymbol{p}}$ its empirical estimator from n i.i.d. samples. Classical analyses measure estimation error by $\|\boldsymbol{p} - \hat{\boldsymbol{p}}\|_1$ (Cohen et al., 2020; Qian et al., 2020), relative entropy (Mourtada, 2026), or $\|\boldsymbol{p} - \hat{\boldsymbol{p}}\|_\infty$ (Kontorovich & Painsky, 2025). For ℓ_1 estimation, standard distribution-free bounds scale as $O(\sqrt{K/n})$, which can be pessimistic when the alphabet is large but the distribution or the downstream quantity of interest has substantially lower effective complexity. This motivates us to study the more general data-dependent weighted deviation $\sum_{i=1}^{K} f_i(p_i - \hat{p}_i)$, where the weights f_i may depend on the observed samples.

Our first contribution is a new concentration theorem for this data-dependent weighted deviation. Existing bounds primarily depend on the ambient alphabet size K or, more recently, the number of observed categories $\mathbb{T}$. In particular, Kawaguchi et al. (2022) obtain a data-dependent bound with leading dependence $O(\sqrt{|\mathbb{T}|(\ln K)/n})$. While this improves over the classical $O(\sqrt{K/n})$ rate (Qian et al., 2020; Xu & Mannor, 2012; Cohen et al., 2020) when $|\mathbb{T}| \ll K$, the improvement can disappear when $|\mathbb{T}| \approx K$, and the heterogeneous structure of the weights is exploited only through coarse summary quantities. Our theorem instead adapts directly to the realized weights and separates fluctuations over observed cells from the probability mass of unseen cells. Since $\sum_i (p_i - \hat{p}_i) = 0$, its centered form naturally leads to the effective residual mass

$$s_{\text{eff}} = \inf_{c \in \mathbb{R}} \sum_{i \in \mathbb{T}} |f_i - c|. \tag{1}$$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

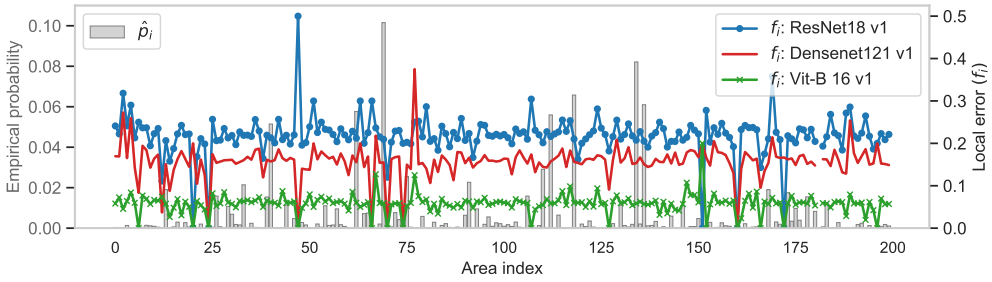

Figure 1: Empirical distribution in ImageNet and local errors (f_i) of three PyTorch pretrained models (ResNet, DenseNet, ViT) in different areas of the input space, where $K = 200$.

We show that the leading deviation scales as $O(s_{\text{eff}}/\sqrt{n})$, together with a separate missing-mass term. Hence, the bound can be essentially independent of K even when the observed support is large. This captures a broad notion of *effective sparsity*: the residual weights may be sparse, small, or rapidly decaying, or may contain a large common component that is removed by centering. Such structure is relevant to modern neural networks and selectively activated architectures, where substantial redundancy and nonuniform computation have been widely observed (Hoefler et al., 2021; Fedus et al., 2022). Figure 1 illustrates this phenomenon on ImageNet, where the weights are local errors of pretrained neural networks. Although these local errors are not exactly sparse, they have remarkably small effective residual mass: $s_{\text{eff}} = 5.27$ for ResNet18, 4.36 for DenseNet121, and 2.63 for ViT-B/16. These values are substantially smaller than both $\sqrt{|\mathbb{T}| \ln K} \approx 32.47$ and $\sqrt{K} \approx 14.14$ in prior bounds, illustrating the potentially significant gain.

Our second contribution is a sharper bound for generalization error of trained models. We apply the theorem to Algorithmic Robustness, where the input space is partitioned into K regions and existing analyses incur a generalization term of order $O(\sqrt{K/n})$ (Xu & Mannor, 2012) or $O(\sqrt{|\mathbb{T}|(\ln K)/n})$ (Kawaguchi et al., 2022). Our bound replaces the dependence on the observed partition size by the effective residual mass of the local errors and the missing mass of the induced input distribution. Consequently, the explicit dependence on $|\mathbb{T}| \ln K$ can disappear when local errors exhibit effective sparsity or small centered residual mass, or when the input distribution is highly structured. This distinction is particularly important for modern high-dimensional models, where K can be large. Empirically, on ImageNet, our data-dependent bound provides substantially tighter estimates of test error across a broad collection of modern neural networks, including vision transformers, than existing robustness-based bounds.

Our third contribution is an improved finite-sample guarantee for off-policy evaluation (OPE) in confounded POMDPs. Partial observability complicates OPE because observations may not identify the underlying state, and the history space for fully history-dependent policies grows exponentially with horizon T (Tennenholtz et al., 2020; Shi et al., 2022; Bennett & Kallus, 2024). For tabular confounded POMDPs with observation space $\mathcal{O}$ and action space $\mathcal{A}$, Kuang et al. (2025) obtain a first-order error of $O\left(Tn^{-1/2}(|\mathcal{O}||\mathcal{A}|)^{T/2}\right)$ for a model-based estimator of fully history-dependent policies. Complementing existing identification and bridge-function frameworks, our theorem controls observed-history fluctuation through centered coefficient variation given a data-dependent multinomial first-order representation, while separating unseen-history and higher-order errors. Under sparse active histories, approximate history sparsity, or near constancy, our bound replaces $(|\mathcal{O}||\mathcal{A}|)^{T/2}$ with polynomial dependence on T. This yields an exponentially tighter guarantee when the unseen and higher-order terms are also polynomially controlled.

Overall, this work develops a general principle for high-dimensional problems: *the complexity relevant to a data-dependent functional can be much smaller than the nominal alphabet or support size*. Our concentration theorem makes this effective-sparsity principle explicit and turns it into concrete improvements in both model generalization and off-policy evaluation.

Roadmap. After a short discussion about related work, Section 3 presents the main bound for distribution estimation under effective sparsity. Section 4 contains generalization analysis and improved bound for the test error of trained models, while Section 5 presents the main result for OPE.

Notations. We will often use the following common notations throughout the paper. $[K]$ denotes the set of natural numbers not larger than K, $|\boldsymbol{A}|$ denotes the cardinality of a set $\boldsymbol{A}$. Denote $\mathbf{1}\{\mathtt{C} : f\}$ as the function that returns value f if condition $\mathtt{C}$ holds and returns 0 otherwise.

2 Related work

Discrete distribution estimation. Estimating a discrete distribution from i.i.d. samples is a classical problem in machine learning, statistics and information theory (Qian et al., 2020; Cohen et al., 2020; Mourtada, 2026). Much of this literature focuses on estimating the distribution itself and obtaining guarantees (with ℓ_1, relative entropy, and ℓ_∞ among the standard loss measures) that depend on the alphabet size, effective support, or properties of the probability vector. Recent results exploit sparsity and missing-mass structure to obtain more adaptive guarantees, particularly for large or heavy-tailed alphabets (Cohen et al., 2020; Mourtada, 2026). Our setting is different: we study a data-dependent functional $\sum_i f_i(\boldsymbol{X})(p_i - \hat{p}_i)$ and show that its estimation error can instead adapt to the effective residual structure of the realized coefficients, even when the underlying distribution has large or dense support.

Generalization analysis of trained models. Classical generalization theory typically provides guarantees in terms of hypothesis-class complexity, stability, or algorithmic properties. Algorithmic Robustness (Xu & Mannor, 2012) offers a complementary model-dependent perspective by partitioning the data space into regions on which the loss is similar. Kawaguchi et al. (2022) made these guarantees data-dependent by exploiting the observed partition structure and improving the dependence on the partition size. More recently, Than et al. (2025) replaced the global robustness parameter by local sensitivity. These bounds, however, still retain a support-dependent term. Our work exploits the effective residual mass of the local errors to provide a dimension-free bound when the local-error weights have effective sparsity structure.

Off-policy evaluation. OPE in partially observable, confounded environments is difficult because the observed process need not be Markovian and latent-state-dependent behavior policies invalidate standard observed-data importance weights (Tennenholtz et al., 2020; Shi et al., 2022; Miao et al., 2022; Bennett & Kallus, 2024). Bridge-function methods provide identification and estimators for confounded POMDPs (Shi et al., 2022; Miao et al., 2022; Bennett & Kallus, 2024), while Kuang et al. (2025) derive finite-sample bounds for memoryless and fully history-dependent target policies, with exponential dependence on T for the latter. We study concentration once the OPE error is represented as a data-dependent multinomial functional. Under effective sparsity of its history-dependent coefficients, we show that the observed-history term depends polynomially on T.

3 Distribution estimation from i.i.d. samples

We first consider the risk when approximating a discrete distribution $\boldsymbol{p}$ by its empirical version $\hat{\boldsymbol{p}}$, which is computed from i.i.d. samples. In particular, our main interest is to bound the weighted risk $\sum_{i=1}^K f_i(p_i - \hat{p}_i)$, where the weights f_i are arbitrary functions which may depend on the observed samples. We obtain the following result, whose proof appears in Appendix A.

Theorem 3.1. *Let $\boldsymbol{X}$ contain n i.i.d. samples from a discrete distribution $\boldsymbol{p} = (p_1, \ldots, p_K)$, supported on a domain $\mathcal{X} = \bigcup_{i=1}^K \mathcal{X}_i$. Let $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_K)$ be the empirical estimate of $\boldsymbol{p}$. Let $f_1, \ldots, f_K$ be any real-valued functions which may depend on $\boldsymbol{X}$. Let $\mathbb{T} = \{j : |\boldsymbol{X} \cap \mathcal{X}_j| > 0\}$, $c^* = \arg\min_{c \in \mathbb{R}} \sum_{i \in \mathbb{T}} |f_i(\boldsymbol{X}) - c|$, $s_{\text{eff}}(\boldsymbol{X}) = \sum_{i \in \mathbb{T}} |f_i(\boldsymbol{X}) - c^*|$, and $e_o = \sum_{i=1}^K p_i(1-p_i)^n$. For any $\delta_1 \in (0,1)$ and function $c(\boldsymbol{X})$, denoting $f_o^c = \mathbf{1}\{|\mathbb{T}| < K : \sum_{k \notin \mathbb{T}}(f_k - c(\boldsymbol{X}))p_k / (\sum_{j \notin \mathbb{T}} p_j)\}$ and $\alpha = \frac{2}{\sqrt{n}} + \sqrt{\frac{\ln(1/\delta_1)}{2n}}$, with probability at least $1 - \delta_1$:*

$$\sum_{i=1}^K f_i(p_i - \hat{p}_i) \quad \leq \quad \alpha \sum_{i \in \mathbb{T}} |f_i - c(\boldsymbol{X})| + \sum_{k \notin \mathbb{T}} (f_k - c(\boldsymbol{X}))p_k, \tag{2}$$

$$\sum_{i=1}^K f_i(p_i - \hat{p}_i) \quad \leq \quad \alpha s_{\text{eff}}(\boldsymbol{X}) + \sum_{k \notin \mathbb{T}} (f_k - c^*)p_k. \tag{3}$$

Furthermore, for any constant $\delta_o \in (0,1)$, with probability at least $1-\delta_1-\delta_o$:

$$\sum_{i=1}^{K} f_i(p_i - \hat{p}_i) \quad \leq \quad \alpha \sum_{i \in \mathbb{T}} |f_i - c(\boldsymbol{X})| + |f_o^c| e_o + |f_o^c| \sqrt{\frac{2\ln(1/\delta_o)}{n}}. \tag{4}$$

Theorem 3.1 gives a distribution-estimation guarantee for a substantially broader class of risks than the usual ℓ_1 error. While $|\boldsymbol{p} - \hat{\boldsymbol{p}}|_1$ measures estimation error uniformly over all categories, we consider the weighted functional $\sum_i f_i(\boldsymbol{X})(p_i - \hat{p}_i)$, where the weights can depend on the observed sample. This formulation is natural when the estimated distribution is used to evaluate a downstream quantity whose sensitivity to different categories is heterogeneous, as illustrated in Figure 1. The theorem shows that such a functional can be estimated much more accurately than a uniform distributional bound: its leading error is governed by the effective residual mass of the realized weights, while the contribution of unobserved categories is isolated through the missing mass e_o.

This leads to a fundamentally different notion of statistical complexity from standard bounds. Classical ℓ_1 bounds typically scale as $O(\sqrt{K/n})$ (Qian et al., 2020; Cohen et al., 2020), while Mourtada (2026) replaces K by the number of observed categories $|\mathbb{T}|$. The closely related weighted bound by Kawaguchi et al. (2022) replaces K by $|\mathbb{T}| \ln K$. Our result can remove both dependencies when the downstream weights have small effective residual mass, as empirically shown in Section 4.3. In particular, even if $|\mathbb{T}| \approx K$, the weighted estimation error can remain of order $O(s_{\mathrm{eff}}(\boldsymbol{X})/\sqrt{n})$. Thus, the difficulty of estimating a distribution depends not only on the complexity of the distribution itself, but also on *which functional of the distribution is being estimated.*

The centered form is especially important in this respect. Because $\sum_i (p_i - \hat{p}_i) = 0$, a common component of the weights does not contribute to the estimation error. The bound therefore automatically ignores weight structure that is irrelevant to the functional and focuses on the residual variation that actually interacts with the empirical distribution. This is the source of the effective-sparsity improvement and distinguishes our result from bounds based on support size or worst-case magnitudes.

The result is also adaptive to distributional structure through the missing-mass term. When the distribution has a light or otherwise regular tail, e_o can decay rapidly with n (McAllester & Ortiz, 2003). Hence, our guarantee combines two distinct forms of adaptivity: it exploits structure in the distribution through unseen mass and structure in the downstream functional through effective residual weight. This combination is particularly useful in high-dimensional applications, where the alphabet may be large but only a limited amount of the estimated distribution is relevant to the quantity of interest.

The following corollaries make explicit the regimes in which our bound improves over dimension- or support-dependent bounds.

Corollary 3.1 (Effective-sparse weights)**.** *Under the assumptions of Theorem 3.1, suppose that there exist $c(\boldsymbol{X}) \in \mathbb{R}$, a set $\boldsymbol{A} \subseteq \mathbb{T}$ with $|\boldsymbol{A}| \leq s$, and $B > 0$ such that $|f_i - c(\boldsymbol{X})| \leq B$ for all $i \in \boldsymbol{A}$, $f_i = c(\boldsymbol{X})$ for all $i \in \mathbb{T} \setminus \boldsymbol{A}$, and $|f_k - c(\boldsymbol{X})| \leq B$ for all $k \notin \mathbb{T}$. Then, with probability at least $1-\delta_1-\delta_o$: $\sum_{i=1}^{K} f_i(p_i - \hat{p}_i) \leq B\big(\alpha s + e_o + \sqrt{\frac{2\ln(1/\delta_o)}{n}}\big)$.*

Remark 3.1 (Improvement over support-size bounds)**.** *Assume the setting of Corollary 3.1, with B bounded independently of K, and suppose that the missing-mass contribution is lower order. Then our leading term is $O(Bs/\sqrt{n})$, whereas the bound of Kawaguchi et al. (2022) is of order $O(B\sqrt{|\mathbb{T}| \ln K/n})$. Consequently, whenever $s = o(\sqrt{|\mathbb{T}| \ln K})$, our bound is asymptotically smaller; in particular, for $s = O(1)$, our leading term is $O(B/\sqrt{n})$ with no explicit dependence on either K or $|\mathbb{T}|$, even when $|\mathbb{T}| \approx K$.*

Corollary 3.1 shows that the relevant complexity can be the number of coordinates carrying substantial *residual* weight rather than the number of observed categories. Remark 3.1 then gives a direct separation from support-dependent bounds: our improvement occurs whenever the effective residual complexity is substantially smaller than $\sqrt{|\mathbb{T}| \ln K}$ and can become independent of the alphabet size in the strongly effective-sparse regime.

The following result readily follows when choosing $f_i = \mathrm{sign}(p_i - \hat{p}_i)$ for $i \in \mathbb{T}$ and $c(\boldsymbol{X})$ as a median of the realized sign weights (i.e., the majority sign).

Corollary 3.2. *Under the assumptions of Theorem 3.1, let* $N_+ = |i \in \mathbb{T} : p_i > \hat{p}_i|$ *and* $N_- = |i \in \mathbb{T} : p_i < \hat{p}_i|$. *Then* $\|\boldsymbol{p} - \hat{\boldsymbol{p}}\|_1 \leq 2\alpha \min\{N_+, N_-\} + 2e_o + 2\sqrt{\frac{2\ln(1/\delta_o)}{n}}$ *with probability at least* $1 - \delta_1 - \delta_o$.

Thus, the ℓ_1 estimation error can be controlled by the smaller number of positive or negative coordinates of the empirical estimation error, rather than directly by the alphabet size K or the observed support size $|\mathbb{T}|$. In particular, when $\min\{N_+, N_-\} = O(1)$ and the missing-mass contribution is lower order, the bound is $O(1/\sqrt{n})$ independently of K.

Remark 3.2 (Practical relevance of effective sparsity)**.** *Effective sparsity can arise from both the data-dependent weights and the underlying distribution. Large-alphabet data are often highly nonuniform, with probability mass concentrated on a relatively small set of frequent categories and a long, decaying tail; this motivates distribution estimators that adapt to empirical or effective sparsity rather than the nominal alphabet size (Cohen et al., 2020; Mourtada, 2026). Such heavy-tailed or Zipf-like structure is well documented in natural language and other categorical data (Piantadosi, 2014). At the same time, modern learning systems often exhibit highly nonuniform component-wise contributions and redundancy, including sparse expert activation and parameter sparsity (Hoefler et al., 2021; Fedus et al., 2022). Our bound exploits both forms of structure: it adapts to the residual weight mass through* $\sum_{i\in\mathbb{T}} |f_i - c(\boldsymbol{X})|$ *while separately controlling unseen probability mass through* e_o. *Thus, a large alphabet or dense observed support need not imply a large estimation error for a structured downstream functional.*

4 Error Bounds via Robustness

This section focuses on generalization ability of a trained model. Generalization concerns performance on unseen data and is a crucial aspect of modern machine learning systems. However, quantifying this ability remains challenging for complex models, e.g., overparameterized ones. We follow the robustness approach which can provide tight bound for test error (Than et al., 2025).

4.1 Setup

We consider a learning problem on an input space $\mathcal{Z}$ with data distribution P. Let $\boldsymbol{D}$ be a training set of n i.i.d. samples from P, and let $\boldsymbol{h} : \mathcal{Z} \mapsto \mathcal{Y}$ be a model that maps from an input space $\mathcal{Z}$ to an output space $\mathcal{Y}$. For a loss function ℓ with maximum value C, denote $\ell(\boldsymbol{h}, \boldsymbol{z})$ as the error/loss of a prediction $\boldsymbol{h}(\boldsymbol{z})$ about $\boldsymbol{z}$. The population and empirical losses of $\boldsymbol{h}$ are $F(P, \boldsymbol{h}) = \mathbb{E}_{\boldsymbol{z}\sim P}[\ell(\boldsymbol{h}, \boldsymbol{z})]$ and $F(\boldsymbol{D}, \boldsymbol{h}) = \frac{1}{n}\sum_{\boldsymbol{s}\in\boldsymbol{D}} \ell(\boldsymbol{h}, \boldsymbol{s})$, respectively. Following Algorithmic Robustness (Xu & Mannor, 2012), partition $\mathcal{Z}$ into K disjoint regions $\mathcal{Z} = \bigcup_{i=1}^{K} \mathcal{Z}_i$, and let $p_i = P(\mathcal{Z}_i)$ and $n_i = |\boldsymbol{D} \cap \mathcal{Z}_i|$. We write $\mathbb{T} = \{i : n_i > 0\}$ for the observed regions and let $f_i(\boldsymbol{h}) = \mathbb{E}_{\boldsymbol{z}\sim P}[\ell(\boldsymbol{h}, \boldsymbol{z}) : \boldsymbol{z} \in \mathcal{Z}_i]$ be the local error at region i.

4.2 Main bound

When studying the generalization ability of a trained model $\boldsymbol{h}$, a natural quantity of interest is its population loss $F(P, \boldsymbol{h})$ or the gap $F(P, \boldsymbol{h}) - F(\boldsymbol{D}, \boldsymbol{h})$. Direct estimation of this quantity is difficult for complex models, and algorithmic robustness provides a useful way to connect local properties of a trained model to its global generalization behavior (Xu & Mannor, 2012). Under the (K, ϵ)-robustness condition of the learner, Xu & Mannor (2012) shows that, with probability at least $1 - \delta$,

$$F(P, \boldsymbol{h}) - F(\boldsymbol{D}, \boldsymbol{h}) \leq C\sqrt{(2K\ln 2 - 2\ln\delta)/n} + \epsilon(\boldsymbol{D}). \tag{5}$$

This result reveals that the generalization gap is bounded by using a robustness term and a statistical uncertainty term of order $O(\sqrt{K/n})$. The dependence on the partition size K can be restrictive when a fine-grained partition is needed to characterize a complex model.

Kawaguchi et al. (2022) sharpened the uncertainty term by exploiting the observed partition regions to bound the data-dependent quantity $\sum_i f_i(\boldsymbol{h})(p_i - n_i/n)$ by Unc_K. In particular, they obtain

$$F(P, \boldsymbol{h}) - F(\boldsymbol{D}, \boldsymbol{h}) \leq \mathrm{Unc}_K + \epsilon(\boldsymbol{D}), \tag{6}$$

where $\mathrm{Unc}_K = \frac{\sqrt{\ln(2K/\delta)}}{n} \sum_{i\in\mathbb{T}} \sqrt{n_i} \left(f_o + \sqrt{2} f_i(\boldsymbol{h})\right) + \frac{2\ln(2K/\delta)}{n} \left(f_o|\mathbb{T}| + \sum_{i\in\mathbb{T}} f_i(\boldsymbol{h})\right)$, and $f_o = \max_{j\notin\mathbb{T}} f_j(\boldsymbol{h})$. It is worth noticing that the uncertainty term Unc_K scales with $\sqrt{|\mathbb{T}| \ln K/n}$. This can be substantially smaller than $\sqrt{K/n}$ when $|\mathbb{T}| \ll K$. This suggests that their bound can exploit the sparsity of the data distribution but scales with $\sqrt{|\mathbb{T}| \ln K/n}$.

Than et al. (2025) further refined the robustness term by replacing the global parameter $\epsilon(\boldsymbol{D})$ with an average local sensitivity. Specifically, defining $\bar{\epsilon}(\boldsymbol{h}) = \sum_{i\in\mathbb{T}} \frac{n_i}{n} \bar{\epsilon}_i(\boldsymbol{h})$ with $\bar{\epsilon}_i(\boldsymbol{h}) = \frac{1}{n_i} \sum_{\boldsymbol{s}\in\boldsymbol{D}_i} \mathbb{E}_{\boldsymbol{z}\sim P}[|\ell(\boldsymbol{h},\boldsymbol{z}) - \ell(\boldsymbol{h},\boldsymbol{s})| : \boldsymbol{z} \in \mathcal{Z}_i]$ for each index i, they showed

$$F(P,\boldsymbol{h}) - F(\boldsymbol{D},\boldsymbol{h}) \le \mathrm{Unc}_K + \bar{\epsilon}(\boldsymbol{h}). \tag{7}$$

This bound is more refined because a small number of regions with large sensitivity need not dominate the guarantee. Nevertheless, its uncertainty term Unc_K still inherits the $\sqrt{|\mathbb{T}| \ln K/n}$ scale. These results therefore leave open *whether the estimation about $\sum_i f_i(\boldsymbol{h})(p_i - n_i/n)$ can itself adapt to the heterogeneous structure of the local errors?*

Our analysis addresses this question by applying Theorem 3.1 to $\sum_i f_i(\boldsymbol{h})(p_i - n_i/n)$. The following result thus follows, whose proof appears in Appendix B.

Theorem 4.1 (Sparsity-aware error bound)**.** *Consider a model $\boldsymbol{h}$ that can depend on a dataset $\boldsymbol{D}$ with n i.i.d. samples from distribution P. Let $\bar{\epsilon}(\boldsymbol{h}) = \sum_{i\in\mathbb{T}} \frac{n_i}{n} \bar{\epsilon}_i(\boldsymbol{h})$, with $\bar{\epsilon}_i(\boldsymbol{h}) = \frac{1}{n_i} \sum_{\boldsymbol{s}\in\boldsymbol{D}_i} \mathbb{E}_{\boldsymbol{z}\sim P}[|\ell(\boldsymbol{h},\boldsymbol{z}) - \ell(\boldsymbol{h},\boldsymbol{s})| : \boldsymbol{z} \in \mathcal{Z}_i]$ for each index i. For any function $c(\boldsymbol{D})$ and any constants $\delta_o, \delta_1 \in (0,1)$, with probability at least $1 - \delta_1 - \delta_o$:*

$$F(P,\boldsymbol{h}) - F(\boldsymbol{D},\boldsymbol{h}) \le \mathit{Unc}_O + \bar{\epsilon}(\boldsymbol{h}), \tag{8}$$

where $\mathit{Unc}_O = \alpha \sum_{i\in\mathbb{T}} |f_i(\boldsymbol{h}) - c(\boldsymbol{D})| + |f_o^c| e_o + |f_o^c| \sqrt{\frac{2\ln(1/\delta_o)}{n}}$, and f_o^c and α are defined as in Theorem 3.1.

This result provides a finer decomposition of the generalization gap than existing robustness-based bounds. Like Than et al. (2025), it uses the average local sensitivity $\sum_{i\in\mathbb{T}}(n_i/n)\bar{\epsilon}_i(\boldsymbol{h})$ rather than the uniform robustness $\epsilon(\boldsymbol{D})$ in (Xu & Mannor, 2012; Kawaguchi et al., 2022). The key improvement is the uncertainty term: while (Xu & Mannor, 2012) incurs an $O(\sqrt{K/n})$ dependence and (Kawaguchi et al., 2022; Than et al., 2025) replace K by the observed support size through a term $\mathrm{Unc}_K = O(\sqrt{|\mathbb{T}| \ln K/n})$, our Unc_O depends on the effective residual mass of the local errors of $\boldsymbol{h}$. In particular, choosing $c(\boldsymbol{D})$ appropriately gives a leading term of order $O(s_{\mathrm{eff}}(\boldsymbol{D})/\sqrt{n})$, together with a separate missing-mass contribution. Hence, the bound can be independent of both K and $|\mathbb{T}|$ when the local errors are effectively sparse, even when the partition itself is large or nearly fully observed. Consequently, our bound can be substantially tighter for modern overparameterized models whose local errors are small or concentrated on a relatively small subset of the input space.

4.3 Empirical evaluation

We empirically evaluate the tightness of our generalization bound (8) against the latest robustness-based bound (7) in (Than et al., 2025), and examine the uncertainty terms introduced in Section 3. Since the bounds share the empirical-loss and local-sensitivity terms, we focus on their uncertainty terms: Unc_O from Theorem 4.1 and Unc_K from Kawaguchi et al. (2022), defined below equation 6.

To make the evaluation realistic, we consider pretrained neural networks (NNs) on ImageNet. For each model $\boldsymbol{h}$, $f_i(\boldsymbol{h})$ represents its local error on region $\mathcal{Z}_i$ under the 0-1 loss, with a partition size of $K = 200$. Since the exact local errors are generally intractable to compute, we approximate them by the empirical local errors $f_i(\boldsymbol{h}) \approx F(\boldsymbol{D}_i,\boldsymbol{h})$ using the ImageNet training set $\boldsymbol{D}$ with $n = 1,281,167$ images. For unseen regions, we set $f_o^c = \hat{a}_o = 1$, corresponding to the worst-case loss. For Unc_O, we choose c^* as the median of the observed weights and use the distribution-free upper bound $e_o = K/(ne)$. To ensure comparable confidence levels, we set $\delta_o = \delta_1 = 0.005$ for Unc_O and $\delta = 0.01$ for Unc_K. Additional experimental details are provided in Appendix C.

Bound comparison. Figure 2 reports the results for 22 pretrained models from the PyTorch model zoo.[1] Our uncertainty term Unc_O is smaller than Unc_K for every model, with Unc_K being approximately twice as large as Unc_O on average. The improvement is particularly pronounced for

[1] https://pytorch.org/vision/stable/models.html

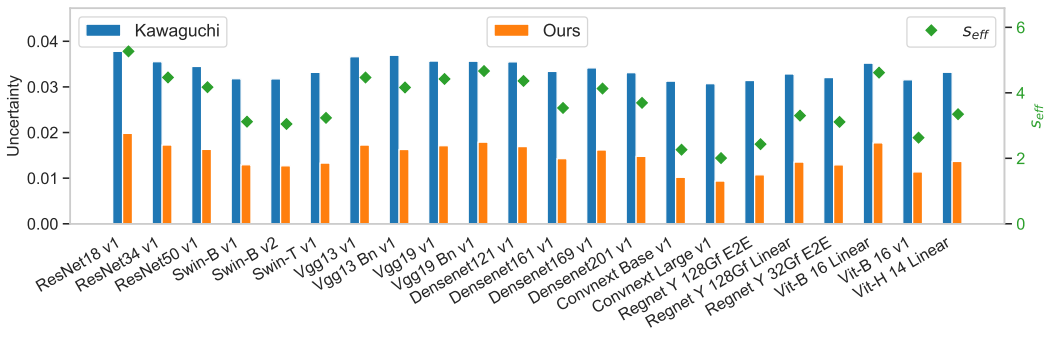

Figure 2: The uncertainty terms: Our Unc_O and Unc_K by Kawaguchi et al. (2022).

accurate models, such as ConvNeXt-Large and ViT-B/16, whose local errors are small over most regions. In these cases, our bound directly exploits the small effective residual mass of the local errors, whereas Unc_K remains tied to the observed support and worst-case local quantities. Models with larger errors, such as ResNet18, yield larger uncertainty terms for both bounds, but Unc_O remains consistently smaller. These results demonstrate that the proposed uncertainty term can substantially tighten robustness-based generalization bounds for modern neural networks.

Effective residual mass. Figure 2 also reports s_{eff} for each model. Despite $K = 200$ and $|\mathbb{T}| = 199$, $s_{\mathrm{eff}} < 5.3$ for all 22 models, and is particularly small for the ConvNeXt and RegNet families. Moreover, we observe a strong correlation between s_{eff} and model accuracy, with better-performing models generally exhibiting smaller effective residual mass. This empirical evidence supports the central premise of our analysis: although the partition is large and almost fully observed, the local-error weights can have a much smaller effective complexity, which our bound is able to exploit.

5 Application to Off-Policy Evaluation

We apply Theorem 3.1 to off-policy evaluation (OPE) in finite-horizon partially observable decision processes. Partial observability creates two related difficulties: the current observation need not predict future outcomes, and proxy-state histories may be exponential in the horizon. Existing work addresses identification and estimator construction through bridge-function methods for latent confounding (Shi et al., 2022; Bennett & Kallus, 2024), future-dependent value functions (Uehara et al., 2023), and model-based reconstruction of latent-state information (Zhang & Jiang, 2025).

In this work, we do not introduce a new bridge equation or a coverage condition. Instead, we show that once the first-order OPE error is represented as a data-dependent linear functional of an empirical multinomial distribution, its observed-cell fluctuation can be controlled without explicit dependence on the ambient observation, action, or history alphabet.

Off-policy evaluation setup. Consider a finite-horizon confounded partially observable Markov decision process (POMDP) with state space $\mathcal{S}$, observation space $\mathcal{O}$, action space $\mathcal{A}$, reward space $\mathcal{R}$, and horizon T. At time t, the unobserved state is $S_t \in \mathcal{S}$, the agent observes $O_t \in \mathcal{O}$, selects $A_t \in \mathcal{A}$, and receives $R_t \in \mathcal{R}$. Let O_0 be a pre-treatment proxy observation and let $H_t = (O_1, A_1, \ldots, O_t, A_t) \in \mathcal{H}_t$ denote the observed history. The offline data consists of n independent trajectories $\mathcal{D} = \{O_0^i, (O_t^i, A_t^i, R_t^i)_{t=1}^T\}_{i=1}^n$ generated by a behavior policy π^b. The target policy $\pi = \{\pi_t\}_{t=1}^T$ may depend on the entire observed history, with value $V(\pi) = \mathbb{E}^\pi[\sum_{t=1}^T R_t]$.

5.1 Fully history-dependent target policies

A fully history-dependent target policy satisfies $\pi_t : \mathcal{O} \times \mathcal{H}_{t-1} \to \Delta(\mathcal{A})$, while its formal history space has cardinality $|\mathcal{H}_{t-1}| = (|\mathcal{O}||\mathcal{A}|)^{t-1}$. Broad classes of model-free estimators for history-dependent policies face information-theoretic hardness, whereas model-based estimators can avoid some barriers under additional structure (Zhang & Jiang, 2025); future-dependent value methods achieve polynomial guarantees for memoryless policies under belief and outcome coverage, but this

coverage degrades exponentially for history-dependent policies (Zhang & Jiang, 2024). We isolate a structured subclass in which the formal history domain is exponentially large but the value is sensitive to a sparse, compressible, or nearly constant collection of history effects.

For each t, let $\mathcal{Y}_t$ be the finite collection of history-outcome tuples entering the first-order Bellman expansion. A typical cell is $y = (O_0, H_{t-1}, O_t, A_t, R_t, O_{t+1})$, possibly with some variables marginalized out. Let $q_{t,y}$ and $\widehat{q}_{t,y}$ be its population and empirical probabilities.

Assumption 5.1 (Trajectory-level first-order representation)**.** *The OPE error decomposes as* $\widehat{V}(\pi) - V(\pi) = \sum_{t=1}^{T} \sum_{y \in \mathcal{Y}_t} F_{t,y}(\widehat{q}_{t,y} - q_{t,y}) + \mathcal{R}_{\mathrm{ho}}$*, where* $F_{t,y}$ *is a possibly data-dependent policy-weighted Bellman coefficient. For each* t, $(n\widehat{q}_{t,y})_{y \in \mathcal{Y}_t} \sim \mathrm{Multinomial}(n; (q_{t,y})_{y \in \mathcal{Y}_t})$.

This follows from linearizing the OPE functional with respect to the empirical trajectory distribution; Appendix D.2 gives a sufficient differentiability condition, which requires every conditioning context to be visited (Appendix D.3).

Let $\mathbb{T}_t = \{y : \widehat{q}_{t,y} > 0\}$. For any data-dependent center c_t, define $U_t^c := \sum_{y \notin \mathbb{T}_t} q_{t,y}(F_{t,y} - c_t)$ and

$$B_t(c_t) := \Big(\sum_{y \in \mathbb{T}_t} (F_{t,y} - c_t)^2 \Big)^{1/2}, \qquad s_{\mathrm{hist},t}(F_t; c_t) := \frac{\big(\sum_{y \in \mathbb{T}_t} |F_{t,y} - c_t|\big)^2}{\sum_{y \in \mathbb{T}_t} (F_{t,y} - c_t)^2}, \tag{9}$$

with $s_{\mathrm{hist},t} = 0$ when its denominator is zero.

Theorem 5.1 (Effective-history OPE bound)**.** *Under Assumption 5.1, let each* c_t *be measurable with respect to the raw trajectory sample* $\mathcal{D}$ *and set* $\alpha_T(\delta) := 2/\sqrt{n} + \sqrt{\log(2T/\delta)/(2n)}$*. Then, with probability at least* $1 - \delta$*,* $|\widehat{V}(\pi) - V(\pi) + \sum_{t=1}^{T} U_t^c - \mathcal{R}_{\mathrm{ho}}| \le \alpha_T(\delta) \sum_{t=1}^{T} B_t(c_t) \sqrt{s_{\mathrm{hist},t}(F_t; c_t)}$*. Consequently,* $|\widehat{V}(\pi) - V(\pi)| \le \alpha_T(\delta) \sum_{t=1}^{T} B_t(c_t) \sqrt{s_{\mathrm{hist},t}(F_t; c_t)} + \Big|\sum_{t=1}^{T} U_t^c\Big| + |\mathcal{R}_{\mathrm{ho}}|$.

The proof (Appendix D.4) applies the two-sided form of Theorem 3.1 to each time-t multinomial vector in Assumption 5.1, then combines the T bounds by a union bound and the triangle inequality. The resulting observed-cell term contains neither $|\mathcal{Y}_t|$ nor $\log |\mathcal{Y}_t|$.

Corollary 5.1 (Uniform effective-history rate)**.** *Suppose the conditions of Theorem 5.1 hold and, in addition,* $B_t(c_t) \le B$ *and* $s_{\mathrm{hist},t}(F_t; c_t) \le s_H$ *for all* t*. Then, with probability at least* $1 - \delta$*,*

$$|\widehat{V}(\pi) - V(\pi)| \lesssim BT \sqrt{\frac{s_H \log(T/\delta)}{n}} + \left|\sum_{t=1}^{T} U_t^c\right| + |\mathcal{R}_{\mathrm{ho}}|. \tag{10}$$

In tabular confounded POMDPs, Kuang et al. (2025) analyze a model-based estimator built from empirical history-conditioned probabilities for fully history-dependent policies. Under identification and regularity assumptions, including bounded confounded concentrability, its first-order error is $O\Big(T\sqrt{(|\mathcal{O}||\mathcal{A}|)^T/n}\Big)$. Corollary 5.1 retains the $T/\sqrt{n}$ dependence but replaces $(|\mathcal{O}||\mathcal{A}|)^{T/2}$ by $B\sqrt{s_H}$. Their exponential factor arises from uniformly controlling proxy inverses over history cells. Because $F_{t,y}$ may contain such an inverse, invertibility and conditioning remain necessary. Whether $B\sqrt{s_H}$ improves upon $(|\mathcal{O}||\mathcal{A}|)^{T/2}$ depends on these coefficients. By Cauchy–Schwarz, $s_{\mathrm{hist},t}(F_t; c_t) \le |\mathbb{T}_t| \le n$ for every t, so s_H may be chosen at most n, irrespective of alphabet size.

5.2 Structural regimes yielding exponential gains

A small effective-history complexity reflects the broader principle that reinforcement-learning complexity can depend on decision-relevant structure rather than observation cardinality. Low-Bellman-rank models can avoid dependence on the distinct observation count, sparse batch OPE can scale with relevant features, and factored decision processes can represent exponentially large tabular spaces via compact local structure (Jiang et al., 2017; Hao et al., 2021; Guestrin et al., 2003). Our conditions instead concern the centered, possibly data-dependent trajectory coefficients in the OPE multinomial functional. Exponential gains can arise across the sparse-to-dense regimes below.

Sparse active histories. Let $I_t \subseteq \mathcal{Y}_t$ contain all cells with $F_{t,y} \neq c_t$, and suppose $|I_t| \le s_H$ for every t. Define $F_{\max} := \max_{t, y \in \mathcal{Y}_t} |F_{t,y} - c_t|$. Then $s_{\mathrm{hist},t} \le s_H$ and $B_t(c_t) \le \sqrt{s_H} F_{\max}$. Applying Corollary 5.1 gives the observed-cell rate $O(s_H F_{\max} T \sqrt{\log(T/\delta)/n})$. Compared with the

first-order rate $O(T\sqrt{(|\mathcal{O}||\mathcal{A}|)^T/n})$ of Kuang et al. (2025), this replaces $(|\mathcal{O}||\mathcal{A}|)^{T/2}$ by $s_H F_{\max}$, yielding the improvement factor $(|\mathcal{O}||\mathcal{A}|)^{T/2}/(s_H F_{\max})$. This gain is exponential in T whenever $s_H F_{\max}$ grows at most polynomially in T. Such polynomial complexity can arise when a policy branches only after a few exceptional observations, when few histories represent failures or threshold crossings, or when the policy is a sparse decision tree or a set of event triggered rules, provided that these histories are the only active ones and the value estimate reacts to them only moderately.

Approximate history sparsity. Let $|F_t - c_t|_{(1)} \geq |F_t - c_t|_{(2)} \geq \cdots$ be the decreasing rearrangement of the observed centered coefficients and define $\tau_t(s) := \sum_{j>s} |F_t - c_t|_{(j)}$. For every s, $\sum_{y\in\mathbb{T}_t} |F_{t,y} - c_t| \leq \sqrt{s}(\sum_{j=1}^{s} |F_t - c_t|^2_{(j)})^{1/2} + \tau_t(s)$. Consequently,

$$\Big|\sum\nolimits_{y\in\mathcal{Y}_t} F_{t,y}(\widehat{q}_{t,y} - q_{t,y}) + U_t^c\Big| \leq \alpha_T(\delta)\Big\{\sqrt{s}\big(\sum\nolimits_{j=1}^{s} |F_t - c_t|^2_{(j)}\big)^{1/2} + \tau_t(s)\Big\}. \tag{11}$$

Thus exponentially many histories are harmless for the observed-cell term if polynomially many coefficients capture the centered variation at polynomial cost. This resembles sparsity-aware batch OPE without requiring sparse linear parameterization or explicit support recovery (Hao et al., 2021).

Near constancy across histories. A dense coefficient vector can have low complexity even when every coefficient is nonzero. Suppose $F_{t,y} = \mu_t + r_{t,y}$ and $\sum_{y\in\mathbb{T}_t} |r_{t,y}| \leq C_t$. Choosing μ_t as the center gives $|\sum_{y\in\mathcal{Y}_t} F_{t,y}(\widehat{q}_{t,y} - q_{t,y}) + U_t^c| \leq \alpha_T(\delta)C_t$. If $C_t \leq C$, the cumulative observed-cell error is $O(CT\sqrt{\log(T/\delta)/n})$, independently of $|\mathcal{H}_{t-1}|$, $|\mathcal{O}|$, and $|\mathcal{A}|$.

Near constancy can arise when histories induce approximately the same predictive states, belief-state regions, finite-controller states, or proxy-adjusted continuation values. Future-dependent and low-Bellman-complexity methods similarly exploit compact decision-relevant information (Uehara et al., 2023; Jiang et al., 2017). Centering adds a separate invariance: components proportional to the all-ones vector cancel against a multinomial deviation. If $|r_{t,y}| \leq \epsilon_t$ on unseen cells, $|U_t^c| = |\sum_{y\notin\mathbb{T}_t} q_{t,y} r_{t,y}| \leq \epsilon_t$, controlling unseen histories even if their total probability is non-negligible.

5.3 Summary and scope

We analyze OPE error without establishing identification or constructing an estimator. Once an estimator's error admits the first-order representation in Assumption 5.1, Theorem 5.1 bounds its observed-cell fluctuation by $\alpha_T(\delta)\sum_{t=1}^{T} B_t(c_t)\sqrt{s_{\mathrm{hist},t}(F_t; c_t)}$, with no explicit history-alphabet factor. Under effective-sparsity regimes, this term can achieve polynomial horizon dependence, whereas Kuang et al. (2025) obtain $O(T\sqrt{(|\mathcal{O}||\mathcal{A}|)^T/n})$ for the first-order error in tabular confounded POMDPs. The polynomial conclusion here applies to observed cells; in particular, when all cells are observed, the full first-order error is polynomial in T. Otherwise, the full first-order error also requires control of unseen histories. Appendix E bounds the unseen and higher-order terms, giving the combined bound equation 88. The complete bound has polynomial horizon dependence when its effective-history, unseen-influence, and higher-order terms are polynomially controlled. As in Kuang et al. (2025), the trajectory-representation argument requires every conditioning context to be visited (Appendix D.3). Weakening these conditions, designing estimators to discover active history classes, and deriving minimax lower bounds are left for future work.

6 Conclusion

We introduced a concentration bound for data-dependent distributional errors that exploits effective sparsity and separately accounts for unseen probability mass. This principle yields sharper guarantees for high-dimensional distribution and error estimation, with applications to generalization analysis of modern neural networks and off-policy evaluation in confounded POMDPs. Our empirical and theoretical results show that effective residual structure can replace nominal dimension or exponential history size with substantially smaller complexity.

This work has two limitations. The effective residual mass may be difficult to characterize before observing the data, and the OPE improvement relies on structural sparsity conditions. Extending the framework to observable effective-sparsity measures and broader distribution classes is an important direction for future work.

AI USE STATEMENT

In this work, we used generative AI tools to refine the writing for clarity and readability, strengthen connections to relevant prior work, improve the presentation of proofs, and suggest possible interpretations of our theoretical results. Generative AI was used as an assistive tool and did not replace the authors' analysis, derivations, or scientific judgment. We carefully reviewed and verified all AI-assisted content, including mathematical arguments, citations, and factual claims. The authors take full responsibility for the final content of this work.

REFERENCES

Andrew Bennett and Nathan Kallus. Proximal reinforcement learning: Efficient off-policy evaluation in partially observed markov decision processes. *Operations Research*, 72(3):1071–1086, 2024.

Daniel Berend and Aryeh Kontorovich. On the concentration of the missing mass. *Electronic Communications in Probability*, 18:1–7, 2013.

Doron Cohen, Aryeh Kontorovich, and Geoffrey Wolfer. Learning discrete distributions with infinite support. *Advances in Neural Information Processing Systems*, 33:3942–3951, 2020.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

I. J. Good. The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3–4):237–264, 1953.

Carlos Guestrin, Daphne Koller, Ronald Parr, and Shobha Venkataraman. Efficient solution algorithms for factored mdps. *Journal of Artificial Intelligence Research*, 19:399–468, 2003.

Botao Hao, Yaqi Duan, Tor Lattimore, Csaba Szepesvári, and Mengdi Wang. Sparse feature selection makes batch reinforcement learning more sample efficient. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 4063–4073. PMLR, 2021.

Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, and Alexandra Peste. Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. *Journal of Machine Learning Research*, 22:1–124, 2021.

Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, and Robert E. Schapire. Contextual decision processes with low bellman rank are pac-learnable. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1704–1713. PMLR, 2017.

Kenji Kawaguchi, Zhun Deng, Kyle Luh, and Jiaoyang Huang. Robustness implies generalization via data-dependent generalization bounds. In *International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 10866–10894. PMLR, 2022.

Aryeh Kontorovich and Amichai Painsky. Distribution estimation under the infinity norm. *Journal of Machine Learning Research*, 26(162):1–30, 2025.

Qi Kuang, Jiayi Wang, Fan Zhou, and Zhengling Qi. Breaking the order barrier: Off-policy evaluation for confounded pomdps. In *Advances in Neural Information Processing Systems*, 2025.

David McAllester and Luis Ortiz. Concentration inequalities for the missing mass and for histogram rule error. *Journal of Machine Learning Research*, 4(Oct):895–911, 2003.

David A. McAllester and Robert E. Schapire. On the convergence rate of good–turing estimators. In *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, pp. 1–6, 2000.

Rui Miao, Zhengling Qi, and Xiaoke Zhang. Off-policy evaluation for episodic partially observable markov decision processes under non-parametric models. *Advances in Neural Information Processing Systems*, 35:593–606, 2022.

Jaouad Mourtada. Estimation of discrete distributions in relative entropy, and the deviations of the missing mass. *Mathematical Statistics and Learning*, 9(1):69–144, 2026.

Steven T Piantadosi. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, 21(5):1112–1130, 2014.

Jian Qian, Ronan Fruit, Matteo Pirotta, and Alessandro Lazaric. Concentration inequalities for multinoulli random variables. *arXiv preprint arXiv:2001.11595*, 2020.

Chengchun Shi, Masatoshi Uehara, Jiawei Huang, and Nan Jiang. A minimax learning approach to off-policy evaluation in confounded partially observable markov decision processes. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 20057–20094. PMLR, 2022.

Guy Tennenholtz, Uri Shalit, and Shie Mannor. Off-policy evaluation in partially observable environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 10276–10283, 2020.

Khoat Than, Dat Phan, and Giang Vu. Gentle local robustness implies generalization. *Machine Learning*, 114(6):142, 2025.

Masatoshi Uehara, Haruka Kiyohara, Andrew Bennett, Victor Chernozhukov, Nan Jiang, Nathan Kallus, Chengchun Shi, and Wen Sun. Future-dependent value-based off-policy evaluation in pomdps. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Huan Xu and Shie Mannor. Robustness and generalization. *Machine Learning*, 86(3):391–423, 2012.

Yuheng Zhang and Nan Jiang. On the curses of future and history in future-dependent value functions for off-policy evaluation. *Advances in Neural Information Processing Systems*, 2024.

Yuheng Zhang and Nan Jiang. Statistical tractability of off-policy evaluation of history-dependent policies in pomdps. In *International Conference on Learning Representations*, 2025.

A PROOFS FOR DATA-DEPENDENT BOUNDS

Proof of Theorem 3.1. Observe that $1 = \sum_{i\in[K]} p_i = \sum_{i\in[K]} \hat{p}_i$. Therefore

$$\sum_{i\in[K]} (f_i - c(\boldsymbol{X}))\,(p_i - \hat{p}_i) = \sum_{i\in[K]} f_i\,(p_i - \hat{p}_i) - c(\boldsymbol{X}) \sum_{i\in[K]} (p_i - \hat{p}_i) = \sum_{i\in[K]} f_i\,(p_i - \hat{p}_i)\,.$$

Using Theorem A.1 to bound $\sum_{i\in[K]}(f_i - c(\boldsymbol{X}))\,(p_i - \hat{p}_i)$ will complete the proof. □

Theorem A.1 (Deviation bound)**.** *Given the notations and assumptions in Theorem 3.1, denote* $f_o = \mathbf{1}\{|\mathbb{T}| < K : \sum_{k\notin\mathbb{T}} f_k p_k/(\sum_{j\notin\mathbb{T}} p_j)\}$*. Then with probability at least* $1-\delta_1$*:*

$$\sum_{i\in[K]} f_i\,(p_i - \hat{p}_i) \quad\leq\quad \alpha \sum_{i\in\mathbb{T}} |f_i| + \sum_{k\notin\mathbb{T}} f_k p_k; \tag{12}$$

and with probability at least $1-\delta_1-\delta_o$*:*

$$\sum_{i\in[K]} f_i\,(p_i - \hat{p}_i) \quad\leq\quad \alpha \sum_{i\in\mathbb{T}} |f_i| + |f_o| e_o + |f_o| \sqrt{\frac{2\ln(1/\delta_o)}{n}}. \tag{13}$$

Proof of Theorem A.1. Denote $\boldsymbol{z} = (z_1, \ldots, z_K)$ with $z_i = n\hat{p}_i$ which counts the number of observed instances in $\mathcal{X}_i$, for each $i \in [K]$. Due to independence, $\boldsymbol{z}$ is in fact a multinomial random variable. Since $z_k = 0$ for all $k \notin \mathbb{T}$, we have

$$\sum_{i=1}^{K} f_i \left(p_i - \frac{z_i}{n}\right) = \sum_{i\in\mathbb{T}} f_i \left(p_i - \frac{z_i}{n}\right) + \sum_{k\notin\mathbb{T}} f_k p_k. \tag{14}$$

Consider the first term $\sum_{i\in\mathbb{T}} f_i \left(p_i - \frac{z_i}{n}\right)$. Hölder's inequality shows that $\sum_{i\in\mathbb{T}} f_i \left(p_i - \frac{z_i}{n}\right) \leq (\sum_{i\in\mathbb{T}} |f_i|)\|\boldsymbol{p} - \boldsymbol{z}/n\|_\infty$. Note that $\boldsymbol{z}/n$ is an empirical estimate of $\boldsymbol{p}$. Therefore, Lemma A.2 gives $\Pr\left(\|\boldsymbol{p} - \boldsymbol{z}/n\|_\infty > \frac{2}{\sqrt{n}} + \sqrt{\frac{\log(1/\delta_1)}{2n}}\right) < \delta_1$. As a result,

$$\Pr\left(\sum_{i\in\mathbb{T}} f_i \left(p_i - \frac{z_i}{n}\right) > (\sum_{i\in\mathbb{T}} |f_i|)\Big(\frac{2}{\sqrt{n}} + \sqrt{\frac{\log(1/\delta_1)}{2n}}\Big)\right) < \delta_1. \tag{15}$$

Combining this with equation 14 gives the first result.

To obtain the second result, it remains to control the unobserved part. Let $p_o = \sum_{k\notin\mathbb{T}} p_k$. Note that $\sum_{k\notin\mathbb{T}} f_k p_k = f_o p_o \leq |f_o| p_o$. The unobserved mass is $p_o = \sum_{k=1}^{K} p_k \mathbf{1}\{z_k = 0\}$. We have

$$\mathbb{E}[p_o] = \sum_{k=1}^{K} p_k (1-p_k)^n = e_o. \tag{16}$$

To tightly bound the upper tail independently of the support size K, we apply the specialized concentration inequality for the missing mass (McAllester & Ortiz, 2003). For any $\epsilon > 0$, the deviation of p_o from its expectation is strictly bounded by

$$\Pr(p_o - e_o > \epsilon) \leq \exp\left(-\frac{n\epsilon^2}{2}\right). \tag{17}$$

Setting the right-hand side to δ_o and solving for ϵ, we obtain $\Pr\left(p_o > e_o + \sqrt{\frac{2\ln(1/\delta_o)}{n}}\right) \leq \delta_o$. Multiplying the terms in this event by $|f_o|$ yields the upper bound for the unobserved component:

$$\Pr\left(f_o p_o > |f_o| e_o + |f_o| \sqrt{\frac{2\ln(1/\delta_o)}{n}}\right) \leq \Pr\left(|f_o| p_o > |f_o| e_o + |f_o| \sqrt{\frac{2\ln(1/\delta_o)}{n}}\right) \leq \delta_o. \tag{18}$$

Combining equation 15, equation 18, and the union bound, we obtain the second result. □

Lemma A.2. *Let $p = (p_i)_{i\in\mathbb{N}}$ be a probability distribution on $\mathbb{N} = \{1, 2, \ldots\}$, and let $X^n = (X_1, \ldots, X_n)$ be an i.i.d. sample from p. Define the empirical distribution by $\widehat{p}_i = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}\{X_j = i\}$ for $i \in \mathbb{N}$. Then, for every $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\sup_{i\in\mathbb{N}} |p_i - \widehat{p}_i| \le \frac{2}{\sqrt{n}} + \sqrt{\frac{\log(1/\delta)}{2n}}. \tag{19}$$

Proof. Define $Z(X^n) = \sup_{i\in\mathbb{N}} |p_i - \widehat{p}_i|$.

We first establish a bounded-difference property for Z. Consider two samples x^n and x'^n that differ only in their j-th coordinates. For every $i \in \mathbb{N}$, $|\widehat{p}_i(x^n) - \widehat{p}_i(x'^n)| \le \frac{1}{n}$. Therefore, $|Z(x^n) - Z(x'^n)| \le \sup_{i\in\mathbb{N}} |\widehat{p}_i(x^n) - \widehat{p}_i(x'^n)| \le \frac{1}{n}$. Thus, Z satisfies the bounded-difference condition. McDiarmid's inequality consequently gives, for every $t > 0$,

$$\mathbb{P}\left(Z - \mathbb{E}Z \ge t\right) \le \exp(-2nt^2). \tag{20}$$

Taking $t = \sqrt{\log(1/\delta)/(2n)}$ yields, with probability at least $1 - \delta$,

$$Z \le \mathbb{E}Z + \sqrt{\frac{\log(1/\delta)}{2n}}. \tag{21}$$

It remains to bound $\mathbb{E}Z$. Let $X'_1, \ldots, X'_n$ be an independent copy of $X_1, \ldots, X_n$. By the standard symmetrization inequality,

$$\mathbb{E}Z \le \frac{1}{n}\mathbb{E}\sup_{i\in\mathbb{N}} \left| \sum_{j=1}^{n} \left(\mathbb{I}\{X_j = i\} - \mathbb{I}\{X'_j = i\}\right) \right|. \tag{22}$$

Let $\sigma_1, \ldots, \sigma_n$ be independent Rademacher random variables. By exchangeability of X_j and X'_j, the right-hand side of equation 22 is bounded by

$$\mathbb{E}Z \le \frac{2}{n}\mathbb{E}\sup_{i\in\mathbb{N}} \left| \sum_{j=1}^{n} \sigma_j \mathbb{I}\{X_j = i\} \right|. \tag{23}$$

For a fixed realization of X^n, define

$$S_i := \sum_{j=1}^{n} \sigma_j \mathbb{I}\{X_j = i\}, \qquad i \in \mathbb{N}. \tag{24}$$

Since $\sup_i |S_i| \le (\sum_i S_i^2)^{1/2}$, Jensen's inequality gives $\mathbb{E}_\sigma \sup_{i\in\mathbb{N}} |S_i| \le \left(\mathbb{E}_\sigma \sum_{i\in\mathbb{N}} S_i^2\right)^{1/2}$. Expanding the square and using $\mathbb{E}_\sigma[\sigma_j\sigma_\ell] = 0$ for $j \ne \ell$ and $\mathbb{E}_\sigma[\sigma_j^2] = 1$, we obtain $\mathbb{E}_\sigma \sum_{i\in\mathbb{N}} S_i^2 = \sum_{i\in\mathbb{N}} \sum_{j=1}^{n} \mathbb{I}\{X_j = i\} = n$. Hence,

$$\mathbb{E}_\sigma \sup_{i\in\mathbb{N}} |S_i| \le \sqrt{n}. \tag{25}$$

Taking expectation with respect to X^n and substituting equation 25 into equation 23, we obtain $\mathbb{E}Z \le \frac{2}{\sqrt{n}}$. Combining this with equation 21 gives equation 19. □

B Proof for the main error bound

Proof of Theorem 4.1. Firstly, we make the following decomposition:

$$F(P, \boldsymbol{h}) - F(\boldsymbol{D}, \boldsymbol{h}) \quad = \quad F(P, \boldsymbol{h}) - \sum_{i=1}^{K} \frac{n_i}{n} f_i(\boldsymbol{h}) + \sum_{i=1}^{K} \frac{n_i}{n} f_i(\boldsymbol{h}) - F(\boldsymbol{D}, \boldsymbol{h}) \tag{26}$$

Secondly, observe that

$$F(P,\boldsymbol{h}) - \sum_{i=1}^{K} \frac{n_i}{n} f_i(\boldsymbol{h}) = \sum_{i=1}^{K} P(\mathcal{Z}_i) f_i(\boldsymbol{h}) - \sum_{i=1}^{K} \frac{n_i}{n} f_i(\boldsymbol{h}) \quad = \quad \sum_{i=1}^{K} f_i(\boldsymbol{h}) \left[p_i - \frac{n_i}{n} \right].$$

Theorem 3.1 shows that, with probability at least $1 - \delta_1 - \delta_o$:

$$\sum_{i=1}^{K} f_i(\boldsymbol{h}) \left[p_i - \frac{n_i}{n} \right] \quad \leq \quad \alpha \sum_{i \in \mathbb{T}} |f_i(\boldsymbol{h}) - c(\boldsymbol{D})| + |f_o^c| e_o + |f_o^c| \sqrt{\frac{2\ln(1/\delta_o)}{n}}. \tag{27}$$

Furthermore,

$$\begin{aligned}
\sum_{i=1}^{K} \frac{n_i}{n} f_i(\boldsymbol{h}) - F(\boldsymbol{D},\boldsymbol{h}) &= \sum_{i=1}^{K} \frac{n_i}{n} \mathbb{E}_{\boldsymbol{z}\sim P}[\ell(\boldsymbol{h},\boldsymbol{z}) | \boldsymbol{z} \in \mathcal{Z}_i] - \frac{1}{n} \sum_{\boldsymbol{s}\in\boldsymbol{D}} \ell(\boldsymbol{h},\boldsymbol{s}) && (28)\\
&= \sum_{i\in\mathbb{T}} \frac{n_i}{n} \mathbb{E}_{\boldsymbol{z}\sim P}[\ell(\boldsymbol{h},\boldsymbol{z}) | \boldsymbol{z} \in \mathcal{Z}_i] - \frac{1}{n} \sum_{i\in\mathbb{T}} \sum_{\boldsymbol{s}\in\boldsymbol{D}_i} \ell(\boldsymbol{h},\boldsymbol{s}) && (29)\\
&= \frac{1}{n} \sum_{i\in\mathbb{T}} \left[n_i \mathbb{E}_{\boldsymbol{z}\sim P}[\ell(\boldsymbol{h},\boldsymbol{z}) | \boldsymbol{z} \in \mathcal{Z}_i] - \sum_{\boldsymbol{s}\in\boldsymbol{D}_i} \ell(\boldsymbol{h},\boldsymbol{s}) \right] && (30)\\
&= \frac{1}{n} \sum_{i\in\mathbb{T}} \sum_{\boldsymbol{s}\in\boldsymbol{D}_i} \mathbb{E}_{\boldsymbol{z}\sim P}[\ell(\boldsymbol{h},\boldsymbol{z}) - \ell(\boldsymbol{h},\boldsymbol{s}) \mid \boldsymbol{z} \in \mathcal{Z}_i] && (31)\\
&\leq \frac{1}{n} \sum_{i\in\mathbb{T}} \sum_{\boldsymbol{s}\in\boldsymbol{D}_i} \mathbb{E}_{\boldsymbol{z}\sim P}[|\ell(\boldsymbol{h},\boldsymbol{z}) - \ell(\boldsymbol{h},\boldsymbol{s})| \mid \boldsymbol{z} \in \mathcal{Z}_i] && (32)\\
&= \sum_{i\in\mathbb{T}} \frac{n_i}{n} \bar{\epsilon}_i(\boldsymbol{h}). && (33)
\end{aligned}$$

Combining the decomposition (26) with (27) and (33) completes the proof. □

C More experimental details

Data preprocessing and space partition. Before computing the bounds, all images are preprocessed according to the PyTorch implementation of the pretrained ViT-B/16 model.[2] Specifically, each image is resized to 256×256 using bilinear interpolation, followed by a central crop to 224×224. The pixel values are then rescaled to the range $[0.0, 1.0]$. These preprocessing steps are required by the pretrained PyTorch models used in our experiments.

To construct a partition $\{\mathcal{Z}_1, \ldots, \mathcal{Z}_K\}$ of the input space, we randomly sample K points from $[0.0, 1.0]^{C\times H\times W}$ as centroids, since every preprocessed image lies in this space. Each centroid defines one region $\mathcal{Z}_i$ of the partition. Given a training image $\boldsymbol{x}$, we assign it to the region whose centroid is closest in Euclidean distance. This assignment allows us to obtain the corresponding subset $\boldsymbol{D}_i$ for each region $\mathcal{Z}_i$ from the training set $\boldsymbol{D}$.

D Proofs for the Off-Policy Evaluation Application

D.1 Two-sided multinomial consequence

We first derive the two-sided form used in the OPE analysis. Use the raw sample $\boldsymbol{X}$, empirical probabilities $\widehat{p}_i$, and visited set $\mathbb{T} = \{i : \widehat{p}_i > 0\}$ from Theorem 3.1, and set $g_i := f_i(\boldsymbol{X}) - c(\boldsymbol{X})$. Define

$$U^c := \sum_{i\notin\mathbb{T}} g_i p_i. \tag{34}$$

[2] https://pytorch.org/vision/0.20/models/generated/torchvision.models.vit_b_16.html

Since p and $\widehat{p}$ are probability vectors,

$$\sum_{i=1}^{K} f_i(\widehat{p}_i - p_i) = \sum_{i=1}^{K} g_i(\widehat{p}_i - p_i). \tag{35}$$

Moreover, since $\widehat{p}_i = 0$ for $i \notin \mathbb{T}$,

$$\sum_{i=1}^{K} g_i(\widehat{p}_i - p_i) = \sum_{i \in \mathbb{T}} g_i(\widehat{p}_i - p_i) - U^c. \tag{36}$$

Lemma D.1 (Two-sided observed-cell deviation)**.** *Under the conditions of Theorem 3.1, define*

$$\alpha_n(\delta) := \frac{2}{\sqrt{n}} + \sqrt{\frac{\log(1/\delta)}{2n}}. \tag{37}$$

For every $\delta \in (0,1)$, with probability at least $1 - 2\delta$,

$$\left| \sum_{i=1}^{K} f_i(\widehat{p}_i - p_i) + U^c \right| \le \alpha_n(\delta) \sum_{i \in \mathbb{T}} |f_i - c(\boldsymbol{X})|. \tag{38}$$

Furthermore, for every $\delta_1, \delta_o \in (0,1)$, with probability at least $1 - 2\delta_1 - 2\delta_o$,

$$\left| \sum_{i=1}^{K} f_i(\widehat{p}_i - p_i) \right| \le \alpha_n(\delta_1) \sum_{i \in \mathbb{T}} |f_i - c(\boldsymbol{X})| + |f_o^c| \left(e_o + \sqrt{\frac{2\log(1/\delta_o)}{n}} \right). \tag{39}$$

Proof. Apply equation 2 to $g_i = f_i(\boldsymbol{X}) - c(\boldsymbol{X})$ with center zero. Since

$$\sum_{i=1}^{K} g_i(p_i - \widehat{p}_i) = -\sum_{i=1}^{K} f_i(\widehat{p}_i - p_i), \tag{40}$$

with probability at least $1 - \delta$,

$$-\sum_{i=1}^{K} f_i(\widehat{p}_i - p_i) \le \alpha_n(\delta) \sum_{i \in \mathbb{T}} |g_i| + U^c. \tag{41}$$

Therefore,

$$\sum_{i=1}^{K} f_i(\widehat{p}_i - p_i) + U^c \ge -\alpha_n(\delta) \sum_{i \in \mathbb{T}} |g_i|. \tag{42}$$

Apply the same theorem to $-g_i$. Its unseen contribution is $-U^c$, and

$$\sum_{i=1}^{K} (-g_i)(p_i - \widehat{p}_i) = \sum_{i=1}^{K} f_i(\widehat{p}_i - p_i). \tag{43}$$

Thus, with probability at least $1 - \delta$,

$$\sum_{i=1}^{K} f_i(\widehat{p}_i - p_i) \le \alpha_n(\delta) \sum_{i \in \mathbb{T}} |g_i| - U^c. \tag{44}$$

Equations equation 42 and equation 44, followed by a union bound, prove equation 38.

For the second statement, apply equation 4 to g and $-g$. The unseen average for $-g$ is $-f_o^c$, whose absolute value is $|f_o^c|$. A union bound proves equation 39. □

D.2 EXACT TRAJECTORY-LEVEL LINEARIZATION

Let $q = (q_1, \dots, q_T)$ and $\widehat{q} = (\widehat{q}_1, \dots, \widehat{q}_T)$. Let Φ be the OPE functional satisfying

$$V(\pi) = \Phi(q), \qquad \widehat{V}(\pi) = \Phi(\widehat{q}) + \mathcal{R}_{\text{emp}}. \tag{45}$$

Lemma D.2 (Exact integral representation)**.** *Suppose Φ is continuously Fréchet differentiable on the segment $\{q + u(\widehat{q} - q) : u \in [0, 1]\}$. Then*

$$\Phi(\widehat{q}) - \Phi(q) = \sum_{t=1}^{T} \sum_{y \in \mathcal{Y}_t} \overline{F}_{t,y}(\widehat{q}_{t,y} - q_{t,y}), \tag{46}$$

where

$$\overline{F}_{t,y} := \int_0^1 D_{t,y}\Phi(q + u(\widehat{q} - q))\, du. \tag{47}$$

Proof. Define $\psi(u) := \Phi(q + u(\widehat{q} - q))$. Continuous Fréchet differentiability gives

$$\psi'(u) = D\Phi(q + u(\widehat{q} - q))[\widehat{q} - q]. \tag{48}$$

The fundamental theorem of calculus implies

$$\Phi(\widehat{q}) - \Phi(q) = \int_0^1 D\Phi(q + u(\widehat{q} - q))[\widehat{q} - q]\, du. \tag{49}$$

Since each alphabet is finite,

$$D\Phi(\widetilde{q})[v] = \sum_{t=1}^{T} \sum_{y \in \mathcal{Y}_t} D_{t,y}\Phi(\widetilde{q})v_{t,y}. \tag{50}$$

Substitution into equation 49 and interchange of a finite sum with the integral prove the result. □

Lemma D.2 produces an exact data-dependent coefficient vector and therefore directly exploits the main multinomial theorem. If the coefficients are instead evaluated at q, their derivative variation forms part of $\mathcal{R}_{\text{ho}}$.

D.3 DERIVATION OF THE TRAJECTORY-LEVEL REPRESENTATION

We derive the trajectory-level representation used in Assumption 5.1. The derivation is formulated in terms of complete trajectories so that every empirical quantity entering the OPE estimator is a function of a single multinomial count vector.

Let Ω denote the finite set of complete observed trajectories

$$\omega = (O_0, O_1, A_1, R_1, \dots, O_T, A_T, R_T). \tag{51}$$

Let $q = (q_\omega)_{\omega \in \Omega}$ be the behavior-policy distribution over Ω, and let

$$\widehat{q}_\omega := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\omega^i = \omega\}. \tag{52}$$

Since the n trajectories are independent,

$$(n\widehat{q}_\omega)_{\omega \in \Omega} \sim \text{Multinomial}\,(n; (q_\omega)_{\omega \in \Omega})\,. \tag{53}$$

Every empirical marginal or conditional probability used by the OPE estimator is a function of $\widehat{q}$. For example, if E and C are trajectory events and $\widehat{q}(C) > 0$, then

$$\widehat{\Pr}(E \mid C) = \frac{\sum_{\omega \in E \cap C} \widehat{q}_\omega}{\sum_{\omega \in C} \widehat{q}_\omega}. \tag{54}$$

Consequently, empirical proxy matrices, empirical inverse-proxy operators, proxy value functions, and the final value estimate can all be regarded as functions of $\widehat{q}$ on a suitable domain.

Let Φ denote the resulting OPE functional, so that

$$V(\pi) = \Phi(q), \qquad \widehat{V}(\pi) = \Phi(\widehat{q}) + \mathcal{R}_{\mathrm{emp}}, \tag{55}$$

where $\mathcal{R}_{\mathrm{emp}}$ contains any final empirical averaging term not already included in $\Phi(\widehat{q})$.

Let $\mathcal{Q}$ be a convex subset of the probability simplex containing the line segment

$$[q, \widehat{q}] := \{q + u(\widehat{q} - q) : u \in [0, 1]\}. \tag{56}$$

We assume that, on an open neighborhood of $\mathcal{Q}$, all conditional probabilities defining the proxy Bellman recursion have positive denominators and each proxy matrix to which the Moore–Penrose inverse is applied has constant rank. These conditions ensure that Φ is well defined and smooth on a neighborhood of $[q, \widehat{q}]$.

Since $\widehat{q} \in \mathcal{Q}$, the positivity condition requires every conditioning context used by the estimator to be visited in the sample. For contexts with positive population mass, this is also implied by the sufficient-visitation condition of Kuang et al. (2025) (their Assumption 2(c)). Their Assumption 2(b) imposes full column rank at the population and empirical matrices; our smoothness argument additionally requires constant rank on a neighborhood of $\mathcal{Q}$. If a conditioning context is unvisited, this sufficient-condition argument is unavailable: assigning a default value does not by itself establish the smoothness required by Lemmas D.2–D.4. Visiting every context at time t requires n to be at least the number of such contexts. The comparison with Kuang et al. (2025) in Section 5 therefore concerns first-order bounds within this common visitation regime, under the respective regularity assumptions and our coefficient conditions; it does not establish polynomial sample complexity. Visiting every context does not imply observing every history–outcome cell, so the unseen-cell terms U_t^c remain relevant.

Lemma D.3 (Exact trajectory-level representation)**.** *Suppose Φ is continuously Fréchet differentiable on an open neighborhood of $[q, \widehat{q}]$. Then*

$$\widehat{V}(\pi) - V(\pi) = \sum_{\omega \in \Omega} \overline{F}_\omega (\widehat{q}_\omega - q_\omega) + \mathcal{R}_{\mathrm{emp}}, \tag{57}$$

where

$$\overline{F}_\omega := \int_0^1 D_\omega \Phi(q + u(\widehat{q} - q))\, du. \tag{58}$$

The coefficient vector $(\overline{F}_\omega)_{\omega \in \Omega}$ may depend on the complete empirical count vector $(n\widehat{q}_\omega)_{\omega \in \Omega}$.

Proof. Define the scalar function

$$\psi(u) := \Phi(q + u(\widehat{q} - q)), \qquad u \in [0, 1]. \tag{59}$$

Continuous Fréchet differentiability of Φ implies that ψ is continuously differentiable and

$$\psi'(u) = D\Phi(q + u(\widehat{q} - q))[\widehat{q} - q]. \tag{60}$$

The fundamental theorem of calculus gives

$$\Phi(\widehat{q}) - \Phi(q) = \int_0^1 D\Phi(q + u(\widehat{q} - q))[\widehat{q} - q]\, du. \tag{61}$$

Because Ω is finite, the derivative has the coordinate representation

$$D\Phi(\widetilde{q})[v] = \sum_{\omega \in \Omega} D_\omega \Phi(\widetilde{q}) v_\omega. \tag{62}$$

Substituting equation 62 into equation 61 and interchanging the finite sum with the integral yield

$$\Phi(\widehat{q}) - \Phi(q) = \sum_{\omega \in \Omega} \left\{ \int_0^1 D_\omega \Phi(q + u(\widehat{q} - q))\, du \right\} (\widehat{q}_\omega - q_\omega). \tag{63}$$

equation 57 follows from equations equation 55 and equation 58. □

Lemma D.3 gives an exact data-dependent multinomial functional. It therefore permits a direct application of Theorem 3.1 with the raw sample $\boldsymbol{X} = \mathcal{D}$, complete-trajectory cells indexed by $\omega \in \Omega$, $p_\omega = q_\omega$, $\widehat{p}_\omega = \widehat{q}_\omega$, and $f_\omega(\mathcal{D}) = \overline{F}_\omega$. No union bound over time or history contexts is needed. In particular, the observed-cell term contains neither $\log|\Omega|$ nor $\log T$.

The exact representation may be less convenient for structural analysis because $\overline{F}_\omega$ is an integrated derivative. We therefore also record the population-linearized form.

Lemma D.4 (Population-linearized trajectory representation)**.** *Let $\|\cdot\|$ be a seminorm on the zero-sum subspace $\mathbb{R}_0^\Omega := \{v \in \mathbb{R}^\Omega : \sum_{\omega\in\Omega} v_\omega = 0\}$, which contains $\widehat{q} - q$ because q and $\widehat{q}$ are probability vectors. Suppose Φ is twice continuously Fréchet differentiable on an open neighborhood of $[q, \widehat{q}]$ and satisfies*

$$|D^2\Phi(\widetilde{q})[u, v]| \le M_T\|u\|\,\|v\| \tag{64}$$

for every $\widetilde{q} \in [q, \widehat{q}]$ and all $u, v \in \mathbb{R}_0^\Omega$. Then

$$\widehat{V}(\pi) - V(\pi) = \sum_{\omega\in\Omega} F_\omega(\widehat{q}_\omega - q_\omega) + \mathcal{R}_{\mathrm{ho}}, \tag{65}$$

where

$$F_\omega := D_\omega\Phi(q) \tag{66}$$

and

$$|\mathcal{R}_{\mathrm{ho}}| \le \frac{M_T}{2}\|\widehat{q} - q\|^2 + |\mathcal{R}_{\mathrm{emp}}|. \tag{67}$$

Proof. Taylor's theorem with integral remainder gives

$$\Phi(\widehat{q}) = \Phi(q) + D\Phi(q)[\widehat{q} - q] + \int_0^1 (1-u)D^2\Phi(q + u(\widehat{q} - q))[\widehat{q} - q, \widehat{q} - q]\,du. \tag{68}$$

Since Ω is finite,

$$D\Phi(q)[\widehat{q} - q] = \sum_{\omega\in\Omega} D_\omega\Phi(q)(\widehat{q}_\omega - q_\omega). \tag{69}$$

By equation 64,

$$\left|\int_0^1 (1-u)D^2\Phi(q + u(\widehat{q} - q))[\widehat{q} - q, \widehat{q} - q]\,du\right| \le \frac{M_T}{2}\|\widehat{q} - q\|^2. \tag{70}$$

Combining equations equation 55, equation 68, equation 69, and equation 70 proves the claim. □

Time-indexed representation. The time-indexed representation used in the main text follows by grouping complete trajectories according to their time-t history-outcome cells. Let $\varphi_t : \Omega \to \mathcal{Y}_t$ be the corresponding projection, and define

$$q_{t,y} := \sum_{\omega:\varphi_t(\omega)=y} q_\omega, \qquad \widehat{q}_{t,y} := \sum_{\omega:\varphi_t(\omega)=y} \widehat{q}_\omega. \tag{71}$$

For every fixed t,

$$(n\widehat{q}_{t,y})_{y\in\mathcal{Y}_t} \sim \mathrm{Multinomial}\left(n; (q_{t,y})_{y\in\mathcal{Y}_t}\right). \tag{72}$$

The vectors $\widehat{q}_t$ need not be independent across t, because they are computed from the same trajectories. Independence is unnecessary when applying the multinomial theorem separately at each time and combining the resulting events through a union bound.

Suppose the derivative coefficients admit the additive decomposition

$$F_\omega = \sum_{t=1}^T F_{t,\varphi_t(\omega)}. \tag{73}$$

Then

$$\sum_{\omega\in\Omega} F_\omega(\widehat{q}_\omega - q_\omega) = \sum_{t=1}^T \sum_{y\in\mathcal{Y}_t} F_{t,y}(\widehat{q}_{t,y} - q_{t,y}). \tag{74}$$

Thus, Assumption 5.1 follows from Lemma D.4 whenever the derivative of the OPE functional decomposes as in equation 73.

Connection with proxy Bellman operators. For a fixed target policy, write the proxy Bellman recursion in affine form

$$b_{V,t}(q) = r_t(q) + P_t(q) b_{V,t+1}(q), \qquad b_{V,T+1}(q) = 0, \tag{75}$$

where $r_t(q)$ is the proxy reward term and $P_t(q)$ is the linear continuation operator, with their conditional probabilities and proxy inverses expressed as functions of q. The OPE functional is

$$\Phi(q) = \mathcal{A}_1(q) b_{V,1}(q), \tag{76}$$

where $\mathcal{A}_1(q)$ is the linear initial value-aggregation operator. Smooth conditional probability maps and proxy matrices of constant rank on a neighborhood of $[q, \widehat{q}]$ give smooth Moore–Penrose inverse maps there, and hence smooth r_t, P_t, and Φ.

For a direction v, set $a_t(q)[v] := Dr_t(q)[v] + DP_t(q)[v] b_{V,t+1}(q)$. Differentiating equation 75 gives

$$Db_{V,t}(q)[v] = a_t(q)[v] + P_t(q) Db_{V,t+1}(q)[v]. \tag{77}$$

Iterating this recursion yields

$$Db_{V,1}(q)[v] = \sum_{t=1}^{T} P_1(q) \cdots P_{t-1}(q) a_t(q)[v], \tag{78}$$

where the empty product is the identity. Therefore,

$$D\Phi(q)[v] = D\mathcal{A}_1(q)[v] b_{V,1}(q) + \sum_{t=1}^{T} \mathcal{A}_1(q) P_1(q) \cdots P_{t-1}(q) a_t(q)[v]. \tag{79}$$

If $Dr_t(q)[v]$ and $DP_t(q)[v]$ depend on v only through the time-t marginal perturbation, and $D\mathcal{A}_1(q)[v]$ depends only on the time-1 marginal perturbation, then equation 79 admits the additive decomposition equation 73.

Data-dependent coefficients. There are two valid ways to use the representation.

First, Lemma D.3 gives coefficients $\overline{F}_\omega$ that depend on $\widehat{q}$. This exact representation fully exploits the data-dependent-weight theorem and has no Taylor remainder beyond $\mathcal{R}_{\mathrm{emp}}$. Its effective support, however, must be controlled for these integrated coefficients, for example uniformly along the segment from q to $\widehat{q}$.

Second, Lemma D.4 uses the population coefficients $F_\omega = D_\omega \Phi(q)$. These coefficients are often easier to interpret through the Bellman recursion, but the second-order term in equation 67 must be retained in $\mathcal{R}_{\mathrm{ho}}$.

A time-indexed coefficient $F_{t,y}$ and its center c_t may depend on the full raw trajectory sample $\mathcal{D}$, including statistics outside the time-t counts. Theorem 3.1 applies using the multinomial law in Assumption 5.1 at each t.

D.4 Proof of the effective-history theorem

Proof of Theorem 5.1. For each t, apply Lemma D.1 to the time-t multinomial vector in Assumption 5.1, with coefficients $F_{t,y}$, center c_t, and $\delta_t = \delta/(2T)$. Independence across t is not required.

A union bound gives, simultaneously for all t,

$$\left| \sum_{y \in \mathcal{Y}_t} F_{t,y}(\widehat{q}_{t,y} - q_{t,y}) + U_t^c \right| \le \alpha_T(\delta) B_t(c_t) \sqrt{s_{\mathrm{hist},t}(F_t; c_t)}. \tag{80}$$

Let $A_t := \sum_{y \in \mathcal{Y}_t} F_{t,y}(\widehat{q}_{t,y} - q_{t,y}) + U_t^c$. Assumption 5.1 gives $\widehat{V}(\pi) - V(\pi) + \sum_t U_t^c - \mathcal{R}_{\mathrm{ho}} = \sum_t A_t$. Thus, $|\sum_t A_t| \le \sum_t |A_t|$ and equation 80 yield

$$\left| \widehat{V}(\pi) - V(\pi) + \sum_{t=1}^{T} U_t^c - \mathcal{R}_{\mathrm{ho}} \right| \le \alpha_T(\delta) \sum_{t=1}^{T} B_t(c_t) \sqrt{s_{\mathrm{hist},t}(F_t; c_t)}. \tag{81}$$

This proves the first inequality in Theorem 5.1; applying the triangle inequality once more gives its displayed bound. □

E COMPLETING THE HISTORY-DEPENDENT OPE GUARANTEE

Section 5.1 controls the observed-cell fluctuation of the first-order OPE error. This appendix supplies the two components a complete guarantee additionally requires: the unseen-cell contribution to the same first-order term (Appendix E.1) and the higher-order remainder $\mathcal{R}_{\mathrm{ho}}$ (Appendix E.2).

E.1 EXACT AND CONCENTRATED CONTROL OF UNSEEN HISTORIES

The separation between observed and unseen cells is essential when $n \ll |\mathcal{Y}_t|$. Classical missing-mass theory controls the total probability assigned to unseen categories through Good–Turing estimation and related concentration inequalities (Good, 1953; McAllester & Schapire, 2000; McAllester & Ortiz, 2003; Berend & Kontorovich, 2013). OPE has an additional structure because unseen histories are multiplied by signed, data-dependent Bellman coefficients.

The exact term

$$\mathfrak{U}_n^{\mathrm{signed}} := \sum_{t=1}^{T} U_t^c \tag{82}$$

preserves cancellation among unseen histories and across time. This form is preferable when structural information gives a direct bound on $\mathfrak{U}_n^{\mathrm{signed}}$.

Alternatively, the concentrated unseen-mass inequality in Theorem 3.1 gives

$$\mathfrak{U}_n^{\mathrm{hist}} := \sum_{t=1}^{T} |F_{o,t}^c| \left(e_{o,t} + \sqrt{\frac{2\log(4T/\delta)}{n}} \right), \tag{83}$$

where $e_{o,t}$ is the expected missing mass and $F_{o,t}^c$ is the average centered coefficient over unseen trajectory cells. This form is easier to use when only coefficient magnitude bounds are available, but it does not preserve sign cancellation.

Sparse observed histories alone are insufficient if a rare unobserved history has large probability and Bellman influence. Near constancy, history aggregation, predictive-state equivalence, and explicit coverage of active histories provide natural mechanisms for controlling both observed and unseen contributions.

E.2 POLYNOMIAL CONTROL OF HIGHER-ORDER ERRORS

We bound $\mathcal{R}_{\mathrm{ho}}$ for the population-linearized construction of Lemma D.4. Its Taylor remainder depends on $D^2\Phi$ throughout $[q, \widehat{q}]$, so effective sparsity of the first-order coefficients at q is insufficient by itself. We give sufficient conditions when the functional uses fixed linear summaries of the time-marginal probabilities.

For $u \in \mathbb{R}^{\Omega}$, write $(\Pi_t u)_y := \sum_{\omega:\varphi_t(\omega)=y} u_\omega$. Then $\Pi_t(\widehat{q} - q) = \widehat{q}_t - q_t$, and $\sum_y (\Pi_t u)_y = 0$ whenever $u \in \mathbb{R}_0^{\Omega}$. For fixed matrices $S_t \in \mathbb{R}^{m_t \times |\mathcal{Y}_t|}$, assume that, on a neighborhood of $[q, \widehat{q}]$,

$$\Phi(\widetilde{q}) = \Psi\big((S_t \Pi_t \widetilde{q})_{t=1}^T\big), \tag{84}$$

where Ψ is twice continuously differentiable on a neighborhood of the corresponding summary vectors. This factorization holds if the proxy reward, continuation, and initial aggregation maps are smooth functions of these summaries. The matrices S_t may depend on the population law but are not selected using $\mathcal{D}$.

Two seminorms for probability errors. The cellwise bound $\mathbb{E}\|\widehat{q}_t - q_t\|_1 \le \min\{2, \sqrt{|\mathcal{Y}_t|/n}\}$ need not shrink when the alphabet is much larger than the sample. We instead measure class-total errors or the largest cell error. Fix a partition $C_{t,1}, \dots, C_{t,d_t}$ of each $\mathcal{Y}_t$, independently of $\mathcal{D}$, and define $(G_t w)_k := \sum_{y \in C_{t,k}} w_y$. Set

$$\|u\|_G := \sum_{t=1}^{T} \|G_t \Pi_t u\|_1, \qquad \|u\|_{\max} := \sum_{t=1}^{T} \|\Pi_t u\|_\infty. \tag{85}$$

A smaller probability error is useful only if the curvature is bounded in the same seminorm. The following lemma supplies this bound for pooling and for summary rows with bounded centered ℓ_1 mass.

Lemma E.1 (Curvature under summary structure)**.** *Assume equation 84. Let M_Ψ be a deterministic upper bound on the absolute values of all second partial derivatives of Ψ at $(S_t\Pi_t\widetilde{q})_{t=1}^T$, for every $\widetilde{q} \in [q, \widehat{q}]$. For $u, v \in \mathbb{R}_0^\Omega$:*

(a) If $S_t = G_t$ for every t, then

$$|D^2\Phi(\widetilde{q})[u,v]| \le M_\Psi \|u\|_G \|v\|_G.$$

(b) If $\sum_y |S_{t,jy} - c_{tj}| \le C$ for some c_{tj} for every row j and time t, then, with $m := \max_t m_t$,

$$|D^2\Phi(\widetilde{q})[u,v]| \le m^2 C^2 M_\Psi \|u\|_{\max} \|v\|_{\max}.$$

Proof. The summary map is linear, so the chain rule gives

$$D^2\Phi(\widetilde{q})[u,v] = \sum_{t,t'=1}^{T} \sum_{j=1}^{m_t} \sum_{l=1}^{m_{t'}} \partial^2_{tj,t'l}\Psi \, (S_t\Pi_t u)_j (S_{t'}\Pi_{t'} v)_l,$$

with derivatives evaluated at $(S_t\Pi_t\widetilde{q})_t$. Hence

$$|D^2\Phi(\widetilde{q})[u,v]| \le M_\Psi \Big(\sum_t \|S_t\Pi_t u\|_1 \Big) \Big(\sum_{t'} \|S_{t'}\Pi_{t'} v\|_1 \Big).$$

Case (a) follows from the definition of $\|\cdot\|_G$. In case (b), the zero-sum identity gives

$$|(S_t\Pi_t u)_j| = \Big| \sum_y (S_{t,jy} - c_{tj})(\Pi_t u)_y \Big| \le C \|\Pi_t u\|_\infty.$$

Thus $\sum_t \|S_t\Pi_t u\|_1 \le mC\|u\|_{\max}$, and the same bound for v proves (b). □

The two seminorms exploit different structures. Bounding class totals by $\|\Pi_t u\|_\infty$ can introduce class-size factors, whereas case (b) exploits cancellation of the constant part of each row.

Lemma E.2 (Concentration in coarse seminorms)**.** *Let $d_T := \max_t d_t$. Each of the following bounds holds with probability at least $1 - \delta$:*

$$\|\widehat{q} - q\|_G \le \rho_n^G := T\Big(\sqrt{\tfrac{d_T}{n}} + \sqrt{\tfrac{2\log(T/\delta)}{n}} \Big), \qquad \|\widehat{q} - q\|_{\max} \le \rho_n^{\max} := T\,\alpha_n(\delta/T). \tag{86}$$

Proof. For fixed t, $G_t\widehat{q}_t$ is the empirical distribution of n independent class labels with law $G_t q_t$. Jensen's and the Cauchy–Schwarz inequalities give $\mathbb{E}\|G_t(\widehat{q}_t - q_t)\|_1 \le \sum_k \sqrt{(G_t q_t)_k / n} \le \sqrt{d_t/n}$. Replacing one trajectory changes this norm by at most $2/n$. McDiarmid's inequality at level δ/T, followed by a union bound over t, proves the first bound. Applying Lemma A.2 at level δ/T to each marginal and taking a union bound proves the second. Independence across times is unnecessary. □

Combining these lemmas with Lemma D.4 gives, with probability at least $1 - \delta$,

$$|\mathcal{R}_{\mathrm{ho}}| \le \frac{M_T}{2}\rho_n^2 + |\mathcal{R}_{\mathrm{emp}}|, \tag{87}$$

where $(M_T, \rho_n) = (M_\Psi, \rho_n^G)$ in case (a) and $(M_T, \rho_n) = (m^2C^2M_\Psi, \rho_n^{\max})$ in case (b). These probability statements assume the smoothness and curvature bounds hold almost surely. If they hold only on a joint event of probability at least $1 - \eta$, the success probability becomes $1 - \delta - \eta$; the same qualification applies to the complete bound below.

Connection with the first-order regimes. The sparsity conditions concern the centered derivative coefficients F_t. To identify them under equation 84, set

$$a_{tj} := \partial_{tj}\Psi\big((S_s q_s)_{s=1}^T\big).$$

The chain rule gives $D\Phi(q)[u] = \sum_{t,j} a_{tj}(S_t \Pi_t u)_j$, so the additive representation holds with $F_{t,y} = \sum_j a_{tj} S_{t,jy}$.

Sparse active histories. In case (a), let $k_t(y)$ index the class containing y. Then $F_{t,y} = a_{t,k_t(y)}$. Use singleton classes for the cells of I_t, with $|I_t| \le s_H$, and one class for the remaining cells. Choosing their common coefficient as c_t (arbitrary if there are no remaining cells) gives

$$F_{t,y} - c_t = 0 \quad (y \notin I_t), \qquad d_t \le |I_t| + 1 \le s_H + 1.$$

By Cauchy–Schwarz, $s_{\mathrm{hist},t}(F_t; c_t) \le |I_t|$. This conclusion assumes the functional uses these class totals throughout a neighborhood, as required by equation 84.

Near constancy. In case (b), write $r_{t,jy} := S_{t,jy} - c_{tj}$ and choose $\mu_t := \sum_j a_{tj} c_{tj}$. Then $F_{t,y} - \mu_t = \sum_j a_{tj} r_{t,jy}$ and

$$\sum_y |F_{t,y} - \mu_t| \le \sum_j |a_{tj}| \sum_y |r_{t,jy}| \le C \sum_j |a_{tj}| \le mC \max_j |a_{tj}|.$$

This bounds the total centered variation over all cells, and hence over the observed cells.

Approximate sparsity. In case (b), take s rows $S_{t,jy} = \mathbf{1}\{y = y_j\}$ and write the remaining rows as $S_{t,jy} = c_{tj} + r_{t,jy}$ with $\sum_y |r_{t,jy}| \le C$. Choosing $\mu_t := \sum_{j>s} a_{tj} c_{tj}$ gives a sparse component plus a residual:

$$F_{t,y} - \mu_t = \sum_{j=1}^{s} a_{tj} \mathbf{1}\{y = y_j\} + e_{t,y}, \qquad e_{t,y} := \sum_{j>s} a_{tj} r_{t,jy}.$$

With center μ_t, the observed approximation tail satisfies $\tau_t(s) \le \sum_y |e_{t,y}| \le C \sum_{j>s} |a_{tj}|$. Indicator rows have ℓ_1 mass one with center zero, so the common row bound in Lemma E.1(b) is $\max\{1, C\}$ for this construction.

These constructions give sufficient summary structure for both the first-order coefficient bounds and the curvature bound. First-order effective sparsity alone does not imply the factorization equation 84.

Complete bound. Assume the coefficient bounds of Corollary 5.1 and either case of Lemma E.1. Allocate failure probability $\delta/3$ to the first-order bound, the unseen bound equation 83, and Lemma E.2. A union bound gives, with probability at least $1 - \delta$,

$$|\widehat{V}(\pi) - V(\pi)| \lesssim BT\sqrt{\frac{s_H \log(2T/\delta)}{n}} + \mathfrak{U}_n^{\mathrm{hist}} + M_T T^2 \frac{d_T + \log(2T/\delta)}{n} + |\mathcal{R}_{\mathrm{emp}}|, \tag{88}$$

where $\mathfrak{U}_n^{\mathrm{hist}}$ is evaluated at $\delta/3$ and d_T is replaced by 1 in case (b). The bound has polynomial horizon dependence if B, s_H, d_T, M_T, and the bounds on the unseen and empirical terms are polynomially controlled in T. In particular, controlling M_Ψ requires controlling any proxy-inverse conditioning in Ψ; small summary dimension alone is insufficient. For fixed T, δ, and an n-independent M_T, the curvature contribution is $O(1/n)$, while the unseen and empirical terms require separate control.