OPTIMAL POSTERIOR SAMPLING FOR FIXED-BUDGET CLUSTER IDENTIFICATION THROUGH PAIRWISE OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study fixed-budget cluster identification from noisy pairwise observations. A collection of items is partitioned into an unknown number of clusters. Querying a pair produces a Bernoulli response with an unknown mean $p > 1/2$ if the items belong to the same cluster and $q < 1/2$ otherwise. We establish an instance-dependent lower bound on the posterior probability of an incorrect clustering that holds for every adaptive sampling rule. Its exponent is given by a max–min allocation problem that maximizes the minimum weighted Bernoulli KL divergence between the true instance and alternatives with incorrect clusterings. We propose an adaptive algorithm that combines posterior sampling over neighboring alternatives, an exponential-weights learner over pairs, and vanishing uniform exploration. We prove an upper bound on the algorithm's posterior error probability that matches the lower bound at the exponential scale. The algorithm therefore achieves the optimal instance-dependent posterior error exponent almost surely. Experiments on synthetic instances and real-world datasets demonstrate competitive cluster identification performance against existing baselines.

1 INTRODUCTION

Pairwise judgments provide a natural source of supervision for clustering when the categories and their number are unknown. Crowdsourcing has long been used to collect annotations and uncover structure in unlabeled data (Sorokin & Forsyth, 2008; Raykar et al., 2010; Welinder et al., 2010). In entity resolution, for example, annotators determine whether two records refer to the same underlying entity (Wang et al., 2012; Vesdapunt et al., 2014; Verroios & Garcia-Molina, 2015; Mazumdar & Saha, 2017c). Such judgments reveal relationships among items without requiring a predefined list of cluster labels. However, responses can be incorrect, and each additional query consumes annotation resources. The resulting allocation problem is central to crowdsourced clustering: which pairs should be queried, and when is another response more useful than a query elsewhere?

Existing approaches address this problem through different observation models and query designs. Work on crowdclustering, graph clustering with missing data, and systematic edge queries studies how partial observations reveal clustering structure (Gomes et al., 2011; Korlakai Vinayak et al., 2014; Korlakai Vinayak & Hassibi, 2016; Ibrahim & Fu, 2021), while side information can further reduce the query cost (Mazumdar & Saha, 2017b). These problems are closely related to community recovery in stochastic block models (Holland et al., 1983; Condon & Karp, 2001; Abbe & Sandon, 2015; Abbe, 2018). Random and adaptive sampling have also been compared within community detection (Yun & Proutiere, 2014). A complementary line of work on clustering with noisy or faulty oracles develops query-complexity guarantees and computationally efficient recovery algorithms under different assumptions on the noise and cluster sizes (Mazumdar & Saha, 2017a; Green Larsen et al., 2020; Peng & Zhang, 2021; Xia & Huang, 2022; Mukherjee et al., 2023).

Allowing repeated queries is particularly useful when the response probabilities are unknown. Chen et al. (2023) construct clusters through sequential pairwise tests, drawing on finite-law-of-the-iterated-logarithm methods from bandits and active ranking (Jamieson et al., 2014; Heckel et al., 2019). Their algorithm does not require the number of clusters or the error probabilities, and their crowdsourcing experiments show that the relative benefits of active and passive querying depend on

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.

cluster sizes. Teo et al. (2026) study instance-dependent optimality in the fixed-confidence setting. They derive a Kullback–Leibler (KL) information lower bound and analyze adaptive allocation with likelihood-ratio stopping, quantifying the queries required to meet a prescribed error tolerance.

We consider the complementary setting in which the query budget is fixed in advance. This distinction is standard in best-arm identification (BAI): fixed-confidence methods minimize the expected sampling cost subject to an error constraint (Garivier & Kaufmann, 2016), whereas fixed-budget methods seek to minimize the probability of an incorrect recommendation after a prescribed number of samples (Karnin et al., 2013). Linear and transductive BAI further show how shared statistical structure shapes the allocation problem (Soare et al., 2014; Fiez et al., 2019; Jedra & Proutiere, 2020; Degenne et al., 2020), and related pure-exploration formulations arise in online clustering with bandit feedback (Yang et al., 2024). For pairwise clustering, each pair is a sampling arm, while the target is an entire partition. Its identification requires distinguishing competing partitions subject to consistency among pairwise relationships.

Specifically, the learner receives T noisy pairwise responses and then returns a partition of N items. A response is positive with probability $p^* > 1/2$ for items in the same cluster and $q^* < 1/2$ otherwise. The partition, its number of clusters, and both response parameters are unknown. We study the regime in which N and the true instance remain fixed while T increases. This budget constraint is natural when the available annotation resources are determined before data collection.

A posterior perspective provides an allocation principle for this setting. Large-deviation analyses relate sampling proportions to the rate at which identification errors vanish (Glynn & Juneja, 2004). Russo (2020) develops Bayesian sampling rules that, with suitable tuning, attain an optimal posterior error exponent, with subsequent work studying top-two methods and their fixed-confidence guarantees (Qin et al., 2017; Shang et al., 2019; Jourdan et al., 2022). For Gaussian linear BAI, Li et al. (2024) connect posterior sampling to a min–max information allocation problem. Their method uses sampling and optimization oracles to obtain an almost-sure posterior error exponent matching the optimal fixed-allocation benchmark. This provides a direct motivation for studying posterior concentration under a fixed clustering budget.

Our objective is to characterize how rapidly posterior uncertainty about the partition decreases as the query budget grows. We measure this uncertainty by the posterior mass assigned to incorrect partitions. When the recommendation is sampled from the posterior, this quantity equals its conditional error probability given the observed history. We study its exponential decay along observation histories, with the number of items and the true instance fixed.

The clustering model introduces several challenges. Candidate partitions pool observations differently when learning the shared response parameters, and an incorrect partition may fit the data using parameters other than (p^*, q^*). The information allocation problem must therefore account jointly for alternative partitions and their response parameters. Bernoulli KL divergences can also be unbounded near the parameter-space boundary, complicating regret arguments based on sampled alternatives. Teo et al. (2026) show that the population KL information benchmark can be evaluated using alternatives obtained by a single cluster merge or split. Motivated by this structure and the posterior exploration approach of Li et al. (2024), we develop a posterior-based min–max method for allocating a fixed query budget.

Our contributions are as follows.

- **A universal posterior error lower bound.** We characterize an instance-dependent information value $\Gamma^*(\theta^*)$ through a max–min Bernoulli KL separation from instances with incorrect partitions. For every adaptive sampling rule, the posterior error exponent is almost surely at most this value. The bound holds for any fixed product prior assigning positive mass to every partition and full support to each response parameter.
- **An adaptive posterior-sampling algorithm.** We propose PSPCI, which combines projection onto valid partitions, likelihood-tempered posterior sampling over merge and split alternatives, and an exponential-weights learner using clipped KL gains. The algorithm incorporates vanishing uniform exploration and draws its terminal recommendation from the ordinary, untempered posterior.
- **A matching posterior error upper bound.** Under the stated priors and parameter choices, we prove that PSPCI attains $\Gamma^*(\theta^*)$ as its almost-sure posterior error exponent, matching the universal

lower bound. The proof combines bounds on pooled marginal likelihoods with regret bounds for both learners to establish that the empirical allocation attains the optimal information value.

2 Problem Setting

We study cluster identification from noisy pairwise observations in the fixed-budget setting. Let $[N] := \{1, \dots, N\}$ be a set of $N \geq 2$ items. A clustering is a surjective map $c : [N] \to [M]$, where $M \in [N]$ is the number of clusters. Let $\mathcal{C}_N$ denote the finite set of all such maps. Since cluster labels are arbitrary, we regard two clusterings as equivalent if they induce the same partition:

$$c \sim c' \quad \iff \quad \big(c(i) = c(j) \iff c'(i) = c'(j)\big) \quad \text{for all } 1 \leq i < j \leq N.$$

The true clustering is $c^* : [N] \to [M^*]$. Neither c^* nor the number of clusters M^* is known to the learner.

The learner can query any pair in $\mathcal{I} := \{(i, j) : 1 \leq i < j \leq N\}$, including pairs queried previously. Each query returns a binary response indicating whether the two items belong to the same cluster. The response is noisy: its mean is p for a within-cluster pair and q for a between-cluster pair, where $0 < q < 1/2 < p < 1$. These two parameters are shared across pairs and are unknown to the learner. Write

$$\mathcal{P} := \left(\tfrac{1}{2}, 1\right), \qquad \mathcal{Q} := \left(0, \tfrac{1}{2}\right), \qquad \Theta := \mathcal{C}_N \times \mathcal{P} \times \mathcal{Q}.$$

An instance is $\theta = (c, p, q) \in \Theta$, and the true instance is $\theta^* := (c^*, p^*, q^*)$. For $e = (i, j) \in \mathcal{I}$, define

$$\mu_e(\theta) := \begin{cases} p, & c(i) = c(j), \\ q, & c(i) \neq c(j), \end{cases} \qquad \mu_e^* := \mu_e(\theta^*).$$

Given a budget $T \in \mathbb{N}$, at each round $t \in [T]$ the learner selects a pair $I_t \in \mathcal{I}$ based on past queries and responses, possibly using internal randomization. Let $\mathcal{F}_t := \sigma(I_1, Y_1, \dots, I_{t-1}, Y_{t-1}, I_t)$ denote the information available after selecting I_t and before observing its response. The observation model is

$$Y_t \mid \mathcal{F}_t \sim \text{Bernoulli}(\mu_{I_t}^*).$$

Thus, a positive response is more likely than a negative response exactly when the queried items belong to the same cluster.

After exactly T queries, the learner returns $\widehat{c}_T \in \mathcal{C}_N$. The fixed-budget objective is to minimize the probability of misidentification

$$\mathbb{P}_{\theta^*}(\widehat{c}_T \nsim c^*), \tag{1}$$

where $\mathbb{P}_{\theta^*}$ accounts for both the observations and the learner's randomization. Our theoretical analysis concerns the exponential decay of the posterior mass assigned to incorrect clusterings, defined next. This posterior quantity is distinct from (1).

Posterior error and information benchmark. Let ρ_0 be a prior on $\mathcal{C}_N$ with $\rho_0(c) > 0$ for every c. Conditionally on c, let p and q have independent priors ν_p and ν_q on $\mathcal{P}$ and $\mathcal{Q}$, respectively. These parameter priors do not depend on c, and all priors are fixed independently of the budget. Each parameter prior assigns positive mass to every nonempty open subset of its domain; no density assumption is required. Let Π_T denote the joint posterior on Θ given $(I_t, Y_t)_{t=1}^T$.

Define the alternative set and posterior error mass by

$$\text{Alt}(\theta^*) := \{(c, p, q) \in \Theta : c \nsim c^*\}, \qquad 1 - \alpha_T := \Pi_T\big(\text{Alt}(\theta^*)\big).$$

The alternative response parameters range freely over $\mathcal{P} \times \mathcal{Q}$. If the terminal recommendation is drawn from the clustering marginal of Π_T, then $1 - \alpha_T$ is its conditional probability of misidentification given the observed history.

For $x, y \in (0, 1)$, write $d(x, y) := x \log(x/y) + (1 - x) \log((1 - x)/(1 - y))$ for the Bernoulli KL divergence, with natural logarithms. Let $\Delta(\mathcal{I})$ be the probability simplex over pairs. Define

$$D(\theta^*, \theta; \lambda) := \sum_{e \in \mathcal{I}} \lambda_e d\big(\mu_e^*, \mu_e(\theta)\big),$$

$$\Gamma^*(\theta^*) := \sup_{\lambda \in \Delta(\mathcal{I})} \inf_{\theta \in \text{Alt}(\theta^*)} D(\theta^*, \theta; \lambda). \tag{2}$$

This benchmark maximizes the minimum information against an incorrect clustering. All limits keep N and θ^* fixed, and almost-sure statements refer to the data-generating law $\mathbb{P}_{\theta^*}$.

3 ALGORITHM

Algorithm 1 couples posterior sampling with Hedge. A min-learner samples tempered-posterior competitors, and a max-learner allocates queries using their KL separation from a reference.

We label clusters in increasing order of their smallest items and write $\mathcal{C}_N^\circ$ for the resulting canonical clusterings. For $e = (i, j)$, let $z_e(c) := \mathbf{1}\{c(i) = c(j)\}$. Let $N_s(e)$ and $S_s(e)$ count the queries and positive responses for pair e through round s, respectively. The neighborhood $\mathcal{A}(c)$ consists of the distinct canonical partitions obtained by merging two clusters of c or splitting one cluster into two nonempty clusters.

Algorithm 1 PSPCI: Posterior sampling for pairwise cluster identification

Require: Budget $T \geq 1$; $\alpha \in (0, 1/2)$ and $c_B > 0$.

Use the fixed uniform priors in (5).
Set $\kappa_T \leftarrow \sqrt{T}$, $\eta_T \leftarrow T^{-1/2}$, $B_T \leftarrow c_B \log(T+1)$, and $u_e \leftarrow 1/|\mathcal{I}|$.
Set $N_0(e) = S_0(e) = \ell_{1,e} = 0$ for every $e \in \mathcal{I}$.
for $t = 1, \ldots, T$ **do**
  Set $M_t \leftarrow \max_e \ell_{t,e}$ and, for every $e \in \mathcal{I}$, $\lambda_{t,e} \leftarrow e^{\ell_{t,e} - M_t} / \sum_{e'} e^{\ell_{t,e'} - M_t}$.
  Set $\gamma_t \leftarrow t^{-\alpha}$ and $\widetilde{\lambda}_t \leftarrow (1 - \gamma_t)\lambda_t + \gamma_t u$.
  Compute $\widehat{\mu}_{t-1}$ using (3), and set $\widehat{\sigma}_t \leftarrow \mathrm{Proj}(\widehat{\mu}_{t-1})$ using Algorithm 2.
  Set $\theta_t^+ \leftarrow (\widehat{\sigma}_t, \widehat{p}_{t-1}(\widehat{\sigma}_t), \widehat{q}_{t-1}(\widehat{\sigma}_t))$ using (4).
  Set $\mathcal{A}_t \leftarrow \mathcal{A}(\widehat{\sigma}_t)$.
  Draw $C_t^- \in \mathcal{A}_t$ according to (6).
  Given C_t^-, draw the competing probabilities (P_t^-, Q_t^-) according to (7).
  Set $\theta_t^- \leftarrow (C_t^-, P_t^-, Q_t^-)$.
  For every $e \in \mathcal{I}$, set

$$g_{t,e} \leftarrow \frac{1}{B_T} \min\left\{ d\big(\mu_e(\theta_t^+), \mu_e(\theta_t^-)\big), B_T \right\}.$$

  Independently draw $I_t \sim \widetilde{\lambda}_t$ and observe Y_t.
  For every $e \in \mathcal{I}$, update

$$N_t(e) \leftarrow N_{t-1}(e) + \mathbf{1}\{I_t = e\}, \quad S_t(e) \leftarrow S_{t-1}(e) + \mathbf{1}\{I_t = e\} Y_t, \quad \ell_{t+1,e} \leftarrow \ell_{t,e} + \eta_T g_{t,e}.$$

end for
Draw $\widehat{c}_T \sim \pi_T^{(1)}$, where $\pi_T^{(1)}$ is defined in (8).
return $\widehat{c}_T$.

Empirical reference. For $s \geq 0$, define

$$\widehat{\mu}_{s,e} := \begin{cases} S_s(e)/N_s(e), & N_s(e) > 0, \\ 1/2, & N_s(e) = 0. \end{cases} \tag{3}$$

Following Teo et al. (2026, Algorithm 1), Algorithm 2 projects thresholded means onto canonical partitions to obtain $\widehat{\sigma}_t$. Let $\mathcal{H}_s$ contain the queries, responses, and algorithmic randomization through round s, before the terminal recommendation.

Algorithm 2 Partition projection

Require: Empirical pair means $\widehat{\mu} \in [0, 1]^{\mathcal{I}}$.

Set $\widehat{z}_e \leftarrow \mathbf{1}\{\widehat{\mu}_e \geq 1/2\}$ for every $e \in \mathcal{I}$.
Select

$$\widehat{\sigma} \in \operatorname*{arg\,min}_{c \in \mathcal{C}_N^\circ} \sum_{e \in \mathcal{I}} |\widehat{z}_e - z_e(c)|$$

using a fixed deterministic tie-breaking rule.
return $\mathrm{Proj}(\widehat{\mu}) := \widehat{\sigma}$.

For a candidate partition c, write $\mathcal{I}^=(c) := \{e : z_e(c) = 1\}$ and $\mathcal{I}^\neq(c) := \mathcal{I} \setminus \mathcal{I}^=(c)$, and pool the counts as

$$N_s^r(c) := \sum_{e \in \mathcal{I}^r(c)} N_s(e), \qquad S_s^r(c) := \sum_{e \in \mathcal{I}^r(c)} S_s(e), \qquad r \in \{=, \neq\}.$$

Set $\epsilon_s := 1/[4(s+1)]$ and write $\mathrm{clip}_{[a,b]}(x) := \min\{b, \max\{a, x\}\}$. We use the smoothed empirical estimates

$$\widehat{p}_s(c) := \mathrm{clip}_{[1/2+\epsilon_s,\, 1-\epsilon_s]}\left(\frac{S_s^{=}(c)+1}{N_s^{=}(c)+2}\right), \qquad \widehat{q}_s(c) := \mathrm{clip}_{[\epsilon_s,\, 1/2-\epsilon_s]}\left(\frac{S_s^{\neq}(c)+1}{N_s^{\neq}(c)+2}\right). \tag{4}$$

Pseudocounts handle empty pools, and clipping keeps the estimates in $\mathcal{P} \times \mathcal{Q}$. Together with $\widehat{\sigma}_t$, they define the reference θ_t^+. Parameters of absent pair types do not affect its pair means.

Min-learner. The min-learner samples competitors from $\mathcal{A}_t$ using the fixed independent priors

$$\varrho_0(c) := \frac{1}{|\mathcal{C}_N^\circ|}, \qquad c \in \mathcal{C}_N^\circ, \qquad \nu_p := \mathrm{Unif}(\mathcal{P}), \qquad \nu_q := \mathrm{Unif}(\mathcal{Q}). \tag{5}$$

These priors are independent of T and define Π_s. A labeled prior induces $\varrho_0(c) = \sum_{\widetilde{c}\sim c} \rho_0(\widetilde{c})$.

For $a, b > 0$ and $J_+ = \mathcal{P}$, $J_- = \mathcal{Q}$, let $\mathrm{TBeta}_\pm(a, b)$ have density

$$f_\pm(x; a, b) := \frac{x^{a-1}(1-x)^{b-1}\mathbf{1}\{x \in J_\pm\}}{Z_\pm(a,b)}, \qquad Z_\pm(a,b) := \int_{J_\pm} z^{a-1}(1-z)^{b-1}\, dz.$$

The likelihood is

$$L_s(c, p, q) := p^{S_s^{=}(c)}(1-p)^{N_s^{=}(c)-S_s^{=}(c)} q^{S_s^{\neq}(c)}(1-q)^{N_s^{\neq}(c)-S_s^{\neq}(c)}.$$

For $\kappa \geq 1$, tempering only the likelihood gives a working distribution proportional to $\varrho_0(c) L_s(c, p, q)^{1/\kappa} \nu_p(dp)\nu_q(dq)$. Its conditional Beta parameters are

$$\begin{aligned} A_{s,\kappa}^p(c) &:= 1 + S_s^{=}(c)/\kappa, & B_{s,\kappa}^p(c) &:= 1 + [N_s^{=}(c) - S_s^{=}(c)]/\kappa, \\ A_{s,\kappa}^q(c) &:= 1 + S_s^{\neq}(c)/\kappa, & B_{s,\kappa}^q(c) &:= 1 + [N_s^{\neq}(c) - S_s^{\neq}(c)]/\kappa. \end{aligned}$$

The conditional parameter law and partition weights are

$$\begin{aligned} \mathsf{Q}_{s,\kappa}(\cdot \mid c) &:= \mathrm{TBeta}_+\big(A_{s,\kappa}^p(c), B_{s,\kappa}^p(c)\big) \otimes \mathrm{TBeta}_-\big(A_{s,\kappa}^q(c), B_{s,\kappa}^q(c)\big), \\ w_s^{(\kappa)}(c) &:= \varrho_0(c) Z_+\big(A_{s,\kappa}^p(c), B_{s,\kappa}^p(c)\big) Z_-\big(A_{s,\kappa}^q(c), B_{s,\kappa}^q(c)\big). \end{aligned}$$

Common prior normalizers cancel. Draw the competing partition as

$$\mathbb{P}(C_t^- = c \mid \mathcal{H}_{t-1}) = \frac{w_{t-1}^{(\kappa_T)}(c)}{\sum_{\widetilde{c}\in\mathcal{A}_t} w_{t-1}^{(\kappa_T)}(\widetilde{c})}, \qquad c \in \mathcal{A}_t, \tag{6}$$

and its response parameters as

$$(P_t^-, Q_t^-) \mid (\mathcal{H}_{t-1}, C_t^-) \sim \mathsf{Q}_{t-1,\kappa_T}(\cdot \mid C_t^-). \tag{7}$$

The tempering factor $\kappa_T = \sqrt{T}$ is fixed throughout the experiment; the empirical reference estimates are untempered.

Max-learner and exploration. Hedge uses the clipped gains $g_{t,e} \in [0, 1]$. Uniform exploration gives every pair sampling probability at least $\gamma_t/|\mathcal{I}|$. The allocation $\widetilde{\lambda}_t$ is $\mathcal{H}_{t-1}$-measurable, and pair selection is independent of the current competitor conditional on $\mathcal{H}_{t-1}$. Write $\widehat{\lambda}_T(e) := N_T(e)/T$ for the empirical allocation.

Recommendation. At termination, draw $\widehat{c}_T$ from the ordinary posterior over all canonical partitions:

$$\pi_T^{(1)}(c) := \frac{w_T^{(1)}(c)}{\sum_{\widetilde{c}\in\mathcal{C}_N^\circ} w_T^{(1)}(\widetilde{c})} = \Pi_T\{(c', p, q) : c' \sim c\}. \tag{8}$$

Fresh recommendation randomness gives

$$\mathbb{P}_{\theta^*}\big(\widehat{c}_T \nsim c^* \mid \mathcal{H}_T\big) = 1 - \alpha_T = \Pi_T(c \nsim c^*).$$

4 THEORETICAL GUARANTEES

We establish a universal lower bound on posterior error and show that Algorithm 1 matches its exponent. Recall that $1 - \alpha_T = \Pi_T(\mathrm{Alt}(\theta^*))$ is the error mass under the ordinary posterior, and that $\Gamma^*(\theta^*)$ is the information benchmark in (2). Throughout, N and θ^* are fixed, and almost-sure statements refer to the data-generating law $\mathbb{P}_{\theta^*}$.

Theorem 1 (Posterior error lower bound)**.** *Let the prior be $\rho_0(c)\nu_p(dp)\nu_q(dq)$, fixed independently of T, with $\rho_0(c) > 0$ for every $c \in \mathcal{C}_N$ and with ν_p, ν_q having full support on $\mathcal{P}, \mathcal{Q}$, respectively. For every $\theta^* \in \Theta$ and every adaptive sampling rule, it almost surely holds that*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \Pi_T\big(\mathrm{Alt}(\theta^*)\big) \le \Gamma^*(\theta^*).$$

The bound requires neither prior densities nor convergence of the empirical allocation. For a given allocation, the inner infimum in $\Gamma^*(\theta^*)$ measures the least KL separation from an incorrect clustering, allowing its response parameters to vary. The outer supremum allocates queries to maximize this separation. Thus, smaller $\Gamma^*(\theta^*)$ indicates a harder instance, with $1/\Gamma^*(\theta^*)$ giving an information complexity scale. As in Russo (2020) and Li et al. (2024), the criterion concerns posterior convergence along realized histories. For a recommendation drawn from the posterior clustering marginal, the frequentist error is $\mathbb{E}_{\theta^*}[1 - \alpha_T]$. An almost-sure posterior exponent does not by itself determine the exponent of this expectation, which also depends on rare histories.

Theorem 2 (Posterior error upper bound)**.** *For every $\theta^* \in \Theta$, Algorithm 1, with the priors in* (5) *and the parameter choices in Section 3, almost surely satisfies*

$$\liminf_{T\to\infty} -\frac{1}{T}\log \Pi_T\big(\mathrm{Alt}(\theta^*)\big) \ge \Gamma^*(\theta^*).$$

The upper bound matches the lower bound at the exponential scale. Hence Algorithm 1 is asymptotically optimal in posterior error: it attains the largest achievable instance-dependent posterior decay exponent. Since its terminal recommendation is a posterior sample, the same exponent governs its conditional error probability given the observed history. Complete proofs are deferred to the appendix.

5 PROOF SKETCH OF THE POSTERIOR ERROR UPPER BOUND

We use the priors and parameter choices in Section 3. Let $K := |\mathcal{I}|$, relabel c^* canonically, and set $\mathcal{A}^* := \mathcal{A}(c^*)$ and $\Theta_{\mathrm{nb}} := \mathcal{A}^* \times \mathcal{P} \times \mathcal{Q}$. Write

$$J_c(\lambda) := \inf_{p,q} D(\theta^*, (c, p, q); \lambda), \qquad J(\lambda) := \min_{c \nsim c^*} J_c(\lambda), \qquad \overline{\lambda}_T := \frac{1}{T}\sum_{t=1}^{T} \widetilde{\lambda}_t.$$

All clustering indices are canonical and all parameter infima are over $\mathcal{P} \times \mathcal{Q}$. Define

$$h_{T,e}(\theta) := \min\{d(\mu_e^*, \mu_e(\theta)), B_T\}, \qquad G_t := B_T g_t, \qquad \overline{h}_t := \mathbb{E}[h_T(\theta_t^-) \mid \mathcal{H}_{t-1}].$$

We use the history $\mathcal{H}_s$ defined in Section 3. All round sums run over $1, \dots, T$, and allocation extrema are over $\Delta(\mathcal{I})$.

Lemma 1 (Good events)**.** *Fix $\theta^* \in \Theta$. Let $v_T := \log(4K(T+1)^4)$ and set*

$$s_T := \max\left\{n_{\mathrm{rec}}(T), \left\lceil (\kappa_T v_T^2)^{1/(1-\alpha)} \right\rceil\right\},$$

where $n_{\mathrm{rec}}(T) = O(v_T^{1/(1-2\alpha)})$ is the explicit recovery threshold derived in the appendix. The following four statements hold simultaneously on an event $\mathcal{E}_T$ with $\mathbb{P}_{\theta^*}(\mathcal{E}_T) \ge 1 - CT^{-3}$ for all sufficiently large T.

Lemma 1.1 (Recovery)**.** *On $\mathcal{E}_T^{\mathrm{rec}}$, for every $t > s_T$,*

$$\widehat{\sigma}_t = c^*, \qquad \mathcal{A}_t = \mathcal{A}^*, \qquad \max_e |\mu_e(\theta_t^+) - \mu_e^*| \le a_T.$$

Lemma 1.2 (Learner regret). *On $\mathcal{E}_T^{\mathrm{reg}}$, the two learners satisfy*

$$\max_{\lambda} \sum_t \langle \lambda, G_t \rangle - \sum_t \langle \lambda_t, G_t \rangle \le R_T^{\max},$$
$$\sum_{t>s_T} \langle \widetilde{\lambda}_t, \overline{h}_t \rangle \le \inf_{\theta \in \Theta_{\mathrm{nb}}} \sum_{t>s_T} \langle \widetilde{\lambda}_t, h_T(\theta) \rangle + R_T^{\min}.$$

Lemma 1.3 (Sampling and tracking). *On $\mathcal{E}_T^{\mathrm{samp}}$, we have*

$$\left\| \frac{1}{T} \sum_{t>s_T} (G_t - \overline{h}_t) \right\|_\infty + \left| \frac{1}{T} \sum_{t>s_T} \langle \lambda_t, G_t - \overline{h}_t \rangle \right| \le r_T^{\mathrm{ps}},$$
$$\|\widehat{\lambda}_T - \overline{\lambda}_T\|_1 \le r_T^{\mathrm{tr}}.$$

Lemma 1.4 (Marginal likelihood). *On $\mathcal{E}_T^{\mathrm{marg}}$, the uninflated weights satisfy*

$$\max_c \left| \frac{1}{T} \log \frac{w_T^{(1)}(c)}{L_T(\theta^*)} + J_c(\widehat{\lambda}_T) \right| \le r_T^{\ell}.$$

Here $\mathcal{E}_T$ is the intersection of the four events above. With $a_T = C\sqrt{\log(T+1)/s_T^{1-\alpha}}$, they satisfy

$$\frac{R_T^{\max} + R_T^{\min}}{T} + r_T^{\mathrm{ps}} + B_T r_T^{\mathrm{tr}} + r_T^{\ell} \le C \left[\frac{B_T s_T}{T} + (1 + B_T) a_T + B_T \sqrt{\frac{\log(T+1)}{T}} \right] \longrightarrow 0. \quad (9)$$

The constant C is independent of T. The cutoff is used only in the analysis: it exceeds the recovery threshold and satisfies $s_T^{1-\alpha}/\kappa_T \ge v_T^2$. Since $\kappa_T = \sqrt{T}$ and $\alpha < 1/2$, $B_T s_T/T \to 0$ and $(1 + B_T)a_T = O(T^{-1/4}\sqrt{\log(T+1)}) \to 0$.

Proofs of the good-event lemmas appear in the appendix.

Proof sketch of Theorem 2. Work on $\mathcal{E}_T$ for sufficiently large T. *First, the neighborhood preserves the information value.* Lemma 3 and (25) give

$$\inf_{\theta \in \Theta_{\mathrm{nb}}} \langle \lambda, h_T(\theta) \rangle = J(\lambda), \qquad \sup_{\lambda} J(\lambda) = \Gamma^*(\theta^*). \quad (10)$$

In particular, $\Gamma^*(\theta^*) \le B_T$.

Second, regret implies information optimality. Write $s = s_T$, $\overline{\gamma}_T = T^{-1}\sum_t \gamma_t$, and $e_T = (R_T^{\max} + R_T^{\min})/T + r_T^{\mathrm{ps}}$. Recovery places the conditional competitor law on Θ_{nb} for $t > s$, so $\sum_{t>s}\langle \lambda, \overline{h}_t \rangle \ge (T-s)J(\lambda)$. Lemmas 1.2 and 1.3, with nonnegative prefix losses added to the comparator, yield

$$\Gamma^*(\theta^*) \le J(\overline{\lambda}_T) + e_T + B_T \left(\overline{\gamma}_T + \frac{2s}{T} \right)$$
$$\le J(\widehat{\lambda}_T) + \varepsilon_T,$$

where $\varepsilon_T = e_T + B_T(\overline{\gamma}_T + r_T^{\mathrm{tr}} + 2s_T/T)$. The two prefix terms arise from replacing $T - s$ by T and applying full-horizon max-learner regret to the remaining rounds. Exploration costs $B_T\overline{\gamma}_T$; tracking uses $|J(\lambda) - J(\lambda')| \le B_T\|\lambda - \lambda'\|_1$, which follows from (10). Equation (9) and $B_T\overline{\gamma}_T = O(T^{-\alpha}\log(T+1))$ give $\varepsilon_T \to 0$. Since $J(\widehat{\lambda}_T) \le \Gamma^*(\theta^*)$, these bounds establish optimality of the information value.

Finally, the information value controls posterior error. Let $J_0 = |\{c \in \mathcal{C}_N^\circ : c \nsim c^*\}|$. Since $J_{c^*}(\widehat{\lambda}_T) = 0$, Lemma 1.4 bounds the alternative weights from above and the denominator's true-partition weight from below, giving

$$\Pi_T\big(\mathrm{Alt}(\theta^*)\big) \le J_0 e^{-T[J(\widehat{\lambda}_T) - 2r_T^{\ell}]} \le e^{-T[\Gamma^*(\theta^*) - \xi_T]},$$

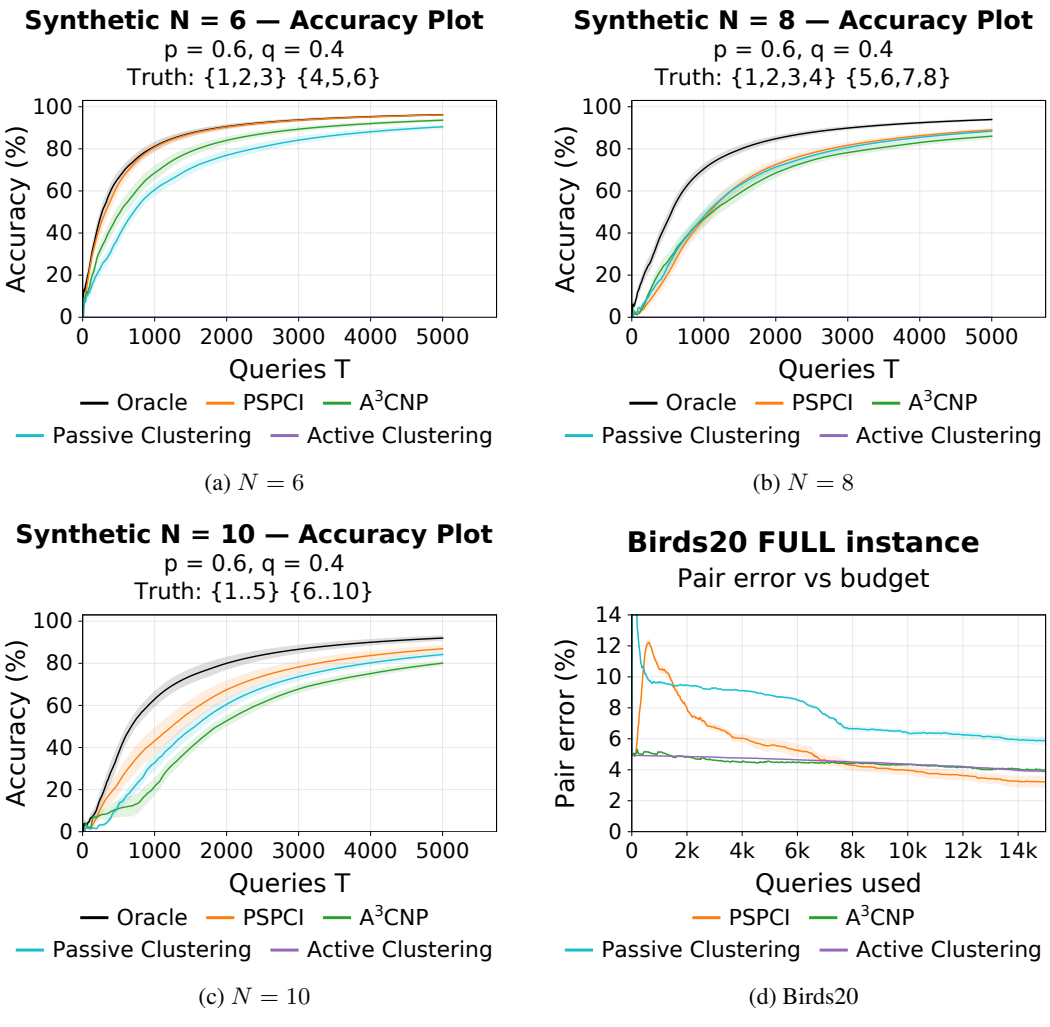
where $\xi_T = \varepsilon_T + 2r_T^{\ell} + T^{-1}\log J_0 \to 0$. On the common probability space for budget-indexed experiments, $\sum_T \mathbb{P}_{\theta^*}(\mathcal{E}_T^c) < \infty$. Borel–Cantelli makes this posterior bound eventual almost surely, proving Theorem 2. Together with Theorem 1, this gives $\lim_{T\to\infty} -T^{-1}\log(1 - \alpha_T) = \Gamma^*(\theta^*)$ almost surely. □

6 EMPRICAL STUDIES

6.1 EXPERIMENTS ON SYNTHETIC DATASET

We evaluate our method on balanced two-cluster instances with $N \in \{6, 8, 10\}$, within- and cross-cluster probabilities $p = 0.6$ and $q = 0.4$. Each run uses $T = 5{,}000$ pairwise queries. We report cumulative exact-recovery accuracy, defined as the fraction of recommendations up to each query that recover the true partition. Results are averaged over 50 runs; shaded regions indicate one standard error. We compare against an oracle allocation computed from the true instance and paired with a GLRT recommendation, A^{3}CNP (Teo et al. (2026)), passive spectral clustering supplied with the true number of clusters (McSherry (2001)), and active clustering of Chen et al. (2023).

Our method nearly matches the oracle at $N = 6$, achieving 96.1% accuracy versus 96.2% at the final budget. At $N = 8$ and $N = 10$, it achieves 89.0% and 87.0%, respectively, trailing the oracle by approximately five percentage points in both cases. The evaluated implementation of Chen et al. (2023) achieves zero exact recovery within the tested budget of (T=5,000), suggesting that more queries are needed for its sequential pairwise tests to recover the entire partition.



(a) $N = 6$ (b) $N = 8$

(c) $N = 10$ (d) Birds20

Figure 1: Recovery accuracy on balanced synthetic two-cluster instances with $p = 0.6$ and $q = 0.4$ (a–c), and pairwise error on the full Birds20 dataset (d).

6.2 Experiments with Real-Data-Calibrated Feedback

We evaluate our method on the full Birds20 dataset of Chen et al. (2023), which contains 125 images from 20 bird species selected from the Caltech–UCSD Birds dataset (Wah et al. (2011)), with 3–10 images per species. The dataset provides crowdsourced judgments of whether two images belong to the same species. We retain all 125 images, giving 7,750 possible item pairs, and evaluate clustering performance over a budget of 15,000 queries. We measure pairwise error: the fraction of item pairs whose predicted same-cluster relation disagrees with the ground truth. Curves report the mean over 10 random seeds, with shaded regions indicating one standard error.

PSPCI initially incurs a higher pairwise error than the other active methods, but its error decreases as more observations are collected. At the final budget, PSPCI achieves approximately 3.2% pairwise error, compared with 3.9%–4.0% for the other two active baselines and 5.9% for the passive spectral baseline. These results show an advantage for PSPCI at larger budgets on the full Birds20 instance, while its early performance leaves room for improvement.

7 Comparison to related works

Chen et al. (2023) establish high-probability recovery guarantees for active crowdsourced clustering, while Teo et al. (2026) characterize instance-dependent query complexity in the fixed-confidence setting. We instead study the almost-sure decay of posterior error under a fixed query budget. Our algorithm builds on the posterior-sampling min–max framework of Li et al. (2024), developed for Gaussian linear pure exploration. Extending this framework to clustering identification presents two challenges. First, sampled alternative response probabilities can lie arbitrarily close to zero or one, making Bernoulli KL gains unbounded and preventing a direct application of bounded-gain Hedge guarantees. Second, posterior sampling must account for a combinatorially large space of competing partitions, together with their unknown response parameters. We address the first challenge by clipping KL gains at a logarithmically growing threshold and normalizing them for Hedge. For the second, we exploit the merge/split reduction of Teo et al. (2026) to restrict competitor sampling to neighbors of the reference partition while preserving the population information benchmark once the reference is correct. Together with our posterior concentration analysis, these modifications yield, to the best of our knowledge, the first matching lower and upper bounds on the posterior error exponent for fixed-budget clustering with unknown within- and between-cluster response probabilities. The achieved exponent equals the universal information bound, with no asymptotic multiplicative gap.

8 Conclusions and Future Work

We studied fixed-budget cluster identification from noisy pairwise observations with unknown clusters and response probabilities. We established a universal lower bound on posterior error and proposed an adaptive posterior-sampling algorithm that matches its exponent almost surely. The algorithm combines sampling over merge/split alternatives with a min–max allocation rule based on clipped KL gains, yielding the optimal instance-dependent posterior error exponent without an asymptotic multiplicative gap. Our experiments evaluated the approach on synthetic instances and feedback calibrated from real crowdsourcing data. Future work includes developing scalable approximations to partition projection and posterior sampling while preserving theoretical guarantees. Another direction is to characterize the unconditional frequentist error and determine under what conditions posterior exponent optimality yields corresponding frequentist guarantees.

AI use statement

We used generative AI tools to assist with manuscript drafting and editing, literature review, and plotting code. The authors take full responsibility for the final manuscript, including its claims, experimental results, and references.

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th annual symposium on foundations of computer science*, pp. 670–688. IEEE, 2015.

Yi Chen, Ramya Korlakai Vinayak, and Babak Hassibi. Crowdsourced clustering via active querying: Practical algorithm with theoretical guarantees. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 11, pp. 27–37, 2023.

Anne Condon and Richard M Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures & Algorithms*, 18(2):116–140, 2001.

Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pp. 2432–2442. PMLR, 2020.

Tanner Fiez, Lalit Jain, Kevin Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pp. 998–1027. PMLR, 2016.

Peter Glynn and Sandeep Juneja. A large deviations perspective on ordinal optimization. In *Proceedings of the 2004 Winter Simulation Conference, 2004.*, volume 1. IEEE, 2004.

Ryan Gomes, Peter Welinder, Andreas Krause, and Pietro Perona. Crowdclustering. *Advances in neural information processing systems*, 24, 2011.

Kasper Green Larsen, Michael Mitzenmacher, and Charalampos Tsourakakis. Clustering with a faulty oracle. In *Proceedings of The Web Conference 2020*, pp. 2831–2834, 2020.

Reinhard Heckel, Nihar B Shah, Kannan Ramchandran, and Martin J Wainwright. Active ranking from pairwise comparisons and when parametric assumptions do not help. 2019.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Shahana Ibrahim and Xiao Fu. Mixed membership graph clustering via systematic edge query. *IEEE Transactions on Signal Processing*, 69:5189–5205, 2021.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pp. 423–439. PMLR, 2014.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020.

Marc Jourdan, Rémy Degenne, Dorian Baudry, Rianne de Heide, and Emilie Kaufmann. Top two algorithms revisited, 2022. URL `https://arxiv.org/abs/2206.05979`.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International conference on machine learning*, pp. 1238–1246. PMLR, 2013.

Ramya Korlakai Vinayak and Babak Hassibi. Crowdsourced clustering: Querying edges vs triangles. *Advances in Neural Information Processing Systems*, 29, 2016.

Ramya Korlakai Vinayak, Samet Oymak, and Babak Hassibi. Graph clustering with missing data: Convex algorithms and analysis. *Advances in Neural Information Processing Systems*, 27, 2014.

Zhaoqi Li, Kevin Jamieson, and Lalit Jain. Optimal exploration is no harder than thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pp. 1684–1692. PMLR, 2024.

Arya Mazumdar and Barna Saha. Clustering with noisy queries. *Advances in Neural Information Processing Systems*, 30, 2017a.

Arya Mazumdar and Barna Saha. Query complexity of clustering with side information. *Advances in Neural Information Processing Systems*, 30, 2017b.

Arya Mazumdar and Barna Saha. A theoretical analysis of first heuristics of crowdsourced entity resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017c.

Frank McSherry. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pp. 529–537. IEEE, 2001.

Chandra Sekhar Mukherjee, Pan Peng, and Jiapeng Zhang. Recovering unbalanced communities in the stochastic block model with application to clustering with a faulty oracle. *Advances in Neural Information Processing Systems*, 36:20540–20561, 2023.

Pan Peng and Jiapeng Zhang. Towards a query-optimal and time-efficient algorithm for clustering with a faulty oracle. In *Conference on Learning Theory*, pp. 3662–3680. PMLR, 2021.

Chao Qin, Diego Klabjan, and Daniel Russo. Improving the expected improvement algorithm, 2017. URL https://arxiv.org/abs/1705.10033.

Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning from crowds. *Journal of machine learning research*, 11(4), 2010.

Daniel Russo. Simple bayesian algorithms for best-arm identification. *Operations Research*, 68(6): 1625–1647, 2020.

Xuedong Shang, Rianne de Heide, Emilie Kaufmann, Pierre Ménard, and Michal Valko. Fixed-confidence guarantees for bayesian best-arm identification, 2019. URL https://arxiv.org/abs/1910.10945.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in neural information processing systems*, 27, 2014.

Alexander Sorokin and David Forsyth. Utility data annotation with amazon mechanical turk. In *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, pp. 1–8. IEEE, 2008.

Rachel SY Teo, PN Karthik, Ramya Korlakai Vinayak, and Vincent YF Tan. Almost asymptotically optimal active clustering through pairwise observations. *arXiv preprint arXiv:2602.05690*, 2026.

Vasilis Verroios and Hector Garcia-Molina. Entity resolution with crowd errors. In *2015 IEEE 31st international conference on data engineering*, pp. 219–230. IEEE, 2015.

Norases Vesdapunt, Kedar Bellare, and Nilesh N Dalvi. Crowdsourcing algorithms for entity resolution. *Proc. VLDB Endow.*, 7(12):1071–1082, 2014.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

Jiannan Wang, Tim Kraska, Michael J Franklin, and Jianhua Feng. Crowder: Crowdsourcing entity resolution. *arXiv preprint arXiv:1208.1927*, 2012.

Peter Welinder, Steve Branson, Pietro Perona, and Serge Belongie. The multidimensional wisdom of crowds. *Advances in neural information processing systems*, 23, 2010.

Jinghui Xia and Zengfeng Huang. Optimal clustering with noisy queries via multi-armed bandit. In *International Conference on Machine Learning*, pp. 24315–24331. PMLR, 2022.

Junwen Yang, Zixin Zhong, and Vincent YF Tan. Optimal clustering with bandit feedback. *Journal of Machine Learning Research*, 25(186):1–54, 2024.

Se-Young Yun and Alexandre Proutiere. Community detection via random and adaptive sampling. In *Conference on learning theory*, pp. 138–175. PMLR, 2014.

A PROOF OF THE POSTERIOR ERROR LOWER BOUND 1

We prove the posterior error lower bound stated in the main text. We first use the Beta–Bernoulli tail sandwich inequality to identify the exponential rate of a pooled marginal likelihood under a reference measure. A local comparison extends this rate to every fixed full-support prior. We then evaluate the posterior error along convergent subsequences of the empirical allocation and use compactness to obtain the bound without assuming that the allocation converges.

For $A \subseteq \mathcal{I}$, write

$$N_T(A) := \sum_{e \in A} N_T(e), \qquad S_T(A) := \sum_{e \in A} S_T(e), \qquad \widehat{\lambda}_T(e) := \frac{N_T(e)}{T}.$$

For $c \in \mathcal{C}_N$, let

$$\begin{aligned} L_T(c,p,q) &:= p^{S_T(\mathcal{I}^=(c))}(1-p)^{N_T(\mathcal{I}^=(c))-S_T(\mathcal{I}^=(c))} \\ &\quad \times q^{S_T(\mathcal{I}^{\neq}(c))}(1-q)^{N_T(\mathcal{I}^{\neq}(c))-S_T(\mathcal{I}^{\neq}(c))}, \\ m_T(c) &:= \int_{\mathcal{P}} \int_{\mathcal{Q}} L_T(c,p,q)\, \nu_q(dq)\nu_p(dp). \end{aligned} \tag{11}$$

Here $m_T(c)$ does not include the clustering prior $\rho_0(c)$. The sampling rule uses only past observations and internal randomization, so its action probabilities contribute no instance-dependent factor to the likelihood. With $\mathcal{C}^- := \{c \in \mathcal{C}_N : c \nsim c^*\}$, Bayes' rule gives

$$\Pi_T\big(\mathrm{Alt}(\theta^*)\big) = 1 - \alpha_T = \frac{\sum_{c \in \mathcal{C}^-} \rho_0(c) m_T(c)}{\sum_{c \in \mathcal{C}_N} \rho_0(c) m_T(c)}. \tag{12}$$

The sums include all labelings, as in the theorem; grouping equivalent labelings gives the same posterior mass.

A.1 POOLED MARGINAL LIKELIHOODS (LEMMA 2)

For $x \in [0,1]$, let

$$h(x) := -x \log x - (1-x)\log(1-x),$$

with $0 \log 0 = 0$. For $J \in \{\mathcal{P}, \mathcal{Q}\}$ and $x \in (0,1)$, define

$$\tau_J(x) := \inf_{u \in J} d(x,u).$$

Since $\partial_u d(x,u) = (u-x)/(u(1-u))$, minimizing over each half-interval gives

$$\tau_{\mathcal{P}}(x) = d(x, \max\{x, 1/2\}), \qquad \tau_{\mathcal{Q}}(x) = d(x, \min\{x, 1/2\}).$$

Both functions are continuous, including at $x = 1/2$.

Lemma 2 (Pooled Bernoulli marginal likelihood)**.** *Let (n_k, s_k) be integer pairs satisfying $0 \le s_k \le n_k$, $n_k \to \infty$, and $s_k/n_k \to x \in (0,1)$. Let $J \in \{\mathcal{P}, \mathcal{Q}\}$, and let ν be any fixed full-support probability measure on J. Then*

$$-\frac{1}{n_k} \log \int_J u^{s_k}(1-u)^{n_k - s_k}\, \nu(du) \longrightarrow h(x) + \tau_J(x).$$

Proof of Lemma 2. Suppress the index k and write $x_n = s/n$. All limits below are along the sequence in the statement. Consider the reference integral

$$H_J(n,s) := \int_J u^s (1-u)^{n-s}\, du.$$

For the auxiliary random variable $X_{n,s} \sim \mathrm{Beta}(s+1, n-s+1)$, the Beta density gives

$$H_J(n,s) = \mathrm{B}(s+1, n-s+1)\, \mathbb{P}(X_{n,s} \in J), \tag{13}$$

where B is the beta function. This Beta distribution represents the reference integral; it is not an assumption on ν.

First, the integer-parameter Beta integral satisfies

$$\mathrm{B}(s+1, n-s+1) = \frac{s!(n-s)!}{(n+1)!} = \frac{1}{(n+1)\binom{n}{s}}.$$

For $0 < s < n$, s is a mode of the $\mathrm{Binomial}(n, s/n)$ distribution. Indeed, if b_j denotes its probability at j, then

$$\frac{b_{j+1}}{b_j} = \frac{n-j}{j+1}\frac{s}{n-s}$$

is greater than one for $j < s$ and less than one for $j \geq s$. The largest of its $n+1$ probabilities therefore lies between $1/(n+1)$ and 1. Since $b_s = \binom{n}{s} e^{-nh(x_n)}$, we obtain

$$\frac{e^{nh(x_n)}}{n+1} \leq \binom{n}{s} \leq e^{nh(x_n)}.$$

These bounds also hold for $s = 0$ and $s = n$ by direct substitution. Consequently,

$$\frac{e^{-nh(x_n)}}{n+1} \leq \mathrm{B}(s+1, n-s+1) \leq e^{-nh(x_n)},$$

and hence

$$h(x_n) \leq -\frac{1}{n}\log \mathrm{B}(s+1, n-s+1) \leq h(x_n) + \frac{\log(n+1)}{n}. \tag{14}$$

As $x_n \to x$, the middle expression converges to $h(x)$.

Next, we apply the Beta–Bernoulli tail sandwich inequality to the tail probabilities in (13). For the present Beta parameters, the corresponding binomial size is

$$m := n+1, \qquad r_n := \frac{s}{n+1}, \qquad \ell_n := \frac{s+1}{n+1}.$$

Thus r_n is the right-tail center and ℓ_n is the left-tail center. Both converge to x, but they differ for finite n. Set

$$y_n := \max\{1/2, r_n\} + m^{-2}, \qquad z_n := \min\{1/2, \ell_n\} - m^{-2}.$$

For every $n \geq 1$, $r_n < y_n < 1$ and $0 < z_n < \ell_n$. Moreover, $y_n > 1/2$ and $z_n < 1/2$. The two lower sandwich bounds therefore give

$$\mathbb{P}(X_{n,s} > 1/2) \geq \mathbb{P}(X_{n,s} \geq y_n) \geq \frac{1}{n+2} e^{-(n+1)d(r_n, y_n)},$$
$$\mathbb{P}(X_{n,s} < 1/2) \geq \mathbb{P}(X_{n,s} \leq z_n) \geq \frac{1}{n+2} e^{-(n+1)d(\ell_n, z_n)}.$$

Since $y_n \to \max\{1/2, x\}$ and $z_n \to \min\{1/2, x\}$, continuity of d implies

$$\limsup -\frac{1}{n}\log \mathbb{P}(X_{n,s} > 1/2) \leq \tau_{\mathcal{P}}(x),$$
$$\limsup -\frac{1}{n}\log \mathbb{P}(X_{n,s} < 1/2) \leq \tau_{\mathcal{Q}}(x). \tag{15}$$

Here we used $(n+1)/n \to 1$ and $\log(n+2)/n \to 0$.

For the reverse inequalities, if $x < 1/2$, then $r_n < 1/2$ for all sufficiently large n. The upper right-tail sandwich bound gives

$$\mathbb{P}(X_{n,s} > 1/2) \leq e^{-(n+1)d(r_n, 1/2)}.$$

Taking the limit inferior of the negative logarithm divided by n gives at least $d(x, 1/2) = \tau_{\mathcal{P}}(x)$. If $x \geq 1/2$, the same lower bound on the exponent follows from $\mathbb{P}(X_{n,s} > 1/2) \leq 1$ and $\tau_{\mathcal{P}}(x) = 0$. For the left tail, if $x > 1/2$, then $\ell_n > 1/2$ eventually, and

$$\mathbb{P}(X_{n,s} < 1/2) \leq e^{-(n+1)d(\ell_n, 1/2)}.$$

This gives a limit inferior of at least $d(x, 1/2) = \tau_{\mathcal{Q}}(x)$. If $x \leq 1/2$, use the bound by one and $\tau_{\mathcal{Q}}(x) = 0$ instead. Together with (15), these arguments show that

$$-\frac{1}{n}\log \mathbb{P}(X_{n,s} \in J) \longrightarrow \tau_J(x). \tag{16}$$

In particular, this includes $x = 1/2$ without requiring either empirical center to stay on one side of $1/2$. Equations (13), (14), and (16) now give

$$-\frac{1}{n}\log H_J(n,s) \longrightarrow h(x) + \tau_J(x). \tag{17}$$

It remains to compare an arbitrary full-support prior with this reference integral. Write

$$Z_\nu(n,s) := \int_J u^s(1-u)^{n-s}\,\nu(du).$$

The identity

$$u^s(1-u)^{n-s} = \exp\{-n[h(x_n) + d(x_n,u)]\}$$

and $\nu(J) = 1$ imply

$$Z_\nu(n,s) \le e^{-n[h(x_n)+\tau_J(x_n)]}.$$

For the lower bound, fix $\varepsilon > 0$. Choose $u_\varepsilon \in J$ with $d(x,u_\varepsilon) < \tau_J(x) + \varepsilon$. By continuity, there is an open interval U_ε containing u_ε, with compact closure in J, such that

$$\sup_{u\in U_\varepsilon} d(x_n,u) \le \tau_J(x) + 2\varepsilon$$

for all sufficiently large n. To see the uniformity, first choose U_ε so that the bound holds with x and a strict margin; then use uniform continuity on a compact neighborhood of $\{x\} \times \overline{U}_\varepsilon$ and $x_n \to x$. Full support gives $\nu(U_\varepsilon) > 0$, so

$$Z_\nu(n,s) \ge \nu(U_\varepsilon)e^{-n[h(x_n)+\tau_J(x)+2\varepsilon]}.$$

Dividing these bounds by $H_J(n,s)$ and taking logarithms yields

$$\frac{1}{n}\log\frac{Z_\nu(n,s)}{H_J(n,s)} \le -h(x_n) - \tau_J(x_n) - \frac{1}{n}\log H_J(n,s),$$
$$\frac{1}{n}\log\frac{Z_\nu(n,s)}{H_J(n,s)} \ge \frac{\log\nu(U_\varepsilon)}{n} - h(x_n) - \tau_J(x) - 2\varepsilon - \frac{1}{n}\log H_J(n,s).$$

The right-hand side of the first inequality converges to zero by (17). The right-hand side of the second converges to -2ε. Letting $\varepsilon \downarrow 0$ proves

$$\frac{1}{n}\log\frac{Z_\nu(n,s)}{H_J(n,s)} \longrightarrow 0. \tag{18}$$

Combining (18) with (17) proves the lemma. The argument uses positive mass near an admissible approximate minimizer even when the infimum is approached at the excluded endpoint $1/2$. □

A.2 Proof of the theorem 1

Proof of the posterior error lower bound (Theorem 1). Fix $\theta^* \in \Theta$, a prior satisfying the theorem's assumptions, and an adaptive sampling rule. If the rule depends on T, realize the budget-indexed experiments on a common probability space and suppress the budget index on within-experiment quantities. The argument below does not require independence between experiments.

Step 1: a simultaneous law for the observation counts. For each $e \in \mathcal{I}$, define

$$X_t(e) := \mathbf{1}\{I_t = e\}(Y_t - \mu_e^*), \qquad \Delta_T(e) := \sum_{t=1}^{T} X_t(e) = S_T(e) - \mu_e^* N_T(e).$$

Since I_t is $\mathcal{F}_t$-measurable,

$$\begin{aligned}\mathbb{E}_{\theta^*}[X_t(e) \mid \mathcal{F}_t] &= \mathbf{1}\{I_t = e\}\big(\mathbb{E}_{\theta^*}[Y_t \mid \mathcal{F}_t] - \mu_e^*\big)\\ &= \mathbf{1}\{I_t = e\}(\mu_{I_t}^* - \mu_e^*) = 0.\end{aligned}$$

Also, $|X_t(e)| \le 1$, so conditional Hoeffding's lemma gives

$$\mathbb{E}_{\theta^*}\Big[e^{uX_t(e)} \mid \mathcal{F}_t\Big] \le e^{u^2/2}, \qquad u \in \mathbb{R}.$$

Since $\Delta_{t-1}(e)$ is $\mathcal{F}_t$-measurable, successive conditioning yields

$$\mathbb{E}_{\theta^*} e^{u\Delta_T(e)} \le e^{Tu^2/2}.$$

For $\varepsilon > 0$, Markov's inequality and optimization over $u > 0$ therefore give

$$\mathbb{P}_{\theta^*}\big(\pm\Delta_T(e) \ge \varepsilon T\big) \le \inf_{u>0} e^{-u\varepsilon T + Tu^2/2} = e^{-\varepsilon^2 T/2}.$$

A union bound over the two signs and all pairs gives

$$\mathbb{P}_{\theta^*}\left(\max_{e\in\mathcal{I}} |\Delta_T(e)| \ge \varepsilon T\right) \le 2|\mathcal{I}| e^{-\varepsilon^2 T/2}.$$

The right-hand side is summable over T. Borel–Cantelli, applied for $\varepsilon = 1, 1/2, 1/3, \ldots$, therefore gives a single event Ω_0 of probability one on which

$$\max_{e\in\mathcal{I}} \frac{|\Delta_T(e)|}{T} \longrightarrow 0.$$

For budget-dependent rules, apply the same martingale argument within each budget-T experiment. The displayed probability bound still holds for every T, and Borel–Cantelli requires no independence between experiments.

Step 2: pooled marginal likelihoods along a convergent subsequence. Fix a path in Ω_0. Let (T_k) be any sequence along which $\widehat{\lambda}_{T_k} \to \lambda \in \Delta(\mathcal{I})$. For a set $A \subseteq \mathcal{I}$, write

$$n_k := N_{T_k}(A), \quad s_k := S_{T_k}(A), \quad \Lambda_A := \sum_{e\in A} \lambda_e, \quad b_A := \sum_{e\in A} \lambda_e \mu_e^*.$$

Then

$$\frac{n_k}{T_k} = \sum_{e\in A} \widehat{\lambda}_{T_k}(e) \longrightarrow \Lambda_A,$$
$$\frac{s_k}{T_k} = \sum_{e\in A} \mu_e^* \widehat{\lambda}_{T_k}(e) + \sum_{e\in A} \frac{\Delta_{T_k}(e)}{T_k} \longrightarrow b_A.$$

Let $J \in \{\mathcal{P}, \mathcal{Q}\}$ and let ν be its fixed full-support prior. Define the pooled integral and the true likelihood on A by

$$Z_T(A, J; \nu) := \int_J u^{S_T(A)} (1-u)^{N_T(A) - S_T(A)}\, \nu(du),$$
$$L_T^*(A) := \prod_{e\in A} (\mu_e^*)^{S_T(e)} (1-\mu_e^*)^{N_T(e) - S_T(e)}.$$

We will show that

$$-\frac{1}{T_k} \log \frac{Z_{T_k}(A, J; \nu)}{L_{T_k}^*(A)} \longrightarrow \inf_{u\in J} \sum_{e\in A} \lambda_e d(\mu_e^*, u). \tag{19}$$

Suppose first that $\Lambda_A > 0$. Then $n_k \to \infty$ and

$$\frac{s_k}{n_k} = \frac{s_k/T_k}{n_k/T_k} \longrightarrow x_A := \frac{b_A}{\Lambda_A} \in [q^*, p^*] \subset (0, 1).$$

Lemma 2 applies and gives

$$-\frac{1}{T_k} \log Z_{T_k}(A, J; \nu) = \frac{n_k}{T_k} \left[-\frac{1}{n_k} \log Z_{T_k}(A, J; \nu)\right] \longrightarrow \Lambda_A [h(x_A) + \tau_J(x_A)]. \tag{20}$$

For each e, Step 1 gives $S_{T_k}(e)/T_k \to \lambda_e \mu_e^*$ and $(N_{T_k}(e) - S_{T_k}(e))/T_k \to \lambda_e(1-\mu_e^*)$. Consequently,

$$\frac{1}{T_k} \log L_{T_k}^*(A) = \sum_{e\in A} \left[\frac{S_{T_k}(e)}{T_k} \log \mu_e^* + \frac{N_{T_k}(e) - S_{T_k}(e)}{T_k} \log(1-\mu_e^*)\right]$$
$$\longrightarrow -\sum_{e\in A} \lambda_e h(\mu_e^*).$$

Adding this limit to (20) shows that the left-hand side of (19) converges to

$$\Lambda_A[h(x_A) + \tau_J(x_A)] - \sum_{e \in A} \lambda_e h(\mu_e^*). \tag{21}$$

For any $u \in (0, 1)$, expand the KL divergences to obtain

$$\begin{aligned}\sum_{e \in A} \lambda_e d(\mu_e^*, u) &= -\sum_{e \in A} \lambda_e h(\mu_e^*) - b_A \log u - (\Lambda_A - b_A) \log(1 - u) \\ &= -\sum_{e \in A} \lambda_e h(\mu_e^*) + \Lambda_A[h(x_A) + d(x_A, u)].\end{aligned}$$

Taking the infimum over $u \in J$ identifies (21) with the right-hand side of (19).

If $\Lambda_A = 0$, the nonnegative weights satisfy $\lambda_e = 0$ for every $e \in A$, and $n_k/T_k \to 0$. Choose a nondegenerate compact interval $V \subset J$. Full support gives $\nu(V) > 0$. With

$$C_V := \max_{u \in V} \max\{-\log u, -\log(1 - u)\} < \infty,$$

we have, for every $0 \le s_k \le n_k$,

$$\nu(V) e^{-C_V n_k} \le Z_{T_k}(A, J; \nu) \le 1.$$

Thus

$$0 \le -\frac{1}{T_k} \log Z_{T_k}(A, J; \nu) \le C_V \frac{n_k}{T_k} - \frac{\log \nu(V)}{T_k} \longrightarrow 0.$$

Let $\eta := \min_{e \in \mathcal{I}} \min\{\mu_e^*, 1 - \mu_e^*\} > 0$. Every true Bernoulli likelihood factor lies between η and one, so $\eta^{n_k} \le L_{T_k}^*(A) \le 1$. It follows that

$$\frac{n_k}{T_k} \log \eta \le \frac{1}{T_k} \log L_{T_k}^*(A) \le 0,$$

and this expression also converges to zero. The limit on the left of (19) is therefore zero, which equals its right-hand side. This includes an empty A or a set that receives only finitely many observations.

Step 3: the exponent for a fixed clustering. For $c \in \mathcal{C}_N$, set $A_c := \mathcal{I}^{=}(c)$ and $B_c := \mathcal{I}^{\neq}(c)$. Prior independence and (11) give

$$m_T(c) = Z_T(A_c, \mathcal{P}; \nu_p) Z_T(B_c, \mathcal{Q}; \nu_q).$$

Since A_c and B_c partition $\mathcal{I}$, $L_T(\theta^*) = L_T^*(A_c) L_T^*(B_c)$. Hence, with $R_T(c) := m_T(c)/L_T(\theta^*)$,

$$\begin{aligned}-\frac{1}{T_k} \log R_{T_k}(c) &= -\frac{1}{T_k} \log \frac{Z_{T_k}(A_c, \mathcal{P}; \nu_p)}{L_{T_k}^*(A_c)} \\ &\quad - \frac{1}{T_k} \log \frac{Z_{T_k}(B_c, \mathcal{Q}; \nu_q)}{L_{T_k}^*(B_c)} \\ &\longrightarrow \inf_{p \in \mathcal{P}} \sum_{e \in A_c} \lambda_e d(\mu_e^*, p) + \inf_{q \in \mathcal{Q}} \sum_{e \in B_c} \lambda_e d(\mu_e^*, q) \\ &=: J_c(\lambda).\end{aligned}$$

The two infima involve independent variables. Their sum equals the infimum of the sum: the inequality in one direction follows by bounding each summand below by its infimum; the reverse follows by choosing a p and a q within any prescribed positive tolerance of their respective infima. Therefore,

$$J_c(\lambda) = \inf_{p \in \mathcal{P},\, q \in \mathcal{Q}} D(\theta^*, (c, p, q); \lambda).$$

This also covers an empty A_c or B_c: its integral is one and its information contribution is zero.

Step 4: combining clusterings and normalizing the posterior. For a nonempty $\mathcal{S} \subseteq \mathcal{C}_N$, define

$$U_T(\mathcal{S}) := \sum_{c \in \mathcal{S}} \rho_0(c) R_T(c), \qquad r_k(c) := -\frac{1}{T_k} \log[\rho_0(c) R_{T_k}(c)].$$

Since $\rho_0(c) > 0$ is fixed,

$$r_k(c) = -\frac{\log \rho_0(c)}{T_k} - \frac{1}{T_k} \log R_{T_k}(c) \longrightarrow J_c(\lambda).$$

Let $r_k^{\min} := \min_{c \in \mathcal{S}} r_k(c)$. Finiteness of $\mathcal{S}$ implies $r_k^{\min} \to \min_{c \in \mathcal{S}} J_c(\lambda)$, because

$$\left| \min_{c \in \mathcal{S}} r_k(c) - \min_{c \in \mathcal{S}} J_c(\lambda) \right| \le \max_{c \in \mathcal{S}} |r_k(c) - J_c(\lambda)| \longrightarrow 0.$$

Bounding the sum by its largest term gives

$$e^{-T_k r_k^{\min}} \le U_{T_k}(\mathcal{S}) \le |\mathcal{S}| e^{-T_k r_k^{\min}}.$$

Taking negative logarithms and dividing by T_k yields

$$r_k^{\min} - \frac{\log |\mathcal{S}|}{T_k} \le -\frac{1}{T_k} \log U_{T_k}(\mathcal{S}) \le r_k^{\min}.$$

Thus

$$-\frac{1}{T_k} \log U_{T_k}(\mathcal{S}) \longrightarrow \min_{c \in \mathcal{S}} J_c(\lambda). \tag{22}$$

The set $\mathcal{C}^-$ is nonempty since $N \ge 2$. Cancelling $L_T(\theta^*)$ in (12) gives

$$\Pi_T\big(\mathrm{Alt}(\theta^*)\big) = 1 - \alpha_T = \frac{U_T(\mathcal{C}^-)}{U_T(\mathcal{C}_N)}. \tag{23}$$

For the numerator,

$$\min_{c \in \mathcal{C}^-} J_c(\lambda) = \inf_{\theta \in \mathrm{Alt}(\theta^*)} D(\theta^*, \theta; \lambda).$$

For the denominator, KL nonnegativity gives $J_c(\lambda) \ge 0$ for all c. Taking $c = c^*$ and $(p, q) = (p^*, q^*)$ gives $J_{c^*}(\lambda) = 0$. Hence $\min_{c \in \mathcal{C}_N} J_c(\lambda) = 0$. Applying (22) to both sums in (23) now gives

$$\begin{aligned} -\frac{1}{T_k} \log(1 - \alpha_{T_k}) &= -\frac{1}{T_k} \log U_{T_k}(\mathcal{C}^-) + \frac{1}{T_k} \log U_{T_k}(\mathcal{C}_N) \\ &\longrightarrow \inf_{\theta \in \mathrm{Alt}(\theta^*)} D(\theta^*, \theta; \lambda) \le \Gamma^*(\theta^*). \end{aligned} \tag{24}$$

Step 5: removing convergence of the allocation. The preceding calculation holds on the fixed path in Ω_0 along every subsequence whose allocations converge. Suppose that

$$\limsup_{T \to \infty} -\frac{1}{T} \log \Pi_T\big(\mathrm{Alt}(\theta^*)\big) > \Gamma^*(\theta^*)$$

on this path. Then there exist $\varepsilon > 0$ and budgets $T_k \to \infty$ such that

$$-\frac{1}{T_k} \log(1 - \alpha_{T_k}) > \Gamma^*(\theta^*) + \varepsilon \qquad \text{for every } k.$$

Compactness of $\Delta(\mathcal{I})$ gives a further subsequence (T_{k_j}) with $\widehat{\lambda}_{T_{k_j}} \to \lambda$. By (24), the exponents on this further subsequence converge to a value at most $\Gamma^*(\theta^*)$, contradicting the displayed inequality. Since $\mathbb{P}_{\theta^*}(\Omega_0) = 1$, we have proved that, almost surely,

$$\limsup_{T \to \infty} -\frac{1}{T} \log \Pi_T\big(\mathrm{Alt}(\theta^*)\big) \le \Gamma^*(\theta^*).$$

Equivalently, on the same event, for every $\varepsilon > 0$ there exists a finite, possibly path-dependent T_ε such that

$$\Pi_T\big(\mathrm{Alt}(\theta^*)\big) \ge \exp\{-T[\Gamma^*(\theta^*) + \varepsilon]\}, \qquad T \ge T_\varepsilon.$$

□

B PROOF OF THE POSTERIOR UPPER BOUND 2

B.0.1 REGRET OF THE TWO LEARNERS (LEMMA 1.2)

Learner regret (Lemma 1.2). Fix the instance θ^* and a positive prior ϱ_0 on canonical partitions. The parameter priors are independent, with laws $\mathrm{TBeta}_+(a_p, b_p)$ and $\mathrm{TBeta}_-(a_q, b_q)$ and fixed positive shape parameters. All logarithms are natural. The algorithm uses $\alpha \in (0, 1/2)$, $c_B > 0$, and

$$K := |\mathcal{I}|, \qquad \eta_T = T^{-1/2}, \qquad \kappa_T = \sqrt{T},$$
$$B_T = c_B \log(T+1), \qquad v_T = \log(4K(T+1)^4).$$

We specify every constant before proving the bounds.

Constants and budget conditions. Set

$$\epsilon := \frac{\min\{q^*, 1-p^*\}}{32}, \qquad L_\epsilon := \log\frac{1}{\epsilon}, \qquad \varrho_{\min} := \min_{c \in \mathcal{C}_N^\circ} \varrho_0(c) > 0.$$

For $x \in \mathbb{R}$, write $x_+ := \max\{x, 0\}$. Define

$$Z_+(a,b) := \int_{1/2}^1 x^{a-1}(1-x)^{b-1}\,dx, \qquad Z_-(a,b) := \int_0^{1/2} x^{a-1}(1-x)^{b-1}\,dx,$$

and the following explicit lower bound for both prior densities:

$$m_\epsilon := \min\left\{1,\ \frac{\epsilon^{(a_p-1)_+ + (b_p-1)_+}}{Z_+(a_p,b_p)},\ \frac{\epsilon^{(a_q-1)_+ + (b_q-1)_+}}{Z_-(a_q,b_q)}\right\} > 0.$$

Indeed, on $[\epsilon, 1-\epsilon]$, a nonnegative exponent gives $x^u \geq \epsilon^u$, whereas a negative exponent gives $x^u \geq 1$. Applying this observation to both x and $1-x$ shows that each normalized prior density is at least m_ϵ on its admissible part of this interval. Put

$$\omega_\epsilon := \varrho_{\min}(\epsilon m_\epsilon)^2, \qquad A_\epsilon := \frac{2e^{1/2}}{\epsilon m_\epsilon}, \qquad c_\epsilon := \frac{9\epsilon^2}{4K}.$$

Thus $0 < \omega_\epsilon < 1$. For a uniform partition prior and uniform truncated Beta priors, $a_p = b_p = a_q = b_q = 1$, $m_\epsilon = 1$, and $\omega_\epsilon = \epsilon^2/|\mathcal{C}_N^\circ|$.

For completeness, the analysis cutoff is the following explicit function of the instance and budget:

$$\Delta_* := \min\{p^* - 1/2, 1/2 - q^*\}, \qquad \eta_* := \min\{\Delta_*, q^*, 1-p^*\},$$
$$n_{\mathrm{rec}}(T) := \left\lceil \max\left\{1, (8K^2 v_T)^{1/(1-2\alpha)}, \left(\frac{4Kv_T}{\Delta_*^2}\right)^{1/(1-\alpha)}, \frac{1}{4\eta_*}\right\}\right\rceil,$$
$$s_T := \max\left\{n_{\mathrm{rec}}(T), \left\lceil (\kappa_T v_T^2)^{1/(1-\alpha)}\right\rceil\right\}.$$

In particular, $s_T^{1-\alpha}/\kappa_T \geq v_T^2$ and $s_T/T \to 0$. For the latter claim, each term defining $n_{\mathrm{rec}}(T)$ grows at most as a fixed power of $\log(T+1)$, whereas the other term is $T^{1/[2(1-\alpha)]} v_T^{2/(1-\alpha)}$ before rounding; $1/[2(1-\alpha)] < 1$.

Finally, define

$$B_* := \max\{d(p^*, q^*), d(q^*, p^*)\}, \qquad T_0 := \left\lceil \max\left\{1, \frac{K^2}{\epsilon^4}, e^{B_*/c_B}\right\}\right\rceil.$$

For every $T \geq T_0$, we will bound the posterior mass outside the compact set S_ϵ by the following quantity, on the event $\mathcal{G}_T$ and for rounds $t > s_T$; both objects are defined below:

$$\zeta_T := \min\left\{1,\ A_\epsilon(T+1)\exp\left(-c_\epsilon \frac{s_T^{1-\alpha}}{\kappa_T}\right)\right\}.$$

We prove the learner-regret inequalities with the explicit choices

$$R_T^{\max} := B_T\sqrt{T}\left(\log K + \frac{1}{8}\right)$$

and

$$R_T^{\min} := L_\epsilon \min\{s_T, T\} + 1 + \kappa_T \left(2\log(T+1) + \log\frac{1}{\omega_\epsilon}\right) + \frac{TL_\epsilon^2}{8\kappa_T} + 2(K+1)L_\epsilon\sqrt{2Tv_T} + (T-s_T)_+ B_T\zeta_T.$$

The max-learner bound holds on every sample path. For $T \geq T_0$, both bounds hold simultaneously with probability at least $1-(T+1)^{-3}$. These are sufficient analysis thresholds; no minimality is claimed. The learner does not need to know θ^*.

1. The max-learner. Let

$$W_t := \sum_{e\in\mathcal{I}} \exp(\ell_{t,e}).$$

The update $\ell_{t+1,e} = \ell_{t,e} + \eta_T g_{t,e}$ gives

$$\frac{W_{t+1}}{W_t} = \sum_e \lambda_{t,e} e^{\eta_T g_{t,e}}.$$

Since $g_{t,e} \in [0,1]$, Hoeffding's lemma implies

$$\log\frac{W_{t+1}}{W_t} \leq \eta_T\langle\lambda_t, g_t\rangle + \frac{\eta_T^2}{8}.$$

On the other hand, $W_1 = K$ and $\ell_{T+1,e} = \eta_T\sum_{t=1}^T g_{t,e}$, so

$$\log\frac{W_{T+1}}{W_1} \geq \eta_T \max_e \sum_{t=1}^T g_{t,e} - \log K.$$

Summing the upper bound, comparing the two displays, and dividing by η_T yields

$$\max_{\lambda\in\Delta(\mathcal{I})} \sum_{t=1}^T \langle\lambda, g_t\rangle - \sum_{t=1}^T \langle\lambda_t, g_t\rangle \leq \frac{\log K}{\eta_T} + \frac{T\eta_T}{8}.$$

Multiplying by B_T, using $G_t = B_T g_t$, proves

$$\max_{\lambda\in\Delta(\mathcal{I})} \sum_{t=1}^T \langle\lambda, G_t\rangle - \sum_{t=1}^T \langle\lambda_t, G_t\rangle \leq R_T^{\max}.$$

This calculation is deterministic and allows adaptive gains.

2. A fixed auxiliary posterior. Let $\mathcal{G}_T$ be the event that, simultaneously for every pair and every $1 \leq s \leq T$,

$$\left|N_s(e) - \sum_{t=1}^s \widetilde{\lambda}_{t,e}\right| \leq \sqrt{2sv_T},$$

$$|\widehat{\mu}_{s,e} - \mu_e^*| \leq \sqrt{\frac{v_T}{2N_s(e)}} \qquad \text{whenever } N_s(e) > 0.$$

For each pair, the count difference is a martingale with increments bounded by one. Azuma's inequality bounds the probability of a violation at a fixed time s by $2e^{-v_T}$. For the response bound, realize each pair's observations as successive entries of an independent Bernoulli sequence. Hoeffding's inequality bounds the probability of a violation at each fixed prefix length by $2e^{-v_T}$. Union bounds over K pairs and T times or prefix lengths give

$$\mathbb{P}_{\theta^*}(\mathcal{G}_T^c) \leq 4KTe^{-v_T} \leq T^{-3}.$$

For $n_{\mathrm{rec}}(T) \leq s \leq T$, forced exploration and the count bound imply

$$N_s(e) \geq \frac{1}{K}\sum_{t=1}^s t^{-\alpha} - \sqrt{2sv_T} \geq \frac{s^{1-\alpha}}{K} - \sqrt{2sv_T} \geq \frac{s^{1-\alpha}}{2K}.$$

The last inequality follows from $s^{1-2\alpha} \geq 8K^2 v_T$. The response bound then gives

$$\max_e |\widehat{\mu}_{s,e} - \mu_e^*| \leq \sqrt{\frac{K v_T}{s^{1-\alpha}}} \leq \frac{\Delta_*}{2},$$

where the last inequality uses $s^{1-\alpha} \geq 4K v_T/\Delta_*^2$. Every empirical sign is therefore correct. The true canonical partition has zero projection loss, while any different partition has positive loss. Hence, on $\mathcal{G}_T$,

$$\widehat{\sigma}_t = c^*, \qquad \mathcal{A}_t = \mathcal{A}^* := \mathcal{A}(c^*), \qquad t > s_T.$$

Only these recovery and concentration properties are used below. All additional martingale bounds will be applied under the original data-generating law, before intersecting their events with $\mathcal{G}_T$.

Set $\Theta_{\mathrm{nb}} = \mathcal{A}^* \times \mathcal{P} \times \mathcal{Q}$ and use the fixed value $\epsilon = \min\{q^*, 1-p^*\}/32$ specified above. Define

$$S_\epsilon := \{\theta \in \Theta_{\mathrm{nb}} : \epsilon \leq \mu_e(\theta) \leq 1-\epsilon \text{ for every } e \in \mathcal{I}\}.$$

This restriction applies only to parameters used by a candidate partition. An unused parameter retains its full prior range. Let Π_0 denote the fixed joint prior and let $\pi^\epsilon = \Pi_0(\cdot \mid S_\epsilon)$, which is well defined because $\Pi_0(S_\epsilon) > 0$. For every round, define

$$Q_t^\epsilon(d\theta) := \frac{L_{t-1}(\theta)^{1/\kappa_T} \pi^\epsilon(d\theta)}{\mathcal{Z}_{t-1}}, \qquad \mathcal{Z}_s := \int_{S_\epsilon} L_s(\theta)^{1/\kappa_T} \pi^\epsilon(d\theta).$$

This is an auxiliary distribution on a fixed set, defined even before the empirical partition is correct. It uses the entire observed history; there is no restart at s_T.

Write $Q_t = \mathcal{L}(\theta_t^- \mid \mathcal{H}_{t-1})$ for the actual min-learner distribution. On $\mathcal{G}_T$, for $t > s_T$, the equality $\mathcal{A}_t = \mathcal{A}^*$ gives

$$Q_t(d\theta) = \frac{\mathbf{1}\{\theta \in \Theta_{\mathrm{nb}}\} L_{t-1}(\theta)^{1/\kappa_T} \Pi_0(d\theta)}{\int_{\Theta_{\mathrm{nb}}} L_{t-1}(\vartheta)^{1/\kappa_T} \Pi_0(d\vartheta)}, \qquad Q_t(\cdot \mid S_\epsilon) = Q_t^\epsilon.$$

3. Posterior mass outside the bounded region. Fix $t > s_T$, a candidate partition, and a nonempty pair group under that partition. Put $s = t-1$, and let n and r be the group's pooled count and empirical mean. On $\mathcal{G}_T$, $n \geq s_T^{1-\alpha}/(2K)$. Because $T \geq T_0$ and $v_T \geq 1$,

$$\frac{K v_T}{s_T^{1-\alpha}} \leq \frac{K}{\sqrt{T}\, v_T} \leq \epsilon^2.$$

Every true pair mean belongs to $[32\epsilon, 1-32\epsilon]$. Thus every empirical pair mean, and hence every pooled mean r, belongs to $[31\epsilon, 1-31\epsilon]$, which is contained in $[4\epsilon, 1-4\epsilon]$. The conditional parameter law on $J \in \{(1/2,1),(0,1/2)\}$ is proportional to

$$e^{(n/\kappa_T)\phi_r(x)} \nu_J(dx), \qquad \phi_r(x) := r\log x + (1-r)\log(1-x),$$

where ν_J is the corresponding fixed truncated Beta prior. Its likelihood maximizer in the closure of J is $x_J = \max\{r, 1/2\}$ for the positive half and $x_J = \min\{r, 1/2\}$ for the negative half. Differentiation gives

$$\phi_r'(x) = \frac{r-x}{x(1-x)}, \qquad \phi_r''(x) = -\frac{r}{x^2} - \frac{1-r}{(1-x)^2} \leq -1.$$

If $x_J = r$, the first derivative vanishes. If $x_J = 1/2$, its sign opposes every feasible displacement. Thus $\phi_r'(x_J)(x - x_J) \leq 0$ for $x \in J$, and Taylor's inequality gives

$$\phi_r(x) \leq \phi_r(x_J) + \phi_r'(x_J)(x-x_J) - \frac{1}{2}(x-x_J)^2 \leq \phi_r(x_J) - \frac{1}{2}(x-x_J)^2.$$

For $x \notin [\epsilon, 1-\epsilon]$, the gap is at least $9\epsilon^2/2$. Consequently, the numerator of the conditional tail probability is at most

$$\exp\left\{\frac{n}{\kappa_T}\left(\phi_r(x_J) - \frac{9}{2}\epsilon^2\right)\right\}.$$

For the denominator, use $\delta_T := \epsilon/(T+1)$ and the interval $(x_J, x_J + \delta_T)$ for $J = (1/2, 1)$, or $(x_J - \delta_T, x_J)$ for $J = (0, 1/2)$. Since $x_J \in [4\epsilon, 1-4\epsilon]$ and $\delta_T \leq \epsilon/2$, the chosen interval lies in $J \cap [2\epsilon, 1-2\epsilon]$. Its prior density is at least m_ϵ. On this interval,

$$|\phi_r'(x)| = \left|\frac{r}{x} - \frac{1-r}{1-x}\right| \leq \frac{1}{2\epsilon}.$$

For each point in the interval,

$$\frac{n}{\kappa_T}\phi_r(x) \geq \frac{n}{\kappa_T}\phi_r(x_J) - \frac{n\delta_T}{2\epsilon\kappa_T} \geq \frac{n}{\kappa_T}\phi_r(x_J) - \frac{1}{2},$$

because $n \leq T$ and $\kappa_T \geq 1$. Integration therefore gives

$$\int_J e^{(n/\kappa_T)\phi_r(x)}\nu_J(dx) \geq \frac{\epsilon m_\epsilon}{T+1}\exp\left\{\frac{n}{\kappa_T}\phi_r(x_J) - \frac{1}{2}\right\}.$$

Dividing the numerator by this lower bound yields

$$Q_t\big(x \notin [\epsilon, 1-\epsilon] \mid c\big) \leq \frac{e^{1/2}(T+1)}{\epsilon m_\epsilon}\exp\left(-\frac{9\epsilon^2 n}{2\kappa_T}\right).$$

Here the conditional law also conditions on the observed history. Each active group has $n \geq s_T^{1-\alpha}/(2K)$. A union bound over the at most two active parameters, followed by averaging over c, gives

$$\begin{aligned} Q_t(S_\epsilon^c) &\leq \frac{2e^{1/2}(T+1)}{\epsilon m_\epsilon}\exp\left(-\frac{9\epsilon^2}{4K}\frac{s_T^{1-\alpha}}{\kappa_T}\right) \\ &= A_\epsilon(T+1)\exp\left(-c_\epsilon\frac{s_T^{1-\alpha}}{\kappa_T}\right). \end{aligned}$$

Combining this with the trivial probability bound of one proves $Q_t(S_\epsilon^c) \leq \zeta_T$. An inactive parameter contributes no tail event because it affects none of the candidate pair means.

4. Regret for the realized log losses. Define the per-observation negative log likelihood by

$$f_t(\theta) := -Y_t \log \mu_{I_t}(\theta) - (1 - Y_t)\log(1 - \mu_{I_t}(\theta)).$$

On S_ϵ, $0 \leq f_t(\theta) \leq L_\epsilon := \log(1/\epsilon)$. The normalizers satisfy the exact recursion

$$\frac{\mathcal{Z}_t}{\mathcal{Z}_{t-1}} = \mathbb{E}_{Q_t^\epsilon} e^{-f_t(\theta)/\kappa_T}.$$

Applying Hoeffding's lemma, multiplying by $-\kappa_T$, and rearranging yields

$$\mathbb{E}_{Q_t^\epsilon} f_t(\theta) \leq -\kappa_T \log\frac{\mathcal{Z}_t}{\mathcal{Z}_{t-1}} + \frac{L_\epsilon^2}{8\kappa_T}.$$

Since $\mathcal{Z}_0 = 1$, summing over all rounds gives

$$\sum_{t=1}^T \mathbb{E}_{Q_t^\epsilon} f_t(\theta) \leq -\kappa_T \log \mathcal{Z}_T + \frac{TL_\epsilon^2}{8\kappa_T}.$$

Write $F_T(\theta) = \sum_{t=1}^T f_t(\theta)$ and $F_T^* = \inf_{\theta \in S_\epsilon} F_T(\theta)$. For each partition, the active parameters range over intervals with compact closures inside $(0, 1)$. The continuous function F_T therefore attains its infimum in one of these closures. Near such a minimizer, use a one-sided interval of length $\delta_T = \epsilon/(T+1)$ for each active parameter. Each retained interval has length $1/2 - \epsilon$, and $2\delta_T \leq \epsilon < 1/2 - \epsilon$, so at least one side of every point can accommodate the chosen interval. Integrate each inactive parameter over its full prior. Denote the resulting box, with its fixed partition, by U_T, and let $d_c \leq 2$ be the number of active parameters. Since π^ϵ is the prior conditioned on S_ϵ,

$$\begin{aligned} \pi^\epsilon(U_T) &= \frac{\Pi_0(U_T)}{\Pi_0(S_\epsilon)} \geq \Pi_0(U_T) \\ &\geq \varrho_{\min}(m_\epsilon\delta_T)^{d_c} \geq \varrho_{\min}(m_\epsilon\delta_T)^2 = \frac{\omega_\epsilon}{(T+1)^2}. \end{aligned}$$

The penultimate inequality uses $m_\epsilon\delta_T \leq 1$. The derivative of a Bernoulli log loss with respect to its mean is

$$-\frac{Y_t}{x} + \frac{1 - Y_t}{1 - x},$$

whose absolute value is at most $1/\epsilon$ on the retained range. Each observation depends on one active parameter, and that parameter moves by at most δ_T. Hence

$$\sup_{\theta \in U_T} F_T(\theta) \le F_T^* + \frac{T\delta_T}{\epsilon} = F_T^* + \frac{T}{T+1} \le F_T^* + 1.$$

At the excluded endpoint $1/2$, the interval is taken on the admissible side. Thus boundary infima cause no loss of prior mass. Integrating over U_T gives

$$\mathcal{Z}_T \ge \frac{\omega_\epsilon}{(T+1)^2} \exp\left(-\frac{F_T^* + 1}{\kappa_T}\right),$$

and therefore

$$-\kappa_T \log \mathcal{Z}_T \le F_T^* + 1 + \kappa_T \left(2\log(T+1) + \log\frac{1}{\omega_\epsilon}\right).$$

Define the explicit potential remainder

$$A_T^{\mathrm{pot}} := 1 + \kappa_T \left(2\log(T+1) + \log\frac{1}{\omega_\epsilon}\right) + \frac{TL_\epsilon^2}{8\kappa_T}.$$

Combining the normalizer bounds proves, on every realized history,

$$\sum_{t=1}^{T} \mathbb{E}_{Q_t^\epsilon} f_t(\theta) \le \inf_{\theta \in S_\epsilon} \sum_{t=1}^{T} f_t(\theta) + A_T^{\mathrm{pot}}.$$

5. From log losses to KL losses. Put

$$z_t(\theta) := f_t(\theta) - f_t(\theta^*), \qquad h_e(\theta) := d(\mu_e^*, \mu_e(\theta)), \qquad V_t(\theta) := \langle \widetilde{\lambda}_t, h(\theta)\rangle.$$

Both Q_t^ϵ and $\widetilde{\lambda}_t$ are $\mathcal{H}_{t-1}$-measurable. The pair is drawn from $\widetilde{\lambda}_t$ using fresh randomness. The observation noise is fresh and independent of the learner's internal randomization, so

$$Y_t \mid \sigma(\mathcal{H}_{t-1}, \theta_t^-, I_t) \sim \mathrm{Bernoulli}(\mu_{I_t}^*).$$

Taking conditional expectations, first over the response and then over the selected pair, gives

$$\mathbb{E}[z_t(\theta) \mid \mathcal{H}_{t-1}] = V_t(\theta).$$

The true pair means also lie in $[\epsilon, 1-\epsilon]$. Consequently, $|z_t(\theta)| \le L_\epsilon$ and $0 \le V_t(\theta) \le L_\epsilon$ on S_ϵ. The variables

$$D_t := \mathbb{E}_{Q_t^\epsilon} z_t(\theta) - \mathbb{E}_{Q_t^\epsilon} V_t(\theta)$$

are martingale differences with $|D_t| \le 2L_\epsilon$. Let $\mathcal{M}_T$ be the event

$$\left|\sum_{t=1}^{T} D_t\right| \le 2L_\epsilon \sqrt{2Tv_T}.$$

Azuma's inequality gives $\mathbb{P}_{\theta^*}(\mathcal{M}_T^c) \le 2e^{-v_T}$. The expectation $\mathbb{E}_{Q_t^\epsilon}$ here integrates the parameter only; the current observation remains random.

For a comparator, define

$$b_e(\theta) := \log \frac{\mu_e^*[1-\mu_e(\theta)]}{\mu_e(\theta)(1-\mu_e^*)}.$$

Expanding the Bernoulli log-likelihood ratio gives the exact identity

$$\begin{aligned}\sum_{t=1}^{T}[z_t(\theta) - V_t(\theta)] = &\sum_e \left[N_T(e) - \sum_{t=1}^{T} \widetilde{\lambda}_{t,e}\right] h_e(\theta) \\ &+ \sum_e [S_T(e) - \mu_e^* N_T(e)] b_e(\theta).\end{aligned}$$

On S_ϵ, $h_e(\theta) \le L_\epsilon$. Also $|\log(x/(1-x))| \le \log((1-\epsilon)/\epsilon) \le L_\epsilon$ for $x \in [\epsilon, 1-\epsilon]$, so $|b_e(\theta)| \le 2L_\epsilon$. On $\mathcal{G}_T$, the count deviation is at most $\sqrt{2Tv_T}$ and

$$|S_T(e) - \mu_e^* N_T(e)| \le \sqrt{\frac{N_T(e)v_T}{2}} \le \sqrt{\frac{Tv_T}{2}},$$

with the first quantity equal to zero when $N_T(e) = 0$. The coefficient bounds hold for every comparator. Summing the two contributions over the K pairs gives

$$\sup_{\theta \in S_\epsilon} \left| \sum_{t=1}^{T} [z_t(\theta) - V_t(\theta)] \right| \le K L_\epsilon \sqrt{2T v_T} + 2K L_\epsilon \sqrt{T v_T / 2} = 2K L_\epsilon \sqrt{2T v_T}.$$

No union bound over the continuous parameter space is needed.

Subtracting $\sum_t f_t(\theta^*)$ from both sides of the log-loss regret bound gives

$$\sum_{t=1}^{T} \mathbb{E}_{Q_t^\epsilon} z_t(\theta) \le \inf_{\theta \in S_\epsilon} \sum_{t=1}^{T} z_t(\theta) + A_T^{\mathrm{pot}}.$$

On $\mathcal{G}_T \cap \mathcal{M}_T$, the learner and comparator concentration bounds therefore imply

$$\begin{aligned} \sum_{t=1}^{T} \mathbb{E}_{Q_t^\epsilon} V_t(\theta) &\le \sum_{t=1}^{T} \mathbb{E}_{Q_t^\epsilon} z_t(\theta) + 2L_\epsilon \sqrt{2T v_T} \\ &\le \inf_{\theta \in S_\epsilon} \sum_{t=1}^{T} z_t(\theta) + A_T^{\mathrm{pot}} + 2L_\epsilon \sqrt{2T v_T} \\ &\le \inf_{\theta \in S_\epsilon} \sum_{t=1}^{T} V_t(\theta) + r_T^0, \end{aligned}$$

where

$$r_T^0 := A_T^{\mathrm{pot}} + 2(K+1) L_\epsilon \sqrt{2T v_T}.$$

6. The actual min-learner and the tail of the experiment. If $s_T \ge T$, both tail sums in the min-learner inequality are empty, and its bound is immediate. Hence assume $s_T < T$. All V_t are nonnegative and at most L_ϵ on S_ϵ. Consequently,

$$\begin{aligned} \sum_{t > s_T} \mathbb{E}_{Q_t^\epsilon} V_t(\theta) &\le \sum_{t=1}^{T} \mathbb{E}_{Q_t^\epsilon} V_t(\theta) \\ &\le \inf_{\theta \in S_\epsilon} \sum_{t=1}^{T} V_t(\theta) + r_T^0 \\ &\le \inf_{\theta \in S_\epsilon} \sum_{t > s_T} V_t(\theta) + s_T L_\epsilon + r_T^0. \end{aligned}$$

This accounts for the discarded prefix without restarting the posterior.

Recall $h_{T,e}(\theta) = \min\{h_e(\theta), B_T\}$ and $\overline{h}_t = \mathbb{E}_{Q_t} h_T(\theta)$. For $t > s_T$, decompose Q_t into its restrictions to S_ϵ and S_ϵ^c. Since $Q_t(\,\cdot \mid S_\epsilon) = Q_t^\epsilon$, $h_T \le h$, and $0 \le h_{T,e} \le B_T$,

$$\langle \widetilde{\lambda}_t, \overline{h}_t \rangle \le \mathbb{E}_{Q_t^\epsilon} V_t(\theta) + B_T \zeta_T.$$

It remains to identify the comparator infimum. For $x \in \{p^*, q^*\}$,

$$\frac{\partial}{\partial y} d(x, y) = \frac{y - x}{y(1-y)}.$$

Thus replacing $p > p^*$ by p^* and $q < q^*$ by q^* cannot increase any pairwise divergence or its clipped value. We may restrict both infima, in their closures, to

$$(p, q) \in [1/2, p^*] \times [q^*, 1/2].$$

This rectangle lies inside the retained range of pair means. For $x = p^*$, the divergence decreases as y moves from q^* to p^*; for $x = q^*$, it increases on that interval. Therefore,

$$\max_{x \in \{p^*, q^*\},\, y \in [q^*, p^*]} d(x, y) = \max\{d(p^*, q^*), d(q^*, p^*)\} = B_*.$$

The condition $T \geq T_0 \geq e^{B_*/c_B}$ gives $B_T = c_B \log(T+1) \geq B_*$. Clipping is consequently inactive on the rectangle. By continuity at $1/2$,

$$\inf_{\theta \in S_\epsilon} \sum_{t > s_T} V_t(\theta) = \inf_{\theta \in \Theta_{\mathrm{nb}}} \sum_{t > s_T} \langle \widetilde{\lambda}_t, h_T(\theta) \rangle.$$

Summing the posterior-tail error over the $T - s_T$ remaining rounds and combining the bounds yields

$$\begin{aligned}\sum_{t > s_T} \langle \widetilde{\lambda}_t, \overline{h}_t \rangle &\leq \sum_{t > s_T} \mathbb{E}_{Q_t^\epsilon} V_t(\theta) + (T - s_T) B_T \zeta_T \\ &\leq \inf_{\theta \in \Theta_{\mathrm{nb}}} \sum_{t > s_T} \langle \widetilde{\lambda}_t, h_T(\theta) \rangle + s_T L_\epsilon + r_T^0 + (T - s_T) B_T \zeta_T.\end{aligned}$$

The last three terms are exactly $R_T^{\min}$ when $s_T < T$. The definition with $\min\{s_T, T\}$ and $(T - s_T)_+$ also covers the empty-tail case.

Take $\mathcal{E}_T^{\mathrm{reg}} = \mathcal{G}_T \cap \mathcal{M}_T$. The max-learner bound is pathwise, and the min-learner bound holds on this intersection. Its failure probability is explicitly bounded by

$$\mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{reg}})^c) \leq (4KT + 2) e^{-v_T} = \frac{T + 1/(2K)}{(T+1)^4} \leq \frac{1}{(T+1)^3}.$$

7. An explicit constant for the simplified rate. The exact remainder above does not hide any multiplicative constant. To recover the shorter rate bound, define

$$\Lambda_K := 4 + \frac{\log(4K)}{\log 2},$$

$$C_{\mathrm{reg}} := \max\left\{ L_\epsilon,\ 2 + \frac{\log(1/\omega_\epsilon) + L_\epsilon^2/8 + 2}{\log 2} + 2(K+1) L_\epsilon \sqrt{\frac{2\Lambda_K}{\log 2}} \right\}.$$

Also set

$$z_{\mathrm{tail}} := \frac{3 + \sqrt{9 + 64 c_\epsilon [\log(A_\epsilon c_B)]_+}}{32 c_\epsilon}, \qquad T_1 := \max\{T_0, \lceil e^{z_{\mathrm{tail}}} \rceil\}.$$

For $z = \log(T+1)$, we have $v_T \geq 4z$, $T(T+1) \leq e^{2z}$, and $z \leq e^z$. Consequently,

$$\begin{aligned}T B_T \zeta_T &\leq A_\epsilon c_B T(T+1) \log(T+1) e^{-c_\epsilon v_T^2} \\ &\leq A_\epsilon c_B \exp(3z - 16 c_\epsilon z^2).\end{aligned}$$

If $T \geq T_1$, then $z \geq z_{\mathrm{tail}}$ and

$$16 c_\epsilon z^2 - 3z \geq [\log(A_\epsilon c_B)]_+,$$

so $T B_T \zeta_T \leq 1$. The quadratic is increasing beyond z_{tail}, since $z_{\mathrm{tail}} \geq 3/(16 c_\epsilon)$.

Next, $z \geq \log 2$ gives

$$v_T = 4z + \log(4K) \leq \Lambda_K z, \qquad \sqrt{z} \leq \frac{z}{\sqrt{\log 2}}.$$

Using $\kappa_T = \sqrt{T}$, the exact remainder therefore satisfies

$$\begin{aligned}R_T^{\min} &\leq L_\epsilon s_T + 2 + \sqrt{T}\left(2z + \log\frac{1}{\omega_\epsilon} + \frac{L_\epsilon^2}{8}\right) \\ &\quad + 2(K+1) L_\epsilon \sqrt{2\Lambda_K T z} \\ &\leq L_\epsilon s_T + \left[2 + \frac{\log(1/\omega_\epsilon) + L_\epsilon^2/8 + 2}{\log 2} + 2(K+1) L_\epsilon \sqrt{\frac{2\Lambda_K}{\log 2}}\right] \sqrt{T} \log(T+1) \\ &\leq C_{\mathrm{reg}} \left(s_T + \sqrt{T} \log(T+1)\right), \qquad T \geq T_1.\end{aligned}$$

In the second inequality, the constant term 2 is bounded by $2\sqrt{T}\log(T+1)/\log 2$. Thus the shorter bound holds with the displayed value of C_{reg} and the displayed finite budget threshold T_1. Since $s_T/T \to 0$, both $R_T^{\max}/T$ and $R_T^{\min}/T$ tend to zero. □

B.0.2 SAMPLING AND TRACKING (LEMMA 1.3)

Sampling and tracking (Lemma 1.3). Fix the instance θ^* and a budget T, and represent c^* canonically. The algorithm uses $\alpha \in (0, 1/2)$, $c_B > 0$, $\kappa_T = \sqrt{T}$, and $B_T = c_B \log(T+1)$. Write $K := |\mathcal{I}|$ and

$$\mathcal{H}_0 := \{\emptyset, \Omega\}, \qquad \mathcal{H}_s := \sigma(I_r, Y_r, \theta_r^- : 1 \le r \le s).$$

Both λ_t and $\widetilde{\lambda}_t = (1 - t^{-\alpha})\lambda_t + t^{-\alpha}u$, where $u_e = 1/K$, are $\mathcal{H}_{t-1}$-measurable. The current competitor is drawn after these allocations are computed. Pair selection uses fresh randomness, and the observation noise is fresh and independent of the learner's internal randomization. In particular, for each $e \in \mathcal{I}$,

$$\mathbb{P}_{\theta^*}(I_t = e \mid \mathcal{H}_{t-1}, \theta_t^-) = \widetilde{\lambda}_{t,e},$$
$$\mathbb{E}_{\theta^*}[Y_t \mid \sigma(\mathcal{H}_{t-1}, \theta_t^-, I_t)] = \mu^*_{I_t}.$$

Taking the conditional expectation of the first identity given $\mathcal{H}_{t-1}$ gives $\mathbb{P}_{\theta^*}(I_t = e \mid \mathcal{H}_{t-1}) = \widetilde{\lambda}_{t,e}$.

Constants and the claimed bounds. Define

$$v_T := \log(4K(T+1)^4), \qquad \epsilon := \frac{\min\{q^*, 1-p^*\}}{32},$$
$$\Delta_* := \min\{p^* - 1/2, 1/2 - q^*\}, \qquad \eta_* := \min\{\Delta_*, q^*, 1-p^*\}.$$

Use the deterministic cutoff

$$n_{\mathrm{rec}}(T) := \left\lceil \max\left\{1, (8K^2 v_T)^{1/(1-2\alpha)}, \left(\frac{4Kv_T}{\Delta_*^2}\right)^{1/(1-\alpha)}, \frac{1}{4\eta_*}\right\}\right\rceil,$$
$$s_T := \max\left\{n_{\mathrm{rec}}(T), \left\lceil (\kappa_T v_T^2)^{1/(1-\alpha)} \right\rceil\right\}, \qquad m_T := (T - s_T)_+,$$

where $x_+ := \max\{x, 0\}$. Set

$$a_T := \sqrt{\frac{Kv_T}{s_T^{1-\alpha}}} + \frac{2K}{s_T^{1-\alpha}}, \qquad T_{\mathrm{samp}} := \left\lceil \frac{K^2}{\epsilon^4} \right\rceil,$$
$$\mathcal{L}_T := \log\frac{1-\epsilon}{\epsilon} + \frac{B_T + \log 2}{\epsilon}.$$

These are sufficient analysis thresholds and constants; the learner does not need to know them. Define the clipped true-mean gains and their conditional expectations by

$$h_{T,e}(\theta) := \min\{d(\mu_e^*, \mu_e(\theta)), B_T\},$$
$$G_{t,e} := \min\{d(\mu_e(\theta_t^+), \mu_e(\theta_t^-)), B_T\},$$
$$\overline{h}_t := \mathbb{E}[h_T(\theta_t^-) \mid \mathcal{H}_{t-1}].$$

Finally, let

$$\widehat{\lambda}_T(e) := \frac{N_T(e)}{T}, \qquad \overline{\lambda}_T := \frac{1}{T}\sum_{t=1}^{T} \widetilde{\lambda}_t.$$

We prove that, for every integer $T \ge T_{\mathrm{samp}}$, there is an event $\mathcal{E}_T^{\mathrm{samp}}$ with probability at least $1 - (T+1)^{-3}$ on which

$$\left\| \frac{1}{T}\sum_{t>s_T} (G_t - \overline{h}_t) \right\|_\infty + \left| \frac{1}{T}\sum_{t>s_T} \langle \lambda_t, G_t - \overline{h}_t \rangle \right| \le r_T^{\mathrm{ps}},$$
$$\|\widehat{\lambda}_T - \overline{\lambda}_T\|_1 \le r_T^{\mathrm{tr}},$$

with the explicit choices

$$r_T^{\mathrm{ps}} := \frac{2m_T}{T}\mathcal{L}_T a_T + \frac{2B_T}{T}\sqrt{2m_T v_T}, \qquad r_T^{\mathrm{tr}} := K\sqrt{\frac{2v_T}{T}}.$$

All sums with $t > s_T$ are restricted to $t \le T$. If $s_T \ge T$, these sums and r_T^{ps} are zero.

1. Simultaneous concentration. Let $\mathcal{G}_T$ be the event that, for every $e \in \mathcal{I}$ and every $1 \le s \le T$,

$$\left| N_s(e) - \sum_{t=1}^{s} \widetilde{\lambda}_{t,e} \right| \le \sqrt{2sv_T},$$

$$|\widehat{\mu}_{s,e} - \mu_e^*| \le \sqrt{\frac{v_T}{2N_s(e)}} \quad \text{whenever } N_s(e) > 0.$$

Here $\widehat{\mu}_{s,e} = S_s(e)/N_s(e)$ when the count is positive. For each pair, define

$$D_{t,e} := \mathbf{1}\{I_t = e\} - \widetilde{\lambda}_{t,e}, \qquad M_{s,e} := \sum_{t=1}^{s} D_{t,e} = N_s(e) - \sum_{t=1}^{s} \widetilde{\lambda}_{t,e}.$$

The conditional query law gives

$$\mathbb{E}[D_{t,e} \mid \mathcal{H}_{t-1}] = \widetilde{\lambda}_{t,e} - \widetilde{\lambda}_{t,e} = 0, \qquad |D_{t,e}| \le 1.$$

Thus $(M_{s,e})_{s=0}^{T}$ is a martingale. For every $r > 0$, Azuma's inequality yields

$$\mathbb{P}_{\theta^*}(|M_{s,e}| > r) \le 2\exp\left(-\frac{r^2}{2s}\right).$$

Substituting $r = \sqrt{2sv_T}$ gives a failure probability of $2e^{-v_T}$ at each (e, s). Taking a union over K pairs and T times bounds the probability of any count violation by $2KTe^{-v_T}$.

For the response bound, let $(Y_{e,j})_{e\in\mathcal{I},j\ge1}$ be independent random variables with $Y_{e,j} \sim \mathrm{Bernoulli}(\mu_e^*)$, independent of the random seeds used by the learner. The learner receives

$$Y_t = Y_{I_t, N_{t-1}(I_t)+1}.$$

The query is selected before this entry is revealed. The unused entry has the required conditional Bernoulli law, so this construction has the same observation law as the algorithm. For every s and e,

$$S_s(e) = \sum_{j=1}^{N_s(e)} Y_{e,j}.$$

At each fixed prefix length $n \in \{1, \dots, T\}$, Hoeffding's inequality gives, for every $r > 0$,

$$\mathbb{P}_{\theta^*}\left(\left| \frac{1}{n} \sum_{j=1}^{n} Y_{e,j} - \mu_e^* \right| > r \right) \le 2e^{-2nr^2}.$$

With $r = \sqrt{v_T/(2n)}$, this becomes

$$\mathbb{P}_{\theta^*}\left(\left| \frac{1}{n} \sum_{j=1}^{n} Y_{e,j} - \mu_e^* \right| > \sqrt{\frac{v_T}{2n}} \right) \le 2e^{-v_T}.$$

To pass to the adaptive count, observe the event inclusion

$$\left\{ \exists s \le T :\ N_s(e) > 0, \quad |\widehat{\mu}_{s,e} - \mu_e^*| > \sqrt{\frac{v_T}{2N_s(e)}} \right\}$$

$$\subseteq \bigcup_{n=1}^{T} \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} Y_{e,j} - \mu_e^* \right| > \sqrt{\frac{v_T}{2n}} \right\}.$$

Indeed, any violation on the left supplies an integer $n = N_s(e) \in \{1, \dots, T\}$ for which the right-hand event occurs. There is no need to condition on the random count. A union over the K pairs and T prefix lengths bounds the probability of any response violation by $2KTe^{-v_T}$. Combining the two bounds gives

$$\mathbb{P}_{\theta^*}(\mathcal{G}_T^c) \le 4KTe^{-v_T}.$$

2. Recovery and reference accuracy. Since $\lambda_{t,e} \geq 0$, forced exploration gives

$$\widetilde{\lambda}_{t,e} = (1 - t^{-\alpha})\lambda_{t,e} + t^{-\alpha}/K \geq t^{-\alpha}/K.$$

For $t \leq s$, $t^{-\alpha} \geq s^{-\alpha}$, and hence $\sum_{t=1}^{s} t^{-\alpha} \geq s^{1-\alpha}$. On $\mathcal{G}_T$, it follows that, for $n_{\mathrm{rec}}(T) \leq s \leq T$,

$$N_s(e) \geq \frac{1}{K}\sum_{t=1}^{s} t^{-\alpha} - \sqrt{2sv_T} \geq \frac{s^{1-\alpha}}{K} - \sqrt{2sv_T} \geq \frac{s^{1-\alpha}}{2K},$$

$$|\widehat{\mu}_{s,e} - \mu_e^*| \leq \sqrt{\frac{Kv_T}{s^{1-\alpha}}} \leq \frac{\Delta_*}{2}.$$

For the last count inequality, the definition of $n_{\mathrm{rec}}(T)$ gives

$$s^{1-2\alpha} \geq 8K^2 v_T \quad \Longrightarrow \quad \frac{s^{2-2\alpha}}{4K^2} \geq 2sv_T \quad \Longrightarrow \quad \frac{s^{1-\alpha}}{2K} \geq \sqrt{2sv_T}.$$

The response bound follows by substituting this count lower bound into $\sqrt{v_T/(2N_s(e))}$ and then using $s^{1-\alpha} \geq 4Kv_T/\Delta_*^2$. For a within-cluster pair,

$$\widehat{\mu}_{s,e} \geq p^* - \Delta_*/2 \geq 1/2 + \Delta_*/2 > 1/2.$$

For a between-cluster pair,

$$\widehat{\mu}_{s,e} \leq q^* + \Delta_*/2 \leq 1/2 - \Delta_*/2 < 1/2.$$

Thus thresholding gives every true pair relation. The partition c^* has zero projection loss. A different canonical partition disagrees with c^* on at least one pair, so it has positive loss. The projection therefore has the unique minimizer c^*. Since s_T is an integer, $t > s_T$ implies $t-1 \geq s_T \geq n_{\mathrm{rec}}(T)$, and hence $\widehat{\sigma}_t = c^*$ at each such round.

Fix such a round and put $s = t - 1$. For a nonempty true pair group A, let x^* be its common mean, and write $n = \sum_{e\in A} N_s(e)$ and $k = \sum_{e\in A} S_s(e)$. Every pair in A has positive count, so $n > 0$. Expanding the pooled mean gives

$$\begin{aligned}\left|\frac{k}{n} - x^*\right| &= \left|\sum_{e\in A} \frac{N_s(e)}{n}(\widehat{\mu}_{s,e} - x^*)\right| \\ &\leq \sum_{e\in A} \frac{N_s(e)}{n}|\widehat{\mu}_{s,e} - x^*| \leq \sqrt{\frac{Kv_T}{s^{1-\alpha}}}.\end{aligned}$$

The smoothing error is

$$\left|\frac{k+1}{n+2} - \frac{k}{n}\right| = \frac{|n-2k|}{n(n+2)} \leq \frac{1}{n+2}.$$

The last inequality uses $0 \leq k \leq n$, which implies $-n \leq n - 2k \leq n$. The algorithm clips the smoothed estimate to $[1/2+\epsilon_s, 1-\epsilon_s]$ for a within-cluster group, and to $[\epsilon_s, 1/2-\epsilon_s]$ for a between-cluster group, where $\epsilon_s = 1/[4(s+1)]$. Since $s \geq n_{\mathrm{rec}}(T) \geq 1/(4\eta_*)$, these intervals contain their respective true parameters. More explicitly,

$$\begin{gathered}\epsilon_s \leq \eta_*, \qquad p^* \geq 1/2 + \eta_*, \qquad p^* \leq 1 - \eta_*, \\ q^* \geq \eta_*, \qquad q^* \leq 1/2 - \eta_*.\end{gathered}$$

For any interval $[a, b]$ containing x^*, clipping a real number z to this interval cannot increase $|z - x^*|$: if $z < a$, the clipped distance is $x^* - a \leq x^* - z$; if $z > b$, it is $b - x^* \leq z - x^*$; otherwise clipping leaves z unchanged. Writing $\widehat{x}_s$ for the clipped smoothed estimate therefore gives

$$\begin{aligned}|\widehat{x}_s - x^*| &\leq \left|\frac{k+1}{n+2} - x^*\right| \\ &\leq \left|\frac{k+1}{n+2} - \frac{k}{n}\right| + \left|\frac{k}{n} - x^*\right| \\ &\leq \frac{1}{n+2} + \sqrt{\frac{Kv_T}{s^{1-\alpha}}}.\end{aligned}$$

Because A contains at least one pair, $n \geq s^{1-\alpha}/(2K)$. Thus $1/(n+2) \leq 1/n \leq 2K/s^{1-\alpha}$. Applying the resulting bound to each nonempty true group, and using $s \geq s_T$, yields

$$\max_e |\mu_e(\theta_t^+) - \mu_e^*| \leq \sqrt{\frac{K v_T}{s^{1-\alpha}}} + \frac{2K}{s^{1-\alpha}} \leq a_T.$$

An empty group contributes no pair mean and requires no parameter accuracy bound.

To place the reference means in a fixed compact interval, first note that $K, T \geq 1$ implies $v_T \geq \log 64 > 1$, and that $0 < \epsilon < 1/64$. The choice of s_T gives $s_T^{1-\alpha} \geq \sqrt{T}\, v_T^2$. Also,

$$T \geq T_{\text{samp}} \quad \Longrightarrow \quad T^{1/4} \geq \frac{\sqrt{K}}{\epsilon}, \qquad \sqrt{T} \geq \frac{K}{\epsilon^2}.$$

Substituting these bounds gives

$$a_T \leq \frac{\sqrt{K}}{T^{1/4}\sqrt{v_T}} + \frac{2K}{\sqrt{T}\, v_T^2} \leq \epsilon + 2\epsilon^2 \leq 2\epsilon.$$

The last inequality follows from $2\epsilon^2 \leq \epsilon$. By definition of ϵ, $32\epsilon \leq q^* < p^* \leq 1 - 32\epsilon$. Hence every true pair mean belongs to $[32\epsilon, 1 - 32\epsilon]$, and

$$30\epsilon \leq \mu_e^* - a_T \leq \mu_e(\theta_t^+) \leq \mu_e^* + a_T \leq 1 - 30\epsilon.$$

Both the true and reference pair means consequently belong to $[\epsilon, 1-\epsilon]$ on $\mathcal{G}_T$ for every $t > s_T$.

3. An explicit Lipschitz bound for clipped KL. For fixed $B > 0$ and $y \in (0,1)$, set $F_{B,y}(x) := \min\{d(x,y), B\}$. We claim that for all $x, x' \in [\epsilon, 1-\epsilon]$,

$$|F_{B,y}(x) - F_{B,y}(x')| \leq \left(\log \frac{1-\epsilon}{\epsilon} + \frac{B + \log 2}{\epsilon}\right) |x - x'|.$$

Write $\mathfrak{h}(x) = -x \log x - (1-x)\log(1-x)$. Its derivative is $\log((1-x)/x)$, so its maximum is attained at $1/2$ and equals $\log 2$. At a point where $d(x,y) < B$, the cross-entropy identity gives

$$\begin{aligned} \epsilon[-\log y - \log(1-y)] &\leq -x \log y - (1-x)\log(1-y) \\ &= d(x,y) + \mathfrak{h}(x) \leq B + \log 2. \end{aligned}$$

Moreover,

$$\begin{aligned} |\partial_x d(x,y)| &= \left|\log \frac{x}{1-x} - \log \frac{y}{1-y}\right| \\ &\leq \log \frac{1-\epsilon}{\epsilon} + [-\log y - \log(1-y)] \\ &\leq \log \frac{1-\epsilon}{\epsilon} + \frac{B + \log 2}{\epsilon}. \end{aligned}$$

On the open set where $d(x,y) > B$, the clipped function has derivative zero. For fixed y, $d(\cdot, y)$ is continuously differentiable with bounded derivative on $[\epsilon, 1-\epsilon]$, and is therefore Lipschitz on this interval. Since $u \mapsto \min\{u, B\}$ is 1-Lipschitz, $F_{B,y}$ is Lipschitz and hence absolutely continuous there. Moreover,

$$\partial_x^2 d(x,y) = \frac{1}{x} + \frac{1}{1-x} > 0.$$

Strict convexity implies that $d(x,y) = B$ at at most two points. These possible clipping points have zero Lebesgue measure, so the displayed derivative bound holds almost everywhere for $F_{B,y}$. For $x < x'$, the fundamental theorem for absolutely continuous functions gives

$$\begin{aligned} |F_{B,y}(x') - F_{B,y}(x)| &= \left|\int_x^{x'} F'_{B,y}(u)\, du\right| \\ &\leq \int_x^{x'} |F'_{B,y}(u)|\, du \\ &\leq \left(\log \frac{1-\epsilon}{\epsilon} + \frac{B + \log 2}{\epsilon}\right)(x' - x). \end{aligned}$$

Interchanging x, x' covers the other order, proving the claim.

For $t > s_T$, define the reference approximation error $A_t := G_t - h_T(\theta_t^-)$. Applying the claim on $\mathcal{G}_T$ with $B = B_T$ gives, for every coordinate,

$$\begin{aligned}|A_{t,e}| &= \left|\min\{d(\mu_e(\theta_t^+), \mu_e(\theta_t^-)), B_T\} - \min\{d(\mu_e^*, \mu_e(\theta_t^-)), B_T\}\right| \\ &\le \mathcal{L}_T |\mu_e(\theta_t^+) - \mu_e^*| \le \mathcal{L}_T a_T.\end{aligned}$$

Taking a maximum, and then using $\lambda_{t,e} \ge 0$ and $\sum_e \lambda_{t,e} = 1$, yields

$$\begin{aligned}\|A_t\|_\infty &\le \mathcal{L}_T a_T, \\ |\langle \lambda_t, A_t \rangle| &\le \sum_e \lambda_{t,e} |A_{t,e}| \le \mathcal{L}_T a_T \sum_e \lambda_{t,e} = \mathcal{L}_T a_T.\end{aligned}$$

The bound is uniform over the competing means in $(0, 1)$. In particular, it does not require a posterior localization event.

4. Concentration of posterior sampling. Define

$$X_t := h_T(\theta_t^-) - \overline{h}_t, \qquad U_t := \langle \lambda_t, X_t \rangle.$$

For every e, $X_{t,e}$ is $\mathcal{H}_t$-measurable. By the definition and $\mathcal{H}_{t-1}$-measurability of $\overline{h}_t$,

$$\mathbb{E}[X_{t,e} \mid \mathcal{H}_{t-1}] = \mathbb{E}[h_{T,e}(\theta_t^-) \mid \mathcal{H}_{t-1}] - \overline{h}_{t,e} = 0.$$

Since $0 \le h_{T,e}(\theta_t^-) \le B_T$, conditional expectation also gives $0 \le \overline{h}_{t,e} \le B_T$. Their difference is therefore in $[-B_T, B_T]$, and $|X_{t,e}| \le B_T$. Because λ_t is $\mathcal{H}_{t-1}$-measurable and belongs to the simplex,

$$\begin{aligned}\mathbb{E}[U_t \mid \mathcal{H}_{t-1}] &= \sum_e \lambda_{t,e} \mathbb{E}[X_{t,e} \mid \mathcal{H}_{t-1}] = 0, \\ |U_t| &\le \sum_e \lambda_{t,e} |X_{t,e}| \le B_T \sum_e \lambda_{t,e} = B_T.\end{aligned}$$

No independence across coordinates is needed. If $m_T > 0$, define the tail processes

$$M_{j,e}^{\mathrm{ps}} := \sum_{t=s_T+1}^{s_T+j} X_{t,e}, \qquad M_j^{\mathrm{w}} := \sum_{t=s_T+1}^{s_T+j} U_t, \qquad 0 \le j \le m_T.$$

The cutoff s_T is deterministic. Thus both processes are martingales with respect to $(\mathcal{H}_{s_T+j})_{j=0}^{m_T}$, start at zero, and have increments bounded by B_T. For each process, Azuma's inequality gives the bound $2\exp(-r^2/(2m_T B_T^2))$ at radius $r > 0$. Substituting $r = B_T\sqrt{2m_T v_T}$ gives

$$\begin{aligned}\mathbb{P}_{\theta^*}\left(\left|\sum_{t>s_T} X_{t,e}\right| > B_T\sqrt{2m_T v_T}\right) &\le 2e^{-v_T}, \\ \mathbb{P}_{\theta^*}\left(\left|\sum_{t>s_T} U_t\right| > B_T\sqrt{2m_T v_T}\right) &\le 2e^{-v_T}.\end{aligned}$$

Let $\mathcal{M}_T$ be the event that all K coordinate bounds and the scalar bound hold. No independence between these events is required: the union bound gives

$$\mathbb{P}_{\theta^*}(\mathcal{M}_T^c) \le 2(K+1)e^{-v_T}.$$

For $m_T = 0$, take $\mathcal{M}_T = \Omega$. These concentration inequalities are established under the original data-generating law, without conditioning on $\mathcal{G}_T$.

On $\mathcal{G}_T \cap \mathcal{M}_T$, the decomposition $G_t - \overline{h}_t = A_t + X_t$ yields

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t>s_T}(G_t - \overline{h}_t)\right\|_\infty &\le \frac{1}{T}\sum_{t>s_T}\|A_t\|_\infty + \frac{1}{T}\left\|\sum_{t>s_T} X_t\right\|_\infty \\ &\le \frac{m_T}{T}\mathcal{L}_T a_T + \frac{B_T}{T}\sqrt{2m_T v_T}, \\ \left|\frac{1}{T}\sum_{t>s_T}\langle \lambda_t, G_t - \overline{h}_t\rangle\right| &\le \frac{1}{T}\sum_{t>s_T}|\langle \lambda_t, A_t\rangle| + \frac{1}{T}\left|\sum_{t>s_T} U_t\right| \\ &\le \frac{m_T}{T}\mathcal{L}_T a_T + \frac{B_T}{T}\sqrt{2m_T v_T}.\end{aligned}$$

Adding the two inequalities gives exactly the stated r_T^{ps}.

5. Tracking and the joint event. On $\mathcal{G}_T$, the count bound at $s = T$ implies

$$\begin{aligned}\|\widehat{\lambda}_T - \overline{\lambda}_T\|_1 &= \frac{1}{T}\sum_{e\in\mathcal{I}}\left|N_T(e) - \sum_{t=1}^{T}\widetilde{\lambda}_{t,e}\right| \\ &\le \frac{K}{T}\sqrt{2Tv_T} = K\sqrt{\frac{2v_T}{T}} = r_T^{\mathrm{tr}}.\end{aligned}$$

Define $\mathcal{E}_T^{\mathrm{samp}} := \mathcal{G}_T \cap \mathcal{M}_T$. Since $K \ge 1$, its failure probability satisfies

$$\begin{aligned}\mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{samp}})^c) &\le \mathbb{P}_{\theta^*}(\mathcal{G}_T^c) + \mathbb{P}_{\theta^*}(\mathcal{M}_T^c) \\ &\le (4KT + 2K + 2)e^{-v_T} \\ &= \frac{4KT + 2K + 2}{4K(T+1)^4} \\ &= \frac{T + 1/2 + 1/(2K)}{(T+1)^4} \le \frac{1}{(T+1)^3}.\end{aligned}$$

The last inequality uses $1/2 + 1/(2K) \le 1$. This proves both inequalities with the claimed probability, including $K = 1$.

6. Vanishing remainders. Set $z_T := \log(T+1)$ and define the fixed constants

$$A_K := \sqrt{K} + 2K, \qquad \Lambda_K := 4 + \frac{\log(4K)}{\log 2},$$
$$L_* := \frac{c_B}{\epsilon} + \frac{\log((1-\epsilon)/\epsilon) + (\log 2)/\epsilon}{\log 2}.$$

Since $z_T \ge \log 2$,

$$4z_T \le v_T = 4z_T + \log(4K) \le \Lambda_K z_T, \qquad \mathcal{L}_T \le L_* z_T.$$

Also, $T, v_T \ge 1$ implies $\sqrt{T}\, v_T^2 \ge T^{1/4}\sqrt{v_T}$. Therefore,

$$\begin{aligned}a_T &\le \frac{\sqrt{K}}{T^{1/4}\sqrt{v_T}} + \frac{2K}{\sqrt{T}\, v_T^2} \\ &\le \frac{A_K}{T^{1/4}\sqrt{v_T}} \le \frac{A_K}{2T^{1/4}\sqrt{z_T}}.\end{aligned}$$

Using $m_T \le T$ in the definition of r_T^{ps} now gives

$$\begin{aligned}r_T^{\mathrm{ps}} &\le 2\mathcal{L}_T a_T + 2B_T\sqrt{\frac{2v_T}{T}} \\ &\le L_* A_K \frac{\sqrt{z_T}}{T^{1/4}} + 2c_B\sqrt{2\Lambda_K}\frac{z_T^{3/2}}{\sqrt{T}}, \\ B_T r_T^{\mathrm{tr}} &= c_B K z_T\sqrt{\frac{2v_T}{T}} \le c_B K\sqrt{2\Lambda_K}\frac{z_T^{3/2}}{\sqrt{T}}.\end{aligned}$$

All coefficients in these bounds are fixed. Finally, $\log z_T / \log T \to 0$, so

$$\begin{aligned}\log\left(\frac{\sqrt{z_T}}{T^{1/4}}\right) &= \tfrac{1}{2}\log z_T - \tfrac{1}{4}\log T \longrightarrow -\infty, \\ \log\left(\frac{z_T^{3/2}}{\sqrt{T}}\right) &= \tfrac{3}{2}\log z_T - \tfrac{1}{2}\log T \longrightarrow -\infty.\end{aligned}$$

Both ratios therefore tend to zero, proving $r_T^{\mathrm{ps}} \to 0$ and $B_T r_T^{\mathrm{tr}} \to 0$. No convergence of the empirical allocation is assumed. □

B.1 Proof of the recovery lemma (Lemma 1.1)

We derive a finite recovery threshold directly from the concentration inequalities. Fix $\theta^* \in \Theta$ and $\alpha \in (0, 1/2)$, and set

$$K := |\mathcal{I}|, \qquad v_T := \log(4K(T+1)^4), \qquad \Delta_* := \min\{p^* - 1/2, 1/2 - q^*\},$$
$$\eta_* := \min\{\Delta_*, q^*, 1 - p^*\} > 0.$$

Define the deterministic threshold

$$n_{\mathrm{rec}}(T) := \left\lceil \max\left\{1, (8K^2 v_T)^{1/(1-2\alpha)}, \left(\frac{4Kv_T}{\Delta_*^2}\right)^{1/(1-\alpha)}, \frac{1}{4\eta_*}\right\}\right\rceil.$$

The threshold is used only in the analysis and may depend on the fixed unknown instance. It need not be computed by the learner. We prove that, on an event of probability at least $1 - T^{-3}$, for every $n_{\mathrm{rec}}(T) < t \le T$,

$$\widehat{\sigma}_t = c^*, \qquad \mathcal{A}_t = \mathcal{A}^*,$$
$$\max_e |\mu_e(\theta_t^+) - \mu_e^*| \le a_{T,t-1}, \qquad a_{T,s} := \sqrt{\frac{Kv_T}{s^{1-\alpha}}} + \frac{2K}{s^{1-\alpha}}.$$

For large enough budgets the range of rounds is nonempty, since $n_{\mathrm{rec}}(T)$ grows only as a power of $\log(T+1)$.

Proof of the recovery lemma (Lemma 1.1). All quantities below refer to the budget-T experiment.

Step 1: simultaneous concentration. Since the queried pair is sampled from $\widetilde{\lambda}_t$,

$$\mathbb{E}[\mathbf{1}\{I_t = e\} \mid \mathcal{H}_{t-1}] = \widetilde{\lambda}_{t,e}.$$

Thus $M_s(e) := N_s(e) - \sum_{t=1}^{s} \widetilde{\lambda}_{t,e}$ is a martingale with increments bounded in absolute value by one. Azuma's inequality gives, for every $e \in \mathcal{I}$ and $s \le T$,

$$\mathbb{P}_{\theta^*}\left(|M_s(e)| > \sqrt{2sv_T}\right) \le 2e^{-v_T}.$$

Taking a union bound over e and s, with probability at least $1 - 2KTe^{-v_T}$,

$$\left|N_s(e) - \sum_{t=1}^{s} \widetilde{\lambda}_{t,e}\right| \le \sqrt{2sv_T}, \qquad e \in \mathcal{I}, \quad 1 \le s \le T.$$

To control the responses at adaptive sample counts, realize the observations of each pair e as successive entries of an independent sequence $(Z_{e,m})_{m \ge 1}$ of $\mathrm{Bernoulli}(\mu_e^*)$ variables. Such a realization has the same data-generating law as the observation model. For each fixed $m \ge 1$, Hoeffding's inequality gives

$$\mathbb{P}_{\theta^*}\left(\left|\frac{1}{m}\sum_{j=1}^{m} Z_{e,j} - \mu_e^*\right| > \sqrt{\frac{v_T}{2m}}\right) \le 2e^{-v_T}.$$

A union bound over $e \in \mathcal{I}$ and $1 \le m \le T$ applies simultaneously to every realized count $N_s(e)$. Consequently, with probability at least $1 - 2KTe^{-v_T}$,

$$|\widehat{\mu}_{s,e} - \mu_e^*| \le \sqrt{\frac{v_T}{2N_s(e)}}, \qquad s \le T, \quad N_s(e) > 0.$$

In particular, no conditioning on a realized adaptive sample count is needed to apply Hoeffding's inequality.

Let $\mathcal{G}_T$ be the event on which the count and response concentration bounds hold simultaneously for all pairs and times. Then

$$\mathbb{P}_{\theta^*}(\mathcal{G}_T^c) \le 4KTe^{-v_T} = \frac{T}{(T+1)^4} \le T^{-3}.$$

We work on $\mathcal{G}_T$ for the remainder of the proof.

Step 2: forced exploration and exact recovery. The exploration mixture gives $\widetilde{\lambda}_{t,e} \geq t^{-\alpha}/K$. Hence

$$\sum_{t=1}^{s} \widetilde{\lambda}_{t,e} \geq \frac{1}{K} \sum_{t=1}^{s} t^{-\alpha} \geq \frac{s^{1-\alpha}}{K}.$$

The count concentration bound gives

$$N_s(e) \geq \frac{s^{1-\alpha}}{K} - \sqrt{2sv_T}.$$

Whenever $s \geq n_{\mathrm{rec}}(T)$, the second term is at most half the first. Indeed, the definition of $n_{\mathrm{rec}}(T)$ gives

$$s^{1-2\alpha} \geq 8K^2 v_T \quad \Longrightarrow \quad \sqrt{2sv_T} \leq \frac{s^{1-\alpha}}{2K}.$$

Therefore,

$$N_s(e) \geq \frac{s^{1-\alpha}}{2K}, \qquad e \in \mathcal{I}, \quad n_{\mathrm{rec}}(T) \leq s \leq T.$$

Combining this count lower bound with response concentration gives

$$\max_e |\widehat{\mu}_{s,e} - \mu_e^*| \leq b_{T,s} := \sqrt{\frac{Kv_T}{s^{1-\alpha}}} \leq \frac{\Delta_*}{2}, \qquad n_{\mathrm{rec}}(T) \leq s \leq T.$$

The last inequality follows from $s^{1-\alpha} \geq 4Kv_T/\Delta_*^2$, which follows from the definition of $n_{\mathrm{rec}}(T)$. Hence thresholding the empirical means gives the true pairwise relation:

$$\mathbf{1}\{\widehat{\mu}_{s,e} \geq 1/2\} = \mathbf{1}\{c^*(i) = c^*(j)\}, \qquad e = (i,j), \quad n_{\mathrm{rec}}(T) \leq s \leq T.$$

The canonical representative of c^* has zero disagreements with this relation. Any different canonical partition disagrees on at least one pair and therefore has strictly positive projection loss. Thus the projection has the unique minimizer c^*. Taking $s = t-1$ shows that

$$\widehat{\sigma}_t = c^*, \qquad \mathcal{A}_t = \mathcal{A}(c^*) = \mathcal{A}^*, \qquad n_{\mathrm{rec}}(T) < t \leq T.$$

Step 3: the empirical reference means. Fix $n_{\mathrm{rec}}(T) \leq s \leq T-1$ and a nonempty true pair group $A = \mathcal{I}^r(c^*)$, where $r \in \{=, \neq\}$. Its common true mean is $x^* = p^*$ if r is $=$ and $x^* = q^*$ otherwise. Write $n = N_s(A)$ and $k = S_s(A)$. All pairs in A have positive counts since $N_s(e) \geq s^{1-\alpha}/(2K)$. Therefore,

$$\begin{aligned}\left|\frac{k}{n} - x^*\right| &= \left|\sum_{e \in A} \frac{N_s(e)}{n} (\widehat{\mu}_{s,e} - x^*)\right| \\ &\leq \sum_{e \in A} \frac{N_s(e)}{n} |\widehat{\mu}_{s,e} - x^*| \leq b_{T,s}.\end{aligned}$$

The pseudocount correction satisfies

$$\left|\frac{k+1}{n+2} - \frac{k}{n}\right| = \frac{|n-2k|}{n(n+2)} \leq \frac{1}{n+2}.$$

The algorithm clips this smoothed estimate to $[1/2 + \epsilon_s, 1-\epsilon_s]$ for p and to $[\epsilon_s, 1/2 - \epsilon_s]$ for q, where $\epsilon_s = 1/[4(s+1)]$. Since $s \geq n_{\mathrm{rec}}(T) \geq 1/(4\eta_*)$, we have $\epsilon_s \leq \eta_*$. Thus each active true parameter lies in its clipping interval. Projection onto an interval containing x^* cannot increase the distance to x^*. The clipped estimate $\widehat{x}_s$ consequently satisfies

$$\begin{aligned}|\widehat{x}_s - x^*| &\leq b_{T,s} + \frac{1}{n+2} \\ &\leq b_{T,s} + \frac{2K}{s^{1-\alpha}} = a_{T,s}.\end{aligned}$$

Here the second inequality uses that A contains at least one pair and hence $n \geq s^{1-\alpha}/(2K)$.

Taking $s = t-1$ and using $\widehat{\sigma}_t = c^*$ yields $\max_e |\mu_e(\theta_t^+) - \mu_e^*| \le a_{T,t-1}$ for every $n_{\mathrm{rec}}(T) < t \le T$. If one true pair group is empty, its parameter is unused in every pair mean and needs no estimation bound. We may take $\mathcal{E}_T^{\mathrm{rec}} = \mathcal{G}_T$, with $\mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{rec}})^c) \le T^{-3}$.

For any deterministic $s_T \ge n_{\mathrm{rec}}(T)$, the bound $a_{T,s}$ is decreasing in s, so the uniform version of the lemma holds with

$$a_T = a_{T,s_T} \le C\sqrt{\frac{\log(T+1)}{s_T^{1-\alpha}}}.$$

This recovers the form used in the proof sketch without imposing a power-law parametrization on s_T. □

B.2 An explicit threshold for the combined good events (Lemma 1)

The min-learner's tempered posterior needs a further growth condition beyond exact recovery. This condition can also be satisfied by an explicit threshold. For $\kappa_T = \sqrt{T}$, set

$$s_T := \max\left\{n_{\mathrm{rec}}(T),\ \left\lceil (\kappa_T v_T^2)^{1/(1-\alpha)} \right\rceil\right\}.$$

This is a deterministic cutoff used only in the proof. For all sufficiently large T, it is smaller than T. Below we verify both posterior localization and the negligibility of the discarded prefix.

Conditional posterior localization. Choose a fixed $0 < \epsilon < \min\{q^*, 1-p^*\}/16$ and set

$$S_\epsilon := \{\theta \in \Theta_{\mathrm{nb}} : \epsilon \le \mu_e(\theta) \le 1-\epsilon \text{ for every } e \in \mathcal{I}\}.$$

Only parameters used by a candidate partition are restricted in this set. Let Q_t be the conditional distribution of θ_t^- given $\mathcal{H}_{t-1}$. We show that, on $\mathcal{G}_T$,

$$Q_t(S_\epsilon^c) \le \zeta_T := C(T+1)\exp\left\{-c\frac{s_T^{1-\alpha}}{\kappa_T}\right\}, \qquad t > s_T,$$

for constants $c, C > 0$ independent of T.

Fix a candidate partition and a nonempty pooled group at time $s = t-1 \ge s_T$, with count n and empirical mean r. Since $\max_e |\widehat{\mu}_{s,e} - \mu_e^*| \le b_{T,s}$ and $b_{T,s_T} \to 0$, every empirical pair mean lies in $[4\epsilon, 1-4\epsilon]$ for sufficiently large T. The pooled mean is a convex combination of these means, so $r \in [4\epsilon, 1-4\epsilon]$ as well. Its tempered conditional law on $J \in \{(1/2,1),(0,1/2)\}$ is

$$Q_{s,J}(dx) = \frac{\exp\{(n/\kappa_T)\phi_r(x)\}\,\nu_J(dx)}{\int_J \exp\{(n/\kappa_T)\phi_r(u)\}\,\nu_J(du)}, \qquad \phi_r(x) := r\log x + (1-r)\log(1-x),$$

where ν_J is the fixed truncated Beta prior. The algorithm's uniform prior is the special case with both shape parameters equal to one.

The constrained maximizer in the closure of J is $x_J = \max\{r, 1/2\}$ for $J = (1/2,1)$ and $x_J = \min\{r, 1/2\}$ for $J = (0,1/2)$. In either case, $x_J \in [4\epsilon, 1-4\epsilon]$. Also,

$$\phi_r''(x) = -\frac{r}{x^2} - \frac{1-r}{(1-x)^2} \le -1, \qquad \phi_r'(x_J)(x - x_J) \le 0 \quad (x \in J).$$

Taylor's inequality therefore gives

$$\phi_r(x) \le \phi_r(x_J) - \frac{1}{2}(x - x_J)^2.$$

If $x \in J \setminus [\epsilon, 1-\epsilon]$, then $|x - x_J| \ge 3\epsilon$, so

$$\phi_r(x) \le \phi_r(x_J) - \frac{9}{2}\epsilon^2.$$

To lower-bound the posterior normalizer, take the interval of width $1/(T+1)$ immediately to the right of x_J when $J = (1/2,1)$, and immediately to its left when $J = (0,1/2)$. For sufficiently large T, this interval lies in $J \cap [2\epsilon, 1-2\epsilon]$. On this compact range the fixed truncated Beta density

is bounded below by a positive constant, and $|\phi_r'|$ is bounded uniformly over r. Since $n/\kappa_T \le T$, integration over this interval gives

$$\int_J e^{(n/\kappa_T)\phi_r(x)}\,\nu_J(dx) \ge \frac{c_1}{T+1} e^{(n/\kappa_T)\phi_r(x_J)-c_2},$$

with $c_1, c_2 > 0$ independent of T, n, r. The numerator over $J \setminus [\epsilon, 1-\epsilon]$ is at most $\exp\{(n/\kappa_T)[\phi_r(x_J) - 9\epsilon^2/2]\}$, because the log-likelihood gap is at least $9\epsilon^2/2$ and $\nu_J(J) = 1$. Dividing by the normalizer lower bound yields

$$Q_{s,J}\big(J \setminus [\epsilon, 1-\epsilon]\big) \le C(T+1)\exp\{-cn/\kappa_T\}.$$

Every nonempty group contains a pair with count at least $s_T^{1-\alpha}/(2K)$, so $n \ge s_T^{1-\alpha}/(2K)$. A union bound over the at most two active parameters, followed by averaging over the candidate partition, gives

$$Q_t(S_\epsilon^c) \le C(T+1)\exp\left\{-c\frac{s_T^{1-\alpha}}{\kappa_T}\right\} = \zeta_T, \qquad t > s_T.$$

Recovery ensures that these candidates lie in Θ_{nb}. No tail control is needed for unused parameters.

Verification of the explicit threshold. The definition of s_T gives

$$\frac{s_T^{1-\alpha}}{\kappa_T} \ge v_T^2.$$

Substituting this inequality into the posterior tail bound yields

$$Q_t(S_\epsilon^c) \le C(T+1)e^{-cv_T^2}, \qquad t > s_T.$$

Since $v_T = \log(4K(T+1)^4)$, the right-hand side decays faster than any inverse power of T. In particular, for every fixed $m > 0$ it is at most T^{-m} for all sufficiently large T.

The definition of the recovery threshold gives

$$n_{\mathrm{rec}}(T) = O\Big(v_T^{1/(1-2\alpha)}\Big),$$

where the constant may depend on the fixed instance. Therefore,

$$s_T = O\Big(v_T^{1/(1-2\alpha)} + T^{1/[2(1-\alpha)]} v_T^{2/(1-\alpha)}\Big).$$

The polynomial exponent $1/[2(1-\alpha)]$ is less than one because $\alpha < 1/2$. Since $B_T = c_B \log(T+1)$, it follows that

$$\frac{B_T s_T}{T} \longrightarrow 0.$$

The uniform reference error also satisfies

$$a_T \le C\sqrt{\frac{\log(T+1)}{s_T^{1-\alpha}}} \le \frac{C}{\sqrt{\kappa_T v_T}}, \qquad (1+B_T)a_T \longrightarrow 0.$$

Thus the explicit choice of s_T supplies the cutoff required by the combined good-event argument directly. The remaining term $B_T\sqrt{\log(T+1)/T}$ also vanishes. No additional power exponent needs to be chosen.

The finite recovery threshold $n_{\mathrm{rec}}(T)$ guarantees recovery on $\mathcal{G}_T$, rather than on every possible realization. Its dependence on T comes from the prescribed failure probability T^{-3}; it is not asserted to be the smallest possible recovery threshold.

B.2.1 Marginal likelihood (Lemma 1.4)

Marginal likelihood (Lemma 1.4). Fix $\theta^* = (c^*, p^*, q^*)$, with $0 < q^* < 1/2 < p^* < 1$. Let ϱ_0 be a probability mass function that is positive on every canonical partition. The parameter priors are independent truncated Beta distributions with fixed positive shape parameters (a_p, b_p, a_q, b_q). All quantities below use the ordinary, uninflated likelihood. In fact, this proof applies to every budget $T \ge 1$ under any nonanticipating sampling rule with fresh Bernoulli observation noise.

Constants and the claimed bound. Let $K := |\mathcal{I}|$ and define

$$v_T := \log(4K(T+1)^4), \qquad \varrho_{\min} := \min_{c \in \mathcal{C}_N^\circ} \varrho_0(c) > 0,$$

$$\tau_p := (a_p - 1)_+ + (b_p - 1)_+, \qquad \tau_q := (a_q - 1)_+ + (b_q - 1)_+, \qquad \tau := \tau_p + \tau_q,$$

where $x_+ := \max\{x, 0\}$. For $a, b > 0$, set

$$Z_+(a,b) := \int_{1/2}^1 x^{a-1}(1-x)^{b-1}\,dx,$$

$$Z_-(a,b) := \int_0^{1/2} x^{a-1}(1-x)^{b-1}\,dx, \qquad Z_0 := Z_+(a_p,b_p)Z_-(a_q,b_q).$$

These integrals are finite and positive. Define the fixed constants

$$L_{\text{span}} := \log \frac{p^*(1-q^*)}{q^*(1-p^*)},$$

$$A_{\text{prior}} := (\tau+3)\log 4 + \log \frac{1}{\varrho_{\min}} + |\log Z_0|.$$

For $\widehat{\lambda}_T(e) := N_T(e)/T$, write

$$J_c(\widehat{\lambda}_T) := \frac{1}{T} \inf_{p \in \mathcal{P},\, q \in \mathcal{Q}} \sum_{e \in \mathcal{I}} N_T(e) d(\mu_e^*, \mu_e(c,p,q)).$$

We prove that an event $\mathcal{E}_T^{\text{marg}}$ with $\mathbb{P}_{\theta^*}(\mathcal{E}_T^{\text{marg}}) \geq 1 - (T+1)^{-3}$ satisfies

$$\max_{c \in \mathcal{C}_N^\circ} \left| \frac{1}{T} \log \frac{w_T^{(1)}(c)}{L_T(\theta^*)} + J_c(\widehat{\lambda}_T) \right| \leq r_T^\ell,$$

where every constant in the remainder is explicit:

$$r_T^\ell := L_{\text{span}} \sqrt{\frac{K v_T}{2T}} + \frac{K v_T}{2q^*(1-p^*)T} + \frac{(\tau+4)\log(T+1) + A_{\text{prior}}}{T}.$$

No lower bound on the realized number of queries to a pair or a pooled group is required.

1. The marginal weights and their normalizers. For $\mathcal{R} \in \{\mathcal{P}, \mathcal{Q}\}$ and integers $0 \leq k \leq n$, define

$$f_{n,k}(x) := x^k(1-x)^{n-k},$$

$$U_{n,k}^{\mathcal{R}} := \int_{\mathcal{R}} f_{n,k}(x)\,dx, \qquad H_{n,k}^{\mathcal{R}} := \sup_{x \in \mathcal{R}} f_{n,k}(x),$$

$$V_{n,k}^{\mathcal{R}}(a,b) := \int_{\mathcal{R}} f_{n,k}(x) x^{a-1}(1-x)^{b-1}\,dx.$$

For $n = 0$, necessarily $k = 0$ and $f_{0,0} \equiv 1$. For a candidate partition c, let n_p, k_p be its within-cluster query and success counts, and let n_q, k_q be its between-cluster counts. Then $n_p + n_q = T$, and the algorithm's weight is

$$w_T^{(1)}(c) = \varrho_0(c) V_{n_p,k_p}^{\mathcal{P}}(a_p,b_p) V_{n_q,k_q}^{\mathcal{Q}}(a_q,b_q).$$

Indeed, these two integrals equal $Z_+(a_p + k_p, b_p + n_p - k_p)$ and $Z_-(a_q + k_q, b_q + n_q - k_q)$, respectively. If the prior densities are normalized, the corresponding joint marginal likelihood is

$$\varrho_0(c) \int L_T(c,p,q)\,\nu_p(dp)\nu_q(dq) = \frac{w_T^{(1)}(c)}{Z_0}.$$

The common factor Z_0 cancels in posterior probabilities. We retain it below because the claimed bound compares the weight to $L_T(\theta^*)$. Since p and q range independently over their half-intervals,

$$H_T(c) := \sup_{p \in \mathcal{P},\, q \in \mathcal{Q}} L_T(c,p,q) = H_{n_p,k_p}^{\mathcal{P}} H_{n_q,k_q}^{\mathcal{Q}}.$$

The product identity holds even when a supremum is approached only at an endpoint: choose an approximating sequence for each factor.

2. A uniform integral bound from the Beta sandwich. We first prove, for every $n \geq 0$ and $0 \leq k \leq n$,

$$\frac{H^{\mathcal{R}}_{n,k}}{2(n+1)(n+2)} \leq U^{\mathcal{R}}_{n,k} \leq \frac{1}{2} H^{\mathcal{R}}_{n,k}.$$

The upper bound follows by integrating the pointwise bound $f_{n,k} \leq H^{\mathcal{R}}_{n,k}$ over an interval of length $1/2$. For $n = 0$, $U^{\mathcal{R}}_{0,0} = 1/2$ and $H^{\mathcal{R}}_{0,0} = 1$, so the lower bound also holds.

Now take $n \geq 1$ and $\mathcal{R} = \mathcal{P}$. Write $r = k/n$ and $h(r) = -r \log r - (1-r)\log(1-r)$, with $0 \log 0 = 0$. The Beta-function identity gives

$$\mathrm{B}(k+1, n-k+1) = \frac{k!(n-k)!}{(n+1)!} = \frac{1}{(n+1)\binom{n}{k}}.$$

For $0 < r < 1$, the binomial probability at k is at most one, so

$$\binom{n}{k} r^k (1-r)^{n-k} \leq 1 \quad \Longrightarrow \quad \binom{n}{k} \leq e^{nh(r)}.$$

The same last inequality holds for $k = 0, n$, when both sides equal one. Consequently,

$$\mathrm{B}(k+1, n-k+1) \geq \frac{e^{-nh(r)}}{n+1}.$$

If $X \sim \mathrm{Beta}(k+1, n-k+1)$, its density gives

$$U^{\mathcal{P}}_{n,k} = \mathrm{B}(k+1, n-k+1)\mathbb{P}(X \geq 1/2).$$

Suppose first that $r \geq 1/2$. The Beta–binomial identity yields

$$\mathbb{P}(X \geq 1/2) = \mathbb{P}(\mathrm{Binomial}(n+1, 1/2) \leq k) \geq \frac{1}{2}.$$

To verify the inequality, $k \geq \lceil n/2 \rceil = \lfloor (n+1)/2 \rfloor$, and a symmetric binomial distribution assigns at least half its mass at or below $\lfloor (n+1)/2 \rfloor$. At this threshold, an odd number of trials gives mass exactly one half; an even number gives one half plus half the central point mass. The likelihood is maximized at r in the closure of $\mathcal{P}$, so $H^{\mathcal{P}}_{n,k} = e^{-nh(r)}$. Therefore,

$$U^{\mathcal{P}}_{n,k} \geq \frac{H^{\mathcal{P}}_{n,k}}{2(n+1)} \geq \frac{H^{\mathcal{P}}_{n,k}}{2(n+1)(n+2)}.$$

Suppose next that $r < 1/2$. The supremum is approached at $1/2$, giving $H^{\mathcal{P}}_{n,k} = 2^{-n}$. The Beta–Bernoulli right-tail sandwich uses the center $k/(n+1)$, not k/n. Since $k/(n+1) < 1/2$, it gives

$$\mathbb{P}(X \geq 1/2) \geq \frac{1}{n+2} \exp\left\{ -(n+1) d\left(\frac{k}{n+1}, \frac{1}{2} \right) \right\}.$$

Here $d(x, y)$ is extended continuously in x to $0, 1$, so $d(0, y) = -\log(1-y)$ and $d(1, y) = -\log y$. Put $r' = k/(n+1)$. Entropy concavity and $h(0) = 0$ imply

$$h(r') = h\left(\frac{n}{n+1} r + \frac{1}{n+1} 0 \right) \geq \frac{n}{n+1} h(r).$$

Using $d(r', 1/2) = \log 2 - h(r')$, we obtain

$$\begin{aligned} U^{\mathcal{P}}_{n,k} &\geq \frac{\exp\{-nh(r) - (n+1)d(r', 1/2)\}}{(n+1)(n+2)} \\ &= \frac{\exp\{(n+1)h(r') - nh(r) - (n+1)\log 2\}}{(n+1)(n+2)} \\ &\geq \frac{2^{-(n+1)}}{(n+1)(n+2)} = \frac{H^{\mathcal{P}}_{n,k}}{2(n+1)(n+2)}. \end{aligned}$$

Finally, substituting $z = 1 - x$ gives

$$U^{\mathcal{Q}}_{n,k} = U^{\mathcal{P}}_{n,n-k}, \qquad H^{\mathcal{Q}}_{n,k} = H^{\mathcal{P}}_{n,n-k}.$$

This proves the same bounds on $\mathcal{Q}$.

3. Comparison with each fixed Beta prior. Fix $a, b > 0$ and set $\tau_{a,b} := (a-1)_+ + (b-1)_+$. For a given n, define

$$\delta_n := \frac{1}{4(n+1)}, \qquad b_n := 1 - 2\delta_n, \qquad F_n(x) := \delta_n + b_n x.$$

Here $1/2 \le b_n < 1$. For $x \in (0,1)$, $F_n(x) \in [\delta_n, 1-\delta_n]$, and

$$F_n(x) - 1/2 = b_n(x - 1/2).$$

Thus F_n maps each admissible half-interval into itself. Moreover,

$$F_n(x) \ge b_n x, \qquad 1 - F_n(x) = \delta_n + b_n(1-x) \ge b_n(1-x),$$

which gives $f_{n,k}(F_n(x)) \ge b_n^n f_{n,k}(x)$.

For $y \in [\delta_n, 1-\delta_n]$, a nonnegative exponent u gives $y^u \ge \delta_n^u$, whereas a negative exponent gives $y^u \ge 1$. Applying the same argument to $1-y$ yields

$$y^{a-1}(1-y)^{b-1} \ge \delta_n^{\tau_{a,b}}.$$

Integrating over the image of F_n and then changing variables gives

$$\begin{aligned} V_{n,k}^{\mathcal{R}}(a,b) &\ge \int_{F_n(\mathcal{R})} f_{n,k}(y) y^{a-1}(1-y)^{b-1}\, dy \\ &\ge \delta_n^{\tau_{a,b}} b_n \int_{\mathcal{R}} f_{n,k}(F_n(x))\, dx \\ &\ge \delta_n^{\tau_{a,b}} b_n^{n+1} U_{n,k}^{\mathcal{R}}. \end{aligned}$$

Bernoulli's inequality $(1-u)^m \ge 1 - mu$ for integers $m \ge 1$ and $u \in [0,1]$ gives

$$b_n^{n+1} = \left(1 - \frac{1}{2(n+1)}\right)^{n+1} \ge \frac{1}{2}.$$

Combining the preceding bounds with the uniform integral sandwich, and using $n+2 \le 2(n+1)$, yields

$$\begin{aligned} V_{n,k}^{\mathcal{R}}(a,b) &\ge \frac{\delta_n^{\tau_{a,b}}}{2} U_{n,k}^{\mathcal{R}} \\ &\ge \frac{H_{n,k}^{\mathcal{R}}}{4^{\tau_{a,b}+1}(n+1)^{\tau_{a,b}+1}(n+2)} \\ &\ge \frac{H_{n,k}^{\mathcal{R}}}{2 \cdot 4^{\tau_{a,b}+1}(n+1)^{\tau_{a,b}+2}}. \end{aligned}$$

The upper bound follows by replacing $f_{n,k}$ by its supremum:

$$V_{n,k}^{\mathcal{R}}(a,b) \le H_{n,k}^{\mathcal{R}} \int_{\mathcal{R}} x^{a-1}(1-x)^{b-1}\, dx.$$

These inequalities include $n = 0$, for which the integral is simply the prior normalizer. No sample-size threshold is needed.

Applying them to both groups of c, and using $n_p + 1, n_q + 1 \le T + 1$ and $\varrho_{\min} \le \varrho_0(c) \le 1$, gives

$$\begin{aligned} \frac{w_T^{(1)}(c)}{H_T(c)} &\ge \frac{\varrho_0(c)}{4^{\tau+3}(n_p+1)^{\tau_p+2}(n_q+1)^{\tau_q+2}} \\ &\ge \frac{\varrho_{\min}}{4^{\tau+3}(T+1)^{\tau+4}}, \\ \frac{w_T^{(1)}(c)}{H_T(c)} &\le \varrho_0(c) Z_0 \le Z_0. \end{aligned}$$

Taking logarithms bounds the log ratio below by $-(\tau+4)\log(T+1) - (\tau+3)\log 4 - \log(1/\varrho_{\min})$ and above by $\log Z_0$. Therefore, for every history and every c,

$$|\log w_T^{(1)}(c) - \log H_T(c)| \le (\tau+4)\log(T+1) + A_{\text{prior}}.$$

4. A simultaneous event for the pairwise residuals. Set $E_e := S_T(e) - N_T(e)\mu_e^*$, and define

$$\mathcal{E}_T^{\mathrm{marg}} := \bigcap_{e\in\mathcal{I}} \left\{ |E_e| \le \sqrt{\frac{N_T(e)v_T}{2}} \right\}.$$

When $N_T(e) = 0$, the residual is zero, so its inequality holds automatically. For positive counts, realize each pair's observations as an independent sequence $(Y_{e,j})_{j\ge1}$ of Bernoulli variables with mean μ_e^*, independent of the learner's random seeds. Querying e reveals the next unused entry. This construction has the stipulated conditional observation law and gives $S_T(e) = \sum_{j=1}^{N_T(e)} Y_{e,j}$. For each fixed prefix length $n \in \{1,\dots,T\}$, Hoeffding's inequality gives

$$\mathbb{P}_{\theta^*}\left(\left| \sum_{j=1}^{n} (Y_{e,j} - \mu_e^*) \right| > \sqrt{\frac{nv_T}{2}} \right) \le 2\exp\left(-\frac{2(nv_T/2)}{n} \right) = 2e^{-v_T}.$$

If the residual bound fails at the random count $N_T(e)$, it fails at one of these deterministic prefix lengths. Taking a union over all K pairs and T lengths therefore gives

$$\mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{marg}})^c) \le 2KTe^{-v_T} = \frac{T}{2(T+1)^4} \le \frac{1}{(T+1)^3}.$$

On this event, the triangle and Cauchy–Schwarz inequalities imply

$$\begin{aligned} \sum_e |E_e| &\le \sqrt{\frac{v_T}{2}} \sum_e \sqrt{N_T(e)} \\ &\le \sqrt{\frac{v_T}{2}} \sqrt{K \sum_e N_T(e)} = \sqrt{\frac{KTv_T}{2}}. \end{aligned}$$

More generally, for any group $A \subseteq \mathcal{I}$ with $n_A := \sum_{e\in A} N_T(e) > 0$, let

$$r_A := \frac{\sum_{e\in A} S_T(e)}{n_A}, \qquad m_A := \frac{\sum_{e\in A} N_T(e)\mu_e^*}{n_A}.$$

Then $m_A \in [q^*, p^*]$ and

$$\begin{aligned} n_A|r_A - m_A| &= \left| \sum_{e\in A} E_e \right| \\ &\le \sqrt{\frac{v_T}{2}} \sum_{e\in A} \sqrt{N_T(e)} \le \sqrt{\frac{|A|n_A v_T}{2}}. \end{aligned}$$

Squaring and dividing by n_A gives $n_A(r_A - m_A)^2 \le |A|v_T/2$. These bounds hold for every group on the same event.

5. A deterministic bound for the profiled log likelihood. For $u \in [0,1]$, define

$$\psi_{\mathcal{R}}(u) := \sup_{x\in\mathcal{R}} \{u\log x + (1-u)\log(1-x)\}.$$

The supremum is understood by continuity at an endpoint. Direct maximization gives

$$\psi_{\mathcal{P}}(u) = \begin{cases} -\log 2, & u \le 1/2, \\ -h(u), & u \ge 1/2, \end{cases} \qquad \psi_{\mathcal{Q}}(u) = \psi_{\mathcal{P}}(1-u).$$

Indeed, the derivative with respect to x is $(u-x)/(x(1-x))$, which changes sign at u whenever u is admissible; otherwise the supremum is approached at $1/2$. For $m \in (0,1)$, set

$$x_{\mathcal{P}}(m) := \max\{m, 1/2\}, \qquad x_{\mathcal{Q}}(m) := \min\{m, 1/2\}, \qquad \mathrm{logit}(x) := \log\frac{x}{1-x}.$$

We establish the scalar remainder bound

$$\begin{aligned} 0 \le b_{\mathcal{R}}(r,m) &:= \psi_{\mathcal{R}}(r) - \psi_{\mathcal{R}}(m) - (r-m)\,\mathrm{logit}(x_{\mathcal{R}}(m)) \\ &\le d(r,m) \le \frac{(r-m)^2}{m(1-m)}, \qquad r \in [0,1], \quad m \in (0,1). \end{aligned}$$

First, evaluating the supremum defining $\psi_{\mathcal{R}}(r)$ at $x_{\mathcal{R}}(m)$ gives

$$\begin{aligned}\psi_{\mathcal{R}}(r) &\geq r \log x_{\mathcal{R}}(m) + (1-r)\log(1-x_{\mathcal{R}}(m)) \\ &= \psi_{\mathcal{R}}(m) + (r-m)\operatorname{logit}(x_{\mathcal{R}}(m)).\end{aligned}$$

If $x_{\mathcal{R}}(m) = 1/2$, take a sequence in $\mathcal{R}$ converging to $1/2$ to obtain the same inequality. This proves $b_{\mathcal{R}}(r,m) \geq 0$.

For the upper bound, first consider $\mathcal{R} = \mathcal{P}$. If $m \geq 1/2$, then $x_{\mathcal{P}}(m) = m$ and $\psi_{\mathcal{P}}(m) = -h(m)$. Since the constrained supremum is at most the unconstrained one, $\psi_{\mathcal{P}}(r) \leq -h(r)$. It follows that

$$\begin{aligned}b_{\mathcal{P}}(r,m) &\leq -h(r) + h(m) - (r-m)\log\frac{m}{1-m} \\ &= -h(r) - r\log m - (1-r)\log(1-m) = d(r,m).\end{aligned}$$

If $m < 1/2$ and $r \leq 1/2$, both profile values are $-\log 2$, and $\operatorname{logit}(x_{\mathcal{P}}(m)) = 0$, so the remainder is zero. If $m < 1/2 < r$, the remainder equals $\log 2 - h(r) = d(r, 1/2)$. For $y \in [m, 1/2]$,

$$\partial_y d(r,y) = \frac{y-r}{y(1-y)} < 0,$$

so $d(r,1/2) \leq d(r,m)$, proving this last case as well.

For $\mathcal{Q}$, the identities

$$\begin{aligned}\psi_{\mathcal{Q}}(u) = \psi_{\mathcal{P}}(1-u), \qquad & x_{\mathcal{Q}}(m) = 1 - x_{\mathcal{P}}(1-m), \\ \operatorname{logit}(1-x) = -\operatorname{logit}(x), \qquad & d(1-r,1-m) = d(r,m)\end{aligned}$$

give $b_{\mathcal{Q}}(r,m) = b_{\mathcal{P}}(1-r,1-m) \leq d(r,m)$.

Finally, for $0 < r < 1$, applying $\log z \leq z - 1$ to both terms in $d(r,m)$ yields

$$\begin{aligned}d(r,m) &\leq r\left(\frac{r}{m} - 1\right) + (1-r)\left(\frac{1-r}{1-m} - 1\right) \\ &= \frac{r^2}{m} + \frac{(1-r)^2}{1-m} - 1 = \frac{(r-m)^2}{m(1-m)}.\end{aligned}$$

Continuity in r gives the same inequalities at $r = 0, 1$.

6. From the profile likelihood to the information value. Fix a canonical partition c. Its two pair groups are $A_p := \mathcal{I}^{=}(c)$ and $A_q := \mathcal{I}^{\neq}(c)$, with domains $\mathcal{R}_p := \mathcal{P}$ and $\mathcal{R}_q := \mathcal{Q}$. For every group A with $n_A > 0$, use the empirical mean r_A and true weighted mean m_A defined above. Groups with $n_A = 0$ contribute zero to all log likelihoods and divergences and are omitted from the following sums. The empirical profile is

$$\log H_T(c) = \sum_{A:\, n_A > 0} n_A \psi_{\mathcal{R}_A}(r_A).$$

Define the true-mean log likelihood with the realized counts by

$$\begin{aligned}\overline{\ell}_T(c,p,q) &:= \sum_e N_T(e)\big[\mu_e^* \log \mu_e(c,p,q) + (1-\mu_e^*)\log(1-\mu_e(c,p,q))\big], \\ C_T^* &:= \sum_e N_T(e)\big[\mu_e^* \log \mu_e^* + (1-\mu_e^*)\log(1-\mu_e^*)\big].\end{aligned}$$

These are functions of the realized counts; no conditional expectation given those adaptive counts is asserted. Expanding the definition of KL gives the exact identity

$$\overline{\ell}_T(c,p,q) = C_T^* - \sum_e N_T(e) d(\mu_e^*, \mu_e(c,p,q)).$$

Taking a supremum and using the separate maximization over the two response parameters yields

$$\sum_{A:\, n_A > 0} n_A \psi_{\mathcal{R}_A}(m_A) = \sup_{p,q} \overline{\ell}_T(c,p,q) = C_T^* - T J_c(\widehat{\lambda}_T).$$

The true likelihood also satisfies

$$\begin{aligned}\log L_T(\theta^*) &= \sum_e \{S_T(e)\log\mu_e^* + [N_T(e) - S_T(e)]\log(1-\mu_e^*)\} \\ &= C_T^* + \sum_e E_e \operatorname{logit}(\mu_e^*).\end{aligned}$$

For each nonempty-count group, the scalar profile identity reads

$$\begin{aligned}n_A\psi_{\mathcal{R}_A}(r_A) = n_A\psi_{\mathcal{R}_A}(m_A) &+ \left(\sum_{e\in A} E_e\right)\operatorname{logit}(x_{\mathcal{R}_A}(m_A)) \\ &+ n_A b_{\mathcal{R}_A}(r_A, m_A),\end{aligned}$$

where we used $n_A(r_A - m_A) = \sum_{e\in A} E_e$. Summing this identity and subtracting the true log likelihood gives

$$\begin{aligned}&\log H_T(c) - \log L_T(\theta^*) + TJ_c(\widehat{\lambda}_T) \\ &\quad= \sum_{A:\, n_A>0}\sum_{e\in A} E_e\big[\operatorname{logit}(x_{\mathcal{R}_A}(m_A)) - \operatorname{logit}(\mu_e^*)\big] + R_T(c), \\ &R_T(c) := \sum_{A:\, n_A>0} n_A b_{\mathcal{R}_A}(r_A, m_A) \geq 0.\end{aligned}$$

Pairs in omitted groups have $N_T(e) = S_T(e) = E_e = 0$.

Since $q^* < 1/2 < p^*$ and $m_A \in [q^*, p^*]$, the constrained maximizer $x_{\mathcal{R}_A}(m_A)$ also belongs to $[q^*, p^*]$. The logit function is increasing, so

$$|\operatorname{logit}(x_{\mathcal{R}_A}(m_A)) - \operatorname{logit}(\mu_e^*)| \leq \operatorname{logit}(p^*) - \operatorname{logit}(q^*) = L_{\mathrm{span}}.$$

On $\mathcal{E}_T^{\mathrm{marg}}$, the linear residual is therefore at most $L_{\mathrm{span}}\sqrt{KTv_T/2}$ in absolute value. For the nonnegative remainder, use $m_A(1-m_A) \geq q^*(1-p^*)$ and the group residual bound to obtain

$$\begin{aligned}n_A b_{\mathcal{R}_A}(r_A, m_A) &\leq \frac{n_A(r_A - m_A)^2}{m_A(1-m_A)} \\ &\leq \frac{|A|v_T}{2q^*(1-p^*)}.\end{aligned}$$

The two groups partition $\mathcal{I}$, so the sum of their cardinalities is K. Omitting groups with zero counts can only decrease this sum. Thus

$$0 \leq R_T(c) \leq \frac{Kv_T}{2q^*(1-p^*)}.$$

Combining the linear and nonnegative terms proves, for every c on the same event,

$$|\log H_T(c) - \log L_T(\theta^*) + TJ_c(\widehat{\lambda}_T)| \leq L_{\mathrm{span}}\sqrt{\frac{KTv_T}{2}} + \frac{Kv_T}{2q^*(1-p^*)}.$$

7. Combining the bounds. For every c, the triangle inequality gives

$$\begin{aligned}&|\log w_T^{(1)}(c) - \log L_T(\theta^*) + TJ_c(\widehat{\lambda}_T)| \\ &\quad\leq |\log w_T^{(1)}(c) - \log H_T(c)| \\ &\qquad + |\log H_T(c) - \log L_T(\theta^*) + TJ_c(\widehat{\lambda}_T)| \\ &\quad\leq (\tau+4)\log(T+1) + A_{\mathrm{prior}} + L_{\mathrm{span}}\sqrt{\frac{KTv_T}{2}} + \frac{Kv_T}{2q^*(1-p^*)}.\end{aligned}$$

Dividing by T and taking the maximum over c proves the claimed bound with r_T^ℓ. The same pairwise residual event controls all partitions, so the probability bound already established applies to this maximum.

For an explicit decay bound, set $\Lambda_K := 4 + \log(4K)/\log 2$. Since $\log(T+1) \geq \log 2$,

$$v_T = 4\log(T+1) + \log(4K) \leq \Lambda_K \log(T+1).$$

Substitution in the remainder gives

$$r_T^\ell \leq L_{\text{span}} \sqrt{\frac{K\Lambda_K}{2}} \sqrt{\frac{\log(T+1)}{T}} + \left[\frac{K\Lambda_K}{2q^*(1-p^*)} + \tau + 4\right] \frac{\log(T+1)}{T} + \frac{A_{\text{prior}}}{T} \longrightarrow 0.$$

All coefficients are fixed, and each displayed function of T tends to zero. This establishes the marginal-likelihood statement. □

B.3 A USEFUL REDUCTION LEMMA (LEMMA 3)

The following reduction is due to Teo et al., stated in our notation. Their appendix establishes the pointwise form in the allocation and includes the two boundary cases for the number of clusters.

Lemma 3 (Reduction to merge and split alternatives)**.** *Fix $N \geq 2$ and $\theta^* = (c^*, p^*, q^*) \in \Theta$. Let $\mathcal{A}(c^*)$ be the set of canonical partitions obtained from c^* by either merging two distinct clusters or splitting one cluster into two nonempty clusters, and define*

$$\Theta_{\text{nb}} := \mathcal{A}(c^*) \times \mathcal{P} \times \mathcal{Q}.$$

For every allocation $\lambda \in \Delta(\mathcal{I})$,

$$\inf_{\theta \in \text{Alt}(\theta^*)} D(\theta^*, \theta; \lambda) = \inf_{\theta \in \Theta_{\text{nb}}} D(\theta^*, \theta; \lambda).$$

Consequently,

$$\Gamma^*(\theta^*) = \sup_{\lambda \in \Delta(\mathcal{I})} \min_{c \in \mathcal{A}(c^*)} \inf_{p \in \mathcal{P},\, q \in \mathcal{Q}} D(\theta^*, (c,p,q); \lambda).$$

The conclusion includes the cases of a single true cluster and all singleton true clusters, for which the neighborhood contains only splits and only merges, respectively.

The reduction is a property of the KL information objective and therefore also applies in the fixed-budget setting. Restricting the response parameters to the interior does not change the infima: with interior true means, a candidate mean of zero or one has infinite divergence on every pair with positive allocation weight, while a parameter assigned only to zero-weight pairs does not affect the objective.

Clipped information. For the clipped gains used by the algorithm, define

$$B_* := \max\{d(p^*, q^*), d(q^*, p^*)\}.$$

For every $B \geq B_*$ and every $\lambda \in \Delta(\mathcal{I})$,

$$\inf_{\theta \in \Theta_{\text{nb}}} \sum_{e \in \mathcal{I}} \lambda_e \min\{d(\mu_e^*, \mu_e(\theta)), B\} = \inf_{\theta \in \text{Alt}(\theta^*)} D(\theta^*, \theta; \lambda). \tag{25}$$

Indeed, both parameter infima can be restricted, in their closures, to $[1/2, p^*] \times [q^*, 1/2]$ by replacing $p > p^*$ by p^* and $q < q^*$ by q^*. These replacements cannot increase any pairwise KL divergence or its clipped value. On this rectangle, every pairwise divergence is at most B_*, so clipping leaves the objective unchanged. The lemma then identifies its infimum with the full alternative information value. This equality concerns the minimized objective; sampled gains remain clipped in the algorithm.

B.4 PROOF OF THE POSTERIOR ERROR UPPER BOUND (THEOREM 2)

Proof of the posterior error upper bound (Theorem 2). Fix $\theta^* = (c^*, p^*, q^*) \in \Theta$ and use the algorithm's fixed priors and parameter choices. Represent all partitions canonically. Write

$$\mathcal{A}^* := \mathcal{A}(c^*), \qquad \Theta_{\text{nb}} := \mathcal{A}^* \times \mathcal{P} \times \mathcal{Q},$$
$$J_c(\lambda) := \inf_{p \in \mathcal{P},\, q \in \mathcal{Q}} D(\theta^*, (c,p,q); \lambda), \quad J(\lambda) := \min_{c \nsim c^*} J_c(\lambda).$$

Thus $\Gamma^*(\theta^*) = \sup_{\lambda\in\Delta(\mathcal{I})} J(\lambda)$. The empirical and average sampling allocations are

$$\widehat{\lambda}_T(e) := \frac{N_T(e)}{T}, \qquad \overline{\lambda}_T := \frac{1}{T}\sum_{t=1}^{T} \widetilde{\lambda}_t.$$

For each budget T, let

$$B_T := c_B \log(T+1), \quad h_{T,e}(\theta) := \min\{d(\mu_e^*, \mu_e(\theta)), B_T\},$$
$$G_t := B_T g_t, \qquad \overline{h}_t := \mathbb{E}_{\theta^*}[h_T(\theta_t^-) \mid \mathcal{H}_{t-1}].$$

In particular, G_t and $\overline{h}_t$ belong to $[0, B_T]^{\mathcal{I}}$. All allocation extrema below are over $\Delta(\mathcal{I})$.

1. The joint good event (Lemma 1). Let s_T be the deterministic cutoff used in the four good-event lemmas, and define

$$\mathcal{E}_T := \mathcal{E}_T^{\mathrm{rec}} \cap \mathcal{E}_T^{\mathrm{reg}} \cap \mathcal{E}_T^{\mathrm{samp}} \cap \mathcal{E}_T^{\mathrm{marg}}.$$

Each of the four failure probabilities is at most T^{-3} for all sufficiently large T. A union bound therefore gives

$$\mathbb{P}_{\theta^*}(\mathcal{E}_T^c) \le \mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{rec}})^c) + \mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{reg}})^c) + \mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{samp}})^c) + \mathbb{P}_{\theta^*}((\mathcal{E}_T^{\mathrm{marg}})^c) \le \frac{4}{T^3}.$$

The cutoff and the deterministic remainders established there satisfy

$$\frac{B_T s_T}{T} \longrightarrow 0,$$
$$\frac{R_T^{\max} + R_T^{\min}}{T} + r_T^{\mathrm{ps}} + B_T r_T^{\mathrm{tr}} + r_T^{\ell} \longrightarrow 0.$$

In particular, $s_T < T$ for all sufficiently large T. We derive a pathwise posterior bound on $\mathcal{E}_T$, and then use its summable failure probabilities to obtain the almost-sure statement. Conditional expectations below remain under the original data-generating law; we do not condition that law on $\mathcal{E}_T$.

2. Neighborhood reduction and clipping. The neighborhood reduction of Teo et al. gives, for every allocation λ,

$$J(\lambda) = \inf_{\theta\in\Theta_{\mathrm{nb}}} D(\theta^*, \theta; \lambda).$$

The pointwise statement and the cases of a single true cluster and all singleton true clusters are included in their appendix. This reduction concerns the KL objective and is independent of the sampling or stopping rule. Their parameter space also allows zero and one. Excluding these endpoints leaves the infima unchanged: an endpoint candidate mean has infinite KL against an interior true mean whenever its group has positive allocation weight; a parameter assigned only to zero-weight pairs does not affect the objective.

We next verify the corresponding identity for clipped KL. For $x, y \in (0,1)$,

$$\frac{\partial}{\partial y} d(x,y) = \frac{y-x}{y(1-y)}.$$

Replacing $p > p^*$ by p^* and $q < q^*$ by q^* cannot increase any pairwise divergence, because each true mean belongs to $\{p^*, q^*\}$. Since clipping is nondecreasing, the same comparison holds for clipped divergences. Both parameter infima may therefore be restricted, in their closures, to

$$\mathcal{R}_* := [1/2, p^*] \times [q^*, 1/2].$$

Values at $p = 1/2$ or $q = 1/2$ are limits of admissible values, and continuity ensures that taking this closure preserves each infimum. Define

$$B_* := \max\{d(p^*, q^*), d(q^*, p^*)\}.$$

For $x = p^*$ the divergence decreases on $[q^*, p^*]$, while for $x = q^*$ it increases on that interval. Thus

$$d(x,y) \le B_*, \qquad x \in \{p^*, q^*\}, \quad y \in [q^*, p^*].$$

For $T \geq \lceil \exp(B_*/c_B) \rceil$, we have $B_T \geq B_*$, so clipping is inactive on $\mathcal{R}_*$. Parameter contraction, continuity, and the cited reduction now give

$$\begin{aligned}\inf_{\theta \in \Theta_{\mathrm{nb}}} \langle \lambda, h_T(\theta) \rangle &= \min_{c \in \mathcal{A}^*} \min_{(p,q) \in \mathcal{R}_*} \sum_e \lambda_e \min\{d(\mu_e^*, \mu_e(c,p,q)), B_T\} \\ &= \min_{c \in \mathcal{A}^*} \min_{(p,q) \in \mathcal{R}_*} D(\theta^*, (c,p,q); \lambda) \\ &= J(\lambda).\end{aligned}$$

The minima exist because the neighborhood is finite and $\mathcal{R}_*$ is compact, with all candidate means in $(0,1)$. This identity concerns the comparator infimum; the algorithm's sampled gains remain clipped. Since the neighborhood is nonempty for $N \geq 2$, boundedness of h_T gives $0 \leq J(\lambda) \leq B_T$ and hence $\Gamma^*(\theta^*) \leq B_T$.

3. From learner regret to information accumulation. Fix a sufficiently large T and a path in $\mathcal{E}_T$, and write $s = s_T$. Recovery gives $\mathcal{A}_t = \mathcal{A}^*$ for $s < t \leq T$. Since $\mathcal{A}_t$ is determined by $\mathcal{H}_{t-1}$, the conditional law of θ_t^- at each such history is supported on Θ_{nb}. It follows that, for every allocation λ,

$$\begin{aligned}\langle \lambda, \overline{h}_t \rangle &= \mathbb{E}_{\theta^*}[\langle \lambda, h_T(\theta_t^-) \rangle \mid \mathcal{H}_{t-1}] \\ &\geq \inf_{\theta \in \Theta_{\mathrm{nb}}} \langle \lambda, h_T(\theta) \rangle = J(\lambda).\end{aligned}$$

Summing and taking the supremum over λ gives

$$(T-s)\Gamma^*(\theta^*) \leq \max_\lambda \sum_{t=s+1}^{T} \langle \lambda, \overline{h}_t \rangle.$$

Consequently,

$$T\Gamma^*(\theta^*) \leq \max_\lambda \sum_{t=s+1}^{T} \langle \lambda, \overline{h}_t \rangle + sB_T.$$

Define the two sampling errors

$$\delta_{T,1} := \left\| \frac{1}{T} \sum_{t=s+1}^{T} (G_t - \overline{h}_t) \right\|_\infty,$$

$$\delta_{T,2} := \left| \frac{1}{T} \sum_{t=s+1}^{T} \langle \lambda_t, G_t - \overline{h}_t \rangle \right|.$$

The sampling lemma gives $\delta_{T,1} + \delta_{T,2} \leq r_T^{\mathrm{ps}}$. Since $\|\lambda\|_1 = 1$,

$$\max_\lambda \sum_{t=s+1}^{T} \langle \lambda, \overline{h}_t \rangle \leq \max_\lambda \sum_{t=s+1}^{T} \langle \lambda, G_t \rangle + T\delta_{T,1}.$$

To apply the full-horizon max-learner bound, use the nonnegativity of G_t and then bound the learner's prefix gains by sB_T:

$$\begin{aligned}\max_\lambda \sum_{t=s+1}^{T} \langle \lambda, G_t \rangle &\leq \max_\lambda \sum_{t=1}^{T} \langle \lambda, G_t \rangle \\ &\leq \sum_{t=1}^{T} \langle \lambda_t, G_t \rangle + R_T^{\max} \\ &\leq \sum_{t=s+1}^{T} \langle \lambda_t, G_t \rangle + sB_T + R_T^{\max} \\ &\leq \sum_{t=s+1}^{T} \langle \lambda_t, \overline{h}_t \rangle + T\delta_{T,2} + sB_T + R_T^{\max}.\end{aligned}$$

Combining these inequalities yields

$$T\Gamma^*(\theta^*) \le \sum_{t=s+1}^{T} \langle \lambda_t, \overline{h}_t \rangle + R_T^{\max} + T r_T^{\mathrm{ps}} + 2sB_T.$$

The two prefix terms have distinct sources: replacing $T - s$ by T, and applying full-horizon regret to the remaining rounds.

The exploration mixture satisfies $\widetilde{\lambda}_t = (1-\gamma_t)\lambda_t + \gamma_t u$. Because $\overline{h}_t \in [0, B_T]^{\mathcal{I}}$, both $\langle \lambda_t, \overline{h}_t \rangle$ and $\langle u, \overline{h}_t \rangle$ lie in $[0, B_T]$. Hence

$$\langle \lambda_t, \overline{h}_t \rangle \le \langle \widetilde{\lambda}_t, \overline{h}_t \rangle + \gamma_t B_T.$$

Apply this inequality and then the min-learner bound to obtain

$$T\Gamma^*(\theta^*) \le \inf_{\theta \in \Theta_{\mathrm{nb}}} \sum_{t=s+1}^{T} \langle \widetilde{\lambda}_t, h_T(\theta) \rangle + R_T^{\max} + R_T^{\min} + T r_T^{\mathrm{ps}} + 2sB_T + B_T \sum_{t=1}^{T} \gamma_t.$$

For every fixed θ, adding the prefix increases the sum because $h_T(\theta)$ is nonnegative. Taking infima preserves this inequality, giving

$$\begin{aligned} \inf_{\theta \in \Theta_{\mathrm{nb}}} \sum_{t=s+1}^{T} \langle \widetilde{\lambda}_t, h_T(\theta) \rangle &\le \inf_{\theta \in \Theta_{\mathrm{nb}}} \sum_{t=1}^{T} \langle \widetilde{\lambda}_t, h_T(\theta) \rangle \\ &= T \inf_{\theta \in \Theta_{\mathrm{nb}}} \langle \overline{\lambda}_T, h_T(\theta) \rangle \\ &= T J(\overline{\lambda}_T). \end{aligned}$$

With $\overline{\gamma}_T := T^{-1} \sum_{t=1}^{T} \gamma_t$, we conclude

$$\Gamma^*(\theta^*) \le J(\overline{\lambda}_T) + \frac{R_T^{\max} + R_T^{\min}}{T} + r_T^{\mathrm{ps}} + B_T \left(\overline{\gamma}_T + \frac{2s_T}{T} \right).$$

4. Tracking the realized allocation. For every $\theta \in \Theta_{\mathrm{nb}}$ and allocations λ, λ', boundedness of the clipped gains gives

$$|\langle \lambda - \lambda', h_T(\theta) \rangle| \le B_T \|\lambda - \lambda'\|_1.$$

In particular, $\langle \lambda, h_T(\theta) \rangle \le \langle \lambda', h_T(\theta) \rangle + B_T \|\lambda - \lambda'\|_1$ for every θ. Taking infima and interchanging λ, λ' proves

$$|J(\lambda) - J(\lambda')| \le B_T \|\lambda - \lambda'\|_1.$$

The tracking lemma therefore yields

$$J(\overline{\lambda}_T) \le J(\widehat{\lambda}_T) + B_T r_T^{\mathrm{tr}}.$$

Define the deterministic quantity

$$\varepsilon_T := \frac{R_T^{\max} + R_T^{\min}}{T} + r_T^{\mathrm{ps}} + B_T \left(\overline{\gamma}_T + r_T^{\mathrm{tr}} + \frac{2s_T}{T} \right).$$

Then, on $\mathcal{E}_T$,

$$\Gamma^*(\theta^*) - \varepsilon_T \le J(\widehat{\lambda}_T) \le \Gamma^*(\theta^*).$$

The right inequality follows directly from the definition of $\Gamma^*(\theta^*)$. To verify that the new exploration term vanishes, use $\gamma_t = t^{-\alpha}$ and monotonicity to obtain

$$\sum_{t=1}^{T} t^{-\alpha} \le 1 + \int_1^T x^{-\alpha}\, dx = \frac{T^{1-\alpha} - \alpha}{1-\alpha} \le \frac{T^{1-\alpha}}{1-\alpha},$$

$$B_T \overline{\gamma}_T \le \frac{c_B \log(T+1)}{(1-\alpha) T^{\alpha}} \longrightarrow 0.$$

Together with the cutoff and remainder limits above, this gives $\varepsilon_T \to 0$. No convergence of the allocation vector is needed.

5. Converting information into posterior error. Let

$$J_0 := |\{c \in \mathcal{C}_N^\circ : c \nsim c^*\}|.$$

Since $N \geq 2$ is fixed, $1 \leq J_0 < \infty$. The marginal-likelihood lemma gives, simultaneously for every canonical partition c,

$$L_T(\theta^*)e^{-T[J_c(\widehat{\lambda}_T)+r_T^\ell]} \leq w_T^{(1)}(c) \leq L_T(\theta^*)e^{-T[J_c(\widehat{\lambda}_T)-r_T^\ell]}.$$

For an incorrect partition, $J_c(\widehat{\lambda}_T) \geq J(\widehat{\lambda}_T)$, so the numerator of the posterior error satisfies

$$\sum_{c \nsim c^*} w_T^{(1)}(c) \leq J_0 L_T(\theta^*)e^{-T[J(\widehat{\lambda}_T)-r_T^\ell]}.$$

For the true partition, nonnegativity of KL and the admissible choice $(p, q) = (p^*, q^*)$ give $J_{c^*}(\widehat{\lambda}_T) = 0$. Thus the posterior denominator satisfies

$$\sum_c w_T^{(1)}(c) \geq w_T^{(1)}(c^*) \geq L_T(\theta^*)e^{-Tr_T^\ell}.$$

All weights are positive. Dividing the last two bounds gives

$$\begin{aligned} 1 - \alpha_T = \Pi_T(\mathrm{Alt}(\theta^*)) &= \frac{\sum_{c \nsim c^*} w_T^{(1)}(c)}{\sum_c w_T^{(1)}(c)} \\ &\leq J_0 e^{-T[J(\widehat{\lambda}_T)-2r_T^\ell]} \\ &\leq e^{-T[\Gamma^*(\theta^*)-\xi_T]}, \end{aligned}$$

where

$$\xi_T := \varepsilon_T + 2r_T^\ell + \frac{\log J_0}{T} \longrightarrow 0.$$

The two copies of r_T^ℓ account separately for the numerator and denominator bounds.

6. The almost-sure exponent. The algorithm depends on the budget through its tuning parameters. Realize the budget-indexed experiments on a common probability space, and suppress this additional index in the notation above. Independence between experiments is not required. The event bounds give

$$\sum_{T=1}^{\infty} \mathbb{P}_{\theta^*}(\mathcal{E}_T^c) < \infty,$$

where the finitely many budgets preceding the good-event bounds contribute a finite sum. By the first Borel–Cantelli lemma, almost every path belongs to $\mathcal{E}_T$ for all sufficiently large T. On each such path the posterior bound is eventually valid, and hence

$$-\frac{1}{T}\log \Pi_T(\mathrm{Alt}(\theta^*)) \geq \Gamma^*(\theta^*) - \xi_T.$$

Taking the lower limit and using $\xi_T \to 0$ proves

$$\liminf_{T\to\infty} -\frac{1}{T}\log \Pi_T(\mathrm{Alt}(\theta^*)) \geq \Gamma^*(\theta^*) \qquad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

This establishes the posterior error upper bound.

To apply the universal posterior error lower bound, first match the partition and labeling conventions. Let $m(c)$ be the number of clusters of a canonical partition c. It has exactly $m(c)!$ distinct labelings. Define a prior on labeled clusterings by

$$\rho_0(\widetilde{c}) := \frac{\varrho_0(c)}{m(c)!}, \qquad \widetilde{c} \sim c, \quad c \in \mathcal{C}_N^\circ.$$

This is a positive probability mass function, since its mass over the labelings of c is $\varrho_0(c)$. The likelihood is invariant under relabeling, so this prior induces exactly the algorithm's posterior on partitions. The fixed truncated Beta priors have full support on their respective half-intervals. Thus

the lower-bound assumptions hold for the same posterior error mass. Applied to the same budget-indexed experiments, that theorem gives

$$\limsup_{T\to\infty} -\frac{1}{T}\log(1-\alpha_T) \le \Gamma^*(\theta^*) \qquad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

Intersecting the two probability-one events yields

$$\lim_{T\to\infty} -\frac{1}{T}\log(1-\alpha_T) = \Gamma^*(\theta^*) \qquad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

Finally, $\mathcal{H}_T$ is the history before the terminal draw. Since the recommendation uses fresh randomness and is sampled from the clustering marginal of the ordinary posterior,

$$\mathbb{P}_{\theta^*}(\widehat{c}_T \nsim c^* \mid \mathcal{H}_T) = \Pi_T(\mathrm{Alt}(\theta^*)) = 1-\alpha_T.$$

Thus the same exponent governs its conditional error probability. □