ONLINE CLASS-SPECIFIC CONFORMAL PREDICTION WITH DILUTED BANDIT FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Conformal prediction is a distribution-free framework that constructs prediction sets with prescribed coverage guarantees by calibrating scoring thresholds. Standard online calibration of these thresholds relies on observing the ground truth after each prediction, yet many deployed systems reveal far less: the learner observes only whether the true label falls in the prediction set without knowing which element is the truth, a setting called *diluted bandit feedback*. Such a binary signal conflates the contributions of every label in the prediction set, making class-specific credit assignment ambiguous when training a base model for scoring and calibrating the thresholds. To address this challenge, we propose an online algorithm for multi-class classification, Diluted-bandit Class-specific Conformal Prediction (**DCCP**), which displays a *randomized* prediction set whose label inclusion probabilities are controlled by the learner. DCCP capitalizes on a designed estimator to recover an unbiased class-specific signal from the binary feedback alone. This signal trains the base model and calibrates the thresholds, dedicating the prediction set to covering the true label at a prescribed rate for each class. We provide a theoretical analysis for coverage and prediction set size, and evaluate both empirically relative to reference methods.

1 INTRODUCTION

A prediction system may report a set of candidate labels rather than a single guess, hedging when a point prediction cannot be trusted. Serving a set, however, changes what the system can observe: the environment reports only whether the truth was covered by the set as a whole, not which label was correct. In automated customer support, an assistant unsure of a query's intent answers for several candidate intents at once, and the logs record the resolution of the issue, with no trace of the intent behind it. Attributing the resolution to one intent would require a human review of the transcript that the online loop does not include. In emergency medicine, a clinical decision system outputs a set of plausible diagnoses for the underlying condition and the physician starts broad-spectrum treatment covering all of them before laboratory cultures return. The patient outcome then tells the system whether the true condition was somewhere in the treated set, and no more than that. In both cases, serving a set is what makes the feedback set-level: the returned signal is a single bit, namely whether the true label is in the set, and it credits no individual element.

To equip such a prediction set with a coverage guarantee, canonical conformal prediction (Vovk et al., 2005; Shafer & Vovk, 2008) embeds calibrated thresholds on a base model's scores into a set-valued predictor, tuning the thresholds on a labeled holdout set (Lei et al., 2018) so that the set contains the true label at a user-prescribed rate. Online variants, led by Adaptive Conformal Inference (Gibbs & Candès, 2021) and its extensions (Zaffran et al., 2022; Bhatnagar et al., 2023; Angelopoulos et al., 2024; Areces et al., 2025), update these thresholds in real time. None of them, however, applies to the set-level feedback of the cases above: each still requires the true label to be revealed after every prediction.

Two lines of work loosen this label requirement, each along a different route. One line targets only marginal coverage and never learns the base model from the feedback: Gollapudi et al. (2026) work from the set-level bit, while Ge et al. (2025) and Yang et al. (2026) use semi-bandit feedback that reveals the label on covered instances. The only thing they learn from the feedback is a single threshold, and its target pools every class into one number, so these methods never face credit assignment,

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.

namely disentangling each class's contribution to the single bit. Freezing the model, however, may not always be feasible: the learner may start with no reliable model to freeze, and even when one is available, the model should keep improving as outcomes accumulate. The other line, BCCP and related work (Wang & Qiao, 2024; 2025), does train the model and provide class-specific coverage. It learns from a probe, a single label chosen independently of the set it reports. That is a mismatch with deployments that act on the whole set and never grade one label. Neither, therefore, trains the base model online from the set-level bit alone. Table 1 summarizes the comparison.

Our methodology closes this gap directly: the learner operates under *diluted bandit feedback* (Batra & Manwani, 2021), i.e., the aforementioned single set-level bit. For this setting we design an online algorithm, Diluted-bandit Class-specific Conformal Prediction (**DCCP**). Our contributions are as follows. i) **Randomized conformal set**: the prediction shown to the user is exactly the experiment the learner runs, so unlike BCCP (Wang & Qiao, 2024), no separate probe is needed. ii) **Augmented Horvitz–Thompson estimator**: the bridge converting the single set-level bit into the class-specific learning signal, unbiased regardless of the current model's quality, with the augmentation term lowering its variance in practice. iii) **Joint learning and calibration**: the procedure, driven by the learning signal, trains the base model jointly with the class-specific thresholds online, rather than taking the score function as given (Gollapudi et al., 2026; Ge et al., 2025; Yang et al., 2026). iv) **Theoretical guarantees**: the class-specific coverage gap closes at rate $O(T^{-1/2})$ under certain conditions, and the prediction set size is bounded relative to a calibrated benchmark, alongside the impossibility result that motivates the prediction rule.

The rest of the paper is organized as follows. Section 2 reviews the background on conformal prediction and limited-feedback set-valued prediction. Section 3 introduces the DCCP algorithm, and Section 4 establishes its coverage and prediction set size guarantees. Section 5 reports the experiments, and Section 6 concludes. Proofs and further details are deferred to the Appendix.

2 Preliminary

2.1 Conformal Prediction

Conformal prediction (Vovk et al., 2005; Shafer & Vovk, 2008) is a distribution-free framework for constructing prediction sets with finite-sample coverage guarantees. Two coverage notions are standard. *Marginal coverage* requires the prediction set $\widehat{\mathcal{C}}(\boldsymbol{X})$ to satisfy $\mathbb{P}(Y \in \widehat{\mathcal{C}}(\boldsymbol{X})) \geq 1-\alpha$ under the joint distribution of $(\boldsymbol{X}, Y) \in \mathcal{X} \times \mathcal{Y}$, with $\mathcal{Y}$ the finite label set and $\alpha \in (0,1)$, and is attainable with a single global threshold, but can sacrifice coverage on classes whose scores are distributed differently. In contrast, *class-specific coverage* (Lei et al., 2013; Lei & Wasserman, 2014; Lei et al., 2018; Sadinle et al., 2019) imposes the stronger requirement $\mathbb{P}(Y \in \widehat{\mathcal{C}}(\boldsymbol{X}) \mid Y = k) \geq 1-\alpha$ for every class $k \in \mathcal{Y}$. The standard approach maintains a separate threshold for each class, calibrated on a holdout set (Sadinle et al., 2019; Wang & Qiao, 2023).

In online settings with full feedback, the true label is revealed after each prediction, enabling direct threshold adaptation. Gibbs & Candès (2021) proposed Adaptive Conformal Inference (ACI), an online threshold-updating method that achieves marginal coverage. Building on ACI, subsequent work extends online conformal prediction to richer settings such as distribution shift and general risk control (Zaffran et al., 2022; Gibbs & Candès, 2024; Bhatnagar et al., 2023; Angelopoulos et al., 2023; 2024; Barber et al., 2023; Areces et al., 2025). Across all these works, the true label must be observed each round to update the threshold.

2.2 Set-valued Prediction with Limited Feedback

Unlike the full-feedback setting in Section 2.1, in many deployments, the learner never observes the true label Y_t (at round t), only a coarse signal about its own prediction. In the set-valued prediction paradigm, Batra & Manwani (2021) formalized this as the *diluted bandit* setting for the multi-class classification task: the learner outputs a set $\widehat{\mathcal{C}}_t$ and observes only the bit $\mathbb{1}\{Y_t \in \widehat{\mathcal{C}}_t\}$. Note that their learner minimizes mistakes with the set size fixed in advance, so it cannot adapt to a prescribed coverage level and hence offers no coverage guarantee.

Table 1: Key literature difference summary

Work	Online model training	Hidden probe	Feedback type	Coverage guarantee
Batra & Manwani (2021)	Yes	No	Diluted bandit	None
Wang & Qiao (2024)	Yes	Yes	Single-arm bandit	Class-specific
Ge et al. (2025); Yang et al. (2026)	No (score supplied)	No	Semi-bandit	Marginal
Gollapudi et al. (2026)	No (score supplied)	No	Diluted bandit	Marginal
DCCP (ours)	**Yes**	**No**	**Diluted bandit**	**Class-specific**

Coverage-oriented methods do exist. Wang & Qiao (2024) (BCCP) attains class-specific coverage, but under single-arm bandit feedback: a hidden probe A_t replaces the set-level bit, so the learner sees only $\mathbb{1}\{A_t = Y_t\}$. This feedback is not available in the motivating deployments: the outcome reflects the whole answer or the whole treatment given, never a separate verdict on a label the user did not receive. Absent the probe, only the set-level bit remains, and it carries no class identity.

Without class identity in the set-level bit, training a model online is the real obstacle: any model update needs a class-specific signal for supervision, and recovering one from the single bit is a credit-assignment problem. The probe-free works of Gollapudi et al. (2026); Ge et al. (2025); Yang et al. (2026) sidestep it by assuming the score function to be supplied, leaving only a threshold to tune, though they differ in how much their feedback reveals. Gollapudi et al. (2026) work from the diluted bit alone, treating each candidate threshold as a bandit arm applied to a fixed score function and updating its statistics from whether the prediction set succeeded. Ge et al. (2025) and Yang et al. (2026) instead receive semi-bandit feedback, which reveals Y_t on covered instances, so the true class is known exactly on successes. Neither route updates the base model from the diluted bit, which is what DCCP must do.

A parallel line studies corrupted rather than partial feedback: Wang et al. (2026) observe the set-level bit $\mathbb{1}\{Y_t \in \widehat{\mathcal{C}}_t\}$ but allow it to be flipped by noise or an adversary. Like the probe-free works above, they fix the base model and target marginal coverage.

3 METHOD

In this section, we introduce our proposal, i.e., Diluted-bandit Class-specific Conformal Prediction (DCCP), which is an online algorithm that simultaneously trains a base model for scoring and calibrates class-specific thresholds from set-level binary feedback alone. The protocol is as follows. At step $t = 1, \ldots, T$, the learner receives a batch of instances $\{(\boldsymbol{X}_{t,i}, Y_{t,i})\}_{i \in \mathcal{B}_t}$ from $\mathcal{X} \times \mathcal{Y}$, indexed by the set $\mathcal{B}_t$ (the true label $Y_{t,i}$ is never revealed). Specifically, for every $i \in \mathcal{B}_t$, the learner observes $\boldsymbol{X}_{t,i}$, outputs a prediction set $\widehat{\mathcal{C}}_{t,i} \subseteq \mathcal{Y}$, and receives only the binary feedback $\mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\}$. Algorithm 1 collects the resulting procedure into a single online loop.

To this end, the learner maintains a neural network $\boldsymbol{f}_{\mathcal{W}^t} : \mathcal{X} \to \mathbb{R}^{|\mathcal{Y}|}$ with weights $\mathcal{W}^t$ and class-specific thresholds $\{\tau_t^k\}_{k \in \mathcal{Y}}$, both updated online from the binary feedback. The class-k logit output is written as $(\boldsymbol{f}_{\mathcal{W}^t}(\cdot))_k$. DCCP has three interleaved components: i) a randomized prediction set whose label inclusion probabilities are known to the learner; ii) an estimator that recovers an unbiased class-specific learning signal from the single set-level bit, supervising the base model training; and iii) threshold updates that drive class-specific coverage to level $1 - \alpha$.

3.1 THE RANDOMIZED CONFORMAL SET

Consider an instance $\boldsymbol{X}_{t,i}$ from step t's batch. The model from the previous step $\boldsymbol{f}_{\mathcal{W}^{t-1}}$ induces a class-specific conformity score function $s_{t-1}^k(\cdot), k \in \mathcal{Y}$, e.g., RAPS (Angelopoulos et al., 2021). Given this score function and the class-k threshold τ_{t-1}^k (its update is deferred to Section 3.3), the canonical conformal prediction set for $\boldsymbol{X}_{t,i}$ would be deterministic, $\widetilde{\mathcal{C}}_{t,i} := \{k \in \mathcal{Y} : s_{t-1}^k(\boldsymbol{X}_{t,i}) \geq \tau_{t-1}^k\}$. In such a set, however, membership is fixed, so two classes both inside or both outside the set produce the same feedback whatever the true label is, and the feedback alone cannot tell them apart (Proposition 1). DCCP instead *randomizes* the membership of every class with a probability it controls, so that each class is included on some instances and left out on others. Including or excluding class k changes whether the truth is covered exactly when k is the truth, so the feedback

now responds to each class on its own, and the learner can see that response because it makes and records every draw.

Formally, in DCCP, the prediction set for the instance $\boldsymbol{X}_{t,i}$ is defined as

$$\widehat{\mathcal{C}}_{t,i} := \{k \in \mathcal{Y} : Z^k_{t,i} = 1\}, \tag{1}$$

where $Z^k_{t,i} \sim \text{Bernoulli}(\rho^k_{t,i}(\tau^k_{t-1}))$ is sampled independently across classes and instances, with class inclusion probability defined by

$$\rho^k_{t,i}(\tau) := \text{clip}\left(\frac{1}{1+\exp((\tau - s^k_{t-1}(\boldsymbol{X}_{t,i}))/h)},\ c,\ 1-c\right), \qquad \tau \in \mathbb{R}. \tag{2}$$

Here $h > 0$ is a softness parameter controlling how sharply the inclusion probability responds to the margin $s^k_{t-1}(\boldsymbol{X}_{t,i}) - \tau^k_{t-1}$. Intuitively the more the score $s^k_{t-1}(\boldsymbol{X}_{t,i})$ exceeds the threshold τ^k_{t-1}, the more likely class k is included. At small h the inclusion probability tracks that margin closely. At large h it is smoothed across all classes, each class included with probability close to 0.5, leading to more ambiguity in the prediction rule. The parameter $c \in (0, \alpha]$ clips the inclusion probability to $[c, 1-c]$, which floors the expected set size at $c|\mathcal{Y}|$. We provide the sensitivity analysis and discussion in Section 5.2 for both parameters.

3.2 The Unbiased Estimator and Model Update

Because each class is included with a known probability $\rho^k_{t,i}(\tau^k_{t-1})$, the set-level feedback can be turned into an unbiased class-specific signal. Specifically, reweighting the feedback by its known inclusion probability, following the Horvitz–Thompson principle (Horvitz & Thompson, 1952) and subtracting a model-based control variate inspired by doubly robust estimation (Bang & Robins, 2005; Dudík et al., 2011), we form the *Augmented Horvitz–Thompson* estimator

$$\Delta^k_{t,i} := \frac{Z^k_{t,i} - \rho^k_{t,i}(\tau^k_{t-1})}{\rho^k_{t,i}(\tau^k_{t-1})\big[1-\rho^k_{t,i}(\tau^k_{t-1})\big]}\left(\mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\} - \widehat{R}_{t,i}\right), \tag{3}$$

to estimate the unobserved one-hot label $\mathbb{1}\{Y_{t,i} = k\}$. Here the augmentation term (see the detailed derivation in Section A.1) is defined as $\widehat{R}_{t,i} := \sum_{j\in\mathcal{Y}} \hat{p}(j \mid \boldsymbol{X}_{t,i})\, \rho^j_{t,i}(\tau^j_{t-1})$, in which $\hat{p}(j \mid \boldsymbol{X}_{t,i}) = \text{softmax}(\boldsymbol{f}_{\mathcal{W}^{t-1}}(\boldsymbol{X}_{t,i}))_j$ is the previous step's model softmax output. Proposition 2 proves that $\Delta^k_{t,i}$ is an unbiased estimator of $\mathbb{1}\{Y_{t,i} = k\}$ given $(\boldsymbol{X}_{t,i}, Y_{t,i})$ and all the information up to step $t-1$. The ablation study in Section 5.2 empirically shows the estimator's variance gets larger and the prediction sets are wider with the augmentation removed.

Together with this estimator, the base model parameters are updated from $\mathcal{W}^{t-1}$ to $\mathcal{W}^t$ by one step of mini-batch gradient descent (or its variants) with rate η_1 on the weighted cross-entropy loss

$$\mathcal{L}_t(\mathcal{W}) = -\frac{1}{|\mathcal{B}_t|}\sum_{i\in\mathcal{B}_t}\sum_{k\in\mathcal{Y}} \Delta^k_{t,i} \cdot \log \frac{\exp\big((\boldsymbol{f}_{\mathcal{W}}(\boldsymbol{X}_{t,i}))_k\big)}{\sum_{j\in\mathcal{Y}}\exp\big((\boldsymbol{f}_{\mathcal{W}}(\boldsymbol{X}_{t,i}))_j\big)}, \tag{4}$$

the standard cross-entropy loss with the unobserved one-hot label $\mathbb{1}\{Y_{t,i} = k\}$ substituted by its unbiased estimate $\Delta^k_{t,i}$. Note that the estimate is formed from $\mathcal{W}^{t-1}$, so it is a constant in this step and the gradient passes through the logarithm alone.

3.3 The Threshold Calibration

The same signal $\Delta^k_{t,i}$ that trains the model also updates the thresholds. We calibrate the class-k threshold so that its coverage meets the target $1-\alpha$. Every expectation and probability below is over the step-t data and the inclusion draws, given the score function s^k_{t-1} and the threshold τ^k_{t-1}:

$$1-\alpha = \mathbb{P}[Y_{t,i} \in \widehat{\mathcal{C}}_{t,i} \mid Y_{t,i} = k] = \mathbb{P}[Z^k_{t,i} = 1 \mid Y_{t,i} = k] = \mathbb{E}\big[\rho^k_{t,i}(\tau^k_{t-1}) \mid Y_{t,i} = k\big], \tag{5}$$

$$\iff \quad \mathbb{E}[[1-\rho^k_{t,i}(\tau^k_{t-1})] - \alpha \mid Y_{t,i} = k] = \frac{\mathbb{E}[\mathbb{1}\{Y_{t,i} = k\}\big([1-\rho^k_{t,i}(\tau^k_{t-1})] - \alpha\big)]}{\mathbb{P}(Y_{t,i} = k)} = 0. \tag{6}$$

We solve Eq. (6) for the threshold online in a data-driven manner (see details in Section A.2), by stepping τ^k_{t-1} with the per-instance term $\mathbb{1}\{Y_{t,i} = k\}\big([1-\rho^k_{t,i}(\tau^k_{t-1})] - \alpha\big)$. The indicator $\mathbb{1}\{Y_{t,i} =$

Algorithm 1: Diluted-bandit Class-specific Conformal Prediction (DCCP)

Input: Target miscoverage α, learning rates $\eta_1, \eta_2 > 0$, softness parameter $h > 0$, inclusion floor $c \in (0, \alpha]$.
Initialize: model parameters $\mathcal{W}^0$; thresholds $\tau_0^k = 0$ for all $k \in \mathcal{Y}$;
for $t = 1, 2, \ldots, T$ **do**
  for $i \in \mathcal{B}_t$ (parameters $\mathcal{W}^{t-1}, \{\tau_{t-1}^k\}$ held fixed for the whole batch) **do**
    1. Observe $\boldsymbol{X}_{t,i}$ and compute inclusion probability $\rho_{t,i}^k(\tau_{t-1}^k)$ for all $k \in \mathcal{Y}$ via Eq. (2);
    2. Sample $Z_{t,i}^k \sim \mathrm{Bernoulli}(\rho_{t,i}^k(\tau_{t-1}^k))$ independently to output set $\widehat{\mathcal{C}}_{t,i}$ as in Eq. (1);
    3. Observe the feedback $\mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\}$ from the environment;
    4. Compute $\Delta_{t,i}^k$ for each $k \in \mathcal{Y}$ via Eq. (3);
  end
  5. Update model parameters $\mathcal{W}^t$ by one step with rate η_1 on the loss in Eq. (4);
  6. Update τ_t^k for all $k \in \mathcal{Y}$ with step size η_2 via Eq. (7);
end

$k\}$ is unobserved and hence substituted with the unbiased estimate $\Delta_{t,i}^k$, and step t averages the resulting terms over its batch, which yields the updating rule

$$\tau_t^k = \tau_{t-1}^k - \eta_2 \cdot \frac{1}{|\mathcal{B}_t|} \sum_{i \in \mathcal{B}_t} \Delta_{t,i}^k \cdot \big([1 - \rho_{t,i}^k(\tau_{t-1}^k)] - \alpha\big). \tag{7}$$

This rule, in expectation, lowers the threshold (covering class k more often) when class k is under-covered (i.e., $\mathbb{E}[\rho_{t,i}^k(\tau_{t-1}^k) \mid Y_{t,i} = k] < 1 - \alpha$) and raises it when over-covered. This gradient-like updating rule may alternatively be taken with a variant such as Adam. In the full-feedback limit where $\Delta_{t,i}^k = \mathbb{1}\{Y_{t,i} = k\}$ and $h, c \to 0$, it reduces to the class-specific analogue of the ACI update (Gibbs & Candès, 2021) with mini-batch averaging.

4 Theoretical Analysis

In this section, the batch size is $|\mathcal{B}_t| = B$ for simplicity, so a step's instances are indexed by $i = 1, \ldots, B$. Additionally, $\mathcal{F}_{t-1}$ denotes the sigma-algebra generated by everything observed up to but not including step t.

4.1 Impossibility for Deterministic Sets

The randomization in Eq. (2) is not an implementation convenience: with membership fixed, no unbiased class-specific signal is available for two classes that fall on the same side of the set.

Proposition 1 (Impossibility for deterministic sets)**.** *Let $\widetilde{\mathcal{C}}_{t,i}$ be the deterministic set mentioned in Section 3.1, which is fixed given $\boldsymbol{X}_{t,i}$ and $\mathcal{F}_{t-1}$. Suppose two distinct classes y, y' satisfy $\mathbb{1}\{y \in \widetilde{\mathcal{C}}_{t,i}\} = \mathbb{1}\{y' \in \widetilde{\mathcal{C}}_{t,i}\}$, with $\mathbb{P}(Y_{t,i} = y \mid \boldsymbol{X}_{t,i}) > 0$ and $\mathbb{P}(Y_{t,i} = y' \mid \boldsymbol{X}_{t,i}) > 0$. Then no family $(\psi_k)_{k \in \mathcal{Y}}$ of estimators formed from the input $\boldsymbol{X}_{t,i}$ and the diluted bandit feedback $\mathbb{1}\{Y_{t,i} \in \widetilde{\mathcal{C}}_{t,i}\}$ alone is unbiased for $\mathbb{1}\{Y_{t,i} = k\}$ at every class $k \in \mathcal{Y}$.*

The set-level bit reports only whether the true label lies inside or outside $\widetilde{\mathcal{C}}_{t,i}$. Two classes on the same side, i.e., classes satisfying the condition $\mathbb{1}\{y \in \widetilde{\mathcal{C}}_{t,i}\} = \mathbb{1}\{y' \in \widetilde{\mathcal{C}}_{t,i}\}$, therefore produce the same bit, leading any estimator formed from $\boldsymbol{X}_{t,i}$ and that bit to return the same value in both cases (see the proof in Section D.1). Randomizing membership breaks the tie. The bit is the true class's inclusion draw, $\mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\} = Z_{t,i}^{Y_{t,i}}$: when class k is the truth, the bit is k's own draw $Z_{t,i}^k$; when it is not, the bit is another class's draw, independent of $Z_{t,i}^k$. Class k's inclusion therefore correlates with the bit exactly when k is the truth, and Eq. (3) turns that correlation into an unbiased estimate of $\mathbb{1}\{Y_{t,i} = k\}$ (see Proposition 2). Eq. (2) clips $\rho_{t,i}^k$ to $[c, 1 - c]$ to hold it away from $\{0, 1\}$, where $Z_{t,i}^k$ would stop being random and Proposition 1 would apply again.

Consequence. Online model training and class-specific coverage both need a class-specific signal that the set-level bit does not carry on its own. Proposition 1 gives the sharpest version: at a given instance a deterministic set admits no such unbiased signal for two classes that fall on the same side of it and could each be the truth there, and DCCP's randomization recovers one. Each existing method gives up something else instead: Batra & Manwani (2021) gives up class-specific coverage, targeting mistake minimization; Gollapudi et al. (2026) and Wang & Qiao (2024) give up learning the model from the set's own feedback, the former by taking the score function as given so nothing is learned, the latter by learning from a hidden probe rather than the set-level feedback.

4.2 Estimator Properties

The estimator $\Delta^k_{t,i}$, the signal that randomization recovers, stands in for the unobserved one-hot label $\mathbb{1}\{Y_{t,i} = k\}$ in the model loss and the threshold update, so its bias and variance affect both.

Proposition 2. *Condition on $\boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}$ (so each $\rho^k_{t,i}(\tau^k_{t-1})$ is fixed) and take expectation over the independent Bernoulli draws $Z^k_{t,i}, k \in \mathcal{Y}$. Then for every class k:*

(i) Unbiasedness: $\mathbb{E}\big[\Delta^k_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}\big] = \mathbb{1}\{Y_{t,i} = k\}$.

(ii) Bounded variance: $\mathrm{Var}\big(\Delta^k_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}\big) \le \frac{1}{\rho^k_{t,i}(\tau^k_{t-1})\big[1-\rho^k_{t,i}(\tau^k_{t-1})\big]} \le \frac{1}{c(1-c)}$.

4.3 Coverage Guarantee

We now bound the realized class-specific coverage of the prediction set over the stream. The proof is given in Section D.3.

Theorem 1 (Class-specific coverage)**.** *Let the inclusion floor satisfy $c \in (0, \alpha]$. Then for any step size $\eta_2 > 0$ and any $\delta \in (0,1)$, with probability at least $1-\delta$, simultaneously for every class $k \in \mathcal{Y}$ whose $\mathcal{T}^k := \{(t,i) : 1 \le t \le T,\ 1 \le i \le B,\ Y_{t,i} = k\}$ is nonempty, we have*

$$\left| \frac{1}{|\mathcal{T}^k|} \sum\nolimits_{(t,i)\in\mathcal{T}^k} \mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\} - (1-\alpha) \right| \le \frac{B\,|\tau^k_T|}{\eta_2\,|\mathcal{T}^k|} + \Big(1 + \frac{2}{c}\Big) \frac{\sqrt{2TB\log(4|\mathcal{Y}|/\delta)}}{|\mathcal{T}^k|}.$$

The bound holds for any base model. For any $\alpha \le 1/2$, the premise $c \in (0, \alpha]$ puts the target coverage inside the range the prediction set can reach, i.e., $1-\alpha \in [c, 1-c]$. The two terms on the right-hand side of the above inequality come from different sources: the first is the threshold's drift from a finite step size, and the second is the estimator's fluctuation. Note that the average runs over the whole horizon rather than simultaneously for every $t \le T$. At $\eta_2 = \Theta(\sqrt{B/T})$ and $|\mathcal{T}^k| = \Theta(TB)$, the coverage gap is $O(\sqrt{\log(|\mathcal{Y}|/\delta)/(TB)})$ whenever $|\tau^k_T| = O(1)$. The condition on $|\mathcal{T}^k|$ ties each class's rate to how often that class appears: a rarer class (smaller $|\mathcal{T}^k|$) has fewer instances to calibrate on and closes its gap more slowly, while a marginal guarantee pools all TB instances and avoids this. The condition on $|\tau^k_T|$ is closed for a uniformly bounded score such as RAPS: Lemma 2 in Section C establishes it with high probability provided $c < \alpha$, and Corollary 2 states the resulting rate.

4.4 Prediction Set Size

The size of the prediction set, unlike its coverage, depends on the quality of the scores.

Theorem 2 (Size regret)**.** *Fix a benchmark score function and threshold pairs $\{(s^k_\star(\cdot), \tau^k_\star)\}_{k\in\mathcal{Y}}$, inducing the benchmark prediction set $\mathcal{C}^\star_{t,i}$ via Eqs. (1) and (2), and let it be calibrated, i.e., $\mathbb{P}(Y_{t,i} \in \mathcal{C}^\star_{t,i} \mid Y_{t,i} = k) = 1-\alpha$ for every $k \in \mathcal{Y}$. Assume the stream is i.i.d., and a non-degeneracy condition: for some $\mu > 0$ there is an interval $\mathcal{J}$ containing $\tau^k_\star$ and every iterate τ^k_t on which the class-k calibration map $G_{t,k}(u) := \mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\big([1-\rho^k_{t,i}(u)] - \alpha\big) \mid \mathcal{F}_{t-1}\big]$ satisfies $\big(G_{t,k}(u) - G_{t,k}(\tau^k_\star)\big)(u - \tau^k_\star) \ge \mu\,(u-\tau^k_\star)^2$ for every t and every $u \in \mathcal{J}$. Then, with $\varepsilon_t := \big(\mathbb{E}\big[\big(\sum_{k\in\mathcal{Y}} |s^k_{t-1}(\boldsymbol{X}_{t,i}) - s^k_\star(\boldsymbol{X}_{t,i})|\big)^2\big]\big)^{1/2}$, we have*

$$\frac{1}{TB}\Big|\sum_{t=1}^{T}\sum_{i=1}^{B} \mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}^\star_{t,i}|\big]\Big| \le \frac{1}{4Th}\sum_{t=1}^{T}\varepsilon_t + \frac{|\mathcal{Y}|}{16h^2\mu}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2} + \frac{1}{4h}\sum_{k\in\mathcal{Y}}\sqrt{\frac{2}{\mu T}\Big(\frac{(\tau^k_\star)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)}\Big)}.$$

The non-degeneracy condition asks that class-k coverage respond to the threshold. Without that response, many thresholds meet the coverage target equally well, and an update driven by coverage alone cannot tell which of them gives the smaller set. The condition is met when each class recurs and keeps part of its mass where the inclusion probability is unclipped (Section C). Gollapudi et al. (2026) impose a related condition, under the name success margin, and argue it is minimal for sublinear size regret.

In the bound of Theorem 2, the first two terms measure how far the run's own score is from the benchmark's, and are 0 when the two agree. The third is a tracking term of order $|\mathcal{Y}|T^{-1/4}$ from the online threshold calibration when $\eta_2 = \Theta(T^{-1/2})$. Only the tracking term survives when the base model is held fixed, as in Gollapudi et al. (2026); Ge et al. (2025): taking $s^k_\star$ to be that fixed score makes $\varepsilon_t = 0$. Theorem 2 is proved in Section D.5.

5 Experiments

In this section, we conduct experiments to evaluate DCCP's performance on CIFAR-10, CIFAR-100 and SVHN with $\alpha = 0.1$. See the detailed configuration in Section B.1.

Metrics. To check the implementation against Proposition 2, we track the estimator's average on the ground-truth class and, separately, on the non-truth classes,

$$\mathrm{Ubias}^+(t) = \frac{1}{|\mathcal{B}_t|}\sum_{i\in\mathcal{B}_t}\Delta^{Y_{t,i}}_{t,i}, \qquad \mathrm{Ubias}^-(t) = \frac{1}{|\mathcal{B}_t|}\sum_{i\in\mathcal{B}_t}\frac{1}{|\mathcal{Y}|-1}\sum_{k\neq Y_{t,i}}\Delta^{k}_{t,i},$$

with targets 1 and 0 respectively: the estimator matches the one-hot label on the true class and on other classes. Coverage and set size are reported as *accumulative* averages over the online stream. Specifically, the accumulative class-k coverage and set size of the prediction through step t are

$$\mathrm{Acov}_k(t) := \frac{\sum_{r\le t}\sum_{i\in\mathcal{B}_r}\mathbb{1}\{Y_{r,i}=k,\ Y_{r,i}\in\widehat{\mathcal{C}}_{r,i}\}}{\sum_{r\le t}\sum_{i\in\mathcal{B}_r}\mathbb{1}\{Y_{r,i}=k\}}, \qquad \mathrm{Asize}(t) := \frac{\sum_{r\le t}\sum_{i\in\mathcal{B}_r}|\widehat{\mathcal{C}}_{r,i}|}{\sum_{r\le t}|\mathcal{B}_r|}.$$

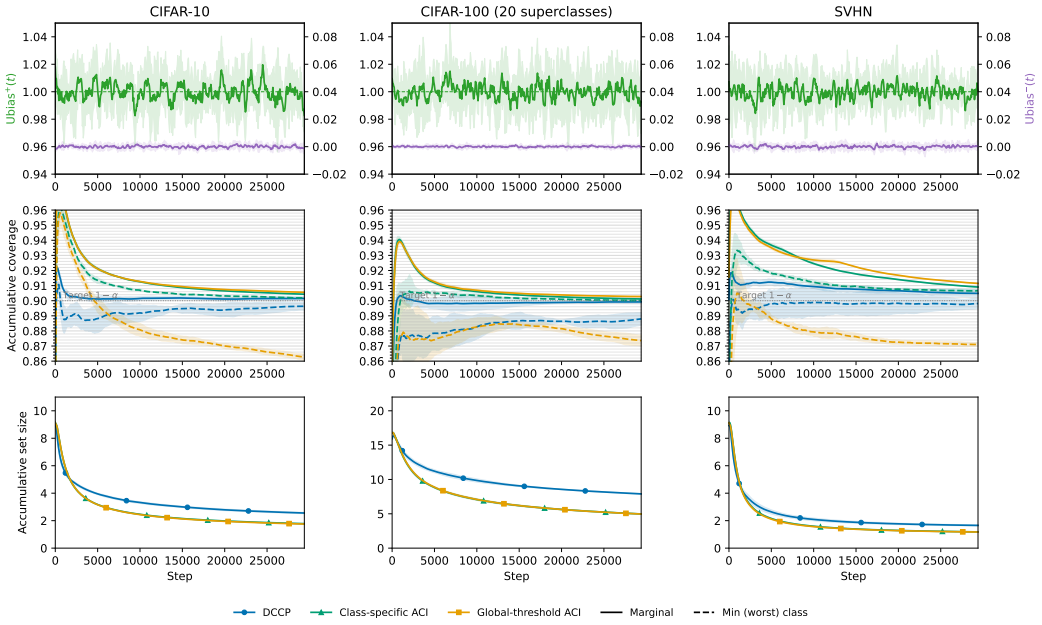

Figure 1: Baseline comparison: one column per dataset. **Estimator average:** $\mathrm{Ubias}^+(t)$ (left axis) and $\mathrm{Ubias}^-(t)$ (right axis). **Accumulative coverage:** marginal (solid) and worst class (dashed), against the target $1-\alpha$ (dotted). **Accumulative set size:** own scale per dataset. Shaded bands span $3\times$ the standard error over runs.

We plot the marginal coverage and the worst class, $\min_k \mathrm{Acov}_k(t)$, against the target coverage $1-\alpha$, and the set size $\mathrm{Asize}(t)$ over the same steps.

Reference methods. No prior method shares our exact setting, so a strictly comparable baseline does not exist. We compare DCCP (diluted set-level bit; class-specific thresholds) with two references adapted to be as close as possible: class-specific ACI (Gibbs & Candès, 2021) (revealed label; class-specific thresholds) and global-threshold ACI in the setting of Gollapudi et al. (2026) (diluted set-level bit; one global threshold). Both references calibrate on the same online scores from DCCP's base model, so they differ from DCCP only in the feedback they consume and the thresholds they maintain. Class-specific ACI therefore shows what the revealed label buys for calibration, and global-threshold ACI what one global threshold costs. The base model's own learning is reported separately in Section B.3.

5.1 Main Results

In Figure 1, the estimator average (first row) holds near 1 on the ground-truth class and near 0 on the others throughout training, as Proposition 2 requires. Global-threshold ACI meets the marginal target but under-covers the worst class on all three datasets (orange lines in middle row), reaching only 0.863 on CIFAR-10, whereas DCCP holds its worst class close to $1-\alpha$ (0.896) and tracks class-specific ACI. The set sizes (bottom row) carry the cost: DCCP's set is larger than either reference on all three datasets. The base model improves over the stream, for instance, on CIFAR-10 its accumulative top-1 accuracy rises from 0.689 at $t/T = 0.2$ to 0.780 at the end (see Section B.3).

5.2 Sensitivity and Ablation Analysis

In this section, we vary the softness h, the inclusion floor c, and the augmentation term in turn, and sweep the threshold step size η_2 in Section B.4.

Softness h. Figure 2(a) tracks each metric over training on CIFAR-10 at the default floor $c = \alpha/2$, sweeping h over a $50\times$ range. Both coverage curves (middle column) converge to $1-\alpha$ for every h. The remaining two columns show the trade-off h controls. Only $h = 0.5$ separates from the rest: it holds the lowest variance and a set size above 6, while $h \in \{0.01, 0.05, 0.1\}$ overlap in variance and cluster below 3 in set size. The default $h = 0.05$ sits at the small-set end.

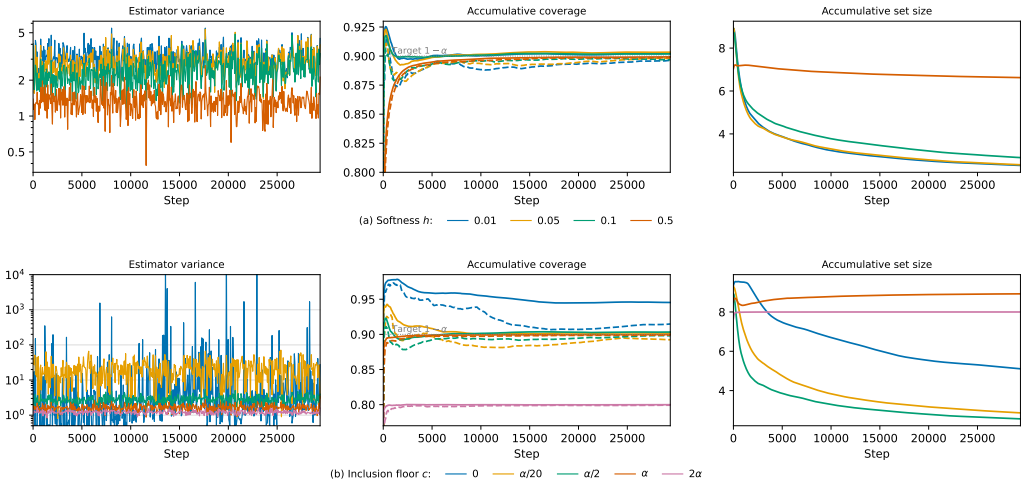

Figure 2: Sensitivity sweeps on CIFAR-10. **(a)** softness $h \in \{0.01, 0.05, 0.1, 0.5\}$ at the default floor $c = \alpha/2$. **(b)** inclusion floor $c \in \{0, \alpha/20, \alpha/2, \alpha, 2\alpha\}$ at the default softness $h = 0.05$. **Estimator variance:** on a log scale, the empirical variance of $\Delta^k_{t,i}$ over the batch $i \in \mathcal{B}_t$ of a single step, averaged over the classes $k \in \mathcal{Y}$. **Accumulative coverage:** marginal (solid) and worst class (dashed), against the target $1-\alpha$ (dotted). **Accumulative set size:** $\mathrm{Asize}(t)$.

Inclusion floor c. Figure 2(b) sweeps c across the admissible range $(0, \alpha]$ and one level on each side of it. The floor confines every inclusion probability to $[c, 1-c]$, which bounds the estimator's variance. Below the range, at $c = 0$, the variance has no bound: most steps look ordinary, but the estimator's variance occasionally spikes by several orders of magnitude. At the other end the binding constraint is the ceiling $1-c$ rather than the variance. At $c = \alpha$ the target $1-\alpha$ coincides with the ceiling $1-c$, so class k meets it only with $\rho^k_{t,i}$ at that ceiling on its own instances, and the set grows to nearly $|\mathcal{Y}|(1-c)$. At $c = 2\alpha$ the target lies outside $[c, 1-c]$ and accumulative coverage stops at $1-2\alpha$. The default $c = \alpha/2$ keeps $1-\alpha$ clear of both ends of $[c, 1-c]$.

Augmentation term $\widehat{R}_{t,i}$. Table 2 reads the stream at five sampled checkpoints and compares the augmented estimator against the alternative with $\widehat{R}_{t,i} = 0$, which reweights the raw bit alone, each entry a mean over seeds with its standard error. Marginal coverage stays close to the target $1-\alpha = 0.9$ at every checkpoint and on both estimators. Worst-class coverage separates the two estimators: with the augmentation it stays closer to the target at every checkpoint. Without it the worst class ends at 0.881. At every checkpoint the estimator's variance is smaller by a factor of $3.1 \sim 3.6$ with the augmentation, and at the end of the stream the set is 1.87 times narrower. Section B.2 also includes the same comparison on CIFAR-100 and SVHN. The augmentation matters for the base model too, which ends the stream $1.1 \sim 2.3$ times as accurate with the term as without it (see Section B.3).

Table 2: Effect of the augmentation on CIFAR-10.

	marginal coverage		worst-class coverage		set size		variance of $\Delta^k_{t,i}$	
t/T	augmented	0	augmented	0	augmented	0	augmented	0
0.2	$0.901_{\pm 0.001}$	$0.895_{\pm 0.002}$	$0.890_{\pm 0.002}$	$0.871_{\pm 0.003}$	$\mathbf{3.79}_{\pm 0.02}$	$5.78_{\pm 0.08}$	$\mathbf{2.85}_{\pm 0.04}$	$10.17_{\pm 0.36}$
0.4	$0.902_{\pm 0.000}$	$0.894_{\pm 0.002}$	$0.893_{\pm 0.001}$	$0.877_{\pm 0.005}$	$\mathbf{3.19}_{\pm 0.02}$	$5.34_{\pm 0.08}$	$\mathbf{2.95}_{\pm 0.04}$	$9.74_{\pm 0.24}$
0.6	$0.901_{\pm 0.000}$	$0.893_{\pm 0.001}$	$0.894_{\pm 0.001}$	$0.877_{\pm 0.006}$	$\mathbf{2.88}_{\pm 0.02}$	$5.08_{\pm 0.08}$	$\mathbf{3.09}_{\pm 0.04}$	$10.27_{\pm 0.64}$
0.8	$0.901_{\pm 0.000}$	$0.894_{\pm 0.002}$	$0.895_{\pm 0.001}$	$0.877_{\pm 0.004}$	$\mathbf{2.69}_{\pm 0.02}$	$4.91_{\pm 0.10}$	$\mathbf{3.11}_{\pm 0.03}$	$9.80_{\pm 0.70}$
1.0	$0.901_{\pm 0.000}$	$0.894_{\pm 0.002}$	$0.896_{\pm 0.001}$	$0.881_{\pm 0.002}$	$\mathbf{2.56}_{\pm 0.01}$	$4.78_{\pm 0.08}$	$\mathbf{3.09}_{\pm 0.04}$	$9.80_{\pm 0.50}$

6 Conclusion

We have studied online class-specific conformal prediction under diluted bandit feedback, where the prediction set shown to the user is the only source of training signal, and that signal is a single binary outcome. That one signal drives both halves of the system: DCCP trains the base model jointly with the thresholds online. The obstacle is that a deterministic set gives the same bit to any two classes on the same side of it, so no estimator reading that bit can distinguish them. Randomizing the set membership recovers an unbiased class-specific signal from that same bit.

The coverage theorem bounds the class-specific coverage gap independently of model quality, closing at the $O(T^{-1/2})$ rate for any class that recurs a constant fraction of the stream and whose threshold stays bounded. The model's quality is felt in the prediction set size, which the size theorem bounds relative to a calibrated benchmark. Across three datasets the experiments show the worst class holding near $1-\alpha$ while the set size falls as the base model learns, and removing the estimator's augmentation term raises its variance and widens the prediction set.

Several directions remain open. The inclusion probability is one design among many: the softness h and the inclusion floor c are single values shared across all classes and steps, and scheduling or adapting them could shrink the prediction set. That may matter more when the label set is large (Ding et al., 2023), since the floor alone contributes $c|\mathcal{Y}|$ to the expected size. Additionally, the Augmented Horvitz–Thompson construction is one unbiased choice among others, and further optimizing its variance reduction could train the base model faster. Finally, we may also extend this diluted feedback to the regression task.

AI use statement

We used generative AI tools to assist with writing: polishing the prose and improving the clarity of exposition. We also used them to produce the figures and tables based on our experimental results. We did not use generative AI tools for research ideation, for the design of the method, for the

derivation or proof of any result, or for generating or altering experimental data. We have reviewed all AI-assisted work and take responsibility for the final content of this paper, including its text, claims and artifacts.

REFERENCES

Anastasios Angelopoulos, Emmanuel J Candès, and Ryan J Tibshirani. Conformal pid control for time series prediction. Advances in Neural Information Processing Systems, 36:23047–23074, 2023.

Anastasios Nikolas Angelopoulos, Stephen Bates, Michael Jordan, and Jitendra Malik. Uncertainty sets for image classifiers using conformal prediction. In International Conference on Learning Representations, 2021. URL https://openreview.net/forum?id=eNdiU_DbM9.

Anastasios Nikolas Angelopoulos, Rina Barber, and Stephen Bates. Online conformal prediction with decaying step sizes. In Forty-first International Conference on Machine Learning, 2024. URL https://openreview.net/forum?id=2XkRIijUKw.

Felipe Areces, Christopher Mohri, Tatsunori Hashimoto, and John Duchi. Online conformal prediction via online optimization. In Forty-second International Conference on Machine Learning, 2025.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4):962–973, 2005.

Rina Foygel Barber, Emmanuel J Candès, Aaditya Ramdas, and Ryan J Tibshirani. Conformal prediction beyond exchangeability. The Annals of Statistics, 51(2):816–845, 2023.

Gaurav Batra and Naresh Manwani. Multiclass classification using dilute bandit feedback. In Pacific Rim International Conference on Artificial Intelligence, pp. 63–75. Springer, 2021.

Aadyot Bhatnagar, Huan Wang, Caiming Xiong, and Yu Bai. Improved online conformal prediction via strongly adaptive online learning. In International Conference on Machine Learning, pp. 2337–2363. PMLR, 2023.

Tiffany Ding, Anastasios Angelopoulos, Stephen Bates, Michael Jordan, and Ryan J Tibshirani. Class-conditional conformal prediction with many classes. Advances in Neural Information Processing Systems, 36:64555–64576, 2023.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In Proceedings of the 28th International Conference on Machine Learning, pp. 1097–1104, 2011.

Haosen Ge, Hamsa Bastani, and Osbert Bastani. Stochastic online conformal prediction with semi-bandit feedback. In Forty-second International Conference on Machine Learning, 2025. URL https://openreview.net/forum?id=IdRrKDsTZ8.

Isaac Gibbs and Emmanuel J Candès. Adaptive conformal inference under distribution shift. Advances in Neural Information Processing Systems, 34:1660–1672, 2021.

Isaac Gibbs and Emmanuel J Candès. Conformal inference for online prediction with arbitrary distribution shifts. Journal of Machine Learning Research, 25(162):1–36, 2024.

Sreenivas Gollapudi, Kostas Kollias, Kamesh Munagala, and Ali Sinop. Efficient online conformal selection with limited feedback. arXiv preprint arXiv:2605.14953, 2026.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(260):663–685, 1952.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. Journal of the Royal Statistical Society Series B: Statistical Methodology, 76(1):71–96, 2014.

Jing Lei, James Robins, and Larry Wasserman. Distribution-free prediction sets. Journal of the American Statistical Association, 108(501):278–287, 2013.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113 (523):1094–1111, 2018.

Mauricio Sadinle, Jing Lei, and Larry Wasserman. Least ambiguous set-valued classifiers with bounded error levels. Journal of the American Statistical Association, 114(525):223–234, 2019.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. Journal of Machine Learning Research, 9:371–421, 2008.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. Algorithmic learning in a random world. Springer, 2005.

Bowen Wang, Matteo Zecchin, and Osvaldo Simeone. Online conformal prediction with corrupted feedback. arXiv preprint arXiv:2605.20515, 2026.

Zhou Wang and Xingye Qiao. Set-valued classification with out-of-distribution detection for many classes. Journal of Machine Learning Research, 24(375):1–39, 2023.

Zhou Wang and Xingye Qiao. Efficient online set-valued classification with bandit feedback. In Forty-first International Conference on Machine Learning, 2024. URL https://openreview.net/forum?id=R1auM3tLPE.

Zhou Wang and Xingye Qiao. Generalized prediction set with bandit feedback. Transactions on Machine Learning Research, 2025.

Junyoung Yang, Kyungmin Kim, and Sangdon Park. Online conformal prediction with adversarial semi-bandit feedback via regret minimization. In The Fourteenth International Conference on Learning Representations, 2026. URL https://openreview.net/forum?id=RMWcdp5IUy.

Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In International conference on machine learning, pp. 25834–25866. PMLR, 2022.

A Derivations

A.1 The Augmentation Term

Eq. (3) subtracts $\widehat{R}_{t,i}$ from the observed bit before reweighting it. For the notation simplicity here and in the proofs, we use

$$w^k_{t,i} := \big(Z^k_{t,i} - \rho^k_{t,i}(\tau^k_{t-1})\big) \big/ \big(\rho^k_{t,i}(\tau^k_{t-1})\big[1 - \rho^k_{t,i}(\tau^k_{t-1})\big]\big)$$

to denote the reweighting factor in Eq. (3), and $R_{t,i} := \mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\}$ the observed bit, and hence $\Delta^k_{t,i} = w^k_{t,i}(R_{t,i} - \widehat{R}_{t,i})$. The unbiasedness argument in the proof of Proposition 2 uses no property of $\widehat{R}_{t,i}$ beyond its being fixed given $\boldsymbol{X}_{t,i}$ and $\mathcal{F}_{t-1}$. Unbiasedness therefore places no constraint on the choice of $\widehat{R}_{t,i}$, which is the property a control variate exploits (Bang & Robins, 2005; Dudík et al., 2011): what remains is a variance question.

Note that the bit is also the truth's own draw, i.e., $R_{t,i} = Z^{Y_{t,i}}_{t,i}$, since $\widehat{\mathcal{C}}_{t,i}$ contains a class exactly when that class's draw is 1. The draws are independent of $Y_{t,i}$ given $\boldsymbol{X}_{t,i}$ and $\mathcal{F}_{t-1}$, so

$$\begin{aligned}
\mathbb{E}\big[R_{t,i} \mid \boldsymbol{X}_{t,i}, \mathcal{F}_{t-1}\big] &= \mathbb{E}\Big[\sum_{j\in\mathcal{Y}} \mathbb{1}\{Y_{t,i} = j\}\, Z^j_{t,i} \;\Big|\; \boldsymbol{X}_{t,i}, \mathcal{F}_{t-1}\Big] \\
&= \sum_{j\in\mathcal{Y}} \mathbb{P}(Y_{t,i} = j \mid \boldsymbol{X}_{t,i})\, \mathbb{E}\big[Z^j_{t,i} \;\big|\; \boldsymbol{X}_{t,i}, \mathcal{F}_{t-1}\big] \\
&= \sum_{j\in\mathcal{Y}} \mathbb{P}(Y_{t,i} = j \mid \boldsymbol{X}_{t,i})\, \rho^j_{t,i}(\tau^j_{t-1}).
\end{aligned}$$

The last sum is unavailable in practice because we do not know $\mathbb{P}(Y_{t,i} = j \mid \boldsymbol{X}_{t,i})$, which is what the base model is trying to learn. Replacing it by the previous step's softmax $\hat{p}(j \mid \boldsymbol{X}_{t,i})$ gives exactly the $\widehat{R}_{t,i}$ of Eq. (3).

A.2 The Threshold Update

In this subsection, we derive Eq. (7), the updating rule for the threshold τ^k_t. Every expectation below is over the step-t data $(\boldsymbol{X}_{t,i}, Y_{t,i})$ and the inclusion draws, conditional on $\mathcal{F}_{t-1}$; where an expectation conditions on more, the extra conditioning is shown. Given a class k, recall from Eq. (2) that $\rho^k_{t,i}(\tau)$ is the inclusion probability at an arbitrary threshold τ, determined by $\boldsymbol{X}_{t,i}$ and $\mathcal{F}_{t-1}$, and nonincreasing in τ.

The calibration condition. Eq. (6) states the requirement at the current threshold τ^k_{t-1}, already in indicator form over the constant denominator $\mathbb{P}(Y_{t,i} = k) > 0$. The threshold we target is the one meeting it at a free τ, and clearing that constant leaves the solution set unchanged, so we seek a root of

$$\mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\,\big([1 - \rho^k_{t,i}(\tau)] - \alpha\big)\big] = 0. \tag{8}$$

An objective with the same solution. Eq. (8) is a condition to solve, not something a learner can descend. We therefore express its left-hand side as the derivative of an objective and solve the equivalent problem

$$\min_{\tau\in\mathbb{R}} \mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\, L^k_{t,i}(\tau)\big], \qquad L^k_{t,i}(\tau) := \int_0^\tau \big([1 - \rho^k_{t,i}(u)] - \alpha\big)\, du. \tag{9}$$

Since $\rho^k_{t,i}$ is nonincreasing, $L^k_{t,i}(\tau)$ is convex, and so is the objective of Eq. (9), which is therefore minimized precisely where its derivative equals 0. That derivative is the left-hand side of Eq. (8), so the minimizers in Eq. (9) are the roots of Eq. (8).

The gradient is not observable. The common way to solve the objective in Eq. (9) is gradient descent. This requires its derivative $\mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}([1 - \rho^k_{t,i}(\tau)] - \alpha)\big]$, which the learner cannot evaluate: it is an expectation over the unknown data distribution, and the label indicator inside it is hidden under diluted feedback. Proposition 2 gives $\mathbb{1}\{Y_{t,i} = k\} = \mathbb{E}[\Delta^k_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}]$, and

$[1-\rho^k_{t,i}(\tau)]-\alpha$ is fixed given $\boldsymbol{X}_{t,i}, \mathcal{F}_{t-1}$, hence we have

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}\{Y_{t,i}=k\}\big([1-\rho^k_{t,i}(\tau)]-\alpha\big)\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[\Delta^k_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}\big]\big([1-\rho^k_{t,i}(\tau)]-\alpha\big)\Big] && \text{Proposition 2} \\
&= \mathbb{E}\big[\Delta^k_{t,i}\big([1-\rho^k_{t,i}(\tau)]-\alpha\big)\big]. && \text{double expectation}
\end{aligned}$$

So $\Delta^k_{t,i}\big([1-\rho^k_{t,i}(\tau)]-\alpha\big)$ is an unbiased estimate of the derivative, and at the current threshold $\tau=\tau^k_{t-1}$ it is $\Delta^k_{t,i}\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)$, which the learner computes from the single bit.

Update. Averaging these estimates over the batch gives one that is still unbiased for the derivative, so the mini-batch gradient step is

$$\tau^k_t = \tau^k_{t-1} - \eta_2 \frac{1}{|\mathcal{B}_t|}\sum_{i\in\mathcal{B}_t}\Delta^k_{t,i}\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big),$$

which is Eq. (7).

B Experimental Details and Additional Results

B.1 Setup

We conduct experiments on CIFAR-10 (10 classes), CIFAR-100 (20 super classes) and SVHN (10 classes) over $T = 29{,}400$ steps with a batch size of 256. Each instance in the stream is augmented as it is randomly drawn. The base model is a ResNet-50 (He et al., 2016), whose backbone starts from ImageNet weights and whose classification head is randomly initialized, and we use $\eta_1 = 10^{-4}$ for the base model and $\eta_2 = 2\times10^{-3}$ (3×10^{-3} on SVHN) for the thresholds. The target miscoverage rate is $\alpha = 0.1$, and the inclusion probabilities are built from the RAPS score (Angelopoulos et al., 2021) with softness $h = 0.05$ and floor $c = \alpha/2$, and RAPS regularization $\lambda = 0.01$, $k_{\text{reg}} = 1$. Figure 1 reports 5 replications with different seeds.

B.2 The Augmentation Term on CIFAR-100 and SVHN

Tables 3 and 4 repeat the comparison of Table 2 on CIFAR-100 and SVHN. The augmentation lowers the estimator's variance by factors of $5.7 \sim 6.6$ and $4.2 \sim 4.6$, ends the stream with a prediction set 1.72 and 1.64 times smaller, and leaves the worst class at 0.888 and 0.898 against 0.859 and 0.876 without it.

Table 3: Effect of the augmentation on CIFAR-100.

	marginal coverage		worst-class coverage		set size		variance of $\Delta^k_{t,i}$	
t/T	augmented	0	augmented	0	augmented	0	augmented	0
0.2	0.898±0.001	0.882±0.001	0.879±0.003	0.831±0.010	**10.89**±0.08	15.39±0.20	**1.96**±0.05	12.93±0.11
0.4	0.898±0.001	0.888±0.001	0.885±0.002	0.857±0.008	**9.53**±0.06	14.88±0.14	**2.01**±0.01	12.46±0.14
0.6	0.899±0.000	0.889±0.002	0.887±0.001	0.864±0.006	**8.77**±0.02	14.28±0.19	**1.98**±0.05	11.99±0.12
0.8	0.899±0.000	0.889±0.001	0.886±0.001	0.857±0.007	**8.27**±0.02	13.91±0.12	**2.05**±0.02	11.89±0.14
1.0	0.899±0.000	0.890±0.002	0.888±0.001	0.859±0.004	**7.90**±0.02	13.58±0.13	**2.03**±0.03	11.52±0.14

B.3 Base Model Accuracy

The base model updates sequentially and the set-level bit is the only supervision the model receives online. Table 5 reports the base model's top-1 accuracy on the stream, accumulated through step t: the fraction of all instances so far whose highest-scoring label is the true one, each scored before its own feedback is used. The checkpoints and the two estimators are those of Section B.2, the augmented one and the alternative with $\widehat{R}_{t,i} = 0$. Accuracy rises over the stream on all three datasets, so the diluted bit trains the base model rather than only calibrating the thresholds. The augmentation term is what makes that training effective. For example, on CIFAR-100, the augmented estimator reaches 0.570 where removing the term leaves the model at 0.250.

Table 4: Effect of the augmentation on SVHN.

	marginal coverage		worst-class coverage		set size		variance of $\Delta^k_{t,i}$	
t/T	augmented	0	augmented	0	augmented	0	augmented	0
0.2	0.911±0.001	0.895±0.002	0.898±0.003	0.870±0.002	**2.48**±0.04	4.41±0.14	**3.10**±0.04	13.01±0.81
0.4	0.909±0.001	0.898±0.003	0.899±0.003	0.876±0.005	**1.99**±0.03	3.59±0.10	**3.24**±0.06	13.56±0.59
0.6	0.908±0.001	0.899±0.002	0.898±0.003	0.878±0.003	**1.81**±0.02	3.15±0.08	**3.16**±0.04	14.50±0.19
0.8	0.906±0.001	0.898±0.002	0.898±0.002	0.875±0.005	**1.71**±0.01	2.87±0.07	**3.16**±0.02	14.34±0.48
1.0	0.906±0.001	0.898±0.002	0.898±0.001	0.876±0.003	**1.65**±0.01	2.71±0.05	**3.21**±0.03	14.62±0.68

Table 5: Accumulative base model top-1 accuracy, with and without the augmentation term.

	CIFAR-10		CIFAR-100		SVHN	
t/T	augmented	0	augmented	0	augmented	0
0.2	**0.689**±0.001	0.501±0.010	**0.437**±0.002	0.155±0.013	**0.851**±0.001	0.691±0.033
0.4	**0.733**±0.000	0.552±0.005	**0.496**±0.002	0.190±0.011	**0.884**±0.001	0.767±0.020
0.6	**0.755**±0.000	0.579±0.004	**0.530**±0.002	0.216±0.009	**0.898**±0.001	0.798±0.014
0.8	**0.769**±0.000	0.598±0.003	**0.553**±0.001	0.235±0.007	**0.907**±0.001	0.816±0.011
1.0	**0.780**±0.000	0.612±0.002	**0.570**±0.001	0.250±0.006	**0.913**±0.000	0.827±0.009

B.4 Sensitivity to the Threshold Step Size

Figure 3 sweeps η_2 over a 256× range on CIFAR-10. Accumulative coverage approaches $1-\alpha$ across the sweep. The smaller the step size, the slower the approach: the smallest settings are still above the target when the run ends. Accumulative set size $\mathrm{Asize}(t)$ is larger at both ends of the sweep than in the middle. The per-step column adds the step-by-step behavior: at the largest step the set size oscillates from step to step (as discussed in Gibbs & Candès (2021)), while the other settings hold a relatively stable set size.

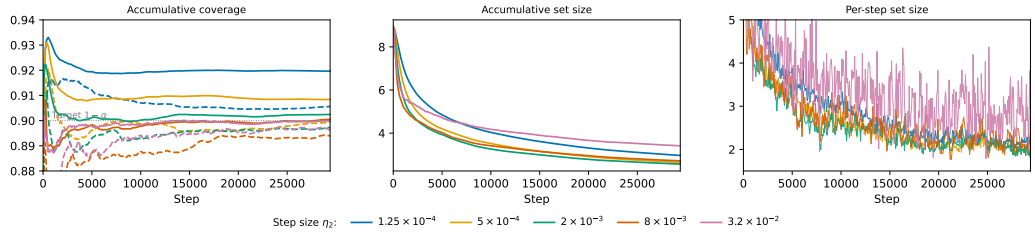

Figure 3: Sensitivity to the threshold step size η_2 on CIFAR-10, all other settings at their defaults. **Accumulative coverage:** marginal (solid) and worst class (dashed), against the target $1-\alpha$ (dotted). **Accumulative set size:** $\mathrm{Asize}(t)$, the running mean of $|\widehat{\mathcal{C}}_{t,i}|$ through step t. **Per-step set size:** $|\widehat{\mathcal{C}}_{t,i}|$ averaged over the step's batch.

C Additional Theorems

Throughout this section and Section D, the batch size is $|\mathcal{B}_t| = B$, so a step's instances are indexed by $i = 1, \ldots, B$. We provide Lemmas 1 and 2, Corollaries 1 and 2, and Proposition 3 below, which are proved in Sections D.4 to D.6. Here $L^k_{t,i}$ is as defined in Eq. (9) of Section A.2, and $\mathcal{T}^k$ is the class-k index set of Theorem 1.

Lemma 1 (Threshold drift bound). *For any comparator $\tau^k_\star \in \mathbb{R}$, define*

$$D_k(T) := \sum_{t=1}^{T} \frac{1}{B} \sum_{i=1}^{B} \Delta^k_{t,i} \left([1 - \rho^k_{t,i}(\tau^k_{t-1})] - \alpha \right) \left(\tau^k_{t-1} - \tau^k_\star \right).$$

For any step size $\eta_2 > 0$, we have

$$\mathbb{E}\big[D_k(T)\big] \le \frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2\, T}{2\, c(1-c)}.$$

Corollary 1 (Calibration regret)**.** *For any comparator $\tau_\star^k \in \mathbb{R}$ and any step size $\eta_2 > 0$,*

$$\mathbb{E}\Big[\frac{1}{BT}\sum_{(t,i)\in\mathcal{T}^k}\big(L_{t,i}^k(\tau_{t-1}^k) - L_{t,i}^k(\tau_\star^k)\big)\Big] \le \frac{(\tau_\star^k)^2}{2\eta_2 T} + \frac{\eta_2}{2\, c(1-c)}.$$

The comparator is arbitrary, so the bound holds in particular for the best fixed threshold. At $\eta_2 = \Theta(T^{-1/2})$ the bound is $O(T^{-1/2})$.

Proposition 3 (Size decomposition)**.** *Fix any benchmark pairs of score function and threshold, i.e., $\{(s_\star^k(\cdot), \tau_\star^k)\}_{k\in\mathcal{Y}}$, which induce the benchmark prediction set $\mathcal{C}_{t,i}^\star$ via Eqs. (1) and (2) with $s_\star^k(\cdot)$ and $\tau_\star^k$ in place of $s_{t-1}^k(\cdot)$ and τ_{t-1}^k. Then, for any $\boldsymbol{X}_{t,i}$,*

$$\Big|\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}_{t,i}^\star| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big]\Big| \le \frac{1}{4h}\sum_{k\in\mathcal{Y}}\Big(\big|s_{t-1}^k(\boldsymbol{X}_{t,i}) - s_\star^k(\boldsymbol{X}_{t,i})\big| + \big|\tau_{t-1}^k - \tau_\star^k\big|\Big).$$

The bound is a comparison, not a guarantee on size: it says the prediction set tracks whichever benchmark it is compared to, and only a calibrated benchmark (with a coverage guarantee) makes that informative.

Lemma 2 (Threshold confinement under a bounded score)**.** *Assume there exists positive constants $\underline{p_k}, k \in \mathcal{Y}$ such that every class recurs, i.e., $\mathbb{P}(Y_{t,i} = k \mid \mathcal{F}_{t-1}) \ge \underline{p_k}$; the scores are uniformly bounded, i.e., $|s_{t-1}^k(\cdot)| \le \bar{s}$; and $0 < c < \alpha \le 1/2$. Define $\gamma := h \ln \frac{1-c}{c}$ and $b_0 := \frac{1-c-\alpha}{c}$. Then for any $\delta \in (0,1)$, with probability at least $1-\delta$, simultaneously for every $t \le T$ and every $k \in \mathcal{Y}$,*

$$\big|\tau_t^k\big| \;\le\; \bar{s} + \gamma + \beta_k, \qquad \textit{where} \quad \beta_k := \eta_2 b_0 + \frac{2\eta_2 b_0^2}{\underline{p_k}(\alpha - c)} \ln \frac{2|\mathcal{Y}|T^2}{\delta}.$$

At $\eta_2 = \Theta(\sqrt{B/T})$, the terms beyond the score range are $O(\sqrt{B/T}\,\log(|\mathcal{Y}|T/\delta))$, so $|\tau_T^k| = O(1)$ with high probability. When we apply a union bound, we can put Theorem 1 and Lemma 2 on one event, and it removes the $|\tau_T^k| = O(1)$ condition from Section 4.3's rate as in the below Corollary 2.

Corollary 2 (Coverage under a bounded score)**.** *Under the assumptions of Lemma 2, for any $\eta_2 > 0$ and $\delta \in (0, 1/2)$, with probability at least $1 - 2\delta$, simultaneously for every class $k \in \mathcal{Y}$,*

$$\left|\frac{1}{|\mathcal{T}^k|}\sum_{(t,i)\in\mathcal{T}^k}\mathbb{1}\{Y_{t,i} \in \widehat{\mathcal{C}}_{t,i}\} - (1-\alpha)\right| \le \frac{B\,(\bar{s} + \gamma + \beta_k)}{\eta_2\,|\mathcal{T}^k|} + \Big(1 + \frac{2}{c}\Big)\frac{\sqrt{2TB\log(4|\mathcal{Y}|/\delta)}}{|\mathcal{T}^k|}.$$

At $\eta_2 = \Theta(\sqrt{B/T})$ and $|\mathcal{T}^k| = \Theta(TB)$, the right-hand side is $O\big(\sqrt{\log(|\mathcal{Y}|/\delta)/(TB)} + \log(|\mathcal{Y}|T/\delta)/T\big)$: the $T^{-1/2}$-type rate with no condition on the threshold.

The RAPS score (Angelopoulos et al., 2021) satisfies the range hypothesis. It subtracts from 1 the sum of a probability mass in $[0,1]$ and the regularization term $\lambda\,(r_k - k_{\mathrm{reg}})_+$, with r_k the rank of class k under $\hat{p}(\cdot \mid \boldsymbol{X}_{t,i})$ in decreasing order and $\lambda > 0, k_{\mathrm{reg}} \ge 1$ its regularization parameters. That sum lies in $[0,\ 1 + \lambda(|\mathcal{Y}| - k_{\mathrm{reg}})]$, so $s_{t-1}^k(\boldsymbol{X}_{t,i}) \in [-\lambda(|\mathcal{Y}| - k_{\mathrm{reg}}),\ 1]$ and $\bar{s} = \max\{1,\ \lambda(|\mathcal{Y}| - k_{\mathrm{reg}})\}$, which is 1 whenever $\lambda \le 1/(|\mathcal{Y}| - k_{\mathrm{reg}})$. At $h = 0.05$ and $c = \alpha/2$ the inclusion probability is then pinned to an endpoint of $[c, 1-c]$ only beyond $\bar{s} + \gamma = 1.147$, and the thresholds observed over training span $[-0.113,\ 0.991]$ on all three datasets.

The non-degeneracy condition of Theorem 2 is met when each class keeps part of its mass in the unclipped band: if $\mathbb{P}\big(c < \sigma((s_{t-1}^k(\boldsymbol{X}_{t,i}) - u)/h) < 1 - c \;\big|\; Y_{t,i} = k, \mathcal{F}_{t-1}\big) \ge q > 0$ for every $u \in \mathcal{J}$ and $\mathbb{P}(Y_{t,i} = k \mid \mathcal{F}_{t-1}) \ge \underline{p_k}$, then $G_{t,k}'(u) \ge \underline{p_k}\, q\, c(1-c)/h$, because $\sigma' > c(1-c)$ on that band. Outside it $\rho_{t,i}^k$ is constant and contributes nothing, so the condition asks about the unclipped share rather than the threshold's range.

D PROOFS

For the notation simplicity, we define:

$$g_{t,i}^k := \Delta_{t,i}^k\big([1-\rho_{t,i}^k(\tau_{t-1}^k)]-\alpha\big),$$
$$\bar{g}_t^k := \frac{1}{B}\sum_{i=1}^{B} g_{t,i}^k,$$

and hence Eq. (7) can be re-written as

$$\tau_t^k = \tau_{t-1}^k - \eta_2 \bar{g}_t^k, \tag{10}$$

and the term defined in Lemma 1 can be re-written as $D_k(T) = \sum_{t=1}^{T} \bar{g}_t^k\,(\tau_{t-1}^k - \tau_\star^k)$.

D.1 IMPOSSIBILITY FOR DETERMINISTIC SETS

Proof of Proposition 1. Let y and y' be the two classes given by the hypothesis. Assume, for contradiction, that a family $(\psi_k)_{k\in\mathcal{Y}}$ formed from $\boldsymbol{X}_{t,i}$ and the feedback bit is unbiased at every $k \in \mathcal{Y}$. Applying the assumption at $k = y$ gives

$$\mathbb{E}\big[\psi_y(\boldsymbol{X}_{t,i}, \mathbb{1}\{Y_{t,i} \in \widetilde{\mathcal{C}}_{t,i}\}) \;\big|\; \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}\big] = \mathbb{1}\{Y_{t,i} = y\}.$$

Given $\boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}$, the set $\widetilde{\mathcal{C}}_{t,i}$ is determined, and with it the bit $\mathbb{1}\{Y_{t,i} \in \widetilde{\mathcal{C}}_{t,i}\}$. Both arguments of ψ_y are then determined, and the conditional expectation collapses to the function value,

$$\psi_y(\boldsymbol{X}_{t,i}, \mathbb{1}\{Y_{t,i} \in \widetilde{\mathcal{C}}_{t,i}\}) = \mathbb{1}\{Y_{t,i} = y\}. \tag{11}$$

Given $\boldsymbol{X}_{t,i}$ and $\mathcal{F}_{t-1}$, the set $\widetilde{\mathcal{C}}_{t,i}$ is determined, while $Y_{t,i}$ remains random. Eq. (11) holds at every value $Y_{t,i}$ can take with positive probability given $\boldsymbol{X}_{t,i}$, in particular at $Y_{t,i} = y$ and at $Y_{t,i} = y'$.

Since $\mathbb{1}\{y \in \widetilde{\mathcal{C}}_{t,i}\} = \mathbb{1}\{y' \in \widetilde{\mathcal{C}}_{t,i}\}$, the bit $\mathbb{1}\{Y_{t,i} \in \widetilde{\mathcal{C}}_{t,i}\}$ takes the same value at $Y_{t,i} = y$ and at $Y_{t,i} = y'$. The function ψ_y therefore returns the same value at both. But Eq. (11) requires that value to be $\mathbb{1}\{y = y\} = 1$ at $Y_{t,i} = y$ and $\mathbb{1}\{y' = y\} = 0$ at $Y_{t,i} = y'$. These contradict each other, so no such family exists. □

D.2 UNBIASED ESTIMATION

Proof of Proposition 2. We use $w_{t,i}^k$, $R_{t,i}$ and $\Delta_{t,i}^k = w_{t,i}^k(R_{t,i} - \widehat{R}_{t,i})$ as they are set up in Section A.1. The augmentation $\widehat{R}_{t,i}$ of Eq. (3) is a convex combination of the $\rho_{t,i}^j(\tau_{t-1}^j) \in [c, 1-c]$, so $\widehat{R}_{t,i} \in [c, 1-c]$, and $R_{t,i} \in \{0, 1\}$ as an indicator, so both lie in $[0, 1]$, and $\rho_{t,i}^k(\tau_{t-1}^k) \in [c, 1-c]$ bounds $|Z_{t,i}^k - \rho_{t,i}^k(\tau_{t-1}^k)|$ by $1-c$ and $\rho_{t,i}^k(\tau_{t-1}^k)\big[1-\rho_{t,i}^k(\tau_{t-1}^k)\big]$ below by $c(1-c)$, so

$$\big|\Delta_{t,i}^k\big| = \big|w_{t,i}^k\big|\,\big|R_{t,i} - \widehat{R}_{t,i}\big| \le \big|w_{t,i}^k\big| \le \frac{1-c}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1-\rho_{t,i}^k(\tau_{t-1}^k)\big]} \le \frac{1-c}{c(1-c)} = \frac{1}{c}. \tag{12}$$

Conditional on $(\boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1})$, both $\rho_{t,i}^k(\tau_{t-1}^k)$ and $\widehat{R}_{t,i}$ are constant, and the draws $Z_{t,i}^k \sim \mathrm{Bernoulli}(\rho_{t,i}^k(\tau_{t-1}^k)), k \in \mathcal{Y}$ are independent. With $R_{t,i} = Z_{t,i}^{Y_{t,i}}$, we have $R_{t,i} = Z_{t,i}^k$ when $k = Y_{t,i}$. When $k \neq Y_{t,i}$, $Z_{t,i}^{Y_{t,i}}$ and $Z_{t,i}^k$ are two different draws, and hence $R_{t,i}$ and $w_{t,i}^k$ are independent.

Unbiasedness Since $\widehat{R}_{t,i}$ is constant and $\mathbb{E}[Z_{t,i}^k - \rho_{t,i}^k(\tau_{t-1}^k) \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] = 0$, we have

$$\mathbb{E}[\Delta_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] = \mathbb{E}[w_{t,i}^k R_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] - \widehat{R}_{t,i}\underbrace{\mathbb{E}[w_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}]}_{=0}$$
$$= \mathbb{E}[w_{t,i}^k R_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}].$$

i) When $k = Y_{t,i}$, $\mathbb{1}\{Y_{t,i} = k\} = 1$. Since $Z_{t,i}^k \in \{0,1\}$ we have $(Z_{t,i}^k)^2 = Z_{t,i}^k$, so

$$w_{t,i}^k R_{t,i} = \frac{\big(Z_{t,i}^k - \rho_{t,i}^k(\tau_{t-1}^k)\big) Z_{t,i}^k}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]} = \frac{Z_{t,i}^k - \rho_{t,i}^k(\tau_{t-1}^k)\, Z_{t,i}^k}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]} = \frac{Z_{t,i}^k}{\rho_{t,i}^k(\tau_{t-1}^k)},$$

and hence

$$\mathbb{E}[w_{t,i}^k R_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] = \frac{\mathbb{E}[Z_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}]}{\rho_{t,i}^k(\tau_{t-1}^k)} = 1 = \mathbb{1}\{Y_{t,i} = k\}.$$

ii) When $k \neq Y_{t,i}$, $\mathbb{1}\{Y_{t,i} = k\} = 0$. Factorizing by the independence of $R_{t,i}$ and $w_{t,i}^k$ gives

$$\begin{aligned}
\mathbb{E}[w_{t,i}^k R_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] &= \mathbb{E}[w_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}]\, \mathbb{E}[R_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] \\
&= \frac{\mathbb{E}[Z_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] - \rho_{t,i}^k(\tau_{t-1}^k)}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]}\, \mathbb{E}[R_{t,i} \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] \\
&= 0 = \mathbb{1}\{Y_{t,i} = k\}.
\end{aligned}$$

In summary, both cases return $\mathbb{E}[\Delta_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] = \mathbb{1}\{Y_{t,i} = k\}$.

Variance bound Since $\mathrm{Var}(\Delta_{t,i}^k \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}) \leq \mathbb{E}[(\Delta_{t,i}^k)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}]$, it suffices to bound this second moment.

i) When $k = Y_{t,i}$, $Z_{t,i}^k \in \{0,1\}$, so $\Delta_{t,i}^k = w_{t,i}^k\big(Z_{t,i}^k - \widehat{R}_{t,i}\big)$ takes exactly two values:

$$Z_{t,i}^k = 1 \implies w_{t,i}^k = \frac{1 - \rho_{t,i}^k(\tau_{t-1}^k)}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]} = \frac{1}{\rho_{t,i}^k(\tau_{t-1}^k)} \implies \Delta_{t,i}^k = \frac{1 - \widehat{R}_{t,i}}{\rho_{t,i}^k(\tau_{t-1}^k)};$$

$$Z_{t,i}^k = 0 \implies w_{t,i}^k = \frac{-\rho_{t,i}^k(\tau_{t-1}^k)}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]} = \frac{-1}{1 - \rho_{t,i}^k(\tau_{t-1}^k)} \implies \Delta_{t,i}^k = \frac{\widehat{R}_{t,i}}{1 - \rho_{t,i}^k(\tau_{t-1}^k)}.$$

With $\widehat{R}_{t,i} \in [0,1]$, both $(1 - \widehat{R}_{t,i})^2 \leq 1$ and $\widehat{R}_{t,i}^2 \leq 1$, so

$$\begin{aligned}
\mathbb{E}[(\Delta_{t,i}^k)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] &= \rho_{t,i}^k(\tau_{t-1}^k) \left(\frac{1 - \widehat{R}_{t,i}}{\rho_{t,i}^k(\tau_{t-1}^k)}\right)^2 + \big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big] \left(\frac{\widehat{R}_{t,i}}{1 - \rho_{t,i}^k(\tau_{t-1}^k)}\right)^2 \\
&= \frac{(1 - \widehat{R}_{t,i})^2}{\rho_{t,i}^k(\tau_{t-1}^k)} + \frac{\widehat{R}_{t,i}^2}{1 - \rho_{t,i}^k(\tau_{t-1}^k)} \\
&\leq \frac{1}{\rho_{t,i}^k(\tau_{t-1}^k)} + \frac{1}{1 - \rho_{t,i}^k(\tau_{t-1}^k)} = \frac{1}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]}.
\end{aligned}$$

ii) When $k \neq Y_{t,i}$, the Bernoulli variance $\mathbb{E}[\big(Z_{t,i}^k - \rho_{t,i}^k(\tau_{t-1}^k)\big)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] = \rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]$ gives

$$\mathbb{E}[(w_{t,i}^k)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] = \frac{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]}{\Big(\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]\Big)^2} = \frac{1}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]},$$

and hence, since $R_{t,i} \in \{0,1\}$ and $\widehat{R}_{t,i} \in [0,1]$ give $(R_{t,i} - \widehat{R}_{t,i})^2 \leq 1$ pointwise,

$$\begin{aligned}
\mathbb{E}[(\Delta_{t,i}^k)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] &= \mathbb{E}[(w_{t,i}^k)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}]\, \mathbb{E}[(R_{t,i} - \widehat{R}_{t,i})^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] \\
&\leq \mathbb{E}[(w_{t,i}^k)^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}] \\
&= \frac{1}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]}.
\end{aligned}$$

In summary, since $\rho_{t,i}^k(\tau_{t-1}^k) \in [c, 1-c]$ gives $\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big] \geq c(1-c)$, both cases return

$$\mathbb{E}\big[(\Delta_{t,i}^k)^2 \,\big|\, \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}\big] \leq \frac{1}{\rho_{t,i}^k(\tau_{t-1}^k)\big[1 - \rho_{t,i}^k(\tau_{t-1}^k)\big]} \leq \frac{1}{c(1-c)}. \tag{13}$$

□

D.3 CLASS-SPECIFIC COVERAGE

Proof of Theorem 1. Fix an arbitrary class $k \in \mathcal{Y}$.

Drift Telescoping Eq. (10) over $t = 1, \ldots, T$ and using the initialization $\tau_0^k = 0$ gives

$$\tau_T^k = \tau_0^k - \eta_2 \sum_{t=1}^{T} \bar{g}_t^k = -\frac{\eta_2}{B} \sum_{t=1}^{T} \sum_{i=1}^{B} g_{t,i}^k.$$

Rearranging gives

$$\Big| \sum_{t=1}^{T} \sum_{i=1}^{B} g_{t,i}^k \Big| = \frac{B\,|\tau_T^k|}{\eta_2}.$$

Estimator fluctuation Define

$$M_{t,i}^k := \big(\Delta_{t,i}^k - \mathbb{1}\{Y_{t,i} = k\}\big)\big([1 - \rho_{t,i}^k(\tau_{t-1}^k)] - \alpha\big).$$

We plan to apply Azuma–Hoeffding to $\{M_{t,i}^k\}$ in streaming order, which needs a filtration indexed by the pair (t, i). Let $\mathcal{G}_{t,i}$ collect three sources: i) the $\mathcal{F}_{t-1}$ as defined in Section 4, ii) the data and inclusion draws of the instances $1, \ldots, i-1$ in step t, and iii) the instance data $(\boldsymbol{X}_{t,i}, Y_{t,i})$, but not its own draws $\{Z_{t,i}^j\}_{j \in \mathcal{Y}}$.

The inclusion draws are independent across instances given $\mathcal{F}_{t-1}$, so Proposition 2's unbiasedness still holds under the finer conditioning, i.e., $\mathbb{E}[\Delta_{t,i}^k \mid \mathcal{G}_{t,i}] = \mathbb{1}\{Y_{t,i} = k\}$. Since $[1 - \rho_{t,i}^k(\tau_{t-1}^k)] - \alpha$ is $\mathcal{G}_{t,i}$-measurable,

$$\mathbb{E}[M_{t,i}^k \mid \mathcal{G}_{t,i}] = \big([1 - \rho_{t,i}^k(\tau_{t-1}^k)] - \alpha\big)\,\mathbb{E}\big[\Delta_{t,i}^k - \mathbb{1}\{Y_{t,i} = k\} \;\big|\; \mathcal{G}_{t,i}\big] = 0$$

so $\{M_{t,i}^k\}$ is a martingale difference sequence with respect to $\{\mathcal{G}_{t,i}\}$.

Additionally, since $\rho_{t,i}^k(\tau_{t-1}^k) \in [c, 1-c]$ and $\alpha \in (0,1)$ put $[1 - \rho_{t,i}^k(\tau_{t-1}^k)] - \alpha$ in $(-1, 1)$,

$$\begin{aligned}
|M_{t,i}^k| &= \big|\Delta_{t,i}^k - \mathbb{1}\{Y_{t,i} = k\}\big|\,\big|[1 - \rho_{t,i}^k(\tau_{t-1}^k)] - \alpha\big| \\
&\le \big|\Delta_{t,i}^k - \mathbb{1}\{Y_{t,i} = k\}\big| \times 1 \\
&\le \frac{1}{c} + 1 \qquad \text{(14a)} \\
&\le \frac{2}{c}, \qquad \text{(14b)}
\end{aligned}$$

where Ineq. (14a) holds by Eq. (12) with $\mathbb{1}\{Y_{t,i} = k\} \le 1$, and Ineq. (14b) holds because $c \le \alpha < 1$ gives $1 \le 1/c$. By applying Azuma–Hoeffding, with probability at least $1 - \delta/(2|\mathcal{Y}|)$, we have

$$\Big| \sum_{t=1}^{T} \sum_{i=1}^{B} M_{t,i}^k \Big| \le \frac{2}{c} \sqrt{2TB \log(4|\mathcal{Y}|/\delta)}.$$

Therefore, with probability at least $1-\delta/(2|\mathcal{Y}|)$, we have

$$
\begin{aligned}
&\Big|\sum_{(t,i)\in\mathcal{T}^k}\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)\Big| \\
=&\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}\mathbb{1}\{Y_{t,i}=k\}\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)\Big| && \text{definition of } \mathcal{T}^k \\
=&\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}\big[\mathbb{1}\{Y_{t,i}=k\}-\Delta^k_{t,i}+\Delta^k_{t,i}\big]\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)\Big| \\
=&\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}g^k_{t,i}\;-\;\sum_{t=1}^{T}\sum_{i=1}^{B}M^k_{t,i}\Big| && \text{definitions of } g^k_{t,i}, M^k_{t,i} \\
\le&\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}g^k_{t,i}\Big|+\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}M^k_{t,i}\Big| \\
\le&\frac{B\,|\tau^k_T|}{\eta_2}+\frac{2}{c}\sqrt{2TB\log(4|\mathcal{Y}|/\delta)}. && \text{drift; Azuma–Hoeffding}
\end{aligned}
$$

Sampling fluctuation Define $V^k_{t,i}:=\mathbb{1}\{Y_{t,i}=k\}\big(Z^k_{t,i}-\rho^k_{t,i}(\tau^k_{t-1})\big)$, thus

$$
\begin{aligned}
\mathbb{E}\big[V^k_{t,i}\;\big|\;\mathcal{G}_{t,i}\big]&=\mathbb{1}\{Y_{t,i}=k\}\,\mathbb{E}\big[Z^k_{t,i}-\rho^k_{t,i}(\tau^k_{t-1})\;\big|\;\mathcal{G}_{t,i}\big]\\
&=\mathbb{1}\{Y_{t,i}=k\}\big(\rho^k_{t,i}(\tau^k_{t-1})-\rho^k_{t,i}(\tau^k_{t-1})\big)=0,
\end{aligned}
$$

so $\{V^k_{t,i}\}$ is a martingale difference sequence with respect to $\{\mathcal{G}_{t,i}\}$, with $|V^k_{t,i}|\le 1$. Therefore, with probability at least $1-\delta/(2|\mathcal{Y}|)$, we have

$$
\begin{aligned}
\Big|\sum_{(t,i)\in\mathcal{T}^k}\big(\mathbb{1}\{Y_{t,i}\in\widehat{\mathcal{C}}_{t,i}\}-\rho^k_{t,i}(\tau^k_{t-1})\big)\Big|&=\Big|\sum_{(t,i)\in\mathcal{T}^k}\big(Z^k_{t,i}-\rho^k_{t,i}(\tau^k_{t-1})\big)\Big|\\
&=\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}V^k_{t,i}\Big|\le\sqrt{2TB\log(4|\mathcal{Y}|/\delta)},
\end{aligned}
$$

where the first equality holds because $\mathbb{1}\{Y_{t,i}\in\widehat{\mathcal{C}}_{t,i}\}=Z^k_{t,i}$ on $\mathcal{T}^k$, and the last inequality follows from Azuma–Hoeffding.

Union bound The $2|\mathcal{Y}|$ failure events have total probability at most $2|\mathcal{Y}|\cdot\delta/(2|\mathcal{Y}|)=\delta$, so with probability at least $1-\delta$, simultaneously for all classes $k\in\mathcal{Y}$,

$$
\begin{aligned}
\Big|\sum_{(t,i)\in\mathcal{T}^k}\big(\mathbb{1}\{Y_{t,i}\in\widehat{\mathcal{C}}_{t,i}\}-(1-\alpha)\big)\Big|&=\Big|\sum_{(t,i)\in\mathcal{T}^k}\Big[\big(Z^k_{t,i}-\rho^k_{t,i}(\tau^k_{t-1})\big)-\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)\Big]\Big|\\
&\le\Big|\sum_{(t,i)\in\mathcal{T}^k}\big(Z^k_{t,i}-\rho^k_{t,i}(\tau^k_{t-1})\big)\Big|+\Big|\sum_{(t,i)\in\mathcal{T}^k}\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)\Big|\\
&\le\sqrt{2TB\log(4|\mathcal{Y}|/\delta)}+\frac{B\,|\tau^k_T|}{\eta_2}+\frac{2}{c}\sqrt{2TB\log(4|\mathcal{Y}|/\delta)}\\
&=\frac{B\,|\tau^k_T|}{\eta_2}+\Big(1+\frac{2}{c}\Big)\sqrt{2TB\log(4|\mathcal{Y}|/\delta)},
\end{aligned}
$$

where the first equality holds because $Z^k_{t,i}=\mathbb{1}\{Y_{t,i}\in\widehat{\mathcal{C}}_{t,i}\}$ on $\mathcal{T}^k$. Dividing by $|\mathcal{T}^k|$, for every class at once,

$$
\Big|\frac{1}{|\mathcal{T}^k|}\sum_{(t,i)\in\mathcal{T}^k}\mathbb{1}\{Y_{t,i}\in\widehat{\mathcal{C}}_{t,i}\}-(1-\alpha)\Big|\le\frac{B\,|\tau^k_T|}{\eta_2\,|\mathcal{T}^k|}+\Big(1+\frac{2}{c}\Big)\frac{\sqrt{2TB\log(4|\mathcal{Y}|/\delta)}}{|\mathcal{T}^k|}.
$$

□

D.4 CALIBRATION REGRET

Proof of Lemma 1. Since $\rho^k_{t,i}(\tau^k_{t-1}) \in [c, 1-c]$ and $\alpha \in (0,1)$ put $[1-\rho^k_{t,i}(\tau^k_{t-1})] - \alpha$ in $(-1,1)$, we bound the second moment of $\bar{g}^k_t$ as follows:

$$\begin{aligned}
\mathbb{E}\big[(\bar{g}^k_t)^2\big] = \mathbb{E}\Big[\Big(\frac{1}{B}\sum_{i=1}^{B} g^k_{t,i}\Big)^2\Big] &\le \mathbb{E}\Big[\frac{1}{B}\sum_{i=1}^{B}(g^k_{t,i})^2\Big] && \text{convexity}\\
&= \frac{1}{B}\sum_{i=1}^{B}\mathbb{E}\big[(g^k_{t,i})^2\big]\\
&= \frac{1}{B}\sum_{i=1}^{B}\mathbb{E}\Big[(\Delta^k_{t,i})^2\big([1-\rho^k_{t,i}(\tau^k_{t-1})]-\alpha\big)^2\Big] && \text{definition of } g^k_{t,i}\\
&\le \frac{1}{B}\sum_{i=1}^{B}\mathbb{E}\Big[\mathbb{E}\big[(\Delta^k_{t,i})^2 \mid \boldsymbol{X}_{t,i}, Y_{t,i}, \mathcal{F}_{t-1}\big]\Big] && \text{second factor } \le 1\\
&\le \frac{1}{c(1-c)}. && \text{Eq. (13)} \qquad (15)
\end{aligned}$$

For any fixed comparator $\tau^k_\star \in \mathbb{R}$, we have

$$\begin{aligned}
(\tau^k_t - \tau^k_\star)^2 &= \big(\tau^k_{t-1} - \tau^k_\star - \eta_2 \bar{g}^k_t\big)^2\\
&= (\tau^k_{t-1} - \tau^k_\star)^2 - 2\eta_2\, \bar{g}^k_t\, (\tau^k_{t-1} - \tau^k_\star) + \eta_2^2 (\bar{g}^k_t)^2,
\end{aligned}$$

which gives

$$\bar{g}^k_t\, (\tau^k_{t-1} - \tau^k_\star) = \frac{(\tau^k_{t-1} - \tau^k_\star)^2 - (\tau^k_t - \tau^k_\star)^2}{2\eta_2} + \frac{\eta_2}{2}(\bar{g}^k_t)^2.$$

Thus,

$$\begin{aligned}
D_k(T) = \sum_{t=1}^{T} \bar{g}^k_t\, (\tau^k_{t-1} - \tau^k_\star) &= \frac{(\tau^k_0 - \tau^k_\star)^2 - (\tau^k_T - \tau^k_\star)^2}{2\eta_2} + \frac{\eta_2}{2}\sum_{t=1}^{T}(\bar{g}^k_t)^2\\
&\le \frac{(\tau^k_0 - \tau^k_\star)^2}{2\eta_2} + \frac{\eta_2}{2}\sum_{t=1}^{T}(\bar{g}^k_t)^2\\
&= \frac{(\tau^k_\star)^2}{2\eta_2} + \frac{\eta_2}{2}\sum_{t=1}^{T}(\bar{g}^k_t)^2,
\end{aligned}$$

where the final equality uses the initialization $\tau^k_0 = 0$. Taking expectations and applying Eq. (15) to each term, we have

$$\mathbb{E}\big[D_k(T)\big] \le \frac{(\tau^k_\star)^2}{2\eta_2} + \frac{\eta_2}{2}\sum_{t=1}^{T}\frac{1}{c(1-c)} = \frac{(\tau^k_\star)^2}{2\eta_2} + \frac{\eta_2\, T}{2\, c(1-c)}.$$

□

Proof of Corollary 1. $L^k_{t,i}(\tau)$ of Eq. (9) is convex by Section A.2, with derivative $L^{k\,\prime}_{t,i}(\tau) = [1-\rho^k_{t,i}(\tau)] - \alpha$. This implies that for any comparator $\tau^k_\star$, we have

$$L^k_{t,i}(\tau^k_{t-1}) - L^k_{t,i}(\tau^k_\star) \le \big([1-\rho^k_{t,i}(\tau^k_{t-1})] - \alpha\big)\,(\tau^k_{t-1} - \tau^k_\star).$$

Therefore, multiplying both sides by $\mathbb{1}\{Y_{t,i} = k\}$, then taking summation and expectations, leads to

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{i=1}^{B}\mathbb{E}\big[\mathbb{1}\{Y_{t,i}=k\}\big(L_{t,i}^k(\tau_{t-1}^k) - L_{t,i}^k(\tau_\star^k)\big)\big] \\
&\qquad \le \sum_{t=1}^{T}\sum_{i=1}^{B}\mathbb{E}\big[g_{t,i}^k\,(\tau_{t-1}^k - \tau_\star^k)\big] && \text{Proposition 2, double expectation} \\
&\qquad = B\sum_{t=1}^{T}\mathbb{E}\big[\bar{g}_t^k\,(\tau_{t-1}^k - \tau_\star^k)\big] && \text{def. of } \bar{g}_t^k \\
&\qquad = B\,\mathbb{E}[D_k(T)] \\
&\qquad \le B\Big(\frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)}\Big) && \text{Lemma 1,}
\end{aligned}$$

and dividing by BT,

$$\mathbb{E}\Big[\frac{1}{BT}\sum_{(t,i)\in\mathcal{T}^k}\big(L_{t,i}^k(\tau_{t-1}^k) - L_{t,i}^k(\tau_\star^k)\big)\Big] \le \frac{(\tau_\star^k)^2}{2\eta_2 T} + \frac{\eta_2}{2\,c(1-c)}.$$

□

D.5 Prediction Set Size

Proof of Proposition 3. By Eq. (1) and $|\widehat{\mathcal{C}}_{t,i}| = \sum_{k\in\mathcal{Y}} Z_{t,i}^k$, we have

$$\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big] = \sum_{k\in\mathcal{Y}}\mathbb{E}\big[Z_{t,i}^k \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big] = \sum_{k\in\mathcal{Y}}\rho_{t,i}^k(\tau_{t-1}^k). \tag{16}$$

Applying the same analysis to $\mathcal{C}_{t,i}^\star$, whose inclusion probability $\rho_{\star,t,i}^k(\cdot)$ is given by Eq. (2) with $s_\star^k$ in place of s_{t-1}^k, gives

$$\mathbb{E}\big[\,|\mathcal{C}_{t,i}^\star| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big] = \sum_{k\in\mathcal{Y}}\rho_{\star,t,i}^k(\tau_\star^k). \tag{17}$$

Define the sigmoid function $\sigma(u) := 1/(1+e^{-u})$, and two margins $m := s_{t-1}^k(\boldsymbol{X}_{t,i}) - \tau_{t-1}^k$ and $m_\star := s_\star^k(\boldsymbol{X}_{t,i}) - \tau_\star^k$. By Eq. (2) and its benchmark counterpart,

$$\begin{aligned}
&\big|\rho_{t,i}^k(\tau_{t-1}^k) - \rho_{\star,t,i}^k(\tau_\star^k)\big| \\
&= \big|\mathrm{clip}(\sigma(m/h), c, 1-c) - \mathrm{clip}(\sigma(m_\star/h), c, 1-c)\big| && \text{Eq. (2)} \\
&\le \big|\sigma(m/h) - \sigma(m_\star/h)\big| && \text{clip is 1-Lipschitz} \\
&\le \frac{1}{4}\Big|\frac{m}{h} - \frac{m_\star}{h}\Big| = \frac{|m-m_\star|}{4h} && \sigma'(u) \le 1/4 \\
&\le \frac{1}{4h}\Big(\big|s_{t-1}^k(\boldsymbol{X}_{t,i}) - s_\star^k(\boldsymbol{X}_{t,i})\big| + \big|\tau_{t-1}^k - \tau_\star^k\big|\Big). && \text{triangle inequality} \qquad (18)
\end{aligned}$$

Combining Eqs. (16) to (18),

$$\begin{aligned}
\Big|\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big] - \mathbb{E}\big[\,|\mathcal{C}_{t,i}^\star| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big]\Big| &= \Big|\sum_{k\in\mathcal{Y}}\big(\rho_{t,i}^k(\tau_{t-1}^k) - \rho_{\star,t,i}^k(\tau_\star^k)\big)\Big| \\
&\le \sum_{k\in\mathcal{Y}}\big|\rho_{t,i}^k(\tau_{t-1}^k) - \rho_{\star,t,i}^k(\tau_\star^k)\big| \\
&\le \frac{1}{4h}\sum_{k\in\mathcal{Y}}\Big(\big|s_{t-1}^k(\boldsymbol{X}_{t,i}) - s_\star^k(\boldsymbol{X}_{t,i})\big| + \big|\tau_{t-1}^k - \tau_\star^k\big|\Big).
\end{aligned}$$

□

Proof of Theorem 2. Fix a step t and an instance i. Conditioning on $\mathcal{F}_{t-1}$ and $\boldsymbol{X}_{t,i}$ brings Proposition 3 to bear on the instance-level size gap,

$$\begin{aligned}
\Big|\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}^\star_{t,i}|\big]\Big| &= \Big|\mathbb{E}\Big[\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}^\star_{t,i}| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big]\Big]\Big| && \text{double expectation}\\
&\le \mathbb{E}\Big[\Big|\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}^\star_{t,i}| \;\big|\; \mathcal{F}_{t-1}, \boldsymbol{X}_{t,i}\big]\Big|\Big] && |\mathbb{E}[\cdot]| \le \mathbb{E}|\cdot|\\
&\le \frac{1}{4h}\,\mathbb{E}\Big[\sum_{k\in\mathcal{Y}}\Big(\big|s^k_{t-1}(\boldsymbol{X}_{t,i}) - s^k_\star(\boldsymbol{X}_{t,i})\big| + \big|\tau^k_{t-1} - \tau^k_\star\big|\Big)\Big]. && \text{Proposition 3}
\end{aligned} \tag{19}$$

Averaging over the TB instances,

$$\begin{aligned}
\frac{1}{TB}\Big|\sum_{t=1}^{T}\sum_{i=1}^{B}\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}^\star_{t,i}|\big]\Big| &\le \frac{1}{TB}\sum_{t=1}^{T}\sum_{i=1}^{B}\Big|\mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}^\star_{t,i}|\big]\Big| && \text{triangle inequality}\\
&\le \frac{1}{4h}\Big[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t + \sum_{k\in\mathcal{Y}}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big|\tau^k_{t-1} - \tau^k_\star\big|\Big], && \text{Eq. (19)}
\end{aligned} \tag{20}$$

where τ^k_{t-1} does not depend on i so its inner sum contributes a factor B, and Cauchy–Schwarz bounds the score-related term by ε_t of Theorem 2:

$$\mathbb{E}\Big[\sum_{k\in\mathcal{Y}}\big|s^k_{t-1}(\boldsymbol{X}_{t,i}) - s^k_\star(\boldsymbol{X}_{t,i})\big|\Big] \le \Big(\mathbb{E}\Big[\Big(\sum_{k\in\mathcal{Y}}\big|s^k_{t-1}(\boldsymbol{X}_{t,i}) - s^k_\star(\boldsymbol{X}_{t,i})\big|\Big)^2\Big]\Big)^{1/2} = \varepsilon_t.$$

Now it remains to bound the threshold-related difference. Define $d^k_t := \tau^k_{t-1} - \tau^k_\star$, $S_{t,i} := \sum_{k\in\mathcal{Y}}\big|s^k_{t-1}(\boldsymbol{X}_{t,i}) - s^k_\star(\boldsymbol{X}_{t,i})\big|$ and $\widetilde{\varepsilon}_t := \mathbb{E}[S_{t,i} \mid \mathcal{F}_{t-1}]$, so that $\varepsilon_t^2 = \mathbb{E}[S^2_{t,i}]$, and let $G_{t,k}$ be the calibration map of Theorem 2. The benchmark is a fixed pair $(s^k_\star, \tau^k_\star)$, so $\rho^k_{\star,t,i}(\tau^k_\star)$ is independent of $\mathcal{F}_{t-1}$, and calibration $\mathbb{P}(Y_{t,i} \in \mathcal{C}^\star_{t,i} \mid Y_{t,i} = k) = 1 - \alpha$ gives

$$\begin{aligned}
\mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\big([1 - \rho^k_{\star,t,i}(\tau^k_\star)] - \alpha\big) \;\big|\; \mathcal{F}_{t-1}\big] &= \mathbb{P}(Y_{t,i} = k)\big(1 - \mathbb{E}[\rho^k_{\star,t,i}(\tau^k_\star) \mid Y_{t,i} = k] - \alpha\big)\\
&= \mathbb{P}(Y_{t,i} = k)\big(1 - (1-\alpha) - \alpha\big) = 0.
\end{aligned} \tag{21}$$

Subtracting Eq. (21) from the definition of $G_{t,k}$ gives

$$\begin{aligned}
\big|G_{t,k}(\tau^k_\star)\big| &= \Big|\mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\big([1 - \rho^k_{t,i}(\tau^k_\star)] - \alpha\big) \;\big|\; \mathcal{F}_{t-1}\big]\\
&\qquad - \mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\big([1 - \rho^k_{\star,t,i}(\tau^k_\star)] - \alpha\big) \;\big|\; \mathcal{F}_{t-1}\big]\Big|\\
&= \Big|\mathbb{E}\big[\mathbb{1}\{Y_{t,i} = k\}\big(\rho^k_{\star,t,i}(\tau^k_\star) - \rho^k_{t,i}(\tau^k_\star)\big) \;\big|\; \mathcal{F}_{t-1}\big]\Big| && \text{Eq. (21)}\\
&\le \mathbb{E}\big[\big|\rho^k_{t,i}(\tau^k_\star) - \rho^k_{\star,t,i}(\tau^k_\star)\big| \;\big|\; \mathcal{F}_{t-1}\big] && \mathbb{1}\{Y_{t,i} = k\} \le 1\\
&\le \frac{1}{4h}\,\mathbb{E}\big[\big|s^k_{t-1}(\boldsymbol{X}_{t,i}) - s^k_\star(\boldsymbol{X}_{t,i})\big| \;\big|\; \mathcal{F}_{t-1}\big] && \text{Eq. (18) at } \tau^k_\star\\
&\le \frac{\widetilde{\varepsilon}_t}{4h}. && \text{definition of } \widetilde{\varepsilon}_t
\end{aligned} \tag{22}$$

By hypothesis $\mathcal{J}$ contains every iterate, so the non-degeneracy condition of Theorem 2 applies at $u = \tau^k_{t-1}$ and gives $\big(G_{t,k}(\tau^k_{t-1}) - G_{t,k}(\tau^k_\star)\big)d^k_t \ge \mu\,(d^k_t)^2$. Hence

$$\begin{aligned}
\mathbb{E}\big[\bar{g}^k_t\, d^k_t \;\big|\; \mathcal{F}_{t-1}\big] &= G_{t,k}(\tau^k_{t-1})\, d^k_t && \text{Proposition 2, double expectation}\\
&= \big[G_{t,k}(\tau^k_{t-1}) - G_{t,k}(\tau^k_\star)\big]d^k_t + G_{t,k}(\tau^k_\star)\, d^k_t\\
&\ge \mu\,(d^k_t)^2 - \big|G_{t,k}(\tau^k_\star)\big|\,|d^k_t| && \text{non-degeneracy}\\
&\ge \mu\,(d^k_t)^2 - \Big(\frac{\mu}{2}(d^k_t)^2 + \frac{\big|G_{t,k}(\tau^k_\star)\big|^2}{2\mu}\Big) && \text{Young, } ab \le \tfrac{\mu}{2}a^2 + \tfrac{b^2}{2\mu}\\
&= \frac{\mu}{2}(d^k_t)^2 - \frac{\big|G_{t,k}(\tau^k_\star)\big|^2}{2\mu}\\
&\ge \frac{\mu}{2}(d^k_t)^2 - \frac{\widetilde{\varepsilon}_t^2}{32h^2\mu}. && \text{Eq. (22)}
\end{aligned} \tag{23}$$

The two score errors are related by

$$\widetilde{\varepsilon}_t^2 = \big(\mathbb{E}[S_{t,i} \mid \mathcal{F}_{t-1}]\big)^2 \le \mathbb{E}\big[S_{t,i}^2 \mid \mathcal{F}_{t-1}\big] \qquad \text{conditional Jensen}$$
$$\Longrightarrow \quad \mathbb{E}\big[\widetilde{\varepsilon}_t^2\big] \le \mathbb{E}\big[\mathbb{E}[S_{t,i}^2 \mid \mathcal{F}_{t-1}]\big] = \mathbb{E}\big[S_{t,i}^2\big] = \varepsilon_t^2. \qquad \text{double expectation}$$

Taking expectations and summing over t,

$$\sum_{t=1}^{T} \mathbb{E}\big[\bar{g}_t^k\, d_t^k\big] \ge \frac{\mu}{2} \sum_{t=1}^{T} \mathbb{E}[(d_t^k)^2] - \frac{1}{32h^2\mu} \sum_{t=1}^{T} \mathbb{E}\big[\widetilde{\varepsilon}_t^2\big] \qquad \text{Eq. (23), summed}$$
$$\ge \frac{\mu}{2} \sum_{t=1}^{T} \mathbb{E}[(d_t^k)^2] - \frac{1}{32h^2\mu} \sum_{t=1}^{T} \varepsilon_t^2. \qquad \mathbb{E}[\widetilde{\varepsilon}_t^2] \le \varepsilon_t^2$$

The left-hand side is $\mathbb{E}[D_k(T)]$, which Lemma 1 bounds from above, so

$$\frac{\mu}{2} \sum_{t=1}^{T} \mathbb{E}[(d_t^k)^2] - \frac{1}{32h^2\mu} \sum_{t=1}^{T} \varepsilon_t^2 \le \sum_{t=1}^{T} \mathbb{E}\big[\bar{g}_t^k\, d_t^k\big] = \mathbb{E}[D_k(T)] \le \frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)},$$

and hence

$$\sum_{t=1}^{T} \mathbb{E}[(d_t^k)^2] \le \frac{2}{\mu}\Big(\frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)}\Big) + \frac{1}{16h^2\mu^2} \sum_{t=1}^{T} \varepsilon_t^2. \tag{24}$$

Furthermore,

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}|\tau_{t-1}^k - \tau_\star^k| = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}|d_t^k| \qquad \text{definition of } d_t^k$$
$$\le \frac{1}{T} \sum_{t=1}^{T} \big(\mathbb{E}[(d_t^k)^2]\big)^{1/2} \qquad \text{Cauchy–Schwarz in } \mathbb{E}$$
$$\le \Big(\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[(d_t^k)^2]\Big)^{1/2} \qquad \text{Cauchy–Schwarz over } t$$
$$\le \Big(\frac{2}{\mu T}\big(\tfrac{(\tau_\star^k)^2}{2\eta_2} + \tfrac{\eta_2 T}{2\,c(1-c)}\big) + \frac{1}{16h^2\mu^2 T} \sum_{t=1}^{T} \varepsilon_t^2\Big)^{1/2} \qquad \text{Eq. (24)}$$
$$\le \sqrt{\frac{2}{\mu T}\Big(\frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)}\Big)} \qquad \sqrt{a+b} \le \sqrt{a} + \sqrt{b}$$
$$+ \frac{1}{4h\mu} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2}.$$

Substituting into Eq. (20), we have

$$\frac{1}{TB}\Big|\sum_{t=1}^{T}\sum_{i=1}^{B} \mathbb{E}\big[\,|\widehat{\mathcal{C}}_{t,i}| - |\mathcal{C}_{t,i}^\star|\big]\Big| \le \frac{1}{4h}\Big[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t + \sum_{k\in\mathcal{Y}}\Big(\sqrt{\frac{2}{\mu T}\Big(\frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)}\Big)}$$
$$+ \frac{1}{4h\mu}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2}\Big)\Big]$$
$$= \frac{1}{4Th}\sum_{t=1}^{T}\varepsilon_t + \frac{|\mathcal{Y}|}{16h^2\mu}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2}$$
$$+ \frac{1}{4h}\sum_{k\in\mathcal{Y}}\sqrt{\frac{2}{\mu T}\Big(\frac{(\tau_\star^k)^2}{2\eta_2} + \frac{\eta_2 T}{2\,c(1-c)}\Big)}.$$

□

D.6 THRESHOLD CONFINEMENT

Proof of Lemma 2. Fix $k \in \mathcal{Y}$ and define $\xi_t := \tau_t^k - \tau_{t-1}^k$, which Eq. (10) makes $-\eta_2 \bar{g}_t^k$. Since $\rho_{t,i}^k \in [c, 1-c]$ and $\alpha \le 1/2$, Eq. (12) gives

$$|\Delta_{t,i}^k| \le \frac{1}{c}, \qquad \big|[1-\rho_{t,i}^k] - \alpha\big| \le \max(\alpha - c,\, 1 - c - \alpha) = 1 - c - \alpha, \qquad \text{so} \quad |\xi_t| \le \eta_2 b_0.$$

For $\tau \ge \bar{s} + \gamma$: $(\tau - s_{t-1}^k(x))/h \ge \ln \frac{1-c}{c}$ for every x, since $s_{t-1}^k(x) \le \bar{s}$, hence $\big(1 + e^{(\tau - s_{t-1}^k(x))/h}\big)^{-1} \le c$ and $\rho_{t,i}^k(\tau) = c$ identically. For $\tau \le -\bar{s} - \gamma$: $(\tau - s_{t-1}^k(x))/h \le -\ln \frac{1-c}{c}$ since $s_{t-1}^k(x) \ge -\bar{s}$, so $\rho_{t,i}^k(\tau) = 1 - c$ identically. On the event $\{\tau_{t-1}^k \ge \bar{s} + \gamma\}$, therefore, $[1 - \rho_{t,i}^k] - \alpha = 1 - c - \alpha$ for every i, and by Proposition 2 and the tower rule, $\mathbb{E}[\Delta_{t,i}^k \mid \mathcal{F}_{t-1}] = \mathbb{E}[\mathbb{1}\{Y_{t,i} = k\} \mid \mathcal{F}_{t-1}] = \mathbb{P}(Y_{t,i} = k \mid \mathcal{F}_{t-1}) \ge \underline{p_k}$, so

$$\mathbb{E}[\xi_t \mid \mathcal{F}_{t-1}] \le -\eta_2 (1 - c - \alpha)\, \underline{p_k} \le -\eta_2 d_k, \qquad d_k := \underline{p_k}(\alpha - c),$$

using $\alpha - c \le 1 - c - \alpha$ at $\alpha \le 1/2$. On $\{\tau_{t-1}^k \le -\bar{s} - \gamma\}$, the bracket is $c - \alpha$ and $\mathbb{E}[\xi_t \mid \mathcal{F}_{t-1}] \ge \eta_2(\alpha - c)\underline{p_k} \ge \eta_2 d_k$. Suppose $\max_{t \le T} \tau_t^k > \bar{s} + \gamma + \beta_k$. Since $\tau_0^k = 0 \in [-\bar{s} - \gamma,\ \bar{s} + \gamma]$ and $\beta_k > \eta_2 b_0$, there exist a first exceedance time t and the last $r < t$ with $\tau_r^k \le \bar{s} + \gamma$, and $t \ge r + 2$. For $r + 2 \le j \le t$ we have $\tau_{j-1}^k > \bar{s} + \gamma$, so the drift bound applies. With $n := t - r - 1 \ge 1$ and $D := \sum_{j=r+2}^{t} \big(\xi_j - \mathbb{E}[\xi_j \mid \mathcal{F}_{j-1}]\big)$,

$$\beta_k - \eta_2 b_0 \;<\; \tau_t^k - \tau_{r+1}^k \;=\; \sum_{j=r+2}^{t} \mathbb{E}[\xi_j \mid \mathcal{F}_{j-1}] + D \;\le\; -n\eta_2 d_k + D,$$

the first inequality as the single step from τ_r^k adds at most $\eta_2 b_0$. Azuma–Hoeffding with martingale increments in $[-2\eta_2 b_0,\, 2\eta_2 b_0]$ and $(a+b)^2 \ge 4ab$ give

$$\mathbb{P}\big(D \ge (\beta_k - \eta_2 b_0) + n\eta_2 d_k\big) \le \exp\Big(-\frac{\big((\beta_k - \eta_2 b_0) + n\eta_2 d_k\big)^2}{8n\eta_2^2 b_0^2}\Big) \le \exp\Big(-\frac{(\beta_k - \eta_2 b_0)\, d_k}{2\eta_2 b_0^2}\Big) = \frac{\delta}{2|\mathcal{Y}|T^2}.$$

The bound is uniform in n, and the event fixes at most T^2 pairs (r, t). The lower barrier is symmetric, and a union bound over both barriers, the pairs, and $k \in \mathcal{Y}$ gives total probability at most δ, so $|\tau_t^k| \le \bar{s} + \gamma + \beta_k$ for every t and every class with probability at least $1 - \delta$. □

Proof of Corollary 2. Apply Theorem 1 and Lemma 2, each at δ. A union bound puts both conclusions on one event of probability at least $1 - 2\delta$. On that event $|\tau_T^k| \le \bar{s} + \gamma + \beta_k$ for every k, and the bound of Theorem 1 is increasing in $|\tau_T^k|$, so the substitution gives the claim. □