ONE-HIT DETECTION IN GENERATIVE SAMPLING: OPTIMAL STOPPING AND CONFORMAL CALIBRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative models can produce candidates for scientific discovery, but most lack the desired property, and determining which ones qualify often requires costly experiments. In practice, predictive models help assess whether at least one *hit*—a candidate satisfying the target property—has been generated, thereby guiding when to stop generation and reducing unnecessary experimental validation. We formulate this task as a finite-horizon optimal stopping problem that maximizes expected reward for early detection while controlling the probability of false alarms, i.e., stopping before any hit appears. Given a sequence of independent p-values, we first construct a monotone family of stopping times based on Lagrangian duality, each optimal for the Lagrangian problem at its corresponding multiplier. We then propose a novel method *Conformalized Optimal Stopping*, applying the monotone family to conformal p-values constructed from screening scores and selecting the multiplier by calibration on simulated global-null trajectories. Under suitable assumptions, the resulting stopping strategy achieves nearly optimal expected reward with exact finite-sample false-alarm control. Experiments on QED–HGraph and TargetDiff candidates demonstrate substantial improvements over competing methods with the same finite-sample false-alarm guarantees.

1 INTRODUCTION

Deep generative models can sequentially propose candidates for scientific discovery, including molecules (Hoogeboom et al., 2022; Qu et al., 2024), proteins (Yim et al., 2023; Watson et al., 2023), and crystalline materials (Xie et al., 2022). However, not all generated candidates possess the target property, and identifying those that do often requires costly experiments or computations. For example, TargetDiff can generate candidate ligands sequentially and independently for a fixed protein binding pocket, but some may fail to meet a prespecified criterion (Guan et al., 2023; Trott & Olson, 2010). A candidate that satisfies the target property is called a *hit*. Determining whether at least one hit has been generated can guide when to stop generation and proceed to validation, reducing unnecessary computations and experiments.

Let $Y_1, \ldots, Y_T$ be independently generated candidates, where T denotes a prespecified generation budget. For each t, let $A_t \in \{0, 1\}$ indicate whether Y_t is a hit; this indicator is unobserved during generation. Define the first-hit time by

$$\nu^* := \inf\{t \geq 1 : A_t = 1\},$$

where $\inf \varnothing := \infty$ throughout the paper. We seek a stopping time $\tau \in \{1, \ldots, T, \infty\}$, based on information available as candidates arrive, at which to signal that a hit has occurred and terminate generation. Here, $\tau = \infty$ means that no signal is raised within the budget. We require the false-alarm rate to satisfy $\mathbb{P}(\tau < \nu^*) \leq \alpha$, for a prespecified $\alpha \in (0, 1)$. In related work, Laghuvarapu et al. (2026) seek a small subset of a completed candidate sequence that can be certified to contain at least one hit. Their procedure selects the subset after all T candidates have been generated. However, the decision to stop must be made as candidates arrive in a sequential generation process.

To guide this decision, we posit a working alternative distribution $P_{\mathcal{A}}$ for $\{(Y_t, A_t)\}_{t=1}^T$, informed by historical data or prior knowledge. Under the independence assumption, the false-alarm rate can be bounded by $\mathbb{P}_0(\tau \leq T)$ (see Proposition A.1), where $\mathbb{P}_0$ denotes probability under the global null $A_1 = \cdots = A_T = 0$. Thus, imposing $\mathbb{P}_0(\tau \leq T) \leq \alpha$ suffices to control the false-alarm rate. We

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

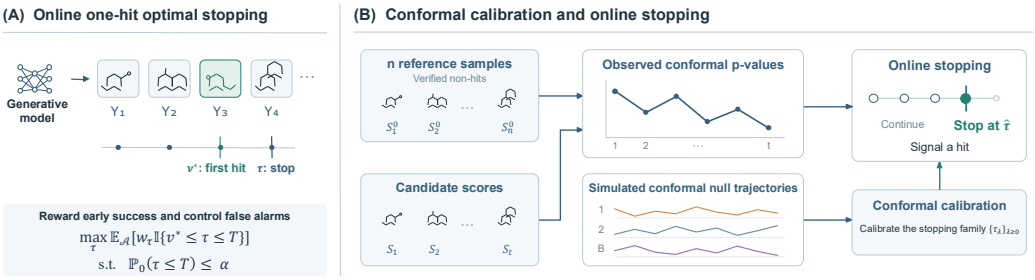


Figure 1: (a) Online stopping to detect at least one hit (green). (b) Candidate scores are converted into conformal p-values using n verified non-hit references. The stopping family is calibrated using B simulated p-value trajectories.

choose τ to maximize expected reward under the working alternative subject to this constraint:

$$\begin{aligned} \max_{\tau} \quad & \mathbb{E}_{\mathcal{A}}[w_\tau \mathbf{1}\{\nu^* \le \tau \le T\}] \\ \text{s.t.} \quad & \mathbb{P}_0(\tau \le T) \le \alpha, \end{aligned} \tag{1}$$

where $\mathbb{E}_{\mathcal{A}}[\cdot]$ denotes the expectation taken over $P_{\mathcal{A}}$ and $w_1 \ge \cdots \ge w_T > 0$ are prespecified rewards. A stop at time τ earns reward w_τ only if at least one candidate generated by that time is a hit. The reward weights favor earlier detection, while the constraint provides uniform false-alarm control over the specified class of hit-generation mechanisms.

We first study optimal stopping when each candidate is accompanied by an independent p-value. Under a two-group working alternative (Efron, 2008), we use Lagrangian duality to construct a family of stopping rules indexed by a Lagrangian multiplier, with each rule solving the corresponding Lagrangian problem. These stopping times are nondecreasing in the multiplier along every trajectory. Under suitable regularity conditions, we identify a multiplier whose stopping rule solves (1), attaining the optimal expected reward among all feasible rules based on the observed p-values. The rules are computed through a Bellman recursion (Bellman, 1966), using the posterior probability that no hit has appeared as the state. To make this computation tractable, we develop a finite-partition approximation and bound the resulting reward loss in terms of the density approximation error.

When oracle p-values are unavailable, we propose *Conformalized Optimal Stopping* (COS), which applies this stopping family to conformal p-values (Vovk et al., 2005; Bates et al., 2023) constructed from screening scores (Qu et al., 2024) using an independent reference sample. Because the shared reference induces dependence among these p-values, the Lagrangian multiplier must be recalibrated under their joint null distribution. Our key idea is to exploit the pathwise monotonicity of the stopping rules to transfer calibration to the Lagrangian multiplier. Specifically, we simulate global-null trajectories that preserve the dependence induced by the shared reference sample and assign each trajectory a critical multiplier: the smallest value that prevents stopping by the horizon. Selecting a calibrated upper quantile of these critical multipliers then determines the final stopping rule.

Our main contributions are as follows.

(1) We formulate sequential hit detection in generative sampling as a constrained optimal stopping problem. The goal is to determine when there is sufficient evidence that at least one hit has been generated, without requiring identification of individual hits. The formulation accounts for both detection success and stopping time under an explicit false-alarm rate constraint. Under the two-group working alternative, we characterize its optimal solution for independent p-values. Figure 1 summarizes the framework.

(2) We develop COS, which uses pathwise monotonicity to calibrate the Lagrangian multiplier, and establish its validity and efficiency. Under exchangeability, COS controls the false-alarm rate in finite samples, even when the working alternative is misspecified. Under additional regularity conditions, we derive explicit bounds on its expected reward loss relative to the oracle, quantifying the effects of the finite reference sample and conformal calibration. These bounds show that COS approaches the oracle optimal reward under the working alternative as both the reference sample size and the number of calibration trajectories grow.

(3) Synthetic experiments and molecular screening studies on two benchmark datasets demonstrate substantial reward gains of COS over competing stopping rules.

2 Related Work

Generative models and candidate scoring. Generative modeling has expanded across scientific domains. CDVAE generates stable crystalline materials while respecting periodic structure (Xie et al., 2022), and RFdiffusion generates proteins for specified design objectives (Watson et al., 2023). For molecular discovery, MolCRAFT generates three-dimensional ligands conditioned on protein binding pockets (Qu et al., 2024). These advances necessitate efficient evaluation prior to experimental validation. AutoDock Vina scores protein–ligand binding poses (Trott & Olson, 2010), and large-scale docking experiments demonstrate that favorable scores enrich experimental binders, albeit imperfectly (Lyu et al., 2019). MolScore further provides a unified framework for evaluating generated molecules using molecular properties and learned predictors (Thomas et al., 2024). These studies motivate the use of candidate scores as informative evidence of the desired property.

Online stopping. Online stopping balances timely detection with sequential error control. Sequential testing methods improve evidence accumulation and stopping efficiency through likelihood ratios, learned betting strategies, and interactive martingale constructions (Fischer & Ramdas, 2024; Shekhar & Ramdas, 2024; Duan et al., 2020). Another line studies latent event times and transient changes, optimizing reaction delay or timely detection subject to false-alarm constraints (Niesen & Tchamkerten, 2009; Heydari & Tajer, 2019; Egea-Roca et al., 2022). With an explicit finite horizon, detection procedures balance latency against false-alarm rate and attain order-optimal dependence on the horizon (Huang & Veeravalli, 2025). Our online one-hit objective differs by rewarding early stopping only after a suitable candidate has appeared. Closer to our reward-based formulation, recent work learns betting policies for testing before a deadline and develops Bellman formulations for optimizing rejection-time rewards (Taga et al., 2026; Clerico et al., 2026). They often use prespecified stopping families or computationally restricted betting strategies, without guarantees of near-global optimality for the reward objective. Our framework characterizes the optimum over all admissible stopping times and achieves nearly optimal expected reward under the working alternative.

Conformal inference and testing. Conformal prediction provides distribution-free, finite-sample marginal coverage under exchangeability (Vovk et al., 2005; Shafer & Vovk, 2008). For sequential testing, conformal test martingales accumulate evidence to detect departures from exchangeability (Fedorova et al., 2012; Vovk et al., 2022). Related work develops conformal change-detection procedures, including CUSUM variants (Vovk et al., 2021; 2025). For outlier and novelty detection, conformal p-values support finite-sample false discovery rate (FDR) control, with adaptive scores and labelled-outlier information improving detection power (Bates et al., 2023; Marandon et al., 2024; Liang et al., 2024). Localized conformal p-values further enable testing of conditional distributions (Wu et al., 2025). Full-conformal and permutation-based methods improve data reuse for score construction and calibration while retaining finite-sample FDR control (Lee et al., 2025; Huo et al., 2026). Further work characterizes the joint distribution and concentration of conformal p-values, accounting for dependence induced by shared calibration data (Gazin et al., 2024). Our work considers online one-hit stopping, combining reward optimization with calibration of complete null trajectories to ensure finite-sample false-alarm control.

3 Optimal Stopping with Independent p-Values

In this section, we assume that each candidate Y_t is associated with a p-value P_t, forming an independent sequence $\mathbf{P} = (P_t)_{1 \le t \le T}$ with $P_t \mid A_t = 0 \sim \mathrm{Unif}(0,1)$. Under a two-group working alternative, we characterize the optimal stopping rule for the primal problem (1).

3.1 Working alternative under the two-group model

The generation mechanism determines the joint distribution of the hit indicators and p-values across the T candidates. We specify a two-group model as our working alternative $\mathbb{P}_{\mathcal{A}}$ for this joint distribution (Efron, 2008):

$$A_t \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(b), \qquad P_t \mid A_t = 0 \sim \mathrm{Unif}(0,1), \qquad P_t \mid A_t = 1 \sim f_1. \tag{2}$$

Here, $b \in (0,1)$ is the probability that a candidate is a hit, and f_1 is the p-value density among hits. The parameters (b, f_1) may be specified using prior knowledge or estimated from independent historical data. Let $\mathcal{F}_t^P = \sigma(P_1, \ldots, P_t)$ and consider a stopping time $\tau \in \{1, \ldots, T, \infty\}$ adapted to the p-value trajectory; that is, $\{\tau \le t\}$ is $\mathcal{F}_t^P$-measurable.

The reward sequence $w_1 \ge \cdots \ge w_T > 0$ controls the preference for early successful certification. Typical choices include (i) $w_t \equiv 1$, which assigns the same reward to every successful detection, so the objective reduces to the probability of detecting the presence of a hit within the generation budget; (ii) $w_t = \kappa - t$ with $\kappa > T$, which reduces the reward by one for each additional step, balancing detection probability against detection time; and (iii) $w_t = \rho^{t-1}$, which discounts the reward by a factor of $\rho \in (0,1)$ per step.

3.2 Optimal stopping rule via Lagrangian duality

In this subsection, we use Lagrangian duality to solve the constrained optimal stopping problem. Given a stopping time τ, let $J_w(\tau) := \mathbb{E}_{\mathcal{A}}[w_\tau \mathbf{1}\{\nu^* \le \tau \le T\}]$ denote its expected reward. For each multiplier $\lambda \ge 0$, define

$$V_0(\lambda) := \sup_{\tau} \left\{ J_w(\tau) - \lambda \mathbb{P}_0(\tau \le T) \right\}, \tag{3}$$

with the associated dual problem $\inf_{\lambda \ge 0}\{V_0(\lambda) + \lambda\alpha\}$. A larger multiplier places a greater penalty on the global-null alarm probability.

Bellman recursion. Let $q_t(\mathbf{P}) := \mathbb{P}_{\mathcal{A}}(\nu^* > t \mid P_1, \ldots, P_t)$ be the *state function*, representing the posterior probability that no hit has appeared by time t. We can rewrite (3) as (see Appendix C.1)

$$\sup_{\tau} \ \mathbb{E}_{\mathcal{A}}\left[\left\{ w_\tau\big(1 - q_\tau(\mathbf{P})\big) - \frac{\lambda q_\tau(\mathbf{P})}{(1-b)^\tau} \right\} \mathbf{1}\{\tau \le T\} \right].$$

This representation expresses both the reward and the penalty in terms of the posterior state. To solve this problem, given a pair (q, λ) with state $q \in [0,1]$ and $\lambda \ge 0$, the Bellman recursion (Bellman, 1966) gives

$$\begin{aligned} R_t(q;\lambda) &= w_t(1-q) - \lambda q/(1-b)^t, \\ C_t(q;\lambda) &= \int_0^1 V_{t+1}\left(\frac{q(1-b)}{m(p)}; \lambda \right) m(p)\, \mathrm{d}p, \\ V_t(q;\lambda) &= \max\{R_t(q;\lambda), C_t(q;\lambda)\}, \end{aligned}$$

for $t = 1, \ldots, T$, with $C_T(q;\lambda) = 0$. Here, $m(p) = 1 - b + b f_1(p)$ is the marginal p-value density under the working alternative. The quantities R_t, C_t, and V_t are, respectively, the penalized payoff from raising an alarm immediately, the expected payoff from continuing optimally, and the optimal remaining value. Starting from $V_T(q;\lambda) = \max\{R_T(q;\lambda), 0\}$, these quantities can be computed by backward induction for $t = T-1, \ldots, 1$; see Appendix C.1 for details.

Since R_t is the immediate stopping payoff and C_t is the best expected payoff achievable by continuing, stopping is optimal when $R_t > C_t$. Therefore, an optimal stopping rule for (3) is

$$\begin{aligned} \tau_\lambda(\mathbf{P}) &:= \inf\left\{ 1 \le t \le T : R_t\big(q_t(\mathbf{P}); \lambda\big) > C_t\big(q_t(\mathbf{P}); \lambda\big) \right\} \\ &= \inf\left\{ 1 \le t \le T : q_t(\mathbf{P}) < c_t(\lambda) \right\}, \end{aligned} \tag{4}$$

where the threshold $c_t(\lambda) \in [0,1]$ is the unique solution of $R_t(q;\lambda) = C_t(q;\lambda)$; see Appendix C.3. We next identify a multiplier whose stopping rule solves the primal problem (1).

Proposition 1. *For any $0 \le \lambda_1 < \lambda_2$, it holds that $\tau_{\lambda_1}(\mathbf{P}) \le \tau_{\lambda_2}(\mathbf{P})$. Moreover, we have*

$$\mathbb{P}_0\left\{\tau_{\lambda_1}(\mathbf{P}) \le T\right\} \ge \mathbb{P}_0\left\{\tau_{\lambda_2}(\mathbf{P}) \le T\right\}, \quad J_w(\tau_{\lambda_1}(\mathbf{P})) \ge J_w(\tau_{\lambda_2}(\mathbf{P})).$$

Increasing λ makes stopping more conservative: it can delay an alarm and lowers both the false-alarm rate and the expected reward under the working alternative.

Theorem 1. *Suppose $\log\{m(P)/(1-b)\}$ is non-atomic for $P \sim \mathrm{Unif}(0,1)$. Then the multiplier*

$$\lambda^* = \inf\{\lambda \ge 0 : \mathbb{P}_0\{\tau_\lambda(\mathbf{P}) \le T\} \le \alpha\} \tag{5}$$

is finite and satisfies $\mathbb{P}_0(\tau_{\lambda^*} \le T) = \alpha$. The stopping time $\tau_{\lambda^*}(\mathbf{P})$ is optimal for the problem (1)*.*

Theorem 1 reduces the population problem to a one-dimensional search for the smallest feasible multiplier. However, exact computation is challenging. Next, we provide an efficient implementation based on a finite partition strategy.

Efficient implementation via finite partition. For a general continuous hit density f_1, the posterior state q_t may take infinitely many values, so exact computation is impractical. We therefore approximate f_1 by a finite piecewise-constant density. For a partition $\{I_\ell\}_{\ell=1}^K$ of $(0,1)$ into intervals of positive length, one convenient choice is

$$f_{1,K}(p) = \sum_{\ell=1}^{K} \left\{ \frac{1}{|I_\ell|} \int_{I_\ell} f_1(u)\,\mathrm{d}u \right\} \mathbf{1}\{p \in I_\ell\}.$$

More generally, any nonnegative, normalized piecewise-constant density may be used. For fixed K, T, the posterior has finitely many reachable states, and the Bellman integral becomes a finite sum. The same construction yields a monotone family $\{\tau_{K,\lambda}(\mathbf{P})\}_{\lambda\geq 0}$ and its population-calibrated rule $\lambda_K^* = \inf\{\lambda \geq 0 : \mathbb{P}_0\{\tau_{K,\lambda}(\mathbf{P}) \leq T\} \leq \alpha\}$. The implementation algorithm is detailed in Algorithm 1 in Appendix B. Under regularity conditions, Appendix C.6 establishes that the expected reward loss of $\tau_{K,\lambda_K^*}(\mathbf{P})$ relative to the original constrained optimum is $O(\|f_{1,K} - f_1\|_\infty)$, which vanishes as the density approximation improves with $K \to \infty$.

Although candidates in practice typically come with scores rather than directly available p-values, the independent-input analysis provides an efficient, computationally tractable monotone stopping family. In the next section, we adapt this stopping family to score-based observations through conformal calibration, yielding finite-sample false-alarm control without requiring the correctness of the working alternative model.

4 Conformalized Optimal Stopping with Finite-Sample Guarantee

In this section, each generated candidate Y_t is assigned a score $S_t = s(Y_t) \in \mathbb{R}$ by a fixed evaluator $s(\cdot)$, with smaller values indicating stronger evidence of a hit (Qu et al., 2024; Thomas et al., 2024). Let $S_t \mid A_t = 0 \sim F_0$. We call $P_t = F_0(S_t)$ for $1 \leq t \leq T$ the "oracle" p-values, which satisfy the independence condition in Section 3. Since the distribution F_0 is usually unknown, we replace these oracle p-values with conformal counterparts (Vovk et al., 2005; Bates et al., 2023).

4.1 Stopping rules under conformal p-values

Suppose we have access to n non-hit reference scores $S_1^0, \ldots, S_n^0 \overset{\text{iid}}{\sim} F_0$, independent of the testing sequence $(S_t)_{1\leq t\leq T}$. Let $(U_t)_{1\leq t\leq T}$ be independent $\mathrm{Unif}(0,1)$ variables. We then compute the conformal p-value trajectory as

$$\widetilde{P}_t = \frac{\sum_{i=1}^n \mathbf{1}\{S_i^0 < S_t\} + U_t}{n+1}, \qquad t = 1, \ldots, T. \tag{6}$$

By exchangeability, we have $\widetilde{P}_t \mid A_t = 0 \sim \mathrm{Unif}(0,1)$. For each multiplier $\lambda \geq 0$, we retain the state function $q_t(\cdot)$ and the thresholds $c_t(\lambda)$ derived under the independent-p-value model in Section 3.2, replacing only the input p-values in (4) with their conformal counterparts, and obtain

$$\tau_\lambda(\widetilde{\mathbf{P}}) := \inf\left\{1 \leq t \leq T : q_t(\widetilde{\mathbf{P}}) < c_t(\lambda)\right\}, \tag{7}$$

where $\widetilde{\mathbf{P}} = (\widetilde{P}_t)_{t=1}^T$. Although each conformal p-value is marginally uniform under the null, the trajectory $\widetilde{\mathbf{P}}$ is dependent because all candidates share the same reference sample. The multiplier defined in (5) is not guaranteed to control the false-alarm rate. The next subsection develops a novel conformal procedure to select the multiplier by generating sequences exchangeable with $\widetilde{\mathbf{P}}$.

4.2 COS: Conformalized optimal stopping

We first generate B conformal p-value trajectories for calibration. For each $j = 1, \ldots, B$, draw a reference set $(W_i^{0,(j)})_{1\leq i\leq n}$ and two trajectories $(W_t^{(j)})_{1\leq t\leq T}$ and $(U_t^{(j)})_{1\leq t\leq T}$ from $\mathrm{Unif}(0,1)$ and then compute a conformal p-value trajectory via

$$\widetilde{P}_t^{(j)} = \frac{\sum_{i=1}^n \mathbf{1}\{W_i^{0,(j)} < W_t^{(j)}\} + U_t^{(j)}}{n+1}, \qquad t = 1, \ldots, T.$$

The same reference set $W_1^{0,(j)}, \ldots, W_n^{0,(j)}$ is used throughout each trajectory, preserving the dependence induced by the shared reference in the observed conformal p-value trajectory (6). Notice that the simulated p-value trajectories $\widetilde{\mathbf{P}}^{(j)} := (\widetilde{P}_t^{(j)})_{1\le t\le T}$ for $1 \le j \le B$ are exchangeable with the observed conformal p-value trajectory $\widetilde{\mathbf{P}}$ under the global non-hit null.

For each simulated trajectory $\widetilde{\mathbf{P}}^{(j)}$, $j = 1, \ldots, B$, the stopping rule in (4) stops and declares that at least one hit has appeared at the first time t for which $q_t(\widetilde{\mathbf{P}}^{(j)}) < c_t(\lambda)$. Define the critical multipliers

$$\begin{aligned}\Gamma_j &:= \min\left\{\lambda \ge 0 : \tau_\lambda\big(\widetilde{\mathbf{P}}^{(j)}\big) > T\right\} \\ &= \min\left\{\lambda \ge 0 : q_t\big(\widetilde{\mathbf{P}}^{(j)}\big) \ge c_t(\lambda) \text{ for all } t = 1, \ldots, T\right\}.\end{aligned}$$

Thus, Γ_j is the smallest multiplier for which continuing has at least as large a value as stopping at every time along the simulated trajectory, so the rule raises no alarm by time T. By the monotonicity in Proposition 1, we know $\tau_\lambda\big(\widetilde{\mathbf{P}}^{(j)}\big) \le T \iff \Gamma_j > \lambda$. Hence, calibrating the probability of stopping by T amounts to selecting an upper quantile of these pathwise critical multipliers.

Choose $B \ge \lceil 1/\alpha \rceil$ and set $k = \lceil (B+1)(1-\alpha) \rceil$. Let $\Gamma_{(k)}$ denote the k-th smallest value among $\Gamma_1, \ldots, \Gamma_B$. We select the multiplier as $\hat{\lambda} = \Gamma_{(k)}$ and apply the resulting rule to the observed conformal p-value trajectory:

$$\hat{\tau} := \tau_{\hat{\lambda}}(\widetilde{\mathbf{P}}) = \inf\left\{1 \le t \le T : q_t(\widetilde{\mathbf{P}}) < c_t(\hat{\lambda})\right\}. \tag{8}$$

We call this procedure *Conformalized Optimal Stopping* (COS). Implementation details are given in Algorithm 2 in Appendix B.

Theorem 2. *Suppose the reference scores and candidate scores under the global null are jointly exchangeable. Then*

$$\mathbb{E}_{\mathrm{cal}}[\mathbb{P}_0(\hat{\tau} \le T)] \le \frac{B+1-k}{B+1} \le \alpha,$$

where $\mathbb{E}_{\mathrm{cal}}$ denotes expectation over the B simulated calibration trajectories.

Together with $\mathbb{P}(\hat{\tau} < \nu^*) \le \mathbb{P}_0(\hat{\tau} \le T)$ in Proposition A.1, we can guarantee the false-alarm rate control $\mathbb{P}(\hat{\tau} < \nu^*) \le \alpha$ in finite samples, even when the working alternative is misspecified.

We now evaluate the reward of the calibrated procedure under the original working alternative (2). The conformal p-values $\widetilde{P}_t$ are dependent since they are computed based on the same reference set. This dependence creates an optimality gap relative to the stopping rule obtained with independent oracle p-values $P_t = F_0(S_t)$ for $1 \le t \le T$. The following theorem bounds this gap.

Theorem 3. *Suppose f_1 is Lipschitz on $[0,1]$ and the original log evidence $\sum_{j=1}^{t} \log\{m(P_j)/(1-b)\}$ has a density under the working alternative uniformly bounded over $t \le T$. For fixed T, α, $b \in (0,1)$,*

$$0 \le J_w(\tau_{\lambda^*}(\mathbf{P})) - \mathbb{E}_{\mathrm{cal}}[J_w(\widehat{\tau})] \le O\left(\sqrt{\frac{\log n}{n}} + \frac{1}{\sqrt{B}}\right),$$

where $\tau_{\lambda^*}(\mathbf{P})$ with $\mathbf{P} = (F_0(S_t))_{1\le t\le T}$ is the oracle optimal stopping rule and λ^* is defined in (5).

The result above shows that the optimality gap vanishes as $n, B \to \infty$, yielding nearly optimal expected reward under the original working alternative (2). The two terms bound the reward loss due to the finite shared reference sample and conformal calibration, respectively. The optimality analysis of the finite-partition implementation is detailed in Appendix F.2.

5 Experiments

In each experimental repeat, we randomly split the dataset into training, reference, and test sets in proportions of 20%, 20%, and 60%, respectively. The training set is used to estimate the working-alternative parameters $\widehat{b}$ and $\widehat{f}_1$. From the reference set, we randomly sample $n = 200$ non-hit candidates to construct the reference data. The test set is partitioned into trajectories of length T for

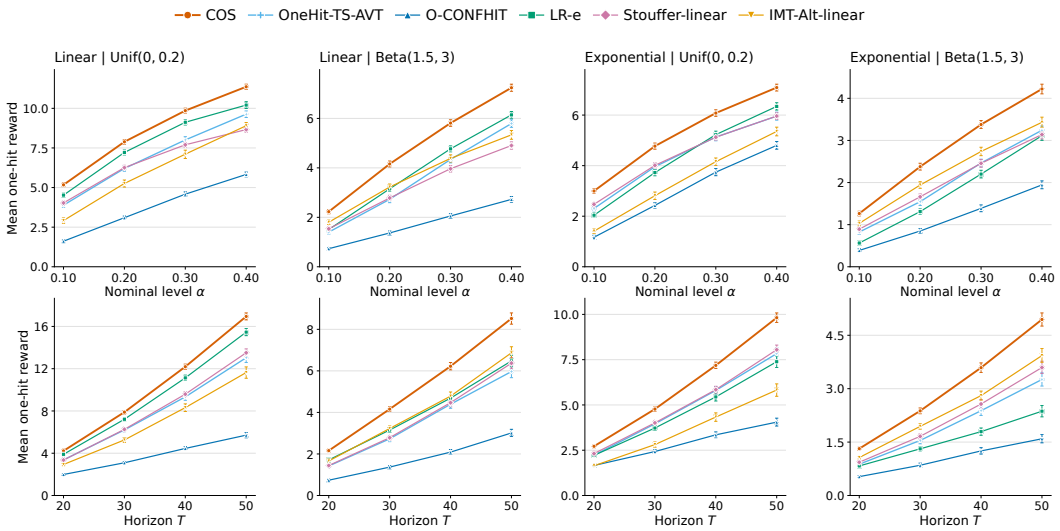


Figure 2: Synthetic one-hit rewards as α varies at $T = 30$ (top row) and as T varies at $\alpha = 0.2$ (bottom row). Columns correspond to the four combinations of two reward schedules and two hit-score alternatives. Points are means over 100 complete repeats; error bars are pointwise 95% complete-repeat percentile-bootstrap intervals.

evaluation. We evaluate the stopping rules using two decreasing reward functions: $w_t^{\mathrm{lin}} = T + 1 - t$, and $w_t^{\mathrm{exp}} = T \exp\{-(t-1)\log T/(T-1)\}$. For each setting, we repeat the full procedure 100 times and report the mean one-hit reward and the false-alarm rate, together with pointwise 95% percentile intervals.

Our proposed **COS** method uses a reward-aware finite-horizon Bellman rule under a fitted five-category working alternative, with its multiplier calibrated on complete null trajectories. **O-CONFHIT** is an online max-pooling adaptation of ConfHit (Laghuvarapu et al., 2026). **LR-e** accumulates the fitted working-alternative likelihood ratio. **Stouffer-linear** accumulates inverse-normal scores against a linear boundary. **IMT-Alt-linear** adapts interactive martingale testing (Duan et al., 2020), using the fitted working alternative to screen masked p-values before updating a linear-boundary signed random walk. **OneHit-TS-AVT** adapts the testing-by-betting framework of time-sensitive anytime-valid testing (Clerico et al., 2026) to the one-hit objective. These five baselines are all recalibrated on complete null trajectories. Appendix G details all implementations.

5.1 Synthetic study

Each repeat contains a finite pool of $N = 10{,}000$ independent candidates. Their labels and non-hit scores follow $A \sim \mathrm{Bernoulli}(0.10)$ and $S \mid A = 0 \sim \mathrm{Unif}(0, 1)$. We consider the two hit-score alternatives: (i) $S \mid A = 1 \sim \mathrm{Unif}(0, 0.2)$; (ii) $S \mid A = 1 \sim \mathrm{Beta}(1.5, 3)$.

Table 1: Synthetic false-alarm rates at $T = 20$ under the exponential reward. Entries are means over 100 complete repeats.

	$S \mid H = 1 \sim \mathrm{Unif}(0, 0.2)$				$S \mid H = 1 \sim \mathrm{Beta}(1.5, 3)$			
Methodα	0.10	0.20	0.30	0.40	0.10	0.20	0.30	0.40
COS	0.062	0.128	0.190	0.246	0.061	0.123	0.187	0.248
OneHit-TS-AVT	0.059	0.120	0.190	0.265	0.056	0.116	0.178	0.243
O-CONFHIT-Recal	0.068	0.145	0.229	0.311	0.068	0.145	0.229	0.311
LR-e-Recal	0.033	0.083	0.142	0.216	0.026	0.064	0.113	0.173
Stouffer-linear	0.059	0.122	0.189	0.260	0.059	0.122	0.189	0.260
IMT-Alt-linear	0.037	0.082	0.130	0.196	0.046	0.106	0.161	0.238

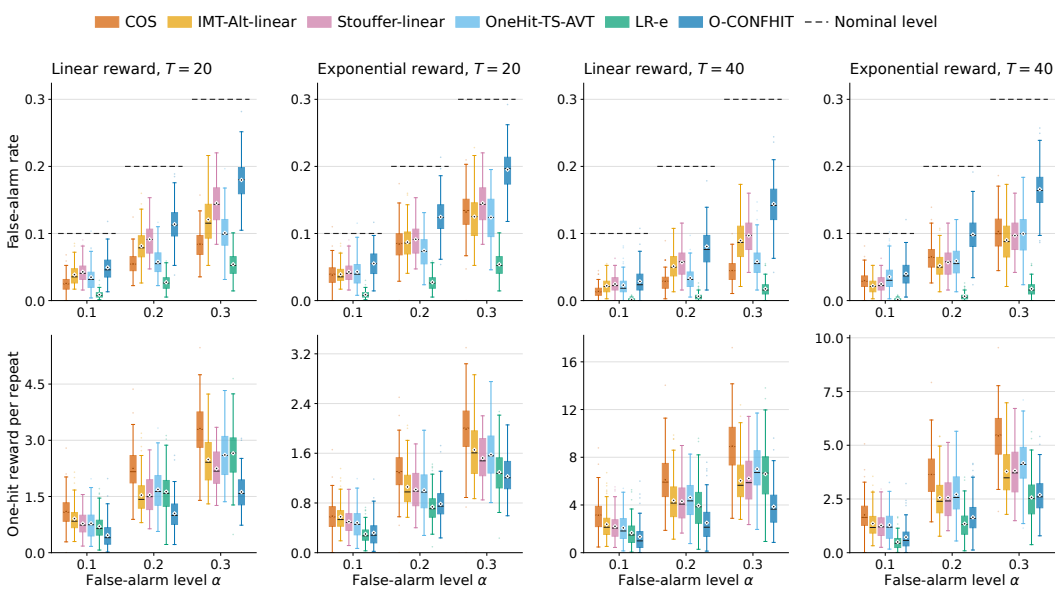


Figure 3: False-alarm rates (top row) and one-hit rewards (bottom row) for the QED–HGraph candidates across two horizons and two reward specifications.

Table 1 reports mean false-alarm rates, corresponding to the event $\tau < \nu^*$, under the exponential reward at $T = 20$. All reported rates are obviously below their nominal levels. This partly reflects the global-null constraint, which provides false-alarm control across a class of admissible generation mechanisms when the true mechanism is unknown and may introduce some conservatism under a particular mechanism. Figure 2 summarizes both reward comparisons. In the top row, the COS method achieves the highest mean reward in all 16 alternative–reward–α settings at $T = 30$, exceeding the strongest baseline in each setting by 8.0%–28.9%. The bottom row shows that this advantage persists as the horizon varies at $\alpha = 0.2$. These gains are consistent with differences in the methods' designs. O-CONFHIT and Stouffer-linear do not use the fitted working alternative, whereas LR-e and IMT-Alt-linear exploit this information but retain prespecified evidence processes and stopping boundaries. OneHit-TS-AVT optimizes the same one-hit reward, but its implementation restricts the available betting actions, without a corresponding near-global optimality guarantee. COS instead derives its stopping family by optimizing the one-hit reward over all admissible stopping times and provides near-global optimality guarantees under the working alternative and the stated assumptions. Additional results appear in Appendix G.2.

5.2 Real data applications

High-QED molecule detection problem. Molecular property optimization seeks candidates that improve a desired chemical property while remaining valid molecular structures. We consider molecular candidates generated by HGraph for QED optimization (Jin et al., 2020). QED is a deterministic score in $[0, 1]$ summarizing several drug-likeness characteristics (Bickerton et al., 2012). The question is whether a generated prefix already contains at least one high-QED candidate. We use 25,457 unique molecules, defining hits by $A = \mathbf{1}\{\text{QED} > 0.93\}$, with a hit rate of 15.87%. The predicted positive-class probability from a pretrained QED classifier serves as the screening score. Exact QED is reserved for evaluation.

Figure 3 summarizes the QED–HGraph results. Across both horizons and reward specifications, all six methods have mean false-alarm rates below the corresponding nominal levels (top row). The COS method achieves the highest mean one-hit reward in all 12 settings (bottom row), with larger gains generally observed at $T = 40$ and at higher values of α.

Molecular docking hit detection problem. Structure-based molecular generation proposes ligands conditioned on protein pockets. We study candidates generated by TargetDiff (Guan et al., 2023),

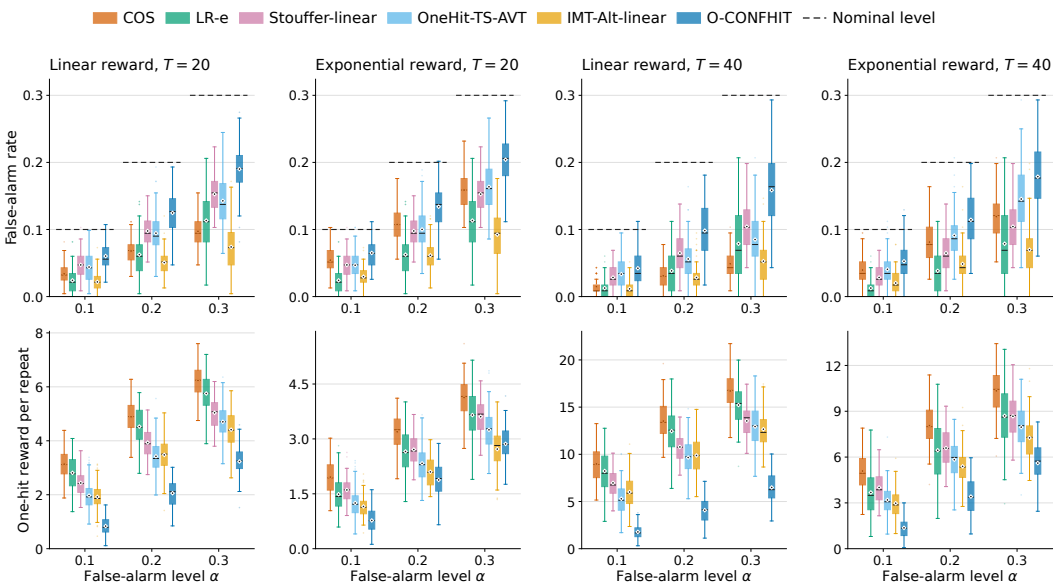


Figure 4: False-alarm rates (top row) and one-hit rewards (bottom row) for the TargetDiff candidates across two horizons and two reward specifications. Dashed segments indicate the nominal false-alarm levels, and all six methods share a common legend.

using 7,781 molecules from 83 protein–pocket contexts. The goal is to use screening scores to detect whether the candidates generated so far contain at least one docking hit. We define hits by $A = \mathbf{1}\{\text{Vina} < -10\,\text{kcal/mol}\}$, using AutoDock Vina energies (Trott & Olson, 2010), yielding a hit rate of 15.50%. Predicted binding affinities from an EGNN trained on PDBBind serve as screening scores, with larger values treated as more hit-like; Vina energies are reserved for evaluation.

Figure 4 summarizes the TargetDiff results. Across both horizons and reward specifications, all six methods have mean false-alarm rates below the corresponding nominal levels (top row). The COS method achieves the highest mean one-hit reward in all 12 settings (bottom row), exceeding the strongest baseline in each setting by 8.0%–26.8%, with particularly pronounced gains at $T = 40$.

6 Discussion

We study one-hit detection in sequential generative sampling, aiming to maximize the expected reward from early successful detection under a false-alarm rate constraint. For independent p-values, Lagrangian duality yields a monotone family of stopping rules, and a suitable Lagrangian multiplier recovers the constrained optimum under the two-group working alternative. We then propose COS, which uses this monotonicity to calibrate the Lagrangian multiplier through quantiles of critical multipliers computed from simulated conformal null trajectories. Under exchangeability, COS provides finite-sample false-alarm control even when the working alternative is misspecified. We also quantify the reward loss caused by the finite reference sample and conformal calibration, showing that COS can approach the oracle optimal reward.

Two directions remain for future research. First, controlling the stopping probability in (1) can be conservative for a particular generation mechanism. Incorporating available information about this mechanism may help reduce conservatism while preserving finite-sample false-alarm control. Second, although validity does not require a correctly specified working alternative, stopping efficiency depends on how well it describes the hit probability and the p-value distribution among hits. Future work could improve estimation of the working alternative from limited historical data and develop stopping rules that achieve high reward across several plausible working alternatives.

AI USAGE STATEMENT

We used generative AI tools for language polishing, assistance with mathematical proofs, literature searches, and code development. The authors take full responsibility for the final manuscript, including the correctness of the theoretical results, the accuracy of the citations, and the code and experimental results.

REFERENCES

Stephen Bates, Emmanuel Candès, Lihua Lei, Yaniv Romano, and Matteo Sesia. Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1):149–178, 2023.

Richard Bellman. Dynamic programming. *science*, 153(3731):34–37, 1966.

G. Richard Bickerton, Gaia V. Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L. Hopkins. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2):90–98, 2012. doi: 10.1038/nchem.1243.

Eugenio Clerico, Tobias Wegel, Iskander Azangulov, and Patrick Rebeschini. Time-sensitive anytime-valid testing. *arXiv preprint arXiv:2605.06521*, 2026.

Boyan Duan, Aaditya Ramdas, Sivaraman Balakrishnan, and Larry Wasserman. Interactive martingale tests for the global null. *Electronic Journal of Statistics*, 14(2):4489–4551, 2020. doi: 10.1214/20-EJS1790.

Bradley Efron. Microarrays, empirical Bayes and the two-groups model. *Statistical Science*, 23(1):1–22, 2008. doi: 10.1214/07-STS236.

Daniel Egea-Roca, Blaise Kévin Guépié, José A. López-Salcedo, Gonzalo Seco-Granados, and Igor V. Nikiforov. Two strategies in transient change detection. *IEEE Transactions on Signal Processing*, 70:1418–1433, 2022. doi: 10.1109/TSP.2022.3158008.

Valentina Fedorova, Alex Gammerman, Ilia Nouretdinov, and Vladimir Vovk. Plug-in martingales for testing exchangeability on-line. *arXiv preprint arXiv:1204.3251*, 2012.

Lasse Fischer and Aaditya Ramdas. Improving Wald's (approximate) sequential probability ratio test by avoiding overshoot. *arXiv preprint arXiv:2410.16076*, 2024. URL https://arxiv.org/abs/2410.16076.

Ulysse Gazin, Gilles Blanchard, and Etienne Roquain. Transductive conformal inference with adaptive scores. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1504–1512. PMLR, 2024.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kJqXEPXMsE0.

Javad Heydari and Ali Tajer. Quickest search for a change point. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 2194–2198. IEEE, 2019.

Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, 2022. URL https://proceedings.mlr.press/v162/hoogeboom22a.html.

Yu-Han Huang and Venugopal V. Veeravalli. Finite-horizon quickest change detection balancing latency with false alarm probability. *arXiv preprint arXiv:2511.12803*, 2025. URL https://arxiv.org/abs/2511.12803.

Yuyang Huo, Xiaoyang Wu, Changliang Zou, and Haojie Ren. Unified conformalized multiple testing with full data efficiency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2026. URL https://arxiv.org/abs/2508.12085. Accepted.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4839–4848. PMLR, 2020. URL https://proceedings.mlr.press/v119/jin20a.html.

Siddhartha Laghuvarapu, Ying Jin, and Jimeng Sun. ConfHit: Conformal generative design with oracle-free guarantees. In *International Conference on Learning Representations*, 2026.

Junu Lee, Ilia Popov, and Zhimei Ren. Full-conformal novelty detection: A powerful and non-random approach. *arXiv e-prints*, pp. arXiv–2501, 2025.

Ziyi Liang, Matteo Sesia, and Wenguang Sun. Integrative conformal p-values for out-of-distribution testing with labelled outliers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(3):671–693, 2024. doi: 10.1093/jrsssb/qkad138.

Jiankun Lyu, Sheng Wang, Trent E. Balius, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L. Roth, and John J. Irwin. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224–229, 2019. doi: 10.1038/s41586-019-0917-9.

Ariane Marandon, Lihua Lei, David Mary, and Etienne Roquain. Adaptive novelty detection with false discovery rate guarantee. *The Annals of Statistics*, 52(1):157–183, 2024. doi: 10.1214/23-AOS2338.

Pascal Massart. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990. doi: 10.1214/aop/1176990746.

Urs Niesen and Aslan Tchamkerten. Tracking stopping times through noisy observations. *IEEE Transactions on Information Theory*, 55(1):422–432, 2009.

Yanru Qu, Keyue Qiu, Yuxuan Song, Jingjing Gong, Jiawei Han, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. MolCRAFT: Structure-based drug design in continuous parameter space. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 41749–41768. PMLR, 2024. URL https://proceedings.mlr.press/v235/qu24a.html.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(12):371–421, 2008. URL https://jmlr.org/papers/v9/shafer08a.html.

Shubhanshu Shekhar and Aaditya Ramdas. Nonparametric two-sample testing by betting. *IEEE Transactions on Information Theory*, 70(2):1178–1203, 2024. doi: 10.1109/TIT.2023.3305867.

Ege Onur Taga, Samet Oymak, and Shubhanshu Shekhar. Learning to bet for horizon-aware anytime-valid testing. *arXiv preprint arXiv:2603.19551*, 2026. URL https://arxiv.org/abs/2603.19551.

Morgan Thomas, Noel M. O'Boyle, Andreas Bender, and Chris De Graaf. MolScore: a scoring, evaluation and benchmarking framework for generative models in de novo drug design. *Journal of Cheminformatics*, 16:64, 2024. doi: 10.1186/s13321-024-00861-w.

Oleg Trott and Arthur J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010.

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

Vladimir Vovk, Ivan Petej, Ilia Nouretdinov, Ernst Ahlberg, Lars Carlsson, and Alex Gammerman. Retrain or not retrain: Conformal test martingales for change-point detection. In *Proceedings of the Tenth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 152 of *Proceedings of Machine Learning Research*, pp. 191–210. PMLR, 2021. URL https://proceedings.mlr.press/v152/vovk21b.html.

Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman. Conformal testing: binary case with Markov alternatives. In *Proceedings of the Eleventh Symposium on Conformal and Probabilistic Prediction with Applications*, volume 179 of *Proceedings of Machine Learning Research*, pp. 207–218. PMLR, 2022. URL https://proceedings.mlr.press/v179/vovk22a.html.

Vladimir Vovk, Ilia Nouretdinov, and Alexander Gammerman. Validity and efficiency of the conformal CUSUM procedure. In *Proceedings of the Fourteenth Symposium on Conformal and Probabilistic Prediction with Applications*, volume 266 of *Proceedings of Machine Learning Research*, pp. 576–594. PMLR, 2025. URL https://proceedings.mlr.press/v266/vovk25b.html.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023. doi: 10.1038/s41586-023-06415-8.

Xiaoyang Wu, Lin Lu, Zhaojun Wang, and Changliang Zou. Conditional testing based on localized conformal p-values. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Ip6UwB35uT.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=03RLpj-tc_.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 40001–40039. PMLR, 2023. URL https://proceedings.mlr.press/v202/yim23a.html.

A GLOBAL-NULL REDUCTION

Controlling false alarms is challenging when the candidate-generation mechanism is unknown. The following proposition reduces this task to controlling the stopping probability under the global null.

Proposition A.1. *Suppose that the pairs (Y_t, A_t) are independent across t. Let $\mathbb{P}_0$ denote the corresponding global-null law, under which $A_1 = \cdots = A_T = 0$. Then every stopping time $\tau \in \{1, \dots, T, \infty\}$ adapted to $\mathcal{F}_t = \sigma(Y_1, \dots, Y_t)$ satisfies*

$$\mathbb{P}(\tau < \nu^*) \le \mathbb{P}_0(\tau \le T),$$

where the same stopping rule is applied under both laws.

Proof. Let $r_t = \mathbb{P}(A_t = 0)$. By independence across t and the definition of $\mathbb{P}_0$, for every measurable set D of candidate prefixes,

$$\mathbb{P}\big(Y_{1:t} \in D,\ A_1 = \cdots = A_t = 0\big) = \left(\prod_{s=1}^{t} r_s\right) \mathbb{P}_0(Y_{1:t} \in D).$$

If any $r_s = 0$ for $s \le t$, both sides of this identity are zero. Since τ is a stopping time, the event $\{\tau = t\}$ depends only on $Y_{1:t}$. Together with $\{\nu^* > t\} = \{A_1 = \cdots = A_t = 0\}$, this gives

$$\mathbb{P}(\tau = t, \nu^* > t) = \left(\prod_{s=1}^{t} r_s\right) \mathbb{P}_0(\tau = t).$$

Summing over $t = 1, \dots, T$ and using $0 \le r_s \le 1$, we obtain

$$\begin{aligned}\mathbb{P}(\tau < \nu^*) &= \sum_{t=1}^{T} \left(\prod_{s=1}^{t} r_s\right) \mathbb{P}_0(\tau = t) \\ &\le \sum_{t=1}^{T} \mathbb{P}_0(\tau = t) = \mathbb{P}_0(\tau \le T).\end{aligned}$$

□

Thus, for any independent hit-generation mechanisms, the constraint in (1) ensures false-alarm control uniformly.

B IMPLEMENTATION ALGORITHMS

We summarize the implementation in two algorithms. Algorithm 1 constructs the stopping family under a piecewise-constant working alternative and uses bisection to select its multiplier based on exactly computed global-null stopping probabilities. Algorithm 2 applies a given stopping family to score-based observations and calibrates its multiplier using an order statistic of simulated conformal-null trajectories. Both algorithms provide finite-sample false-alarm control under the corresponding null assumptions.

Finite-partition stopping with independent p-values. For a piecewise-constant density, the posterior has finitely many reachable states. For each queried multiplier, we evaluate the Bellman recursion at these states and compute the global-null stopping probability by finite-state recursion. Bisection then approximates the smallest feasible multiplier to a prescribed numerical tolerance. The resulting rule stops when the immediate stopping payoff exceeds the continuation payoff, with continuation at equality.

Conformalized optimal stopping from scores. Algorithm 2 takes the stopping family as given; for example, it may use the family constructed in Part I of Algorithm 1. The state functions and decision rules remain unchanged, while their inputs are replaced by conformal p-values. Because these values share a reference sample, the multiplier is recalibrated using simulated null trajectories that preserve this dependence. During deployment, each incoming score is converted into a conformal p-value using the same reference set, and the calibrated rule is applied to the observed prefix.

Algorithm 1 Finite-partition optimal stopping with null calibration

Require: Piecewise-constant density $f_{1,K}$ on $\{I_\ell\}_{\ell=1}^K$, hit probability b, horizon T, rewards $w_{1:T}$, level α, numerical tolerance $\eta > 0$.

Ensure: Multiplier $\hat{\lambda}_K$ and stopping rule $\hat{\tau}_K = \tau_{K,\hat{\lambda}_K}$.

Set $r = 1 - b$, $m_K(p) = r + bf_{1,K}(p)$, and let $m_{K,\ell}$ denote the value of m_K on I_ℓ.
Initialize $\mathcal{Q}_{K,0} = \{1\}$.
for $t = 1, \dots, T$ **do**
    Construct the reachable states

$$\mathcal{Q}_{K,t} = \left\{ \frac{qr}{m_{K,\ell}} : q \in \mathcal{Q}_{K,t-1},\ \ell = 1, \dots, K \right\}.$$

end for
Part I: Construct the stopping family.
For a queried multiplier λ, perform the following steps.
for $t = T, T-1, \dots, 1$ **do**
    for all $q \in \mathcal{Q}_{K,t}$ **do**
        Set $R_t(q;\lambda) = w_t(1-q) - \lambda q / r^t$.
        if $t = T$ **then**
            Set $C_{K,T}(q;\lambda) = 0$.
        else
            Compute

$$C_{K,t}(q;\lambda) = \sum_{\ell=1}^{K} |I_\ell| m_{K,\ell} V_{K,t+1}\left(\frac{qr}{m_{K,\ell}};\lambda\right).$$

        end if
        Set $V_{K,t}(q;\lambda) = \max\{R_t(q;\lambda), C_{K,t}(q;\lambda)\}$.
    end for
end for
Define

$$q_{K,t}(\mathbf{P}) = \frac{r^t}{\prod_{s=1}^t m_K(P_s)},$$
$$\tau_{K,\lambda}(\mathbf{P}) = \inf\left\{1 \le t \le T : R_t\big(q_{K,t}(\mathbf{P});\lambda\big) > C_{K,t}\big(q_{K,t}(\mathbf{P});\lambda\big)\right\}.$$

Part II: Find the smallest feasible multiplier.
For each queried multiplier λ, construct $\tau_{K,\lambda}$ using Part I and compute $\mathbb{P}_0\{\tau_{K,\lambda}(\mathbf{P}) \le T\}$ exactly by forward propagation over the finite states, using transition probabilities $|I_\ell|$ and accumulating the probability mass of paths that stop.
if $\mathbb{P}_0\{\tau_{K,0}(\mathbf{P}) \le T\} \le \alpha$ **then**
    Set $\widehat{\lambda}_K = 0$.
else
    Initialize $\lambda_- = 0$ and $\lambda_+ = w_1/\alpha$.
    while $\lambda_+ - \lambda_- > \eta$ **do**
        Set $\lambda = (\lambda_- + \lambda_+)/2$.
        Compute $\mathbb{P}_0\{\tau_{K,\lambda}(\mathbf{P}) \le T\}$ using the finite-state recursion.
        if $\mathbb{P}_0\{\tau_{K,\lambda}(\mathbf{P}) \le T\} \le \alpha$ **then**
            Set $\lambda_+ = \lambda$.
        else
            Set $\lambda_- = \lambda$.
        end if
    end while
    Set $\widehat{\lambda}_K = \lambda_+$.
end if
return $\widehat{\lambda}_K$ and $\widehat{\tau}_K = \tau_{K,\widehat{\lambda}_K}$, with $\inf \varnothing = \infty$.

C Optimal Stopping with Independent p-Values and Density Approximation

This appendix provides details for Section 3: posterior reduction, Bellman optimality, monotonicity in the multiplier, recovery of the constrained optimum, and density approximation. The approximation

Algorithm 2 Conformalized Optimal Stopping (COS)

Require: Non-hit reference scores $S_1^0, \dots, S_n^0$; incoming candidate scores $S_1, \dots, S_T$, with smaller values indicating stronger evidence of a hit; a fixed pathwise-monotone stopping family $\{\tau_\lambda\}_{\lambda \in \Lambda}$ on a finite increasing grid Λ, independent of calibration and deployment data; level α; and $B \geq \lceil 1/\alpha \rceil$.

Ensure: Calibrated multiplier $\hat{\lambda}$ and stopping time $\hat{\tau}$.

Part I: Calibrate using conformal-null trajectories.

Set $\tau_\infty \equiv \infty$.

for $j = 1, \dots, B$ **do**

Independently draw

$$W_1^{0,(j)}, \dots, W_n^{0,(j)}, W_1^{(j)}, \dots, W_T^{(j)}, U_1^{(j)}, \dots, U_T^{(j)} \overset{\text{iid}}{\sim} \text{Unif}(0,1).$$

Construct

$$\widetilde{P}_t^{(j)} = \frac{\sum_{i=1}^n \mathbf{1}\{W_i^{0,(j)} < W_t^{(j)}\} + U_t^{(j)}}{n+1}, \qquad t = 1, \dots, T.$$

Compute

$$\Gamma_j = \min\left\{\lambda \in \Lambda \cup \{\infty\} : \tau_\lambda(\widetilde{\mathbf{P}}^{(j)}) > T\right\}, \qquad \widetilde{\mathbf{P}}^{(j)} = (\widetilde{P}_t^{(j)})_{t=1}^T.$$

end for

Set $k = \lceil (B+1)(1-\alpha) \rceil$ and $\hat{\lambda} = \Gamma_{(k)}$, where $\Gamma_{(k)}$ is the k-th smallest critical multiplier.

Part II: Apply the calibrated rule to incoming scores.

if $\hat{\lambda} = \infty$ **then**

return $\hat{\tau} = \infty$.

end if

for $t = 1, \dots, T$ **do**

Observe S_t and independently draw $U_t \sim \text{Unif}(0,1)$.

Compute

$$\widetilde{P}_t = \frac{\sum_{i=1}^n \mathbf{1}\{S_i^0 < S_t\} + U_t}{n+1}.$$

if $\tau_{\hat{\lambda}}$ signals an alarm given $\widetilde{P}_{1:t}$ **then**

return $\hat{\tau} = t$.

end if

end for

return $\hat{\tau} = \infty$.

result compares expected rewards under the original working alternative after each family selects its own feasible multiplier.

Standing setup for independent p-values (I). We use the working alternative (2), with independent pairs (A_t, P_t) across time, $b \in (0,1)$, a fixed finite horizon T, and rewards $w_1 \geq \cdots \geq w_T > 0$. Under $\mathbb{P}_0$, the p-values are i.i.d. $\text{Unif}(0,1)$. Stopping times take values in $\{1, \dots, T, \infty\}$, are adapted to $\mathcal{F}_t^P = \sigma(P_1, \dots, P_t)$, and earn zero at infinity, with $w_\infty = 0$. All Bellman calculations are exact, with continuation at ties. Additional smoothness or atomlessness is imposed only where stated.

For the proofs, write

$$r := 1 - b, \qquad m(p) = r + b f_1(p), \qquad \ell(p) := \log\{m(p)/r\}, \qquad Z_t := \sum_{j=1}^t \ell(P_j), \quad Z_0 = 0.$$

Thus $Z_t = -\log q_t(\mathbf{P})$, where q_t is the state function defined in Section 3.2; its explicit form is derived below in (C.1). We set $q_0(\mathbf{P}) = 1$. For a full input path $z = (z_1, \dots, z_T)$, the expressions $q_t(z)$ and $\tau_\lambda(z)$ evaluate the same state function and stopping rule on that path. The state and the stopping decision at time t depend only on its first t entries. Using the expected reward J_w defined in Section 3.2, write the constrained optimum as

$$J_w^*(\alpha) := \sup_{\tau:\, \mathbb{P}_0(\tau \leq T) \leq \alpha} J_w(\tau).$$

The supremum is over the stopping times just specified. The fixed-multiplier value remains $V_0(\lambda) = \sup_\tau \{J_w(\tau) - \lambda \mathbb{P}_0(\tau \leq T)\}$.

C.1 Posterior reduction and the two terms in the Lagrangian

Define $\psi(p) := r/m(p)$. Since $m \geq r > 0$, the distributions of $P_{1:t}$ under the global null and the working alternative are mutually absolutely continuous for every $t \leq T$. Under $\mathbb{P}_{\mathcal{A}}$ its density is $\prod_{j=1}^{t} m(P_j)$, while its joint density with $\{\nu^* > t\}$ is r^t. Bayes' rule yields

$$q_t(\mathbf{P}) = \frac{r^t}{\prod_{j=1}^{t} m(P_j)}, \qquad q_{t+1}(\mathbf{P}) = q_t(\mathbf{P})\psi(P_{t+1}), \qquad t < T. \tag{C.1}$$

Given $q_t(\mathbf{P})$, the next observation has density m independently of the past; therefore $q_t(\mathbf{P})$ is a sufficient Markov state for the design problem. In particular, $0 \leq \psi \leq 1$ and

$$\int_0^1 \psi(p)m(p)\,\mathrm{d}p = r. \tag{C.2}$$

For a stopping time τ, the event $\{\tau = t\}$ is measurable at time t. Conditioning on the observed history gives

$$\begin{aligned} J_w(\tau) &= \sum_{t=1}^{T} \mathbb{E}_{\mathcal{A}}[w_t \mathbf{1}\{\nu^* \leq t\}\mathbf{1}\{\tau = t\}] \\ &= \sum_{t=1}^{T} \mathbb{E}_{\mathcal{A}}[w_t(1 - q_t(\mathbf{P}))\mathbf{1}\{\tau = t\}]. \end{aligned} \tag{C.3}$$

Define the null-to-working-alternative likelihood ratio on the observed prefix by

$$\Lambda_t := \frac{\mathrm{d}\mathbb{P}_0|_{\mathcal{F}_t^P}}{\mathrm{d}\mathbb{P}_{\mathcal{A}}|_{\mathcal{F}_t^P}} = \frac{q_t(\mathbf{P})}{r^t} = \frac{1}{\prod_{j=1}^{t} m(P_j)}.$$

Consequently

$$\begin{aligned} \mathbb{P}_0(\tau \leq T) &= \sum_{t=1}^{T} \mathbb{E}_{\mathcal{A}}[\Lambda_t \mathbf{1}\{\tau = t\}] \\ &= \mathbb{E}_{\mathcal{A}}[\Lambda_\tau \mathbf{1}\{\tau \leq T\}]. \end{aligned} \tag{C.4}$$

Expressions evaluated at $\tau = \infty$ are assigned zero when multiplied by $\mathbf{1}\{\tau \leq T\}$. Equations (C.3) and (C.4) prove

$$J_w(\tau) - \lambda \mathbb{P}_0(\tau \leq T) = \mathbb{E}_{\mathcal{A}}[R_\tau(q_\tau(\mathbf{P}); \lambda)\mathbf{1}\{\tau \leq T\}], \qquad R_t(q; \lambda) = w_t - \left(w_t + \frac{\lambda}{r^t}\right) q. \tag{C.5}$$

C.2 Fixed-multiplier Bellman optimality

For states $q \in [0, 1]$, use $C_T(q; \lambda) = 0$, $V_T(q; \lambda) = \max\{R_T(q; \lambda), 0\}$, and the backward recursion for $t = T - 1, \ldots, 1$. At time T, stopping gives $R_T(q; \lambda)$ and never raising an alarm gives zero. This proves the terminal recursion. Suppose that V_{t+1} dominates the conditional remaining value of every policy at time $t + 1$. Any policy at time t either stops, obtaining $R_t(q; \lambda)$, or continues. In the latter case its conditional value is at most

$$C_t(q; \lambda) = \int_0^1 V_{t+1}(q\psi(p); \lambda)m(p)\,\mathrm{d}p.$$

Conversely, choose a maximizing action at time t, taking continuation when the two Bellman payoffs are equal, and apply the same convention at all future states. This attains the maximum by the induction hypothesis. Backward induction proves fixed-multiplier optimality over all stopping times adapted to the observed history. The pre-observation value is

$$V_0(\lambda) = \int_0^1 V_1(\psi(p); \lambda)m(p)\,\mathrm{d}p; \tag{C.6}$$

there is no time-zero stop.

C.3 A DIRECT QUANTITATIVE PROOF OF THRESHOLD REGULARITY

For this Bernoulli one-hit problem, threshold regularity can be established directly. Define the stopping advantage $D_t(q;\lambda) := R_t(q;\lambda) - C_t(q;\lambda)$ for every $t = 1, \dots, T$. The following lemma is a consequence of the stated model.

Lemma C.1 (Lipschitz values and a strictly decreasing stopping advantage)**.** *For $\lambda \geq 0$, set $L_t(\lambda) = w_t + \lambda r^{-t}$. Then $V_t(\cdot;\lambda)$ is nonnegative, nonincreasing, convex, and $L_t(\lambda)$-Lipschitz on $[0,1]$. For $t < T$, put $d_t = w_t - rw_{t+1} > 0$. For $q_2 > q_1$,*

$$-L_t(\lambda)(q_2 - q_1) \leq D_t(q_2;\lambda) - D_t(q_1;\lambda) \leq -d_t(q_2 - q_1). \tag{C.7}$$

At T, the advantage is affine with slope $-L_T(\lambda)$.

Proof. Fix λ and suppress it in the arguments within this proof. The terminal value is the maximum of zero and the affine decreasing function $w_T - L_T q$, so it has all the stated properties. Suppose they hold at $t+1$. Composition with $q \mapsto q\psi(p)$ and integration preserve convexity and monotonicity. Moreover, for $q_2 > q_1$,

$$-rL_{t+1}(q_2 - q_1) \leq C_t(q_2) - C_t(q_1) \leq 0,$$

by (C.2). Because $rL_{t+1} = rw_{t+1} + \lambda r^{-t} \leq L_t$, taking the maximum of C_t and $R_t(\cdot;\lambda)$ again gives a decreasing, convex, L_t-Lipschitz function. Nonnegativity follows by induction from the option never to raise an alarm. Subtracting the displayed bounds for C_t from $R_t(q_2;\lambda) - R_t(q_1;\lambda) = -L_t(q_2 - q_1)$ proves (C.7). Strict positivity of d_t follows from $w_t \geq w_{t+1} > 0$ and $r < 1$. □

For $t < T$, at $q = 0$ the posterior update stays zero and the largest future reward is w_{t+1}, so $V_t(0;\lambda) = w_t$ and $D_t(0;\lambda) = w_t - w_{t+1} \geq 0$. For $\lambda > 0$, $R_t(1;\lambda) = -\lambda r^{-t} < 0$, whereas $C_t(1;\lambda) \geq 0$. At $\lambda = 0$, the policy that raises an alarm at time $t+1$ gives

$$C_t(1;0) \geq w_{t+1} \int_0^1 \left(1 - \frac{r}{m(p)}\right) m(p)\,\mathrm{d}p = bw_{t+1} > 0,$$

while $R_t(1;0) = 0$. Thus $D_t(1;\lambda) < 0$ for every $\lambda \geq 0$. Lemma C.1 gives a unique zero $c_t(\lambda) \in [0,1)$ and the strict stopping region $[0, c_t(\lambda))$. For $t < T$, the cutoff lies in $(0,1)$ when $w_t > w_{t+1}$ and equals zero when $w_t = w_{t+1}$. At the terminal time,

$$c_T(\lambda) = \frac{w_T r^T}{w_T r^T + \lambda}, \qquad \lambda \geq 0, \tag{C.8}$$

so $c_T(0) = 1$. This establishes the cutoff representation in (4) for every nonnegative multiplier.

C.4 MONOTONICITY IN THE MULTIPLIER

Proof of Proposition 1. Let $0 \leq \lambda_1 < \lambda_2$. By the sufficient-state reduction, we may restrict attention, without loss of optimality, to policies whose decisions at time s depend only on $(s, q_s(\mathbf{P}))$. Fix a time $t < T$ and such a continuation policy π, which continues at time t, with subsequent stopping time $\sigma \in \{t+1, \dots, T, \infty\}$.

The reverse likelihood ratio on $\mathcal{F}_s^P$ is $\Lambda_s = q_s(\mathbf{P})/r^s$. The conditional change-of-measure identity gives, for $s > t$,

$$\mathbb{E}_{\mathcal{A}}[\Lambda_s \mathbf{1}\{\sigma = s\} \mid \mathcal{F}_t^P] = \Lambda_t \mathbb{P}_0(\sigma = s \mid \mathcal{F}_t^P).$$

The prefix likelihood ratio is a function of $q_t(\mathbf{P})$, so the two laws have the same conditional distribution of the prefix given $q_t(\mathbf{P})$. Summing over $s = t+1, \dots, T$ and conditioning on this state therefore gives

$$B_\pi := \mathbb{E}_{\mathcal{A}}[\Lambda_\sigma \mathbf{1}\{\sigma \leq T\} \mid q_t(\mathbf{P}) = q] = \frac{q}{r^t}\mathbb{P}_0(\sigma \leq T \mid q_t(\mathbf{P}) = q), \qquad 0 \leq B_\pi \leq \frac{q}{r^t}. \tag{C.9}$$

For the fixed continuation policy π, its conditional payoff is $u_\pi - \lambda B_\pi$, where u_π is its expected one-hit reward. Thus $C_t(q;\lambda)$ is the supremum of affine functions with slopes in $[-q/r^t, 0]$. Therefore, for $\lambda_2 > \lambda_1$,

$$-(\lambda_2 - \lambda_1)q/r^t \leq C_t(q;\lambda_2) - C_t(q;\lambda_1) \leq 0.$$

The immediate stopping payoff decreases by exactly $(\lambda_2 - \lambda_1)q/r^t$. Consequently,

$$-(\lambda_2 - \lambda_1)q/r^t \le D_t(q;\lambda_2) - D_t(q;\lambda_1) \le 0. \tag{C.10}$$

The terminal case follows by setting $C_T = 0$. Under the continue-on-tie convention, the strict stopping region at the larger multiplier is contained in that at the smaller multiplier. Since the state update does not depend on the multiplier, $\tau_{\lambda_1}(\mathbf{P}) \le \tau_{\lambda_2}(\mathbf{P})$ on every potential full path. The stopping probability is therefore nonincreasing under any fixed input law.

Expected reward monotonicity follows from fixed-multiplier optimality. Put $J_i = J_w(\tau_{\lambda_i}(\mathbf{P}))$ and $F_i = \mathbb{P}_0(\tau_{\lambda_i}(\mathbf{P}) \le T)$. Under the same working alternative,

$$\begin{aligned} J_1 - \lambda_1 F_1 &\ge J_2 - \lambda_1 F_2, \\ J_2 - \lambda_2 F_2 &\ge J_1 - \lambda_2 F_1. \end{aligned}$$

Rearranging gives

$$\lambda_1(F_1 - F_2) \le J_1 - J_2 \le \lambda_2(F_1 - F_2).$$

Since $F_1 \ge F_2$ and $\lambda_1 \ge 0$, also $J_1 \ge J_2$. This proves Proposition 1. □

C.5 Continuity and deterministic constrained recovery

Proof of Theorem 1. Combining (C.7) and (C.10) gives, for $t < T$ and positive multipliers,

$$|c_t(\lambda_2) - c_t(\lambda_1)| \le \frac{r^{-t}}{d_t}|\lambda_2 - \lambda_1|, \qquad d_t = w_t - r w_{t+1} > 0. \tag{C.11}$$

Suppose $\ell(P)$ is non-atomic for $P \sim \mathrm{Unif}(0,1)$. Under the null, $-\log q_t(\mathbf{P}) = \sum_{j=1}^{t} \ell(P_j)$, so independence makes every $q_t(\mathbf{P})$ non-atomic. If $\lambda_j \to \lambda > 0$, boundary continuity ensures that every stop/continue comparison agrees eventually outside $\bigcup_{t \le T}\{q_t(\mathbf{P}) = c_t(\lambda)\}$, a null event. Bounded convergence therefore gives

$$\mathbb{P}_0(\tau_{\lambda_j}(\mathbf{P}) \le T) \longrightarrow \mathbb{P}_0(\tau_{\lambda}(\mathbf{P}) \le T).$$

Both endpoints follow from the model. Non-atomicity implies $\mathbb{P}_0\{\ell(P) = 0\} = 0$, hence $q_T(\mathbf{P}) < 1$ almost surely. Since $c_T(\lambda) \uparrow 1$ as $\lambda \downarrow 0$ and every path with $q_T(\mathbf{P}) < c_T(\lambda)$ stops by T, we have $\mathbb{P}_0(\tau_\lambda(\mathbf{P}) \le T) \to 1$. At the other endpoint, the never-stop option gives

$$0 \le V_0(\lambda) = J_w(\tau_\lambda(\mathbf{P})) - \lambda \mathbb{P}_0(\tau_\lambda(\mathbf{P}) \le T).$$

Because $J_w(\tau_\lambda(\mathbf{P})) \le w_1$, it follows that $\mathbb{P}_0(\tau_\lambda(\mathbf{P}) \le T) \le w_1/\lambda \to 0$. Continuity and monotonicity now yield $0 < \lambda^* < \infty$ and $\mathbb{P}_0(\tau_{\lambda^*}(\mathbf{P}) \le T) = \alpha$ for λ^* as defined in Theorem 1.

To establish constrained optimality, let τ be any feasible stopping time. Fixed-multiplier optimality gives

$$\begin{aligned} J_w(\tau) &\le V_0(\lambda^*) + \lambda^* \mathbb{P}_0(\tau \le T) \\ &\le V_0(\lambda^*) + \lambda^* \alpha = J_w(\tau_{\lambda^*}(\mathbf{P})). \end{aligned} \tag{C.12}$$

This proves Theorem 1 directly. □

For example, $f_1(p) = \eta p^{\eta - 1}$ with $0 < \eta < 1$ gives a continuous, strictly decreasing $\ell(p)$ on $(0,1)$ and therefore satisfies the non-atomicity condition.

C.6 Piecewise-constant approximation of the continuous oracle

We give the finite recursion used in Section 3.2 and bound its reward loss after each family selects its own population-calibrated multiplier. Both rewards are evaluated under the original working alternative.

Uniform density approximation (D). The target density f_1 is continuous on $[0,1]$. Each deterministic $f_{1,K}$ is a nonnegative probability density with finite constant heights on a partition $\{I_\ell\}_{\ell=1}^K$ of $(0,1)$ into intervals of positive length, and

$$\varepsilon_K := \|f_{1,K} - f_1\|_\infty \longrightarrow 0. \tag{C.13}$$

Both models use the same b, T, w, uniform null distribution, and exact Bellman recursion. Define $m_K(p) := r + bf_{1,K}(p)$ and $\ell_K(p) := \log\{m_K(p)/r\}$. Then

$$\|m_K - m\|_\infty = b\varepsilon_K \le \varepsilon_K.$$

For example, cell averages on partitions with mesh $h_K := \max_{1\le\ell\le K} |I_\ell| \to 0$ satisfy

$$f_{1,K}(p) = |I_\ell|^{-1} \int_{I_\ell} f_1(u)\,\mathrm{d}u, \quad p \in I_\ell, \qquad \varepsilon_K \le \omega_{f_1}(h_K) \to 0, \tag{C.14}$$

where $\omega_{f_1}(h) := \sup_{|u-v|\le h} |f_1(u) - f_1(v)|$ is the modulus of continuity on $[0,1]$. Values at partition endpoints are assigned consistently and do not affect the integrals. Exact cell masses are not needed for (D).

Finite recursion and separate calibration. Let $\mathbb{E}_{\mathcal{A},K}$ denote expectation under the working alternative obtained by replacing f_1 by $f_{1,K}$, and set

$$J_{w,K}^{\mathrm{app}}(\tau) := \mathbb{E}_{\mathcal{A},K}[w_\tau \mathbf{1}\{\nu^* \le \tau \le T\}].$$

Let $m_{K,\ell}$ denote the constant value of m_K on I_ℓ. Define $V_{K,t}$ and $C_{K,t}$ by the same Bellman recursion as before with m replaced by m_K. The induced state function and continuation payoff are

$$q_{K,t}(\mathbf{P}) = \frac{r^t}{\prod_{j=1}^t m_K(P_j)}, \qquad q_{K,0}(\mathbf{P}) = 1, \tag{C.15}$$

and

$$C_{K,t}(q;\lambda) = \sum_{\ell=1}^K |I_\ell| m_{K,\ell} V_{K,t+1}(qr/m_{K,\ell};\lambda), \qquad 1 \le t < T. \tag{C.16}$$

The immediate stopping payoff remains $R_t(q;\lambda)$, with $C_{K,T} = 0$ and $V_{K,T} = \max\{R_T, 0\}$. Finite K, T give finitely many reachable states, although exact enumeration can be large. Define $D_{K,t}(q;\lambda) := R_t(q;\lambda) - C_{K,t}(q;\lambda)$. Its unique zero is denoted by $c_{K,t}(\lambda)$; the preceding threshold argument applies with f_1 replaced by $f_{1,K}$. Thus

$$\tau_{K,\lambda}(\mathbf{P}) := \inf\{1 \le t \le T : q_{K,t}(\mathbf{P}) < c_{K,t}(\lambda)\}.$$

For separate population calibration, define

$$\begin{aligned} F_K(\lambda) &= \mathbb{P}_0(\tau_{K,\lambda}(\mathbf{P}) \le T), \\ \lambda_K^* &= \inf\{\lambda \ge 0 : F_K(\lambda) \le \alpha\}. \end{aligned} \tag{C.17}$$

Unlike the original family with atomless evidence, this discrete family may jump across α; the following bound accounts for the resulting unused budget.

Theorem C.1. *Suppose (D) holds and, under the original working alternative, the log evidence Z_t has a density $g_{\mathcal{A},t}$ satisfying*

$$\sup_{t\le T} \|g_{\mathcal{A},t}\|_\infty \le \mathsf{M} < \infty.$$

Then there exists a constant $C < \infty$, independent of K, such that, for all sufficiently large K,

$$0 \le J_w(\tau_{\lambda^*}(\mathbf{P})) - J_w(\tau_{K,\lambda_K^*}(\mathbf{P})) \le C\varepsilon_K. \tag{C.18}$$

For a fixed target model, horizon, and reward schedule, the loss is $O(\varepsilon_K) \to 0$. Finer approximations trade a larger finite-state calculation for a reward closer to the original optimum.

Proof of Theorem C.1. The density assumption on Z_1 and equivalence of the original working-alternative and global-null observation laws imply that $\ell(P)$ is non-atomic under the Uniform

null. Theorem 1 therefore gives the original optimum $J_w^*(\alpha) = J_w(\tau_{\lambda^*}(\mathbf{P}))$. We first prove a concentration-function bound and then apply the density assumption.

1. Feasibility and eventual positive multipliers. Multiplier monotonicity applies to every approximate density. For a fixed input path, the posterior is independent of λ, and each advantage is continuous and nonincreasing in λ. Thus, for $\lambda_j \downarrow \lambda$,

$$\mathbf{1}\{D_{K,t}(q;\lambda_j) > 0\} \longrightarrow \mathbf{1}\{D_{K,t}(q;\lambda) > 0\}.$$

With finite T, bounded convergence makes F_K right-continuous and nonincreasing. Bellman optimality and the never-stop option give

$$0 \le J_{w,K}^{\mathrm{app}}(\tau_{K,\lambda}(\mathbf{P})) - \lambda F_K(\lambda), \qquad J_{w,K}^{\mathrm{app}}(\tau_{K,\lambda}(\mathbf{P})) \le w_1.$$

Hence $F_K(\lambda) \le w_1/\lambda$ for $\lambda > 0$. The multiplier w_1/α is feasible, and right-continuity makes the feasibility infimum attainable. In particular,

$$0 \le \lambda_K^* \le w_1/\alpha$$

for every K.

Let $\mathcal{Z}_K := \{p \in (0,1) : f_{1,K}(p) = 0\}$. At multiplier zero, every path with $q_{K,T}(\mathbf{P}) < 1$ stops by T. If $q_{K,T}(\mathbf{P}) = 1$, all earlier states are also one and their stopping payoffs are zero, so the rule continues throughout. Therefore,

$$F_K(0) = 1 - \mathrm{Leb}(\mathcal{Z}_K)^T, \qquad \mathrm{Leb}(\mathcal{Z}_K) \le \mathrm{Leb}\{f_1 \le \varepsilon_K\} \longrightarrow 0. \tag{C.19}$$

The limit follows from non-atomicity, which implies $\mathrm{Leb}\{f_1 = 0\} = 0$. Thus $F_K(0) > \alpha$ for all sufficiently large K; right-continuity then implies

$$0 < \lambda_K^* \le w_1/\alpha.$$

Fix such a K for the rest of the proof.

2. The unused budget is bounded by boundary mass. Put $s_K := \alpha - F_K(\lambda_K^*) \ge 0$, write $F_K(\lambda^-) := \lim_{u\uparrow\lambda} F_K(u)$, and define

$$E_{K,\lambda} := \bigcup_{t=1}^{T} \{D_{K,t}(q_{K,t}(\mathbf{P});\lambda) = 0\}.$$

Minimality of the positive multiplier gives

$$s_K \le F_K((\lambda_K^*)^-) - F_K(\lambda_K^*) \le \mathbb{P}_0(E_{K,\lambda_K^*}).$$

For the second inequality, take a path that stops at multipliers approaching λ_K^* from below but does not stop at λ_K^*. A subsequence stops at one common time. Continuity makes the advantage there nonnegative in the limit, while non-stopping makes it nonpositive; hence it is zero.

Since $m, m_K \ge r$ and $b \le 1$,

$$\|\ell_K - \ell\|_\infty \le \varepsilon_K/r, \qquad \left|\sum_{j=1}^{t} \ell_K(P_j) - Z_t\right| \le t\varepsilon_K/r. \tag{C.20}$$

At a positive multiplier, a reachable tie is equivalent to $q_{K,t}(\mathbf{P}) = c_{K,t}(\lambda_K^*) > 0$. Whenever $c_{K,t}(\lambda_K^*) > 0$, define $B_{K,t} := -\log c_{K,t}(\lambda_K^*)$. A tie then implies

$$|Z_t - B_{K,t}| \le t\varepsilon_K/r.$$

A zero cutoff cannot be reached by a positive posterior. Define

$$\omega_{0,t}(u) = \sup_{x\in\mathbb{R}} \mathbb{P}_0(|Z_t - x| \le u), \qquad \chi_K = \sum_{t=1}^{T} \omega_{0,t}(t\varepsilon_K/r). \tag{C.21}$$

The union bound yields $s_K \le \chi_K$, and hence

$$0 \le \lambda_K^* s_K \le (w_1/\alpha)\chi_K. \tag{C.22}$$

3. Density-design error at an arbitrary multiplier. The original and approximate one-step joint laws Q, Q_K of a single pair (A, P) have densities

$$Q(0,p) = Q_K(0,p) = r, \qquad Q(1,p) = bf_1(p), \qquad Q_K(1,p) = bf_{1,K}(p).$$

Using $d_{\mathrm{TV}}(\mu,\nu) := \sup_D |\mu(D) - \nu(D)|$, where the supremum is over measurable sets, we obtain

$$d_{\mathrm{TV}}(Q, Q_K) = \tfrac{1}{2}\|m - m_K\|_1 \le \varepsilon_K/2, \qquad d_{\mathrm{TV}}(Q^{\otimes T}, Q_K^{\otimes T}) \le T\varepsilon_K/2.$$

Every realized one-hit reward lies in $[0, w_1]$. Let $\eta_K := w_1 T\varepsilon_K/2$, and let π range over the stopping times specified in the standing setup. Then

$$\sup_\pi |J_w(\pi) - J_{w,K}^{\mathrm{app}}(\pi)| \le \eta_K.$$

The joint laws, rather than just the P marginals, are needed because the reward includes the latent hit event. At any $\lambda \ge 0$, the objectives

$$L_\lambda(\pi) = J_w(\pi) - \lambda \mathbb{P}_0(\pi \le T)$$

and

$$L_{K,\lambda}^{\mathrm{app}}(\pi) = J_{w,K}^{\mathrm{app}}(\pi) - \lambda \mathbb{P}_0(\pi \le T)$$

have identical null penalties. Their suprema differ by at most η_K. Since $\tau_{K,\lambda}(\mathbf{P})$ maximizes the latter objective,

$$\begin{aligned} 0 &\le V_0(\lambda) - L_\lambda(\tau_{K,\lambda}(\mathbf{P})) \\ &= \left\{ V_0(\lambda) - \sup_\pi L_{K,\lambda}^{\mathrm{app}}(\pi) \right\} \\ &\quad + \left\{ L_{K,\lambda}^{\mathrm{app}}(\tau_{K,\lambda}(\mathbf{P})) - L_\lambda(\tau_{K,\lambda}(\mathbf{P})) \right\} \\ &\le 2\eta_K = w_1 T\varepsilon_K. \end{aligned} \tag{C.23}$$

4. Combine design error and unused budget. Fixed-multiplier optimality gives

$$J_w^*(\alpha) \le V_0(\lambda_K^*) + \lambda_K^* \alpha.$$

Thus,

$$\begin{aligned} &J_w^*(\alpha) - J_w(\tau_{K,\lambda_K^*}(\mathbf{P})) \\ &\quad \le V_0(\lambda_K^*) - J_w(\tau_{K,\lambda_K^*}(\mathbf{P})) \\ &\qquad + \lambda_K^* \mathbb{P}_0(\tau_{K,\lambda_K^*}(\mathbf{P}) \le T) + \lambda_K^* s_K \\ &\quad \le w_1 T\varepsilon_K + (w_1/\alpha)\chi_K. \end{aligned} \tag{C.24}$$

The approximate rule is feasible under the same original null, giving nonnegativity. Bounded rewards also give

$$0 \le J_w(\tau_{\lambda^*}(\mathbf{P})) - J_w(\tau_{K,\lambda_K^*}(\mathbf{P})) \le w_1 \min\{1, T\varepsilon_K + \chi_K/\alpha\}. \tag{C.25}$$

5. Apply the evidence-density bound. The prefix likelihood ratio of the original working alternative relative to the null is $r^t e^{Z_t}$, so the original null density of Z_t satisfies

$$g_{0,t}(z) = r^{-t} e^{-z} g_{\mathcal{A},t}(z) \le \mathsf{M} r^{-t}, \qquad z \ge 0.$$

Consequently,

$$\chi_K \le \frac{2\mathsf{M}\varepsilon_K}{r} \sum_{t=1}^{T} t r^{-t}.$$

Substitution in (C.25) proves (C.18), with

$$C = w_1 \left[T + \frac{2\mathsf{M}}{\alpha r} \sum_{t=1}^{T} t r^{-t} \right].$$

Under atomlessness alone, each Z_t is non-atomic under the null, so $\omega_{0,t}(u) \to 0$ as $u \downarrow 0$. The same proof then gives reward convergence without a specified rate. □

D SCORE-TO-p CONSTRUCTION AND ITS NULL LAW

Standing score-reference setup (S). Condition on design information that is independent of calibration and deployment. For the reward comparisons, use the score version of the working alternative in (2): the pairs (A_t, S_t) are independent, $A_t \sim \text{Bernoulli}(b)$, the non-hit score CDF F_0 is continuous and common across times, and $P_t = F_0(S_t)$ has density f_1 conditional on $A_t = 1$. The reference scores $S_1^0, \ldots, S_n^0$ are i.i.d. from F_0 and independent of all candidate pairs. The rank randomizers $U_1, \ldots, U_T$ are i.i.d. $\text{Unif}(0,1)$ and independent of all scores and hit indicators. Define $\widetilde{P}_t$ by (6); ideal and conformal inputs are coupled using the same scores and hit indicators. Write $\mathbf{P} = (P_t)_{t=1}^T$ and $\widetilde{\mathbf{P}} = (\widetilde{P}_t)_{t=1}^T$. The state function $q_t(\cdot)$ and thresholds $c_t(\lambda)$ are those derived under the independent-input model; only their input path is replaced when defining $\tau_\lambda(\widetilde{\mathbf{P}})$ in (7). Under the global null, all reference and candidate scores are i.i.d. from F_0.

These independence assumptions are used for reward comparisons. The rank-based null law also holds under joint exchangeability of the reference and null candidate scores, provided there are almost surely no ties, as shown below.

D.1 MARGINAL VALIDITY

If $(S_1^0, \ldots, S_n^0, S_t)$ are jointly exchangeable conditional on $A_t = 0$ and have no ties, the rank count $\sum_{i=1}^n \mathbf{1}\{S_i^0 < S_t\}$ is uniform on $\{0, \ldots, n\}$. It is independent of U_t, so (6) gives $\widetilde{P}_t \mid A_t = 0 \sim \text{Unif}(0,1)$ (Vovk et al., 2005).

D.2 CONDITIONAL STRUCTURE AND OBSERVATION COUPLING

Under setup (S), conditional on the reference, the conformal null values are i.i.d. with conditional CDF

$$H_n(u) := \mathbb{P}_0(\widetilde{P}_t \le u \mid S_1^0, \ldots, S_n^0), \qquad u \in [0,1].$$

Averaging over this shared reference induces dependence. Let

$$F_{0,n}(s) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{S_i^0 \le s\}, \qquad r_n(\delta) := \sqrt{\frac{\log(2/\delta)}{2n}} + \frac{1}{n+1}, \quad \delta \in (0,1).$$

Because reference–candidate ties have probability zero, $\widetilde{P}_t = \{nF_{0,n}(S_t) + U_t\}/(n+1)$ almost surely, and

$$|\widetilde{P}_t - P_t| \le \|F_{0,n} - F_0\|_\infty + \frac{1}{n+1}.$$

The Dvoretzky–Kiefer–Wolfowitz (DKW) inequality (Massart, 1990) gives a reference event of probability at least $1 - \delta$ on which

$$\max_{t \le T} |\widetilde{P}_t - P_t| \le r_n(\delta). \tag{D.1}$$

The bound holds under both the global null and the working alternative. Under the global null, P_t remains Uniform conditional on the reference. Hence, on the same reference event,

$$\sup_{u \in [0,1]} |H_n(u) - u| \le r_n(\delta). \tag{D.2}$$

D.3 THE DISTRIBUTION-FREE TRAJECTORY SIMULATOR

Under the i.i.d. global null, transforming all $n + T$ scores by F_0 gives independent Uniform variables and preserves every reference–candidate comparison almost surely. Applying (6) to Uniform pseudo-scores therefore reproduces the complete law of $\widetilde{\mathbf{P}}$ for every continuous F_0 (Vovk et al., 2005; Gazin et al., 2024). More generally, joint exchangeability with no ties makes all strict orderings of the $n + T$ labelled scores equally likely. Their reference–candidate rank comparisons thus have the same joint law as those from independent Uniform pseudo-scores. Independent rank randomizers then reproduce the same conformal-null path law. Use one shared pseudo-reference within each simulated trajectory, and redraw all variables independently for each new trajectory.

E Distribution-Free Calibration and Implementation

This section proves Theorem 2 and records the same calibration argument for finite multiplier grids and numerical implementations. The argument requires a fixed monotone stopping family and exchangeable complete null paths; it does not require a correctly specified working alternative or an exact numerical solver.

E.1 Calibrating a fixed monotone stopping family

Let $\mathcal{D}$ denote the design information used for model fitting and numerical construction, independent of the calibration and deployment data, including the deployment reference. Conditional on $\mathcal{D}$, fix a measurable implemented family $\{\tau_\lambda^{\mathrm{num}}\}_{\lambda\in\Lambda}$, where Λ is either $[0,\infty)$ or a finite increasing multiplier grid. For a complete input path $z=(z_1,\dots,z_T)$, each $\tau_\lambda^{\mathrm{num}}(z)$ is a stopping time in $\{1,\dots,T,\infty\}$ and the family satisfies

$$\lambda\le\lambda' \quad\Longrightarrow\quad \tau_\lambda^{\mathrm{num}}(z)\le\tau_{\lambda'}^{\mathrm{num}}(z).$$

Set $\tau_\infty^{\mathrm{num}}\equiv\infty$. The superscript distinguishes this general implementation from the optimal family $\{\tau_\lambda\}$ in Section 3.2. Its instances include the exact Bellman family evaluated on an input path, and the piecewise-constant family $\{\tau_{K,\lambda}\}$ from Appendix C.6. In particular, evaluating the exact family on $\widetilde{\mathbf{P}}$ gives $\tau_\lambda(\widetilde{\mathbf{P}})$ in (7).

E.2 Two exact null-path generators

Independent p-values. Generate $\mathbf{P}\sim\mathrm{Unif}(0,1)^{\otimes T}$, matching the independent p-value null model.

Shared-reference conformal p-values. For each trajectory j, draw

$$W_1^{0,(j)},\dots,W_n^{0,(j)},W_1^{(j)},\dots,W_T^{(j)},U_1^{(j)},\dots,U_T^{(j)}\overset{\mathrm{iid}}{\sim}\mathrm{Unif}(0,1)$$

and construct

$$\widetilde{P}_t^{(j)}=\frac{\sum_{i=1}^n\mathbf{1}\{W_i^{0,(j)}<W_t^{(j)}\}+U_t^{(j)}}{n+1},\qquad t=1,\dots,T.$$

Write $\widetilde{\mathbf{P}}^{(j)}=(\widetilde{P}_t^{(j)})_{t=1}^T$. The reference $W_{1:n}^{0,(j)}$ is shared within a trajectory; all variables are redrawn independently between trajectories and independently of deployment. Under the continuous score model, this is the complete deployment conformal-null law for every F_0. More generally, the same identity holds whenever the reference and global-null candidate scores are jointly exchangeable and have no ties: their ordering is uniformly distributed, and the construction depends only on that ordering and the independent rank randomizers. The rank argument is given in Appendix D.3.

E.3 Critical multipliers

For each input path, define its critical multiplier by

$$\Gamma(z):=\inf\{\lambda\in\Lambda:\tau_\lambda^{\mathrm{num}}(z)>T\},\qquad \inf\varnothing:=\infty.$$

For a finite grid, this is the minimum over $\Lambda\cup\{\infty\}$. For $\Lambda=[0,\infty)$, require that the no-alarm set is closed, so that a finite infimum is attained. This property holds for the Bellman families used here: the stopping advantages are continuous in λ, and equality leads to continuation. Consequently, for every $\lambda\in\Lambda\cup\{\infty\}$,

$$\tau_\lambda^{\mathrm{num}}(z)\le T\quad\Longleftrightarrow\quad\Gamma(z)>\lambda. \tag{E.1}$$

Thus $\Gamma(z)$ is the smallest admissible multiplier that prevents an alarm on the path, allowing ∞ if no finite multiplier does so. Atoms in its distribution are permitted.

Let $z^{(1)},\dots,z^{(B)}$ denote the complete calibration null paths, each generated by the appropriate simulator above, and compute $\Gamma_j=\Gamma(z^{(j)})$. In the conformal case, $z^{(j)}=\widetilde{\mathbf{P}}^{(j)}$, matching the main-text critical multipliers. Define

$$k=\lceil(B+1)(1-\alpha)\rceil,\qquad \hat\lambda=\begin{cases}\Gamma_{(k)}, & k\le B,\\ \infty, & k=B+1,\end{cases}$$

where $\Gamma_{(k)}$ is the k-th smallest calibration value. The main text assumes $B\ge\lceil1/\alpha\rceil$, which ensures $k\le B$ and recovers the calibrated rule in Section 4.2. The implemented calibrated stopping time is $\hat\tau^{\mathrm{num}}:=\tau_{\hat\lambda}^{\mathrm{num}}(z^{(B+1)})$, with $z^{(B+1)}$ denoting the deployment input path.

E.4 Proof of Theorem 2

Proof of Theorem 2. Conditional on $\mathcal{D}$, suppose that $z^{(1)}, \dots, z^{(B+1)}$ are exchangeable complete null input paths, with the last used for deployment. The null generators above supply this property under the corresponding deployment null model. Since the family is fixed, the critical multipliers $\Gamma_1, \dots, \Gamma_{B+1}$ are also exchangeable.

For $k \le B$, (E.1) gives

$$\{\hat{\tau}^{\mathrm{num}} \le T\} = \{\Gamma_{B+1} > \Gamma_{(k)}\}.$$

This identity includes $\hat{\lambda} = \infty$, when both events are empty. By exchangeability,

$$\begin{aligned}\mathbb{P}_0(\Gamma_{B+1} > \Gamma_{(k)} \mid \mathcal{D}) &= \frac{1}{B+1}\mathbb{E}_0\left[\sum_{j=1}^{B+1} \mathbf{1}\left\{\sum_{i\neq j} \mathbf{1}\{\Gamma_i < \Gamma_j\} \ge k\right\} \,\middle|\, \mathcal{D}\right] \\ &\le \frac{B+1-k}{B+1} \le \alpha.\end{aligned}$$

Indeed, at most $B+1-k$ observations can have at least k strictly smaller peers; ties can only make the bound conservative. If $k = B+1$, the selected rule never stops.

For the exact Bellman family on conformal inputs, $\hat{\tau}^{\mathrm{num}} = \tau_{\hat{\lambda}}(\widetilde{\mathbf{P}}) = \hat{\tau}$. Writing the joint null probability as an expectation over the calibration bank gives

$$\mathbb{E}_{\mathrm{cal}}[\mathbb{P}_0(\hat{\tau} \le T)] \le \frac{B+1-k}{B+1} \le \alpha,$$

where the inner probability averages over the deployment scores, reference, and rank randomizers for the selected multiplier. The bound also holds after averaging over $\mathcal{D}$, completing the proof. □

The guarantee applies directly to the implemented family, including piecewise-constant approximations and numerical solvers, provided its pathwise monotonicity and the critical-event identity are preserved.

F Reward of the Calibrated Policy

We first prove Theorem 3 for the original density and calibration over the nonnegative half-line. We then extend the result to a piecewise-constant density and a finite multiplier grid. All rewards are evaluated under the original working alternative. Rewards of calibrated procedures also average over the calibration randomness, the deployment reference, and rank uniforms.

F.1 Exact density and continuous multiplier calibration

Condition on the design information, which is independent of all calibration and deployment data. Under the working alternative, the pairs (A_t, S_t) are independent across time, $A_t \overset{\mathrm{iid}}{\sim} \mathrm{Bernoulli}(b)$, and the non-hit score CDF F_0 is continuous. The ideal coordinates $P_t = F_0(S_t)$ obey (2). We retain the notation $r = 1-b$, $m(p) = r + bf_1(p)$, $\ell(p) = \log\{m(p)/r\}$, and $Z_t = \sum_{s=1}^{t} \ell(P_s)$ from Appendix C. The horizon T, level $\alpha \in (0,1)$, and rewards $w_1 \ge \cdots \ge w_T > 0$ are fixed, with $w_\infty = 0$.

Throughout this section, f_1 is Lipschitz on $[0,1]$, with Lipschitz constant L_f. Let $g_{\mathcal{A},t}$ denote the density of Z_t under the original working alternative, and assume

$$\sup_{t \le T} \|g_{\mathcal{A},t}\|_\infty \le \mathsf{M} < \infty.$$

In particular, m is bounded, and ℓ is Lipschitz with constant $L_\ell = bL_f/r$.

The deployment reference $S_1^0, \dots, S_n^0$ consists of independent draws from F_0, independent of the candidate pairs. Construct $\widetilde{P}_t$ as in (6), with the U_t's independent of all scores, hit indicators, design information, and calibration randomness.

To distinguish a stopping rule from its inputs, we write $\tau_\lambda(z)$ for the evaluation of the same Bellman rule on a complete deterministic path $z = (z_1, \dots, z_T) \in [0,1]^T$:

$$q_t(z) = \frac{r^t}{\prod_{s=1}^t m(z_s)}, \qquad \tau_\lambda(z) = \inf\{1 \le t \le T : q_t(z) < c_t(\lambda)\}.$$

Thus the ideal stopping time is $\tau_\lambda(\mathbf{P})$, whereas $\tau_\lambda(\widetilde{\mathbf{P}})$ is the conformal-input rule in (7). Both evaluations use the same state function $q_t(\cdot)$ and thresholds $c_t(\lambda)$. The strict comparison implements continuation at equality, and $\tau_\lambda(z)$ is nondecreasing in λ for every path. Choose a fixed constant $\bar{m} > \max\{1, \|m\|_\infty\}$ and set

$$\bar{\lambda} = \max_{t \le T} w_t \bar{m}^t.$$

This multiplier never stops: on a reachable state, $q_t(z)/r^t \ge \bar{m}^{-t}$, so the immediate penalized payoff is at most $w_t - \bar{\lambda}\bar{m}^{-t} \le 0$, whereas continuation is nonnegative. Hence calibration over $[0, \infty)$ has finite critical multipliers. For a complete input path z, define

$$\Gamma(z) = \inf\{\lambda \ge 0 : \tau_\lambda(z) > T\}.$$

The Bellman advantages are continuous in λ. Together with pathwise monotonicity and continuation at equality, this makes the no-alarm set a closed upper interval. The infimum is therefore attained, agreeing with the minimum in Section 4.2. Thus

$$0 \le \Gamma(z) \le \bar{\lambda}, \qquad \tau_\lambda(z) \le T \quad \Longleftrightarrow \quad \Gamma(z) > \lambda.$$

Generate B independent complete conformal-null paths $\widetilde{\mathbf{P}}^{(1)}, \dots, \widetilde{\mathbf{P}}^{(B)}$, independently of deployment, as in Section 4.2. Each path uses its own independent reference of size n, shared within that path, and independent rank uniforms. For $j = 1, \dots, B$, let $\Gamma_j = \Gamma(\widetilde{\mathbf{P}}^{(j)})$. Choose $B \ge \lceil 1/\alpha \rceil$, as in the main text, so that $k = \lceil (B+1)(1-\alpha) \rceil \le B$. The calibrated quantities in Section 4.2 are

$$\hat{\lambda} = \Gamma_{(k)}, \qquad \hat{\tau} = \tau_{\hat{\lambda}}(\widetilde{\mathbf{P}}), \tag{F.1}$$

where $\Gamma_{(k)}$ is the k-th smallest calibration value. The convention $\tau_\infty(z) \equiv \infty$ applies if a calibration rule returns the never-stop option. Write the calibrated expected reward as

$$\widehat{J}_{n,B} = \mathbb{E}_{\mathrm{cal},\mathcal{A}}\left[w_{\hat{\tau}} \mathbf{1}\{\nu^* \le \hat{\tau} \le T\}\right]. \tag{F.2}$$

The calibration bank follows its null simulation law, while deployment follows the original working alternative. The expectation also averages over the deployment reference and rank uniforms.

For the proof, it is useful to distinguish the two input laws. Write $\mathbb{P}_{0,\mathrm{I}}$ for the ideal Uniform path law and $\mathbb{P}_{0,\mathrm{C},n}$ for the conformal-null path law. For any deterministic stopping map σ, define

$$\begin{aligned} F_{\mathrm{I}}(\sigma) &= \mathbb{P}_{0,\mathrm{I}}\{\sigma(\mathbf{P}) \le T\}, \\ F_{\mathrm{C},n}(\sigma) &= \mathbb{P}_{0,\mathrm{C},n}\{\sigma(\widetilde{\mathbf{P}}) \le T\}, \\ J_{\mathrm{C},n}(\sigma) &= \mathbb{E}_{\mathcal{A}}\Big[w_{\sigma(\widetilde{\mathbf{P}})} \mathbf{1}\{\nu^* \le \sigma(\widetilde{\mathbf{P}}) \le T\}\Big]. \end{aligned}$$

The ideal-input reward is $J_w(\sigma(\mathbf{P}))$, with J_w as defined in Section 3.2. The conformal-input reward $J_{\mathrm{C},n}(\sigma)$ also averages over the deployment reference and rank uniforms. We use the scalar functions

$$F(\lambda) = F_{\mathrm{I}}(\tau_\lambda), \qquad G_n(\lambda) = F_{\mathrm{C},n}(\tau_\lambda)$$

only within this section. Recall that $V_0(\lambda) = \sup_\sigma\{J_w(\sigma(\mathbf{P})) - \lambda F_{\mathrm{I}}(\sigma)\}$, and that $J_w^*(\alpha)$ is the constrained optimal reward for the ideal independent coordinates. The notation $\mathbb{P}_{0,\mathrm{cal}}$ below denotes null probability averaged over the independent calibration bank, deployment reference, and rank uniforms.

Proof of Theorem 3. The preceding construction gives $0 \le \hat{\lambda} \le \bar{\lambda}$. We control the two input laws, the calibration error, and the resulting reward difference in turn.

1. Evidence densities and the ideal optimum. The working-alternative-to-null prefix likelihood ratio is $r^t e^{Z_t}$. Let $g_{0,t}$ denote the density of Z_t under the ideal null. Since $Z_t \ge 0$, it satisfies

$$g_{0,t}(z) = r^{-t} e^{-z} g_{\mathcal{A},t}(z) \le \mathsf{M} r^{-t}, \qquad z \ge 0. \tag{F.3}$$

In particular, $Z_1 = \ell(P_1)$ is non-atomic under the ideal null. Theorem 1 gives

$$J_w^*(\alpha) = J_w(\tau_{\lambda^*}(\mathbf{P})) = V_0(\lambda^*) + \lambda^* \alpha.$$

2. Null-law comparison. Put $d = n + 1$ and define the rank categories $\mathsf{C}_t = \lfloor d\widetilde{P}_t \rfloor$. Under the conformal null,

$$\begin{aligned}\mathbb{P}_0(\mathsf{C}_t = j \mid \mathsf{C}_{1:t-1}) &= \frac{1 + N_{t-1,j}}{d + t - 1} \\ &= \frac{d}{d+t-1}\frac{1}{d} + \frac{t-1}{d+t-1}\frac{N_{t-1,j}}{t-1}, \qquad t \geq 2,\end{aligned}$$

where $j = 0, \ldots, d-1$ and $N_{t-1,j} = \sum_{s=1}^{t-1} \mathbf{1}\{\mathsf{C}_s = j\}$. Indeed, the transformed reference spacings have a $\mathrm{Dirichlet}(1, \ldots, 1)$ law, and conditioning on the categories adds their counts to its parameters. Each new category is therefore either a fresh uniform category or a uniformly selected previous one. Couple the fresh draws to independent uniform categories and use common independent within-category uniforms. The resulting complete paths disagree with probability at most

$$\epsilon_{n,T} := \min\left\{1, \sum_{t=2}^{T} \frac{t-1}{n+t}\right\} \leq \min\left\{1, \frac{T(T-1)}{2(n+2)}\right\}. \tag{F.4}$$

The empty sum is zero. Consequently, uniformly over deterministic stopping maps σ,

$$|F_1(\sigma) - F_{\mathrm{C},n}(\sigma)| \leq \epsilon_{n,T}. \tag{F.5}$$

3. Continuous-multiplier calibration. The conformal-null law has a density on $(0,1)^T$: conditional on its rank categories, the inputs are independent uniforms within their respective rank intervals. It is therefore absolutely continuous with respect to the ideal Uniform path law.

Let $\Gamma = \Gamma(\widetilde{\mathbf{P}})$ under the conformal null. For a fixed $\lambda > 0$, a path satisfying $\Gamma(z) = \lambda$ must have a tied Bellman advantage at some time. A reachable tie with positive cutoff implies

$$\sum_{s=1}^{t} \ell(z_s) = -\log c_t(\lambda).$$

Under the ideal input law, this is the event $\{Z_t = -\log c_t(\lambda)\}$, which has probability zero by (F.3). A zero cutoff cannot be reached by a positive posterior. Absolute continuity transfers these zero-probability path events to the conformal null. At $\lambda = 0$, never stopping requires $f_1(z_t) = 0$ for all t. Non-atomicity of $\ell(P_1)$ implies $\mathrm{Leb}\{f_1 = 0\} = 0$, so $\mathbb{P}_{0,\mathrm{C},n}(\Gamma = 0) = 0$ as well. Thus the conformal-null distribution of Γ is continuous.

By the probability integral transform for order statistics,

$$G_n(\hat{\lambda}) \sim \mathrm{Beta}(B + 1 - k, k).$$

Its mean is $(B+1-k)/(B+1)$, within $1/(B+1)$ of α, and its variance is at most $1/[4(B+2)]$. Hence

$$\mathbb{E}_{\mathrm{cal}}|\alpha - G_n(\hat{\lambda})| \leq \frac{1}{B+1} + \frac{1}{2\sqrt{B+2}}. \tag{F.6}$$

This bound does not require a positive density at the multiplier quantile or convergence of the multiplier itself.

4. The reward change from conformal observations. Use the reference approximation radius $r_n(\delta)$ from (D.1), with $0 < \delta < 1$. Couple ideal and conformal inputs through the same scores, reference, and hit indicators. The reference DKW event (Massart, 1990) has probability at least $1 - \delta$ and gives $\max_{t \leq T} |\widetilde{P}_t - P_t| \leq r_n(\delta)$. Since ℓ is L_ℓ-Lipschitz, on this event,

$$\left|\sum_{s=1}^{t} \ell(\widetilde{P}_s) - Z_t\right| \leq t L_\ell r_n(\delta).$$

Both executions use the same cutoffs. Their first different action therefore requires

$$|Z_t + \log c_t(\lambda)| \leq t L_\ell r_n(\delta)$$

at a time with positive cutoff. Outside these boundary events and the bad-reference event, the realized rewards agree. The density bound and bounded rewards yield, uniformly in λ,

$$\begin{aligned}|J_w(\tau_\lambda(\mathbf{P})) - J_{\mathrm{C},n}(\tau_\lambda)| &\leq R_n(\delta), \\ R_n(\delta) &:= w_1\left[\delta + \mathsf{M}T(T+1)L_\ell r_n(\delta)\right].\end{aligned} \tag{F.7}$$

5. Apply the bounds at the selected multiplier. Exact Bellman optimality gives $V_0(\lambda) = J_w(\tau_\lambda(\mathbf{P})) - \lambda F(\lambda)$. By weak duality, for every realized calibration bank,

$$\begin{aligned} J_w^*(\alpha) - J_{\mathrm{C},n}(\tau_{\hat{\lambda}}) &\le \hat{\lambda}\{\alpha - F(\hat{\lambda})\} + R_n(\delta) \\ &\le \bar{\lambda}\left\{|\alpha - G_n(\hat{\lambda})| + \epsilon_{n,T}\right\} + R_n(\delta). \end{aligned}$$

Taking expectation over calibration proves

$$J_w^*(\alpha) - \widehat{J}_{n,B} \le R_n(\delta) + \bar{\lambda}\left\{\epsilon_{n,T} + \frac{1}{B+1} + \frac{1}{2\sqrt{B+2}}\right\}. \tag{F.8}$$

6. Nonnegativity and the rate. The deployment null critical multiplier and the calibration critical multipliers are exchangeable. Their strict rank event gives

$$\mathbb{P}_{0,\mathrm{cal}}(\hat{\tau} \le T) = \mathbb{E}_{\mathrm{cal}} G_n(\hat{\lambda}) \le \frac{B+1-k}{B+1} \le \alpha.$$

Independent uniform reference coordinates and rank uniforms causally transform the ideal path into the conformal path under both laws. Fixing this auxiliary randomness and the independent calibration bank turns the calibrated procedure into an ideal-history stopping map. Its Lagrangian payoff is at most $V_0(\lambda)$; averaging gives

$$\widehat{J}_{n,B} - \lambda \mathbb{P}_{0,\mathrm{cal}}(\hat{\tau} \le T) \le V_0(\lambda).$$

At $\lambda = \lambda^*$, exact dual recovery therefore implies $\widehat{J}_{n,B} \le J_w^*(\alpha)$. This argument also applies to any calibrated stopping map with the same independent auxiliary construction and marginal null guarantee.

Finally, take $\delta = n^{-1}$ in (F.8). For fixed T, $\epsilon_{n,T} = O(n^{-1})$ and $R_n(n^{-1}) = O(\sqrt{\log n/n})$. Together with (F.6), this proves Theorem 3 and the stated convergence. □

F.2 Piecewise-constant density and a finite multiplier grid

We now retain the same score model and account for numerical approximation of both the density and the multiplier domain.

Let $f_{1,K}$ be the nonnegative, normalized piecewise-constant density on the partition $\{I_\ell\}_{\ell=1}^K$ from Appendix C.6. Recall the corresponding mixture density and log evidence, and write

$$m_K = r + b f_{1,K}, \qquad \varepsilon_K = \|f_{1,K} - f_1\|_\infty, \qquad \ell_K(p) = \log\{m_K(p)/r\}.$$

Then $\|m_K - m\|_\infty = b\varepsilon_K \le \varepsilon_K$. As above, $\tau_{K,\lambda}(z)$ denotes evaluation of the Bellman rule from Appendix C.6 on the path z:

$$q_{K,t}(z) = \frac{r^t}{\prod_{s=1}^t m_K(z_s)}, \qquad \tau_{K,\lambda}(z) = \inf\{1 \le t \le T : q_{K,t}(z) < c_{K,t}(\lambda)\}.$$

Here $c_{K,t}(\lambda)$ is the cutoff computed by exact Bellman recursion under $f_{1,K}$. Continuation at equality makes the map nondecreasing in λ on every input path.

Use the fixed constants $\bar{m}$ and $\bar{\lambda}$ from the continuous-density analysis. Uniform approximation ensures $\|m_K\|_\infty \le \bar{m}$ for all sufficiently large K. Choose a finite grid

$$\begin{aligned} \Lambda &= \{0 = \lambda_0 < \lambda_1 < \cdots < \lambda_M = \bar{\lambda}\}, \\ h_\Lambda &= \max_{1 \le j \le M} (\lambda_j - \lambda_{j-1}). \end{aligned} \tag{F.9}$$

The family and grid are fixed independently of calibration and deployment. The endpoint $\bar{\lambda}$ never stops by the same payoff bound used for the original family.

Generate B independent complete conformal-null paths $\widetilde{\mathbf{P}}^{(1)}, \ldots, \widetilde{\mathbf{P}}^{(B)}$, independently of deployment. Each path uses its own independent reference of size n, shared within that path. In this

subsection, $\hat{\lambda}$ and $\hat{\tau}$ denote the grid-calibrated implementation. With the same B and k as above, define, for $j = 1, \dots, B$,

$$\begin{aligned} \Gamma_j &= \min\{\lambda \in \Lambda : \tau_{K,\lambda}(\widetilde{\mathbf{P}}^{(j)}) > T\}, \\ \hat{\lambda} &= \Gamma_{(k)}, \qquad \hat{\tau} = \tau_{K,\hat{\lambda}}(\widetilde{\mathbf{P}}). \end{aligned} \tag{F.10}$$

The choice $B \geq \lceil 1/\alpha \rceil$ ensures $k \leq B$. Let

$$\widehat{J}_{n,K,B,\Lambda} = \mathbb{E}_{\mathrm{cal},\mathcal{A}}[w_{\hat{\tau}} \mathbf{1}\{\nu^* \leq \hat{\tau} \leq T\}]. \tag{F.11}$$

The calibration bank always follows its null simulation law, while deployment follows the original working alternative. The expectation also averages over the deployment reference and rank uniforms. The benchmark remains $J_w^*(\alpha) = J_w(\tau_{\lambda^*}(\mathbf{P}))$, the constrained optimum under the original working alternative.

Theorem F.1 (Reward of the calibrated piecewise-constant policy)**.** *Suppose f_1 is Lipschitz on $[0,1]$, $\varepsilon_K \to 0$, and the original evidence Z_t has a density $g_{\mathcal{A},t}$ satisfying $\sup_{t \leq T} \|g_{\mathcal{A},t}\|_\infty \leq \mathsf{M} < \infty$. For fixed T, α, $b \in (0,1)$, reward schedule, and the common bound $\bar{m}$, the construction above admits a constant $C < \infty$, independent of n, K, B, and Λ, such that, for all $n \geq 2$ and all sufficiently large K, B,*

$$0 \leq J_w^*(\alpha) - \widehat{J}_{n,K,B,\Lambda} \leq C \left\{ \sqrt{\frac{\log n}{n}} + \varepsilon_K + B^{-1/2} + h_\Lambda \right\}. \tag{F.12}$$

Consequently, $\widehat{J}_{n,K,B,\Lambda} \to J_w^*(\alpha)$ whenever $n \to \infty$, $\varepsilon_K \to 0$, $B \to \infty$, and $h_\Lambda \to 0$.

The result concerns the actual order-statistic multiplier $\hat{\lambda}$. Independence of the simulated trajectories is used for the $B^{-1/2}$ term; exchangeability alone suffices for rank-based error control but does not imply this concentration rate. Both the nonnegative reward gap and exact error control average over calibration.

Proof of Theorem F.1. Use the ideal and conformal probabilities and rewards defined above. The ideal-null function F_K agrees with (C.17); for comparison, write

$$F_K(\lambda) = F_{\mathrm{I}}(\tau_{K,\lambda}), \qquad G_{n,K}(\lambda) = F_{\mathrm{C},n}(\tau_{K,\lambda}).$$

Then $\widehat{J}_{n,K,B,\Lambda} = \mathbb{E}_{\mathrm{cal}} J_{\mathrm{C},n}(\tau_{K,\hat{\lambda}})$. The null-law comparison (F.5) applies uniformly to this family, so

$$\sup_{\lambda \geq 0} |F_K(\lambda) - G_{n,K}(\lambda)| \leq \epsilon_{n,T}.$$

1. Error from approximating the design density. Recall that $J_{w,K}^{\mathrm{app}}(\sigma(\mathbf{P}))$ evaluates the reward under the approximate working alternative with hit density $f_{1,K}$, as defined in Appendix C.6. The one-step joint laws of (A_t, P_t) under the original and approximate alternatives have total variation distance

$$\frac{b}{2} \|f_1 - f_{1,K}\|_1 \leq \frac{\varepsilon_K}{2}.$$

Tensorization and bounded rewards give, uniformly over all stopping maps,

$$|J_w(\sigma(\mathbf{P})) - J_{w,K}^{\mathrm{app}}(\sigma(\mathbf{P}))| \leq \eta_K := w_1 T \varepsilon_K / 2.$$

The two design models have the same ideal null law. Since $\tau_{K,\lambda}(\mathbf{P})$ maximizes the corresponding approximate Lagrangian, comparison of its value and the supremum yields

$$0 \leq V_0(\lambda) - \{J_w(\tau_{K,\lambda}(\mathbf{P})) - \lambda F_K(\lambda)\} \leq 2\eta_K = w_1 T \varepsilon_K \quad \text{for every } \lambda \geq 0. \tag{F.13}$$

2. The probability lost between adjacent grid points. Set $w_{T+1} = 0$ and $d_t = w_t - r w_{t+1} > 0$. Recall the Bellman stopping advantage $D_{K,t}(q;\lambda) = R_t(q;\lambda) - C_{K,t}(q;\lambda)$. It satisfies

$$D_{K,t}(q_2;\lambda) - D_{K,t}(q_1;\lambda) \leq -d_t(q_2 - q_1), \qquad q_2 > q_1,$$

and, for $\lambda' \geq \lambda$,

$$-(\lambda' - \lambda) q r^{-t} \leq D_{K,t}(q;\lambda') - D_{K,t}(q;\lambda) \leq 0.$$

These are the Bellman state and multiplier inequalities, which apply equally to the normalized piecewise-constant model. For two positive roots $c_1 = c_{K,t}(\lambda) \geq c_2 = c_{K,t}(\lambda')$, they imply

$$d_t(c_1 - c_2) \leq D_{K,t}(c_2; \lambda) \leq (\lambda' - \lambda)c_2 r^{-t}.$$

Therefore, writing $B_{K,t}(\lambda) = -\log c_{K,t}(\lambda)$,

$$0 \leq B_{K,t}(\lambda') - B_{K,t}(\lambda) \leq \frac{\lambda' - \lambda}{d_t r^t}. \tag{F.14}$$

For $t < T$ with $w_t = w_{t+1}$ the root is identically zero and no stopping is possible at that time; otherwise it is positive at every finite multiplier. The terminal root is positive as well. Thus the zero-root cases can be omitted from subsequent boundary events.

Since $m, m_K \geq r$,

$$\|\ell_K - \ell\|_\infty \leq \varepsilon_K / r, \qquad \left| \sum_{s=1}^{t} \ell_K(P_s) - Z_t \right| \leq t\varepsilon_K / r. \tag{F.15}$$

The original ideal-null evidence density obeys (F.3). If a path stops by T at λ_{j-1} but not at λ_j, its approximate evidence at some time lies between the two corresponding log cutoffs. By (F.14) and (F.15), the original evidence then lies in an interval of length at most $h_\Lambda/(d_t r^t) + 2t\varepsilon_K/r$. Integrating (F.3), summing over times, and transferring this single path event to the conformal null gives

$$\begin{aligned} \Delta_{n,K,\Lambda} &:= \max_{1 \leq j \leq M} \{G_{n,K}(\lambda_{j-1}) - G_{n,K}(\lambda_j)\} \\ &\leq C_{\mathrm{grid}} h_\Lambda + \zeta_K + \epsilon_{n,T}, \\ C_{\mathrm{grid}} &:= \mathsf{M} \sum_{t=1}^{T} \frac{r^{-2t}}{d_t}, \qquad \zeta_K := \frac{2\mathsf{M}\varepsilon_K}{r} \sum_{t=1}^{T} t r^{-t}. \end{aligned} \tag{F.16}$$

The constant C_{grid} does not depend on the grid.

3. The actual empirical quantile. Define

$$\widehat{G}_B(\lambda) = B^{-1} \sum_{j=1}^{B} \mathbf{1}\{\Gamma_j > \lambda\}, \qquad \alpha_B = (B-k)/B, \qquad s_B = \alpha - \alpha_B \in [0, 2/B).$$

Pathwise monotonicity and continuation at the selected critical grid point give

$$\hat{\lambda} = \min\{\lambda \in \Lambda : \widehat{G}_B(\lambda) \leq \alpha_B\}.$$

Let $D_B = \sup_{\lambda \in \Lambda} |\widehat{G}_B(\lambda) - G_{n,K}(\lambda)|$. The DKW inequality (Massart, 1990), applied to the scalar critical multipliers, whose atoms are allowed, gives

$$\mathbb{P}_{\mathrm{cal}}(D_B > u) \leq 2e^{-2Bu^2}, \qquad \mathbb{E}_{\mathrm{cal}} D_B \leq \sqrt{\frac{\pi}{2B}}. \tag{F.17}$$

In particular, there is no factor involving $\log|\Lambda|$. If $\hat{\lambda} = \lambda_j > 0$, the previous grid point has $\widehat{G}_B(\lambda_{j-1}) > \alpha_B$, so

$$G_{n,K}(\hat{\lambda}) \geq \alpha_B - D_B - \Delta_{n,K,\Lambda}.$$

Together with (F.5), this implies

$$\hat{\lambda}\{\alpha - F_K(\hat{\lambda})\} \leq \bar{\lambda}\{s_B + D_B + \Delta_{n,K,\Lambda} + \epsilon_{n,T}\}. \tag{F.18}$$

The same inequality holds at $\hat{\lambda} = 0$ because its left-hand side is zero. Thus no positivity or budget-saturation property of the selected multiplier is required.

4. The reward change from conformal observations. Use the constants L_ℓ and $r_n(\delta)$ introduced above. Couple ideal and conformal inputs through the same scores, reference, and hit indicators. The reference DKW event has probability at least $1-\delta$ and implies $\max_{t\le T}|\widetilde{P}_t - P_t| \le r_n(\delta)$. On this event,

$$\left|\sum_{s=1}^{t} \ell_K(\widetilde{P}_s) - Z_t\right| \le t\{L_\ell r_n(\delta) + \varepsilon_K/r\},$$

while the ideal approximate evidence satisfies (F.15). Both executions of the same map use the same cutoffs. Their first different action therefore requires

$$|Z_t - B_{K,t}(\lambda)| \le t\{L_\ell r_n(\delta) + \varepsilon_K/r\}$$

at a time with positive cutoff. Outside these boundary events and the bad-reference event the realized rewards agree. Bounded rewards and the original working-alternative evidence densities give, uniformly in λ,

$$\begin{aligned} |J_w(\tau_{K,\lambda}(\mathbf{P})) - J_{\mathrm{C},n}(\tau_{K,\lambda})| &\le R_{n,K}(\delta), \\ R_{n,K}(\delta) &:= w_1\left[\delta + \mathsf{M}T(T+1)\{L_\ell r_n(\delta) + \varepsilon_K/r\}\right]. \end{aligned} \tag{F.19}$$

This argument uses smoothness of the original evidence, not of the piecewise-constant approximation.

5. Apply the bounds at the selected multiplier. Weak duality gives $J_w^*(\alpha) \le V_0(\lambda) + \lambda\alpha$ for every $\lambda \ge 0$. Thus, for every realized calibration bank,

$$\begin{aligned} J_w^*(\alpha) - J_{\mathrm{C},n}(\tau_{K,\hat\lambda}) \le V_0(\hat\lambda) &- \{J_w(\tau_{K,\hat\lambda}(\mathbf{P})) - \hat\lambda F_K(\hat\lambda)\} \\ &+ \hat\lambda\{\alpha - F_K(\hat\lambda)\} + J_w(\tau_{K,\hat\lambda}(\mathbf{P})) - J_{\mathrm{C},n}(\tau_{K,\hat\lambda}). \end{aligned}$$

The design and observation inequalities hold uniformly in the multiplier; (F.18) controls the signed budget term. Taking expectation over calibration yields the explicit estimate

$$\begin{aligned} J_w^*(\alpha) - \widehat{J}_{n,K,B,\Lambda} \le w_1 T\varepsilon_K + R_{n,K}(\delta) \\ + \bar\lambda\left\{s_B + \sqrt{\frac{\pi}{2B}} + C_{\mathrm{grid}}h_\Lambda + \zeta_K + 2\epsilon_{n,T}\right\}. \end{aligned} \tag{F.20}$$

6. Nonnegativity and the rate. The deployment null critical multiplier and the B simulated critical multipliers are exchangeable. The strict rank event therefore gives

$$\mathbb{P}_{0,\mathrm{cal}}(\hat\tau \le T) = \mathbb{E}_{\mathrm{cal}}G_{n,K}(\hat\lambda) \le \frac{B+1-k}{B+1} \le \alpha,$$

including ties. The auxiliary-randomization argument in the proof of Theorem 3 applies without change: conditional on the calibration bank and auxiliary randomness, the implemented procedure is an ideal-history stopping map. Thus

$$\widehat{J}_{n,K,B,\Lambda} - \lambda\mathbb{P}_{0,\mathrm{cal}}(\hat\tau \le T) \le V_0(\lambda).$$

At the original dual optimizer λ^*, marginal null validity gives $\widehat{J}_{n,K,B,\Lambda} \le J_w^*(\alpha)$. No population multiplier is selected by the implemented procedure.

Finally, set $\delta = n^{-1}$ in (F.20). For the fixed model and horizon, $\epsilon_{n,T} = O(n^{-1})$, $\zeta_K = O(\varepsilon_K)$, $s_B = O(B^{-1})$, and $R_{n,K}(n^{-1}) = O(\sqrt{\log n/n} + \varepsilon_K)$. All remaining constants are independent of n, K, B, and the grid. This proves (F.12) and its convergence statement. □

G Additional Experimental Details

G.1 Protocol, implementation, and evaluation

We use A_t for the hit indicator of candidate t, as in the theoretical sections.

Data roles and conformalization. Each complete repeat starts with a label-blind permutation and a $20\%/20\%/60\%$ split into working-alternative training, reference, and evaluation data. The working alternative is fitted only on the first split. From the second split, we sample $n = 200$ verified non-hits and keep this panel fixed across every evaluation trajectory in that repeat. Denote their scores by $S_1^0, \ldots, S_n^0$. For the two molecular data sets, larger screening scores are more hit-like, so a score S_t is mapped to the randomized upper-tail conformal value

$$\widetilde{P}_t = \frac{\sum_{i=1}^n \mathbf{1}\{S_i^0 > S_t\} + U_t\big(1 + \sum_{i=1}^n \mathbf{1}\{S_i^0 = S_t\}\big)}{n+1}, \qquad U_t \overset{\text{iid}}{\sim} \text{Unif}(0,1),$$

where the rank randomizers are independent of all scores and labels. In the absence of ties, this upper-tail orientation agrees with the main-text construction applied to negated molecular scores; the displayed formula also randomizes ties. The synthetic scores use the lower-tail analogue because smaller values are more hit-like. We preserve the dependence induced by the shared reference panel both when calibrating the stopping rules and when evaluating them.

Working alternative. Let $\widehat{b}$ denote the fitted hit probability and $\widehat{f}_1$ the fitted hit density on the conformal scale. The training split is converted to the $n+1 = 201$ natural randomized-rank cells. We analytically average over the size-n non-hit panel and the rank randomizer, rather than conditioning the density estimate on the operational reference selected later. A label-blind two-fold procedure chooses the uniform-prior strength from $\{0.1, 0.3, 1, 3, 10, 30\}$ by held-out hit log density, and the hit prevalence is Jeffreys-smoothed. The COS method and OneHit-TS-AVT share an adaptive contiguous $K = 5$ projection selected by fitted decision-mixture Kullback–Leibler loss. LR-e and IMT-Alt-linear retain the uncompressed 201-cell estimate. Neither O-CONFHIT nor Stouffer-linear uses the fitted alternative density.

Final calibration. All model fitting, action selection, and design-stage tuning are frozen before the final calibration bank is generated. We use $B = 9{,}999$ complete global-null trajectories, each with an independently generated reference panel shared over its full horizon. For a process whose evidence increases with signal, the crossing threshold is the strict upper-rank threshold with $k = \lceil (B+1)(1-\alpha) \rceil$; O-CONFHIT uses the corresponding strict lower-tail rank. The COS multiplier is bracketed and then found by bisection, with the feasible upper endpoint deployed. Thus the calibration acts on complete paths, not on marginal conformal values. No evaluation trajectory is used to select a density, action, tuning constant, or threshold.

Compared stopping rules. The six implementations are as follows.

- **COS.** Conditional on the fitted $K = 5$ partition, we solve the finite-horizon Bellman recursion exactly on the categorical count states. The stopping payoff incorporates the requested w_t and the Lagrange penalty for an alarm under the global null. We stop only when the stopping value is strictly larger than the continuation value; a tie continues. A separate calibrated multiplier is used for each (T, w, α) setting.
- **O-CONFHIT.** This is an online adaptation of ConfHit (Laghuvarapu et al., 2026), rather than the paper's offline design procedure. At time t, it pools the non-hit reference panel with the observed raw-score prefix and computes the exact randomized permutation p-value $p_{\text{CH},t}$ of the prefix maximum. With prespecified temporal weights $a_t > 0$ proportional to w_t, it stops when $p_{\text{CH},t}/a_t$ crosses a calibrated lower threshold. Object-level density-ratio weights and backward monotonization are not used.
- **LR-e.** Using the fitted hit probability $\widehat{b}$ and the 201-cell hit density $\widehat{f}_1$, LR-e accumulates the log of the working-mixture likelihood-ratio increments $(1-\widehat{b}) + \widehat{b}\widehat{f}_1(\widetilde{P}_t)$ and applies a complete-path rank calibration to the maximum cumulative process.
- **Stouffer-linear.** This rule accumulates $\Phi^{-1}(1-\widetilde{P}_t)$, where Φ is the standard normal distribution function, and divides the cumulative sum by a Gaussian linear uniform boundary with tuning parameter $T/4$. We then recalibrate the maximum normalized process on the same type of complete null trajectories used for the other rules.
- **IMT-Alt-linear.** Following the masked-information idea of interactive martingale testing (Duan et al., 2020), the rule first observes the masked value $g(p) = \min(p, 1-p)$ for

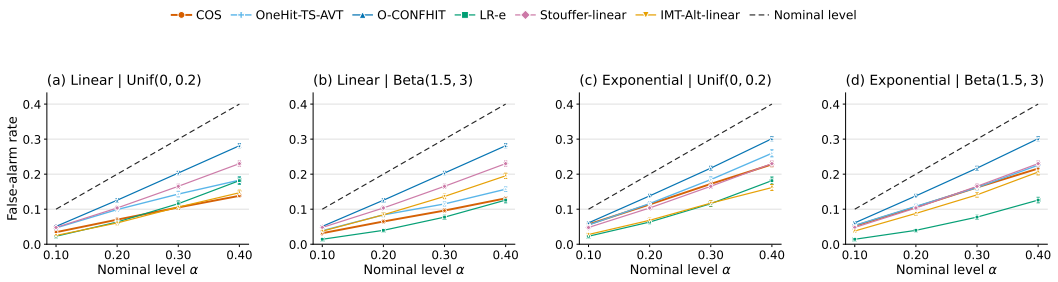


Figure 5: Synthetic false-alarm rate $\Pr(\tau < \nu^*)$ at $T = 30$. Columns correspond to the two hit-score laws and the panels cover both reward schedules. Points are means over 100 complete repeats; bars are pointwise 95% complete-repeat percentile-bootstrap intervals. The dashed line is the nominal level $y = \alpha$.

$p = \widetilde{P}_t$. The 201-cell working alternative supplies $\widehat{\mathbb{P}}_{\mathcal{A}}(A_t = 1 \mid g(\widetilde{P}_t))$, the fitted posterior hit probability given the masked value; only observations above a screening threshold enter a signed walk, with an increment of $+1$ for $p < 1/2$ and -1 otherwise. The walk is compared with a linear boundary indexed by the number of included observations. The screening threshold is selected from $\{0, 0.05, \ldots, 0.95\}$ using 4,999 design-null and 4,999 fitted working-mixture paths, then frozen before the independent final-null calibration.

- **OneHit-TS-AVT.** We adapt the testing-by-betting framework of time-sensitive anytime-valid testing (Clerico et al., 2026) to the one-hit reward. Its Bellman state records time, the state function computed under the fitted independent-input model, and scaled betting wealth. The five actions are coarsened power bets whose averages within each of the five null bins have null mean one. The formal solver uses a 201-point posterior grid, a 1,601-point log-wealth grid, and a minimum scaled wealth of 10^{-20}. The reward-dependent policy is shared across α and is subsequently recalibrated through the maximum absorbed log wealth on complete null paths.

Evaluation. Use the first-hit time ν^* and the convention $\tau = \infty$ for no alarm by T, as in the main text. An alarm is correct when $\nu^* \le \tau \le T$ and false when $\tau < \nu^*$. A correct path earns w_τ and every other path earns zero. If a numerical implementation stores no alarm as $T + 1$, that sentinel is interpreted as ∞ when computing these events. The reported one-hit reward is the average over all mixed evaluation paths. The reported false-alarm rate is likewise the mixed-path frequency of $\tau < \nu^*$, rather than the stopping rate of the global-null calibration bank. Within a repeat, the held-out candidates are shuffled without replacement and divided into disjoint length-T trajectories.

G.2 Synthetic study

The $20\%/20\%/60\%$ split gives 2,000 training observations, 2,000 reference observations, and 6,000 test observations in each repeat. Consequently, the test split forms 200 disjoint paths when $T = 30$ and 120 when $T = 50$; every test observation is used. Common random numbers couple the two hit-score laws and the two horizons within a repeat: they share hit indicators, non-hit scores, the split, the sampled reference, the flat test ordering, and the conformal randomizers. Only the inverse-CDF map for hit scores and the path reshaping change. This coupling reduces Monte Carlo noise in paired comparisons but does not alter any method's marginal experiment.

The supplemental run extends the $T = 30$ grid to $\alpha = 0.40$ and supplies $T \in \{20, 40\}$ at $\alpha = 0.20$. Figure 5 shows the ordinary mixed-path false-alarm rate at $T = 30$. Across both hit-score laws and both reward schedules, every plotted mean and every pointwise 95% upper endpoint is below its nominal level.

As a numerical diagnostic, the adaptive $K = 5$ projection retains 0.9485 of the decision-mixture information in the continuous Beta setting, slightly below the prespecified 0.95 threshold. We therefore do not describe that density-compression diagnostic as passed. LR-e and IMT-Alt-linear continue to use the full 201-cell estimate in this setting.

G.3 QED–HGRAPH DATA DETAILS

The QED experiment uses the candidate file and pretrained binary classifier released with CONFHIT. The raw file contains 31,918 records from 798 initial molecules, with 39–40 optimized candidates per initial molecule. We found no missing property values or invalid optimized SMILES. Canonical isomeric-SMILES deduplication retains the first occurrence of each generated molecule, leaving 25,457 unique candidates and a mapping for the 6,461 removed duplicate records. QED values recomputed with RDKit agree with `PROPERTY_opt` to a maximum absolute difference of 2.22×10^{-16}.

The screening score is the cached positive-class probability from the released classifier, evaluated on the pair $(\texttt{SMILES_ori}, \texttt{SMILES_opt})$; exact QED is used only to form the evaluation label. Candidates from all initial molecules are pooled, and `SIMILARITY_ori_opt` is retained but not used as a filter.

G.4 TARGETDIFF DOCKING DATA DETAILS

This study is a secondary sequential analysis of the TargetDiff candidate table released with CONFHIT; it does not regenerate molecules, rerun docking, or reproduce the offline CONFHIT design algorithm. The table contains 7,781 candidates from 83 protein–pocket contexts, together with precomputed predictor features and a screening score. Its `PROPERTY_opt` field is sign-negated AutoDock Vina energy, so the frozen rule `PROPERTY_opt`> 10 is exactly equivalent to $\mathrm{Vina} < -10\,\mathrm{kcal/mol}$ (Trott & Olson, 2010). This produces 1,206 hits and 6,575 non-hits. The released `label` column uses a different cutoff, so every experimental repeat recomputes the hit indicator A_t for candidate t from `PROPERTY_opt`.

All 83 contexts are pooled before the label-blind split. Each repeat therefore contains 1,556 training, 1,556 reference, and 4,669 evaluation candidates, from which $\lfloor 4669/T \rfloor$ complete paths are formed. The estimand is performance for this pooled released-candidate distribution. Larger affinity-predictor scores are treated as more hit-like, while Vina is withheld for evaluation labels.