Time-sensitive anytime-valid testing

Eugenio Clerico*
University of Oxford
eugenio.clerico@stats.ox.ac.uk

Tobias Wegel*
ETH Zurich
tobias.wegel@inf.ethz.ch

Iskander Azangulov
University of Oxford
iskander.azangulov@spc.ox.ac.uk

Patrick Rebeschini
University of Oxford
patrick.rebeschini@stats.ox.ac.uk

Abstract

Anytime-valid tests allow evidence to be checked during data collection: one can either continue testing or stop and reject the null while still controlling type-I error. Yet, in many applications rejection is useful only if it comes soon enough. We introduce a time-sensitive testing-by-betting framework that favours early rejection by assigning rewards to rejection times and maximising their expected value under a given alternative. This encompasses hard deadlines and softer time preferences. The resulting optimal control problem admits a Bellman representation in terms only of time and evidence against the null, rather than the full history. For hard deadlines, we show that the simple-vs-simple case reduces to a finite-horizon Neyman–Pearson problem and identify the corresponding canonical optimal e-process. Moreover, we show that exponentially decaying rewards admit a stationary approximation, leading to the exponential-decay-optimal (EDO) criterion: a finite-time-scale counterpart to the classical growth-rate-optimal (GRO) viewpoint.

1 Introduction

Classical hypothesis testing in statistics is built around fixed-sample designs: one specifies a sample size in advance, then collects the data and computes a p-value. However, this framework is often too rigid for modern settings, where data arrive sequentially and may be monitored continuously. Practitioners increasingly need the freedom to adaptively stop or continue testing based on real-time evidence. As a consequence, recent literature has focused on anytime-valid inference, developing tools, such as e-values and e-processes, that provide statistical guarantees which remain valid under optional stopping and continuation [Grünwald et al., 2024, Ramdas and Wang, 2025]. A concrete and intuitive instantiation of anytime-valid inference is the *testing-by-betting* game [Ramdas et al., 2023], where a gambler starts with unit wealth and, at each round, places a bet against the null hypothesis. The game is structured to be fair (or disadvantageous to the player) when the null is true, meaning that the gambler is highly unlikely to become rich. Consequently, accumulating significant wealth constitutes evidence against the null. The appeal of this perspective is that it turns sequential testing into an algorithmic problem: based on currently available observations, one needs to decide where and how aggressively to bet. In this sense, the statistical question naturally admits a sequential-control interpretation, creating a bridge towards dynamic programming and reinforcement learning.

To guide betting decisions in the testing-by-betting game, the statistical literature typically fixes an alternative hypothesis and chooses bets that maximise asymptotic wealth when that alternative holds. This is formalised by the growth-rate-optimal (GRO) criterion [Grünwald et al., 2024], which in simple cases can recover classical ideas such as the Kelly criterion [Kelly Jr., 1956] and Wald's

*Contributed equally

Code available at tobiaswegel/time-sensitive-anytime-valid.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

sequential likelihood-ratio test (SPRT) [Wald, 1945]. For composite nulls, the GRO criterion is closely related to the numeraire e-variable and reverse information projections [Larsson et al., 2025b]. This approach is natural and theoretically justified, especially in idealised settings where no time or cost constraints affect the experiment. Yet, from a control perspective, long-run growth is only one possible objective, and in many practical applications arguably not the most relevant one. Often, a guarantee of eventual rejection is insufficient, with practical considerations requiring the test to reject as early as possible. It is thus natural to think of the *value* of a rejection as depreciating over time. Recent work has tackled this or related issues by focusing on finite-horizon settings, showing that when a fixed deadline is imposed, asymptotically optimal strategies can be strictly suboptimal [Voráček and Orabona, 2025, Taga et al., 2026].

The starting point of our work is that the finite-horizon formulation addresses only one case within a fundamentally broader control problem. Most sequential experiments do not naturally come with a single hard deadline, yet there are natural preferences over rejection times: earlier rejection is better, but later rejection may still retain some value. This observation suggests shifting the focus from a fixed horizon to a reward assigned directly to the rejection time. In this framework, the hard-deadline scenario is recovered in the special case where the reward is an indicator that vanishes after a given step. Alternative reward choices allow softer and more expressive formulations of time preference. Among valid betting strategies, we therefore propose to consider those that maximise the expected reward associated with the rejection time under the given alternative. Fig. 1 shows how, in a simple Gaussian testing problem, changing the reward yields different optimal betting strategies, hence inducing different distributions of rejection times. We dub our approach *time-sensitive anytime-valid testing*, and view it as the natural control formulation of sequential testing when rejection times matter. In summary, the main conceptual contribution of this paper is to formulate a reward-based sequential decision problem that naturally encodes preferences over the duration of the experiment, all while preserving anytime validity. For the sake of clarity, we develop our formulation for i.i.d. testing-by-betting, although we expect it to extend to more general statistical frameworks.

We remark that group sequential methods also address the timing of rejection through alpha-spending functions [Lan and DeMets, 1983]. While these methods control type I error by allocating it across interim analyses, our framework directly optimises the betting strategy for a reward assigned to the rejection time, allowing different time preferences to be incorporated into the objective while retaining anytime validity.

The reward-based viewpoint that we introduce brings the control perspective to the forefront. Although testing by betting is already sequential in nature, assigning rewards to rejection times makes the choice of a betting strategy an explicit control problem over time and wealth. This viewpoint is in line with recent reinforcement-learning approaches to horizon-aware anytime-valid testing [Taga et al., 2026], but differs in that we optimise against a prescribed alternative and allow general rewards over rejection times rather than only a fixed deadline. The same target-reaching structure also relates our work to a longer tradition of optimal gambling, including red-and-black and Dubins–Savage gambling, where stakes are chosen as a function of current fortune in order to reach a target wealth, with finite-time and discounted variants studied in later work [Dubins and Savage, 1965, Yao, 2007]. Our setting has a similar target-reaching flavour, yet also fundamental differences, as both the admis-

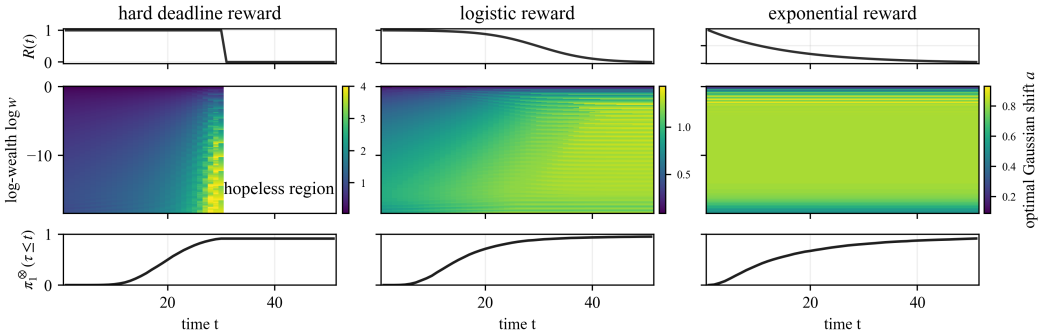


Figure 1: Heatmap of the optimal action a, representing the bet $E_{[a]}(x) = \exp\{ax - a^2/2\}$, as a function of time t and log-wealth $\log w$ when testing $\pi_0 = \mathcal{N}(0, 1)$ against $\pi_1 = \mathcal{N}(0.6, 1)$ with $\alpha = .05$, under different rewards. The corresponding stopping-time cdfs are shown below. Further details in Appendix D.1.

sible bets and the utility are dictated by the statistical testing problem, with validity under the null constraining the betting rules and the prescribed alternative determining the reward objective.

Our main contributions are threefold. First, we show that, among all predictable betting strategies that can depend on the full observation history, it is enough to consider policies depending only on the current wealth and round, which yields a simpler Bellman formulation. Second, in the simple-versus-simple setting, we show that restricting attention to testing-by-betting incurs no loss compared with optimising over more general anytime-valid procedures (testing via e-processes). Third, we analyse hard-deadline and exponentially decaying rewards. The hard-deadline reward recovers a finite-horizon Neyman–Pearson structure in the simple-versus-simple setting. For exponentially decaying rewards, we derive the exponential-decay optimal (EDO) criterion, which selects a stationary constant bet linked to power utilities and Rényi divergences, and recovers the GRO log-utility criterion in the large-time-scale limit. We illustrate our framework with Gaussian and Bernoulli examples.

General notation and conventions used throughout the paper are collected in Appendix A.

2 Testing-by-betting and time-sensitive formulation

We let $\mathcal{X}$ be a standard Borel space and write $\mathcal{P}_{\mathcal{X}}$ for the set of Borel probability measures on $\mathcal{X}$. We fix a null hypothesis $\mathcal{P}_0 \subseteq \mathcal{P}_{\mathcal{X}}$ and a *simple* alternative $\pi_1 \in \mathcal{P}_{\mathcal{X}} \setminus \mathcal{P}_0$. We consider the problem of testing $\mathcal{P}_0$ against π_1 in the i.i.d. setting. More precisely, we sequentially observe a stream of observations $x_1, x_2, \dots$ and wish to test whether these are i.i.d. draws from some unknown $\pi \in \mathcal{P}_0$, against the alternative that they are i.i.d. draws from π_1. It is convenient to formulate the problem directly on the path space $\mathcal{X}^\infty$ (endowed with the usual product σ-algebra). For $\pi \in \mathcal{P}_{\mathcal{X}}$ we write $\pi^\otimes$ for the corresponding infinite product measures, and set $\mathcal{P}_0^\otimes = \{\pi^\otimes : \pi \in \mathcal{P}_0\}$. Our goal is therefore to test the sequential null $\mathcal{P}_0^\otimes$ against the simple alternative $\pi_1^\otimes$. Throughout the paper, we fix a significance level $\alpha \in (0, 1)$: we are interested in procedures whose type-I error is at most α.

We call an *e-variable* for $\mathcal{P}_0$ any Borel function $E : \mathcal{X} \to \mathbb{R}_+$ such that $\mathbb{E}_\pi[E] \leq 1$ for every $\pi \in \mathcal{P}_0$. E-variables serve as building blocks of modern anytime-valid testing [Ramdas et al., 2023, Ramdas and Wang, 2025] and as the core ingredient of the following *testing-by-betting* framework, where they represent bets against the null. Fix a collection $\mathcal{E}$ of e-variables. A player starts with wealth $W_0 = 1$ and, at each round $t \geq 1$, chooses predictably (based on past observations) a bet $E_t \in \mathcal{E}$. Then x_t is revealed and the wealth updated as $W_t = W_{t-1} E_t(x_t)$. Under the null, $(W_t)_{t \geq 0}$ is a non-negative supermartingale. So, regardless of the betting strategy, the player is unlikely to accumulate substantial wealth, and rejecting the null once $W_t \geq \alpha^{-1}$ yields a valid level-α test. This is formalised in the next proposition (Proposition 1 in Clerico, 2025), a direct consequence of Ville's inequality.

Proposition 1. *For any predictable betting strategy in the testing-by-betting game, the procedure that rejects when the wealth first reaches α^{-1} is a valid sequential test at level α. That is, when i.i.d. draws from some $\pi \in \mathcal{P}_0$ are observed, the procedure rejects with probability at most α.*

A central question is how to choose the bets E_t. When an alternative π_1 is specified (i.e., under π_1 one wants the test to reject efficiently) the standard recommendation is the growth-rate-optimal (GRO) criterion [Grünwald et al., 2024]: choose an e-variable maximising the *e-power* $\mathbb{E}_{\pi_1}[\log E]$. This objective is the testing-by-betting analogue of the Kelly criterion [Kelly Jr., 1956], and provides a standard notion of asymptotic power under optional continuation [Ramdas and Wang, 2025]. For a *simple* null (i.e., $\mathcal{P}_0 = \{\pi_0\}$ is a single measure), the GRO bet coincides with the likelihood ratio $\mathrm{d}\pi_1/\mathrm{d}\pi_0$, recovering Wald's sequential probability ratio test (SPRT) [Wald, 1945]. Its statistical justification is fundamentally asymptotic: under a constant strategy $E_t = E$, the law of large numbers gives $\log W_t \approx t\, \mathbb{E}_{\pi_1}[\log E]$ for large t under $\pi_1^\otimes$. Hence, if $\mathbb{E}_{\pi_1}[\log E] > 0$, wealth grows exponentially and the null is rejected in finite time almost surely. Moreover, Agrawal and Ramdas [2025] showed that, as $\alpha \to 0$, the rejection time has first-order asymptotic scale $\log(1/\alpha)/\mathbb{E}_{\pi_1}[\log E]$. Thus, maximising e-power minimises the first-order asymptotic scale of the rejection time.

Since the GRO criterion controls the rejection time only through its first-order asymptotic scale, it need not reflect preferences over when rejection occurs. In many applications, a test that rejects eventually may still be unsatisfactory if it does so too late. This suggests formulating the problem directly in terms of the *rejection time* $\tau = \inf\{t \geq 0 : W_t \geq \alpha^{-1}\}$, with $\tau = \infty$ if the threshold is never reached. To express preferences over rejection times, we fix a non-increasing *reward function* $R : \mathbb{N} \cup \{\infty\} \to \mathbb{R}_+$, with $R(\infty) = 0$. We evaluate the performance of a betting strategy in the

testing-by-betting game by $\mathbb{E}_{\pi_1^\otimes}[R(\tau)]$: we aim at a betting strategy that maximises this expected reward, thus replacing the asymptotic GRO criterion with an objective directly accounting for rejection time. A few natural examples of reward functions are the following. $R(t) = \mathbf{1}_{\{t \le T\}}$ corresponds to maximising the probability of rejection within the first T rounds, as in the hard-deadline setting of Taga et al. [2026]. A softer version is the logistic $R(t) = 1/(1 + e^{(t-T)/\beta})$, with $\beta > 0$, which smooths the transition around T. Another option is $R(t) = e^{-t/T}$, assigning an exponentially decaying value to rejection, with time scale $T > 0$. These examples are illustrated in Fig. 1.

3 Sequential control formalisation

We now formalise the testing-by-betting procedure as a sequential control problem. We consider an *action space* $\mathcal{A}$ standard Borel and a Borel map $\phi : \mathcal{A} \times \mathcal{X} \to \mathbb{R}_+$ satisfying the validity condition

$$\mathbb{E}_\pi[\phi(a, \cdot)] \le 1\,,$$

for every $a \in \mathcal{A}$ and $\pi \in \mathcal{P}_0$. Each $a \in \mathcal{A}$ indexes an e-variable $E_{[a]} = \phi(a, \cdot)$: in a testing-by-betting game with $\mathcal{E} = \{E_{[a]} : a \in \mathcal{A}\}$, "*choosing action* a" means "*playing bet* $E_{[a]}$".

In simple-null examples $\mathcal{P}_0 = \{\pi_0\}$, a natural construction is to let a index a measure $\nu_a \ll \pi_0$ and set $\phi(a, x) = \mathrm{d}\nu_a/\mathrm{d}\pi_0(x)$ (cf. Section 5). For Bernoulli testing with $\pi_0 = \mathrm{Ber}(p_0)$, for $\mathcal{A} = [0, 1]$ and $\nu_a = \mathrm{Ber}(a)$, $\phi(a, 1) = a/p_0$ and $\phi(a, 0) = (1 - a)/(1 - p_0)$. For Gaussian testing with $\pi_0 = \mathcal{N}(0, \sigma^2)$, taking $\mathcal{A} = \mathbb{R}$ and $\nu_a = \mathcal{N}(a, \sigma^2)$ gives $\phi(a, x) = \exp(ax/\sigma^2 - a^2/(2\sigma^2))$. Beyond simple nulls, natural parametrisations are often available. For instance, for $\mathcal{P}_0$ the class of measures supported on $[-1, 1]$ with mean $m \in (-1, 1)$, one obtains the coin-betting class [Orabona and Jun, 2024, Clerico, 2025] by taking $\mathcal{A} = [(m - 1)^{-1}, (m + 1)^{-1}]$ and $\phi(a, x) = 1 + a(x - m)$. More generally, when the null is specified through constraints, the corresponding dual representation of e-variables yields a natural parametrised action space [Clerico, 2024, Larsson et al., 2025a].

We call a *policy* any sequence $A^\infty = (A_t)_{t \ge 1}$, where each $A_t : \mathcal{X}^{t-1} \to \mathcal{A}$ is a Borel map. At round t, after observing $x^{1:t-1}$, the policy plays the e-variable $E_t = E_{[A_t(x^{1:t-1})]} = \phi(A_t(x^{1:t-1}), \cdot)$. Define

$$\widehat{W}_t^{A^\infty}(x^{1:t}) = \widehat{W}_{t-1}^{A^\infty}(x^{1:t-1})\, \phi\big(A_t(x^{1:t-1}), x_t\big) = \prod_{k=1}^{t} \phi\big(A_k(x^{1:k-1}), x_k\big) = \prod_{k=1}^{t} E_k(x_k)\,,$$

recursively, with $\widehat{W}_0^{A^\infty}(x^{1:0}) = 1$. This is the *wealth* at time t under the policy A^∞. We define $\tau^{A^\infty} : \mathcal{X}^\infty \to \mathbb{N} \cup \{\infty\}$ as the first time the wealth under A^∞ reaches the threshold α^{-1}, namely

$$\tau^{A^\infty}(x^{1:\infty}) = \inf\{t \ge 1 : \widehat{W}_t^{A^\infty}(x^{1:t}) \ge \alpha^{-1}\}\,,$$

where $\inf \emptyset = \infty$. τ^{A^∞} is a stopping time with respect to the natural filtration induced by the sequence of observations. With these definitions, we have reformulated the testing-by-betting game as a sequential control problem with action space $\mathcal{A}$, where the available bets are precisely $\{E_{[a]} : a \in \mathcal{A}\}$. Hence, by Proposition 1, the procedure that rejects at time τ^{A^∞} is valid at level α.

Recall a *reward function* is any non-increasing map $R : \mathbb{N} \cup \{\infty\} \to \mathbb{R}_+$ with $R(\infty) = 0$. Evaluating a policy through the expected reward of its rejection time under $\pi_1^\otimes$ leads to the objective

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})]\,. \tag{1}$$

4 Markovian reduction and Bellman formulation

The control formulation that we introduced allows very general policies: the action chosen at time $t+1$ may depend on the entire past $x^{1:t}$. This raises a natural question: is such full history dependence needed? Our first main structural result shows that it is not. In fact, for the reward objective that we consider, there is no loss in restricting to policies whose decision at time $t + 1$ depends on the past only through the wealth. We call $a^\infty = (a_t)_{t \ge 0}$ a *Markov policy* if each $a_t : [0, \alpha^{-1}) \to \mathcal{A}$ is Borel. We identify such a sequence with any ordinary policy that chooses the action $a_t(w)$ at round t if the current wealth is $w < \alpha^{-1}$. Notably, once the threshold is reached the choice of action is irrelevant. With slight abuse of notation, we write τ^{a^∞} for the rejection time induced by the Markov policy.

Theorem 1. *For any non-increasing reward function $R : \mathbb{N} \cup \{\infty\} \to \mathbb{R}_+$ we have*

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] = \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})] .$$

Although intuitive, the Markovian reduction is not straightforward at this level of generality. Yet, this generality matters: the infinite-dimensional class of e-variables might not admit a natural finite-dimensional parametrisation without loss of power (e.g., Section 5 shows that for simple-vs-simple tests the canonical action space is the space of all probability measures dominated by the null). Theorem 1's proof embeds the problem in a positive Borel dynamic programme whose state records time, wealth, and rejection, then concludes via a stationary-policy result [Schäl and Sudderth, 1987].

Since the restriction of a Markov policy to future time steps remains Markov, it is natural to consider continuation problems starting from an arbitrary wealth $w \in [0, \alpha^{-1})$. For a Markov policy a^∞, let

$$\widetilde{W}_t^{w,a^\infty}(x^{1:t}) = \widetilde{W}_{t-1}^{w,a^\infty}(x^{1:t-1})\, \phi\big(a_{t-1}(\widetilde{W}_{t-1}^{w,a^\infty}(x^{1:t-1})), x_t\big) ,$$

where formally $\widetilde{W}_0^{w,a^\infty}(x^{1:0}) = w$. We also define $\tau_w^{a^\infty}(x^{1:\infty}) = \inf\{t \geq 1 : \widetilde{W}_t^{w,a^\infty}(x^{1:t}) \geq \alpha^{-1}\}$, the number of additional rounds needed for rejection when the current wealth is $w \in [0, \alpha^{-1})$ and future actions follow a^∞. This continuation viewpoint naturally leads to the value function

$$V_t(w) = \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}\big[R(t + \tau_w^{a^\infty})\big] .$$

In words, $V_t(w)$ is the maximal expected reward that can still be achieved when t rounds have already been played, the current wealth is w, and the threshold has not yet been reached. Now (1) reads

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] = \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})] = V_0(1) .$$

For $t \geq 0$ and $w \in [0, \alpha^{-1})$, we obtain formally the following Bellman recursion:

$$V_t(w) = \sup_{a \in \mathcal{A}} \int_{\mathcal{X}} \Big(R(t+1)\mathbf{1}_{\{w\phi(a,x) \geq \alpha^{-1}\}} + V_{t+1}(w\phi(a,x))\mathbf{1}_{\{w\phi(a,x) < \alpha^{-1}\}} \Big) \mathrm{d}\pi_1(x) . \quad (2)$$

The first term accounts for rejection at the next round, while the second accounts for continuation with updated wealth. This is the usual dynamic programming decomposition [Bertsekas and Shreve, 1978]. We remark that, at this level of generality, the Bellman recursion should be interpreted with some care. Since V_t is defined as a supremum, there is no reason a priori for V_t itself to be Borel. Nevertheless, V_t is at least universally measurable, which is enough to ensure that the integral in (2) is well defined. However, in general, there is no guarantee of either a closed-form expression for V_t or an optimal Borel policy attaining the supremum in (2). We therefore view the Bellman formulation primarily as a guiding principle. On the one hand, it provides a natural starting point for numerical approximation schemes, such as value-iteration type methods. On the other hand, it helps identify simpler but still informative approximations of the control problem itself, as we will see in Section 7.

5 Simple-vs-simple case: equivalence with e-processes

Before turning to concrete reward functions, it is worth relating testing-by-betting to the broader class of sequential tests based on e-processes of which testing-by-betting is a special case. Given a null hypothesis on $\mathcal{X}^\infty$, an *e-process* is a sequence $U^\infty = (U_t)_{t \geq 0}$ of non-negative Borel maps $U_t : \mathcal{X}^t \to \mathbb{R}_+$ such that $\mathbb{E}_{\Pi_0}[U_\tau] \leq 1$ for every stopping time τ and every Π_0 in the null. Such a process induces a valid sequential test by rejecting when U_t first reaches α^{-1} [Ramdas and Wang, 2025]. Testing by betting constructs evidence processes by multiplying one-round bets, whereas general e-processes need not admit such a product representation [Ramdas et al., 2022]. It is natural to ask whether, for our reward-based objective, restricting to testing-by-betting can lose power.

We now specialise to the case of a simple null $\pi_0^\otimes$ and simple alternative $\pi_1^\otimes$, where $\pi_0, \pi_1 \in \mathcal{P}_\mathcal{X}$ and $\pi_1 \ll \pi_0$. In this setting, there is a canonical choice of action space $\mathcal{A}$ that entails no loss of power. First, note that every e-variable E for π_0 satisfies $\mathbb{E}_{\pi_0}[E] \leq 1$, and since the reward is non-increasing, any non-zero such E may be normalised to satisfy $\mathbb{E}_{\pi_0}[E] = 1$, which only increases the resulting wealth pathwise. Hence, without loss of generality, one may restrict attention to those e-variables

satisfying $\mathbb{E}_{\pi_0}[E] = 1$. These are precisely the Radon–Nikodym densities of probability measures dominated by π_0. Accordingly, in the simple-null case we may take as action space $\mathcal{A}$ the set

$$\mathcal{P}^{\pi_0} = \{\nu \in \mathcal{P}_{\mathcal{X}} \,:\, \nu \ll \pi_0\}\,.$$

Lemma 2 in Appendix B shows that $\mathcal{P}^{\pi_0}$ is a standard Borel space and that versions of Radon–Nikodym densities can be chosen so that $(\nu, x) \mapsto \phi^{\pi_0}(\nu, x) = \mathrm{d}\nu/\mathrm{d}\pi_0(x)$ is Borel on $\mathcal{P}^{\pi_0} \times \mathcal{X}$.[2] Thus, for simple nulls we may take $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$. This is the full canonical version of the likelihood-ratio parametrisation used in the examples at the beginning of Section 3. In the Bernoulli case with $p_0 \in (0,1)$, $\mathcal{A} = [0,1]$ coincides with $\mathcal{P}^{\pi_0}$. By contrast, when $\pi_0 = \mathcal{N}(0, \sigma^2)$, setting $\mathcal{A} = \mathbb{R}$ as in Section 3 only parametrises the Gaussian-shift subfamily $\{\mathcal{N}(a, \sigma^2) : a \in \mathbb{R}\} \subsetneq \mathcal{P}^{\pi_0}$.

Let $\tau^{U^\infty} = \inf\{t \geq 0 \,:\, U_t \geq \alpha^{-1}\}$ be the rejection time of U^∞. The following equivalence holds.

Theorem 2. *Fix $\pi_0, \pi_1 \in \mathcal{P}_{\mathcal{X}}$, with $\pi_1 \ll \pi_0$, and let $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$. Then, we have*

$$\sup_{U^\infty \text{ e-process}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{U^\infty})] = \sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] = \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})]\,.$$

The proof leverages that every e-process for a simple null is dominated by a martingale [Ramdas et al., 2020]. Thus, every e-process procedure can be realised as a testing-by-betting strategy with history-dependent one-round bets, and the Markovian reduction of Theorem 1 then yields the claim. Theorem 2 shows that, in the simple-vs-simple i.i.d. setting, the control formulation loses no power from using one-round bets, provided $(\mathcal{A}, \phi)$ is chosen canonically. It also recovers the classical Wald's SPRT under GRO (whose bet is $d\pi_1/d\pi_0$ each round), while our reward-based approach can yield more general policies that may depend on time and wealth.

6 Hard-deadline reward

A natural choice for the reward is the hard-deadline $R(t) = \mathbf{1}_{\{t \leq T\}}$. This yields the objective (1)

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] = \sup_{A^\infty} \pi_1^\otimes(\tau^{A^\infty} \leq T) = \sup_{a^\infty \text{ Markov}} \pi_1^\otimes(\tau^{a^\infty} \leq T)\,,$$

namely maximising the probability of rejecting the null under the alternative before a fixed horizon. Related finite-horizon ideas appear in Voráček and Orabona [2025] in the context of coin-betting for mean estimation (see, e.g., Orabona and Jun, 2024, Waudby-Smith and Ramdas, 2024). There, the deadline affects the betting strategy by inducing more aggressive bets as the horizon approaches, a behaviour one also expects in our formulation. The key difference is that their approach lacks a prescribed alternative, but implicitly generates adaptive surrogate alternatives from observed deviations from the null. The horizon-aware formulation of Taga et al. [2026] is closer to ours in its control objective: bounded mean testing under a fixed deadline is formulated as a finite-horizon control problem over time and wealth, aiming to maximise rejection probability before the deadline while preserving anytime-validity. However, unlike our formulation, they do not fix a single simple alternative. Instead, their reinforcement learning implementation uses several synthetic alternatives as oracle training instances. As solving the dynamic programme exactly is generally hard, they mainly use its structure to describe regimes of betting aggressiveness and motivate heuristic betting policies.

Even in our framework, with a fixed simple alternative π_1, the hard-deadline control problem remains intractable in general. Yet, in the simple-vs-simple case it admits a canonical optimal Doob e-process (Fig. 2), whose construction is closely related to the induction of Koning and van Meer [2026].

Proposition 2. *Consider testing $\pi_0^\otimes$ against $\pi_1^\otimes$, with $\pi_1 \ll \pi_0$. Then there exists a Borel set $\Gamma^\star \subseteq \mathcal{X}^T$, viewed as an event depending only on the first T observations, such that $\pi_0^\otimes(\Gamma^\star) \leq \alpha$ and*

$$\sup_{U^\infty \text{ e-process}} \pi_1^\otimes\big(\tau^{U^\infty} \leq T\big) = \pi_1^\otimes(\Gamma^\star) = \max\big\{\pi_1^\otimes(\Gamma) \,:\, \Gamma \subseteq \mathcal{X}^T \text{ Borel}\,,\, \pi_0^\otimes(\Gamma) \leq \alpha\big\}\,.$$

Fixing Borel versions of the conditional probabilities, the supremum is achieved by the $\pi_0^\otimes$-martingale $U_t^\star(x^{1:t}) = \pi_0^\otimes(\Gamma^\star \mid x^{1:t})/\pi_0^\otimes(\Gamma^\star)$ if $\pi_0^\otimes(\Gamma^\star) > 0$, and by the constant e-process $U_t \equiv 1$ otherwise.

[2] Since $\pi_1 \ll \pi_0$, the particular choice of version does not affect either the null constraint or the objective under $\pi_1^\otimes$.

Here $\Gamma^\star$ is a Neyman–Pearson rejection region for the fixed-sample problem at horizon T [Neyman and Pearson, 1933, Lehmann and Romano, 2005]. Notably, the hard-deadline reward does not favour rejection earlier than the deadline. This can make the induced Doob test effectively non-sequential, as in Gaussian mean testing, where it coincides with the fixed-horizon Neyman–Pearson test (cf. Section 8). In Bernoulli testing, earlier rejection is possible, but the procedure is still conservative until close to the deadline (see Appendix E for further discussion). This motivates considering other rewards, such as the exponential-decay reward of the next section.

Proposition 2 shows that for the simple-vs-simple, the hard-deadline problem can be tackled directly at the e-process level, without solving the Bellman recursion. The optimal level-α sequential test rejects the null as soon as $U^{\star,\infty}$ crosses α^{-1}. For $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$, $U^{\star,\infty}$ is the wealth of any predictable policy satisfying $\nu_t^\star(x^{1:t-1}) = \pi_0^\otimes(x_t \in \cdot \mid x^{1:t-1}, \Gamma^\star)$ for $t \leq T$. This canonical policy implementation need not be Markovian, as it may depend on the full history. This is consistent with Theorem 1, which does not require every optimiser to be Markovian. Concrete examples are discussed in Section 8 and Appendix C, where the canonical induced policy is Markovian for Gaussian mean testing, not necessarily so in the Bernoulli case.

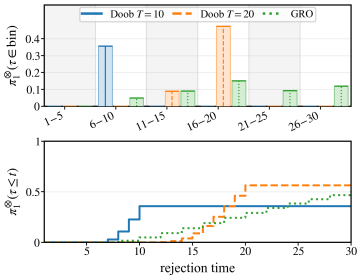


Figure 2: Rejection-time distributions for $\pi_0 = \text{Ber}(0.4)$ vs $\pi_1 = \text{Ber}(0.6)$ under the Doob e-process, compared with GRO. Details in Appendix D.2.

7 Exponentially decaying reward

We now specialise to the exponentially decaying reward $R(t) = e^{-t/T}$, with $T > 0$ the time scale. It is convenient to factor out the explicit time dependence from the value function and rescale the wealth. Thus, we let $z = \alpha w$ and $V_t(z/\alpha) = e^{-t/T} v(z)$. The Bellman recursion (2) now reads

$$v(z) = e^{-1/T} \sup_{a \in \mathcal{A}} \int_{\mathcal{X}} \Big(\mathbf{1}_{\{z\phi(a,x) \geq 1\}} + v(z\phi(a,x)) \mathbf{1}_{\{z\phi(a,x) < 1\}} \Big) \mathrm{d}\pi_1(x) \,.$$

The problem is therefore stationary (t no longer appears) and the only remaining inhomogeneity is the rejection boundary at $z = 1$. In general, this stationary Bellman equation does not admit a closed-form solution, but can be tackled numerically by standard dynamic-programming methods, after discretising wealth and actions, or by truncating the horizon and applying backward induction (as for Fig. 1). However, our focus here is different. We derive an analytically tractable approximation which, like GRO, selects a constant betting rule and admits an explicit large-threshold interpretation.

Since the test starts from $z = \alpha$, typically much smaller than 1, it is natural to start by studying the behaviour far from the threshold. In that regime, the truncation at 1 should have only a limited effect, suggesting the approximate homogeneous equation

$$v(z) \approx e^{-1/T} \sup_{a \in \mathcal{A}} \int_{\mathcal{X}} v(z\phi(a,x)) \mathrm{d}\pi_1(x) \,. \tag{3}$$

Under this approximation, the homogeneous equation becomes scale invariant in the wealth, suggesting the multiplicative ansatz $v(z) = cz^\eta$, with $c > 0$ and $\eta > 0$. Substituting this into (3) factors out the dependence on z. For the ansatz to yield a solution, there must exist $a_\star \in \mathcal{A}$ and $\eta_\star > 0$ such that

$$\sup_{a \in \mathcal{A}} \mathbb{E}_{\pi_1}\big[\phi(a, \cdot)^{\eta_\star}\big] = \mathbb{E}_{\pi_1}\big[\phi(a_\star, \cdot)^{\eta_\star}\big] = e^{1/T} \,. \tag{4}$$

When this holds, $z \mapsto cz^{\eta_\star}$ is a fixed point of the homogeneous Bellman equation (3) and we define the exponential-decay-optimal (EDO) criterion: play the action $a_\star$ from Equation (4), independently of wealth. For T large enough, the existence of a pair satisfying Equation (4) is natural when the action space is sufficiently rich: if $e^{1/T}$ is close to 1, it is enough to have actions with a small amount of moment growth under π_1. This becomes explicit in Proposition 4 for the simple-vs-simple case.

The threshold boundary effect has deliberately been ignored. Near $z = 1$, the stationary action should only be viewed as a baseline: once some outcomes already trigger rejection, extra payoff beyond the threshold has no value, and a wealth-dependent action may improve performance by reallocating betting mass to non-rejection outcomes. For instance, in Bernoulli testing, if observing 1 already

rejects, one can cap the payoff assigned to 1 at the level needed to cross the threshold and shift the remaining mass to increase the payoff assigned to 0. Such boundary-aware corrections can be applied to baselines such as EDO or GRO, and are related in spirit to Fischer and Ramdas [2026]. Numerically, in Fig. 3 (larger version in Appendix D.3), EDO already provides a close approximation to the Bellman optimum, and the boundary-aware capped variant nearly matches it.

The following bound shows that the policy suggested by the homogeneous equation has the right large-threshold order. We write τ^a for the rejection time under the policy that always plays $a \in \mathcal{A}$.

Proposition 3. *Let $(a_\star, \eta_\star)$ satisfy (4), with $\eta_\star \in (0,1)$ and $\phi(a_\star, \cdot) \le \Phi$, π_1-almost surely. Then*

$$(\alpha/\Phi)^{\eta_\star} \le \mathbb{E}_{\pi_1^\otimes}[e^{-\tau^{a_\star}/T}] \le \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[e^{-\tau^{a^\infty}/T}] \le \alpha^{\eta_\star}\,.$$

Thus EDO provides a first-order proxy: it identifies a constant policy with optimal α-exponent, in the sense that both the optimal Markov value and the value achieved by the constant action $a_\star$ have logarithm $\eta_\star \log(\alpha) + O(1)$ as $\alpha \to 0$. Hence, like GRO, EDO gives an α-agnostic constant-bet criterion whose optimality is first-order in the rejection threshold [Agrawal and Ramdas, 2025]. The boundedness assumption in Proposition 3 is mainly technical, as it allows us to control the overshoot at rejection. It could be relaxed under suitable integrability conditions, at the cost of a more involved analysis.

To make things more concrete, we consider the simple-vs-simple case, with $\pi_1 \ll \pi_0$ and $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$. Then, for T large enough, a closed-form solution of (4) typically exists: $a^\star$ is a likelihood-ratio tilt of π_0, and $\eta^\star$ is expressed through a Rényi divergence. For $\eta \in (0,1)$, let $L_\eta = (\mathrm{d}\pi_1/\mathrm{d}\pi_0)^{1/(1-\eta)}$ and $D_\xi(\pi\|\pi') = \frac{1}{\xi-1}\log \mathbb{E}_{\pi'}[(\mathrm{d}\pi/\mathrm{d}\pi')^\xi]$ denote the Rényi divergence of order $\xi > 1$.

Proposition 4. *Let $G = \{(1-\eta)\log \mathbb{E}_{\pi_0}[L_\eta] : \eta \in [0,1), \mathbb{E}_{\pi_0}[L_\eta] < \infty\}$. If $1/T \in (0, \sup G)$, then there exist $\eta_\star \in (0,1)$ and $\pi_\star \ll \pi_0$ satisfying (4). Moreover, $\mathrm{d}\pi_\star/\mathrm{d}\pi_0 = L_{\eta_\star}/\mathbb{E}_{\pi_0}[L_{\eta_\star}]$ and $\log(\sup_{\pi \ll \pi_0} \mathbb{E}_{\pi_1}[(\mathrm{d}\pi/\mathrm{d}\pi_0)^{\eta_\star}]) = \eta^\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = 1/T$.*

Under the mild integrability condition $G \neq \{0\}$, (4) admits a solution for all sufficiently large T. EDO then selects the e-variable $\mathrm{d}\pi_\star/\mathrm{d}\pi_0 = L_{\eta_\star}/\mathbb{E}_{\pi_0}[L_{\eta_\star}]$, a power tilt of π_0 by the likelihood ratio, with $\eta_\star$ chosen so that $\eta_\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = 1/T$. We can compare EDO with the GRO criterion. In the simple-vs-simple case, GRO recovers the constant action π_1, achieving $\sup_{\pi \ll \pi_0} \mathbb{E}_{\pi_1}\big[\log(\mathrm{d}\pi/\mathrm{d}\pi_0)\big] = \mathrm{KL}(\pi_1||\pi_0)$. This KL divergence is also the quantity governing the first-order expected stopping-time scale under GRO: for the likelihood-ratio strategy, $\mathbb{E}_{\pi_1^\otimes}[\tau^{\pi_1}] \sim \log(1/\alpha)/\mathrm{KL}(\pi_1||\pi_0)$ as $\alpha \to 0$ [Agrawal and Ramdas, 2025]. By contrast, EDO plays $\pi^\star$ every round. In this case, the large-threshold relation reads $-T\log \mathbb{E}_{\pi_1^\otimes}[e^{-\tau^{\pi_\star}/T}] \sim \log(1/\alpha)/D_{1/(1-\eta_\star)}(\pi_1\|\pi_0)$. The two pictures are consistent, as $T \to \infty$ the objective approaches the linear regime $e^{-\tau/T} \sim 1 - \tau/T$, while $\eta_\star \to 0$ and $D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) \to \mathrm{KL}(\pi_1||\pi_0)$ (cf. Fig. 4).

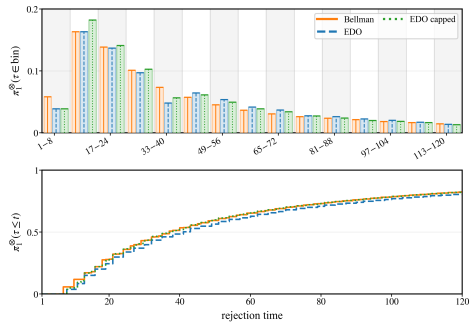


Figure 3: Comparison of EDO with the numerical Bellman optimum for $\pi_0 = \mathrm{Ber}(1/2)$ vs $\pi_1 = \mathrm{Ber}(2/3)$ in terms of rejection times. With boundary-aware capping, the approximation is nearly indistinguishable. Details in Appendix D.3.

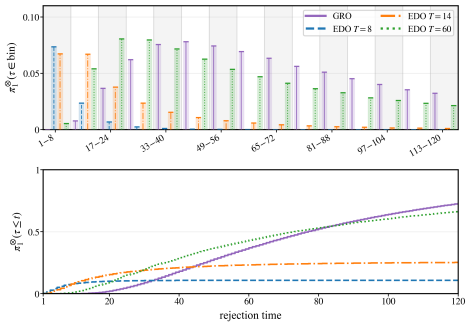


Figure 4: Effect of the reward time scale on EDO rejection-time distributions for $\pi_0 = \mathcal{N}(0,1)$ vs $\pi_1 = \mathcal{N}(.25,1)$: smaller time scales favour earlier rejection, with a visible power trade-off. GRO is shown for comparison. Details in Appendix D.4.

There is another comparison with GRO worth making. GRO favours actions with large expected log-growth under the alternative, and positive log-growth leads to almost sure eventual rejection. Conversely, the exponentially decaying reward values earlier rejection, and for small time scales may select an action with non-positive log-drift. Thus early rejection can come at the cost of power. The next result records this trade-off for constant policies in the general action-space formulation.

Proposition 5. *Fix $a \in \mathcal{A}$, assume that $\phi(a, \cdot) > 0$, π_1-almost surely, and that $\gamma(a) = \mathbb{E}_{\pi_1}[\log \phi(a, \cdot)] \in [-\infty, \infty]$ is well-defined. Then, $\pi_1^{\otimes}(\tau_w^a < \infty) = 1$ for every $w \in (0, \alpha^{-1})$ if and only if either $\gamma(a) > 0$, or $\gamma(a) = 0$ and $\pi_1(\log \phi(a, \cdot) > 0) > 0$. Moreover, when $\gamma(a) < 0$, if $\phi(a, \cdot) \leq \Phi$, π_1-almost surely, and there exists $\kappa > 0$ such that $\mathbb{E}_{\pi_1}[\phi(a, \cdot)^\kappa] = 1$, then $(\alpha w/\Phi)^\kappa \leq \pi_1^{\otimes}(\tau_w^a < \infty) \leq (\alpha w)^\kappa$ for every $w \in (0, \alpha^{-1})$.*

Proposition 5 essentially says that a strategy betting the same action each round has power 1 against $\pi_1^{\otimes}$ if and only if its log-wealth has non-negative drift under π_1. In the simple-vs-simple setting, under the exponentially decaying reward this criterion becomes especially explicit for the constant action $\pi_\star$ selected by Proposition 4, for which we get $\gamma(\pi_\star) = (\mathrm{KL}(\pi_1||\pi_0) - 1/T)/(1 - \eta_\star)$. In such case for a test to have power 1 we need $\mathrm{KL}(\pi_1||\pi_0)$ to be at least $1/T$. The trade-off between T and the power is illustrated in Fig. 5 for two Bernoulli distributions.

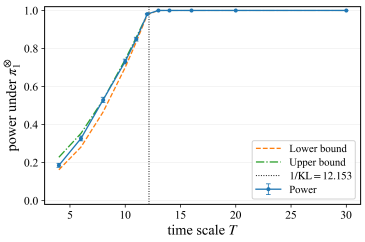


Figure 5: Power trade-off in $\pi_0 = \mathrm{Ber}(.5)$ vs $\pi_1 = \mathrm{Ber}(.7)$ for EDO criterion. Details in Appendix D.5.

8 Two illustrative examples

We discuss two simple-vs-simple examples. The full calculations are deferred to Appendix C.

Testing Gaussians. We consider the Gaussian case, where $\pi_0 = \mathcal{N}(0, \sigma^2)$ and $\pi_1 = \mathcal{N}(\mu, \sigma^2)$, with $\mu > 0$ and $\sigma > 0$. We fix $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$. For the hard-deadline reward $R(t) = \mathbf{1}_{t \leq T}$, Proposition 2 gives the event $\Gamma^\star = \{S_T \geq c\}$, where $c = \sigma\sqrt{T}F^{-1}(1-\alpha)$, $S_T = \sum_{t=1}^T x_t$ and F is the cdf of a standard normal. The optimal e-process is the Doob martingale

$$U_t^\star = \alpha^{-1}\pi_0^{\otimes}(S_T \geq c \mid S_t) = \tfrac{1}{\alpha}\left(1 - F\left(\tfrac{c - S_t}{\sigma\sqrt{T-t}}\right)\right),$$

for $t < T$, with terminal convention $U_T^\star = \alpha^{-1}\mathbf{1}_{[c,\infty)}(S_T)$. The induced strategy is Markovian, since the same conditional probability is strictly increasing in S_t, hence in the current wealth (see Appendix C.1). Note that S_T given S_t is Gaussian with full support, so the conditional probability above is strictly below one for every $t < T$. Thus the exact Gaussian hard-deadline test never rejects before the deadline: it rejects at T precisely on $\Gamma^\star$, and hence effectively coincides with the fixed-sample Neyman–Pearson test. This contrasts with Fig. 1, where the action space is restricted to Gaussian-shift bets, so the displayed policy is the Bellman optimum within that smaller class rather than the unrestricted Doob strategy. Finally, the magnitude of μ affects the power but not the rejection region $\Gamma^\star$, only its sign does.

For the exponentially decaying reward $R(t) = e^{-t/T}$, the EDO action from Proposition 4 is obtained solving $(T\eta^\star)^{-1} = D_\xi(\pi_1\|\pi_0) = \xi\mu^2/(2\sigma^2)^{-1}$, which gives $\eta^\star = 2\sigma^2/(\mu^2 T + 2\sigma^2) \in (0, 1)$. The EDO action is $\pi_\star = \mathcal{N}\left(2\sigma^2/(\mu T) + \mu,\ \sigma^2\right)$, corresponding to the e-variable

$$E_\star(x) = \tfrac{\mathrm{d}\pi_\star}{\mathrm{d}\pi_0}(x) = \exp\left(\left(\tfrac{\mu}{\sigma^2} + \tfrac{2}{\mu T}\right)(x - \mu) + \tfrac{\mu^2}{2\sigma^2} - \tfrac{2\sigma^2}{\mu^2 T^2}\right)$$

in the testing-by-betting game. Compared with the GRO strategy, which plays $\pi_1 = \mathcal{N}(\mu, \sigma^2)$, EDO is more aggressive, shifting further away from the null. As $T \to \infty$, $\pi_\star$ recovers the GRO π_1. By Proposition 5, this aggressiveness preserves power one if and only if $T \geq \mathrm{KL}(\pi_1||\pi_0)^{-1} = 2\sigma^2/\mu^2$.

Testing Bernoullis. Consider the case $\pi_0 = \mathrm{Ber}(p_0)$ and $\pi_1 = \mathrm{Ber}(p_1)$, with $0 < p_0 < p_1 < 1$. Set $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$. For the hard-deadline $R(t) = \mathbf{1}_{t \leq T}$, identifying $\Gamma^\star$ is more delicate than in the Gaussian case, since Bernoulli is atomic. This can still be done exactly through a finite integer programme (see Appendix C.2). However, the canonical Doob strategy may fail to be Markovian in the wealth (cf. the discussion after Proposition 2). For large T, a simpler and explicit upper-tail approximation holds, only depending on $S_T = \sum_{t=1}^T x_t$ (see Appendix C.2).

For $R(t) = e^{-t/T}$, to determine the EDO action, $\eta_\star$ does not admit a closed-form but can be easily computed numerically solving a one-dimensional equation. The resulting action $\pi_\star = \mathrm{Ber}(p_\star)$ has a closed-form expression in terms of $\eta_\star$ (see Appendix C.2), from which it follows that EDO is more aggressive than GRO: $p_\star > p_1$. As $T \to \infty$, we have $\eta_\star \to 0$, and $\pi_\star$ recovers the GRO action π_1. By Proposition 5, EDO has power one if and only if $T \geq \mathrm{KL}(\mathrm{Ber}(p_1)||\,\mathrm{Ber}(p_0))^{-1}$.

9 Discussion

We close by discussing some limitations of the present formulation and possible open directions. First, the current work only focuses on the case of a simple alternative π_1. This choice makes the control problem transparent, but it is restrictive when the alternative is composite. A natural extension would be to replace the expectation under $\pi_1^{\otimes}$ by a Bayesian, worst-case, or relative worst-case criterion over a class of alternatives, in the spirit of the GROW and REGROW criteria from safe testing [Grünwald et al., 2024]. The resulting problem would be a robust version of our time-sensitive control formulation. A related open direction concerns composite nulls. Our explicit approximation results are derived in the simple-vs-simple setting, while under GRO the extension to a composite null often admits a least-favourable-null interpretation through a KL projection of the alternative onto the null model. Interestingly, Larsson et al. [2025b] extend reverse information projections to power utilities through reverse Rényi projections. Given the connection between our EDO criterion and Rényi divergences, investigating whether analogous projections characterise optimal time-sensitive strategies under composite nulls is a natural direction for future work.

Another point concerns the role of the significance level. Throughout our analysis, we fix a significance level α, making the objective transparent: the reward is assigned to the hitting time of α^{-1}. However, the same wealth process determines hitting times for all rejection boundaries, and one may want a strategy that performs well at several values of α simultaneously. A natural extension would be to replace the single-threshold objective by an aggregate, worst-case, or relative worst-case criterion over a family of significance levels, in analogy with the composite-alternative criteria discussed above. In any case, for the exponential decay reward, EDO is already α-agnostic.

While we have focused on two particular reward functions to illustrate the main ideas, many other choices remain to be explored. A particularly interesting direction concerns finite-horizon rewards that gradually decay to zero, rather than remaining constant until a hard deadline, such as the linearly decaying reward $R(t) = (1 - t/T)_+$. Such problems can be tackled numerically through backward induction. In the simple-vs-simple setting that we have considered, one may also seek analytical characterisations in the spirit of the Doob-martingale construction that we used in Section 6, although identifying the optimal rejection event would require a more involved sequential optimisation and is unlikely to yield the same simple closed-form expressions as in the hard-deadline case.

Finally, our analysis is restricted to non-negative rewards. This is natural when $R(t)$ represents the *value* of a rejection at t, and it allows us to invoke a result of Schäl and Sudderth [1987] to prove the Markovian reduction. However, it excludes unbounded loss-type objectives, such as directly minimising $\mathbb{E}_{\pi_1^{\otimes}}[\tau]$ (which would correspond to $R(t) = -t$). Extending the theory to suitable unbounded rewards remains an interesting direction. However, the exponentially decaying reward already gives a *proxy* for minimising the expected rejection time: as $T \to \infty$, we enter the linear regime $e^{-t/T} \approx 1 - t/T$ and EDO recovers GRO.

LLM usage. We acknowledge the use of LLMs to improve exposition, explore proof ideas, and generate code. We take full responsibility for the content of the paper.

References

S. Agrawal and A. Ramdas. On stopping times of power-one sequential tests: Tight lower and upper bounds. *arXiv:2504.19952*, 2025.

D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, New York, 1978.

E. Clerico. Optimal e-value testing for properly constrained hypotheses. *arXiv:2412.21125*, 2024.

E. Clerico. On the optimality of coin-betting for mean estimation. *International Journal of Approximate Reasoning*, 187:109550, 2025.

L. E. Dubins and L. J. Savage. *How to Gamble If You Must: Inequalities for Stochastic Processes*. McGraw-Hill, New York, 1965.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, New York, 2 edition, 1971.

L. Fischer and A. Ramdas. Improving Wald's (Approximate) Sequential Probability Ratio Test by Avoiding Overshoot. *IEEE Transactions on Information Theory*, 72(4):2457–2471, 2026.

P. Grünwald, R. de Heide, and W. M. Koolen. Safe testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5):1091–1128, 2024.

O. Kallenberg. Stationary and invariant densities and disintegration kernels. *Probability Theory and Related Fields*, 160(3–4):567–592, 2014.

A. S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer, New York, 1995.

J. L. Kelly Jr. A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926, 1956.

N. W. Koning and S. van Meer. Anytime validity is free: Inducing sequential tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkag050, 2026.

K. K. G. Lan and D. L. DeMets. Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3): 659–663, 1983.

K. Lange. Borel sets of probability measures. *Pacific Journal of Mathematics*, 48(1):141–161, 1973.

M. Larsson, A. Ramdas, and J. Ruf. Testing hypotheses generated by constraints. *arXiv:2504.02974*, 2025a.

M. Larsson, A. Ramdas, and J. Ruf. The numeraire e-variable and reverse information projection. *The Annals of Statistics*, 53(3):1015–1043, 2025b.

E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer, New York, 3 edition, 2005.

A. A. Liapounoff. Sur les fonctions-vecteurs complètement additives. *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya*, 4(6):465–478, 1940.

J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231:289–337, 1933.

F. Orabona and K.-S. Jun. Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*, 70(1):436–455, 2024.

A. Ramdas and R. Wang. Hypothesis testing with e-values. *Foundations and Trends in Statistics*, 1 (1–2):1–390, 2025.

A. Ramdas, J. Ruf, M. Larsson, and W. M. Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv:2009.03167*, 2020.

A. Ramdas, J. Ruf, M. Larsson, and W. M. Koolen. Testing exchangeability: Fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning*, 141:83–109, 2022.

A. Ramdas, P. Grünwald, V. Vovk, and G. Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601, 2023.

M. Schäl and W. Sudderth. Stationary policies and markov policies in borel dynamic programming. *Probability Theory and Related Fields*, 74(1):91–111, 1987.

E. O. Taga, S. Oymak, and S. Shekhar. Learning to bet for horizon-aware anytime-valid testing. *arXiv:2603.19551*, 2026.

V. Voráček and F. Orabona. STaR-Bets: Sequential target-recalculating bets for tighter confidence intervals. In *Advances in Neural Information Processing Systems*, 2025.

A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2): 117–186, 1945.

I. Waudby-Smith and A. Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27, 2024.

D. Williams. *Probability with martingales*. Cambridge university press, 1991.

Y.-C. Yao. On optimality of bold play for discounted Dubins–Savage gambling problems with limited playing times. *Journal of Applied Probability*, 44(1):212–225, 2007.

A Notation and conventions

We use the following notation throughout. We write $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{R}_+ = [0, \infty)$. For a statement B, $\mathbf{1}_B$ denotes its indicator. We use the convention that empty products are equal to one and that $\inf \emptyset = \infty$. All logarithms are natural logarithms.

For a sequence $x_1, x_2, \dots$, we write $x^{1:t} = (x_1, \dots, x_t)$ for the first t coordinates, with $x^{1:0}$ denoting the empty history. We write $x^{1:\infty} = (x_1, x_2, \dots)$ for the full infinite sequence. For sequences indexed from zero, we analogously write $x^{0:t} = (x_0, \dots, x_t)$.

All measurable spaces are assumed to be standard Borel unless explicitly stated otherwise. A standard Borel space is a measurable space that is measurably isomorphic to a Borel subset of a Polish space. This convention is used to avoid repeated measurability qualifications and to ensure that regular conditional probabilities and measurable kernels are available.

If $\mathcal{X}$ is a measurable space, $\mathcal{P}_{\mathcal{X}}$ denotes the set of Borel probability measures on $\mathcal{X}$. For $\pi \in \mathcal{P}_{\mathcal{X}}$, $\mathbb{E}_\pi$ denotes expectation under π. If π is a one-step law, then $\pi^{\otimes}$ denotes the infinite product measure on $\mathcal{X}^\infty$, corresponding to an i.i.d. sequence with common law π.

For probability measures ν and π on the same measurable space, $\nu \ll \pi$ means that ν is absolutely continuous with respect to π, and $\mathrm{d}\nu/\mathrm{d}\pi$ denotes a Radon–Nikodym derivative. Statements involving such derivatives are understood up to almost-sure equivalence unless a particular version is fixed.

Unless otherwise specified, filtrations are the natural filtrations generated by the observed sequence. A random time is called a stopping time with respect to this filtration. All maps and kernels are taken to be Borel measurable whenever this is required by the context.

Measurability is always meant with respect to the Borel σ-fields determined by the spaces under consideration. Maps, policies, kernels, and selections are taken to be Borel measurable unless explicitly stated otherwise.

B Omitted proofs

B.1 Proof of Theorem 1

We prove here the Markovian reduction stated in Theorem 1:

Theorem (Markovian reduction)**.** *For any reward function R we have*

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] = \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})]\,.$$

First, we define a *reduced* control model. This is a positive Borel control model whose state records the current time and wealth, until the threshold is reached. We set

$$\mathcal{S} = \left(\mathbb{N} \times [0, \alpha^{-1})\right) \cup \{\circledast\}\,,$$

where $\circledast$ is a terminal state. The interpretation is that $(t, w) \in \mathbb{N} \times [0, \alpha^{-1})$ means that t observations have been seen, the current wealth is w, and the threshold has not yet been reached. The state $\circledast$ is absorbing and represents the threshold having been reached.

Define $\Phi : \mathcal{S} \times \mathcal{A} \times \mathcal{X} \to \mathcal{S}$ by

$$\Phi((t, w), a, x) = \begin{cases} \circledast & \text{if } w\phi(a, x) \geq \alpha^{-1}\,, \\ (t + 1, w\phi(a, x)) & \text{if } w\phi(a, x) < \alpha^{-1}\,, \end{cases}$$

for $(t, w) \in \mathbb{N} \times [0, \alpha^{-1})$, and $\Phi(\circledast, a, x) = \circledast$. Since ϕ is jointly Borel, Φ is Borel. The alternative π_1 induces a transition kernel q, defined as

$$q(B \mid s, a) = \pi_1(\{x \in \mathcal{X} : \Phi(s, a, x) \in B\})\,,$$

for any Borel $B \subseteq \mathcal{S}$. This is a Borel measurable stochastic kernel on $\mathcal{S}$, by the Borelness of Φ. We now define the one-step reward $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}_+$ as

$$r((t, w), a, s') = R(t + 1)\mathbf{1}_{\{s' = \circledast\}}$$

for $(t, w) \in \mathbb{N} \times [0, \alpha^{-1})$, and $r(\circledast, a, s') = 0$ for every $a \in \mathcal{A}$ and $s' \in \mathcal{S}$. A reward is received exactly when the process first enters $\circledast$, namely when the testing-by-betting procedure rejects.

A *reduced policy* is a sequence $S^\infty = (S_0, S_1, \dots)$, where, for each $n \geq 0$, S_n is a measurable kernel from $\mathcal{S}^{n+1}$ to $\mathcal{A}$. This means that for each $(s_0, \dots, s_n) \in \mathcal{S}^{n+1}$, $S_n(\cdot \mid s_0, \dots, s_n)$ is a Borel probability measure on $\mathcal{A}$, and the map $(s_0, \dots, s_n) \mapsto S_n(B \mid s_0, \dots, s_n)$ is Borel for every Borel set $B \subseteq \mathcal{A}$. Note that the reduced model allows arbitrary state-history-dependent randomised policies, and deterministic reduced policies are included as special cases, when each S_n is a Dirac mass. A *stationary* reduced policy is a Borel map $f : \mathcal{S} \to \mathcal{A}$.

For a given reduced policy S^∞ and an initial state $s \in \mathcal{S}$, let $I(S^\infty, s)$ be its expected total reward

$$I(S^\infty, s) = \mathbb{E}\left[\sum_{n=0}^{\infty} r(Y_n, A_n, Y_{n+1})\right]$$

where $(Y_n, A_n)_{n \geq 0}$ denotes the controlled process of states and actions with $Y_0 = s$, A_n conditioned on $(Y_0, \dots, Y_n)$ with law $S_n(\cdot \mid Y_0, \dots, Y_n)$, and Y_{n+1} conditioned on $(Y_0, \dots, Y_n, A_n)$ with law $q(\cdot \mid Y_n, A_n)$. We define

$$v(s) = \sup_{S^\infty} I(S^\infty, s)\,,$$

where the supremum is over all reduced policies.

The reduced model defined above is a positive Borel dynamic-programming model. The following is the only non-elementary ingredient we will need to prove Theorem 1. This is based on [Schäl and Sudderth, 1987, Corollary 1.1].

Lemma 1. *For every initial state $s \in \mathcal{S}$,*

$$v(s) = \sup_{f \text{ stationary}} I(f, s)\,,$$

where the supremum on the right-hand side is over all Borel stationary policies $f : \mathcal{S} \to \mathcal{A}$.

Proof. Let us first consider the larger class of full-history reduced policies. Such a policy is a sequence $\bar{S}^\infty = (\bar{S}_0, \bar{S}_1, \dots)$, where, for each $n \geq 0$, $\bar{S}_n$ is a Borel measurable kernel from $\mathcal{S} \times (\mathcal{A} \times \mathcal{S})^n$ to $\mathcal{A}$. Thus $\bar{S}_n$ may depend on the full reduced history $(s_0, a_0, s_1, \dots, a_{n-1}, s_n)$. Let $\bar{v}(s) = \sup_{\bar{S}^\infty} I(\bar{S}^\infty, s)$, where the supremum is over all full-history reduced policies. Then $\sup_{f \text{ stationary}} I(f, s) \leq v(s) \leq \bar{v}(s)$, because stationary policies are included in the reduced policies used to define v, and every such reduced policy can be viewed as a full-history reduced policy which ignores the past actions.

We now apply [Schäl and Sudderth, 1987, Corollary 1.1] to the model with full-history policies. Indeed, $\mathcal{S}$ and $\mathcal{A}$ are standard Borel spaces, the feasible action set is constant and equal to $\mathcal{A}$, the transition kernel q is Borel measurable, and the one-step reward r is non-negative and Borel measurable. Since the process can receive a non-zero reward at most once and R is bounded, $\bar{v}(s) < \infty$ for every $s \in \mathcal{S}$. Hence the result applies. In the notation of Schäl and Sudderth [1987], our $\bar{v}$ is the value function of the positive model and Corollary 1.1 says that stationary policies are persistently ε-optimal. Applied to the initial distribution concentrated at s, this means that for every $\varepsilon > 0$ there exists a stationary policy $f_\varepsilon : \mathcal{S} \to \mathcal{A}$ such that $I(f_\varepsilon, s) \geq (1 - \varepsilon)\bar{v}(s)$, with the case $\bar{v}(s) = 0$ being trivial by non-negativity. Therefore $\bar{v}(s) = \sup_{f \text{ stationary}} I(f, s)$. Combining this equality with the previous inequalities yields $v(s) = \sup_{f \text{ stationary}} I(f, s)$. □

Proof of the Markovian reduction. We compare the original problem with the reduced control model. First, fix an arbitrary policy $A^\infty = (A_1, A_2, \dots)$ for the original problem. Let $s_0 = (0, 1)$ and define recursively

$$\widehat{S}_0^{A^\infty}(x^{1:0}) = s_0$$

and

$$\widehat{S}_{t+1}^{A^\infty}(x^{1:t+1}) = \Phi(\widehat{S}_t^{A^\infty}(x^{1:t}), A_{t+1}(x^{1:t}), x_{t+1})\,.$$

These are all Borel maps. By induction over t, for any $x^{1:\infty} \in \mathcal{X}^\infty$,

$$\widehat{S}_t^{A^\infty}(x^{1:t}) = \begin{cases} (t, \widehat{W}_t^{A^\infty}(x^{1:t})) & \text{if } \tau^{A^\infty}(x^{1:\infty}) > t\,, \\ \circledast & \text{if } \tau^{A^\infty}(x^{1:\infty}) \leq t\,. \end{cases}$$

Now, for any $x^{1:\infty} \in \mathcal{X}^\infty$,

$$\sum_{t=0}^{\infty} r\left(\widehat{S}_t^{A^\infty}(x^{1:t}), A_{t+1}(x^{1:t}), \widehat{S}_{t+1}^{A^\infty}(x^{1:t+1})\right) = R(\tau^{A^\infty}(x^{1:\infty}))\,.$$

Indeed, if $\tau^{A^\infty}(x^{1:\infty}) = T < \infty$, then the path enters $\circledast$ at step T, and the only non-zero term in the sum is for $t = T - 1$, where it is equal to $R(T)$. On the other hand, if $\tau^{A^\infty}(x^{1:\infty}) = \infty$, then the path never enters $\circledast$ and the sum is $0 = R(\infty)$.

Now, let us construct a reduced policy starting from A^∞. For each $n \geq 0$, consider the joint law under $X^{1:\infty} \sim \pi_1^{\otimes}$ of

$$\left(\widehat{S}_0^{A^\infty}(X^{1:0}), \dots, \widehat{S}_n^{A^\infty}(X^{1:n}), A_{n+1}(X^{1:n})\right).$$

Since $\mathcal{S}$ and $\mathcal{A}$ are standard Borel spaces, there exists a regular conditional distribution of the last coordinate given the first $n + 1$ ones. Let $S_n^{A^\infty}$ be such a kernel, that is,

$$S_n^{A^\infty}(B \mid s_0, \dots, s_n) = \pi_1^{\otimes}\left(A_{n+1}(X^{1:n}) \in B \mid \widehat{S}_0^{A^\infty} = s_0, \dots, \widehat{S}_n^{A^\infty} = s_n\right)$$

for every Borel set $B \subseteq \mathcal{A}$. We claim that the state process generated by $(S_n^{A^\infty})_{n \geq 0}$ under the transition kernel q has the same law as the process $(\widehat{S}_t^{A^\infty}(X^{1:t}))_{t \geq 0}$ generated by A^∞ in the original model, under $X^{1:\infty} \sim \pi_1^{\otimes}$. Indeed, by construction the two processes start from the same state. Moreover, conditionally on $\widehat{S}_0^{A^\infty}, \dots, \widehat{S}_n^{A^\infty}$, the conditional law of $A_{n+1}(X^{1:n})$ is $S_n^{A^\infty}(\cdot \mid \widehat{S}_0^{A^\infty}, \dots, \widehat{S}_n^{A^\infty})$, and X_{n+1} is independent of $X^{1:n}$ with law π_1. Hence, for every Borel set $C \subseteq \mathcal{S}$,

$$\pi_1^{\otimes}\left(\widehat{S}_{n+1}^{A^\infty} \in C \mid \widehat{S}_0^{A^\infty}, \dots, \widehat{S}_n^{A^\infty}\right) = \int_{\mathcal{A}} q(C \mid \widehat{S}_n^{A^\infty}, a)\, \mathrm{d}S_n^{A^\infty}(a \mid \widehat{S}_0^{A^\infty}, \dots, \widehat{S}_n^{A^\infty})\,,$$

which is precisely the transition rule in the reduced model. It follows by induction on n that, for every $n \geq 0$, the law of the first $n + 1$ states in the reduced model, started from $(0, 1)$, controlled by $S^{A^\infty, \infty}$, and evolved according to q, is the same as the law of $(\widehat{S}_0^{A^\infty}, \dots, \widehat{S}_n^{A^\infty})$ under $X^{1:\infty} \sim \pi_1^{\otimes}$.

Let $\widehat{\tau}^{A^\infty} = \inf\{t \geq 1 : \widehat{S}_t^{A^\infty}(X^{1:t}) = \circledast\}$. Since $\widehat{S}_t^{A^\infty}(x^{1:t}) = \circledast$ if and only if $\tau^{A^\infty}(x^{1:\infty}) \leq t$, we have $\widehat{\tau}^{A^\infty} = \tau^{A^\infty}$. Since the total reward in the reduced model is equal to R evaluated at the first entrance time into $\circledast$, the equality in law just proved gives

$$I(S^{A^\infty,\infty}, (0,1)) = \mathbb{E}_{\pi_1^\otimes}[R(\widehat{\tau}^{A^\infty})] = \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})]\,.$$

Taking the supremum over original policies, we obtain

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] \leq \sup_{S^\infty} I(S^\infty, (0,1)) = v(0,1)\,. \tag{5}$$

Now, we show that every stationary reduced policy gives rise to a Markov policy in the original model achieving the same expected total reward. Fix a stationary reduced policy $f : \mathcal{S} \to \mathcal{A}$. Define $a^{f,\infty} = (a_t^f)_{t\geq 0}$ by

$$a_t^f(w) = f(t,w)$$

for $t \geq 0$ and $w \in [0, \alpha^{-1})$. This is a Markov policy. Let $A^{f,\infty}$ be the corresponding ordinary policy, defined recursively by

$$A_{t+1}^{f,\infty}(x^{1:t}) = \begin{cases} a_t^f(\widehat{W}_t^{A^{f,\infty}}(x^{1:t})) & \text{if } \widehat{W}_t^{A^{f,\infty}}(x^{1:t}) < \alpha^{-1}\,, \\ f(\circledast) & \text{if } \widehat{W}_t^{A^{f,\infty}}(x^{1:t}) \geq \alpha^{-1}\,. \end{cases}$$

Under $X^{1:\infty} \sim \pi_1^\otimes$, the state process induced by $A^{f,\infty}$ starts from $(0,1)$ and, before entering $\circledast$, moves from (t,w) to $\circledast$ if $w\phi(a_t^f(w), X_{t+1}) \geq \alpha^{-1}$, and to $(t+1, w\phi(a_t^f(w), X_{t+1}))$ otherwise. This is exactly the transition rule of the reduced model under the stationary policy f. Hence, by induction on n, the law of the first $n+1$ induced states is the same as the law of the first $n+1$ states in the reduced model, started from $(0,1)$, controlled by f, and evolved according to q. Therefore,

$$\mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^{f,\infty}})] = I(f, (0,1))\,.$$

Taking the supremum over stationary policies f, and using Lemma 1, we obtain

$$\sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})] \geq \sup_{f \text{ stationary}} I(f, (0,1)) = v(0,1)\,.$$

Together with (5), this gives

$$\sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})] \geq v(0,1) \geq \sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})]\,.$$

The reverse inequality follows because Markov policies induce a subclass of admissible ordinary policies in the original problem. This proves the claim. □

B.2 Proof of Theorem 2

We now prove Theorem 2, which we restate here:

Theorem (Simple-vs-simple e-process equivalence)**.** *Fix $\pi_0, \pi_1 \in \mathcal{P}_\mathcal{X}$, with $\pi_1 \ll \pi_0$, and let $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$. Then, we have*

$$\sup_{U^\infty \textit{ e-process}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{U^\infty})] = \sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] = \sup_{a^\infty \textit{ Markov}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{a^\infty})]\,.$$

First, we state a lemma that ensures that the pair $(\mathcal{P}^{\pi_0}, \phi^{\pi_0})$, defined in Section 5, is a valid control framework for the simple-vs-simple case.

Lemma 2. *Fix $\pi_0 \in \mathcal{P}_\mathcal{X}$ and define $\mathcal{P}^{\pi_0} = \{\nu \in \mathcal{P}_\mathcal{X} : \nu \ll \pi_0\}$. Then, $\mathcal{P}^{\pi_0}$ is a standard Borel space. Moreover, there exists a Borel map $\phi^{\pi_0} : \mathcal{P}^{\pi_0} \times \mathcal{X} \to \mathbb{R}_+$ such that, for every $\nu \in \mathcal{P}^{\pi_0}$, the map $x \mapsto \phi^{\pi_0}(\nu, x)$ is a version of the Radon–Nikodym derivative $\mathrm{d}\nu/\mathrm{d}\pi_0$. Equivalently, for every Borel set $B \subseteq \mathcal{X}$,*

$$\nu(B) = \int_B \phi^{\pi_0}(\nu, x)\,\mathrm{d}\pi_0(x)\,.$$

Proof. We endow $\mathcal{P}_{\mathcal{X}}$ with the σ-field generated by the evaluation maps $\nu \mapsto \nu(B)$, where $B \subseteq \mathcal{X}$ is Borel. Since $\mathcal{X}$ is standard Borel, $\mathcal{P}_{\mathcal{X}}$ is standard Borel [Kechris, 1995, Theorem 17.23][3]. Moreover, since $\mathcal{X}$ admits a Polish topology generating the same Borel σ-field, [Lange, 1973, Theorem 3.5] implies that $\mathcal{P}^{\pi_0}$ is a Borel subset of $\mathcal{P}_{\mathcal{X}}$. Therefore $\mathcal{P}^{\pi_0}$ is standard Borel.

For $\nu \in \mathcal{P}^{\pi_0}$ and any Borel set $B \subseteq \mathcal{X}$, define $Q(\nu, B) = \nu(B)$. Then Q is a probability kernel from $\mathcal{P}^{\pi_0}$ to $\mathcal{X}$, since $Q(\nu, \cdot) = \nu$ and the evaluation maps $\nu \mapsto \nu(B)$ are Borel. Moreover, $Q(\nu, \cdot) \ll \pi_0$ for every $\nu \in \mathcal{P}^{\pi_0}$. By [Kallenberg, 2014, Lemma 2.1], there exists a Borel map $\rho : \mathcal{P}^{\pi_0} \times \mathcal{X} \to [0, \infty)$ such that

$$Q(\nu, B) = \int_B \rho(\nu, x) \, \mathrm{d}\pi_0(x)$$

for all $\nu \in \mathcal{P}^{\pi_0}$ and every Borel set $B \subseteq \mathcal{X}$. Taking $\phi^{\pi_0} = \rho$ gives the claim. □

Proof of Theorem 2. Let $(\mathcal{A}, \phi) = (\mathcal{P}^{\pi_0}, \phi^{\pi_0})$ as in Lemma 2. The proof proceeds by first establishing $\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] \leq \sup_{U^\infty \text{ e-process}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{U^\infty})]$ and then $\sup_{U^\infty \text{ e-process}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{U^\infty})] \leq \sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})]$. Taken together, this proves the first equality of Theorem 2. The second equality follows from Theorem 1.

We first show that every policy in the canonical testing-by-betting model induces an e-process. Fix a policy A^∞. Under $\pi_0^\otimes$, the associated wealth process satisfies

$$\widehat{W}_t^{A^\infty} = \widehat{W}_{t-1}^{A^\infty} \phi^{\pi_0}(A_t(X^{1:t-1}), X_t) \,,$$

with $\widehat{W}_0^{A^\infty} = 1$. Since $A_t(X^{1:t-1})$ is $\sigma(X^{1:t-1})$-measurable and

$$\int_{\mathcal{X}} \phi^{\pi_0}(\nu, x) \, \mathrm{d}\pi_0(x) = 1$$

for every $\nu \in \mathcal{P}^{\pi_0}$, we have

$$\mathbb{E}_{\pi_0^\otimes}\left[\widehat{W}_t^{A^\infty} \mid X^{1:t-1}\right] = \widehat{W}_{t-1}^{A^\infty} \,.$$

Thus $\widehat{W}^{A^\infty}$ is a non-negative $\pi_0^\otimes$-martingale with initial value one, and hence an e-process. Therefore,

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{A^\infty})] \leq \sup_{U^\infty \text{ e-process}} \mathbb{E}_{\pi_1^\otimes}[R(\tau^{U^\infty})] \,. \tag{6}$$

We now prove the reverse inequality. Fix an arbitrary e-process $U^\infty = (U_t)_{t \geq 0}$ under the point null $\pi_0^\otimes$. By [Ramdas et al., 2020, Lemma 6], there exists a non-negative $\pi_0^\otimes$-martingale $M^\infty = (M_t)_{t \geq 0}$, which we may take to satisfy $M_0 = 1$, such that $U_t \leq M_t$ for every $t \geq 0$, $\pi_0^\otimes$-almost surely. Since $\pi_1 \ll \pi_0$, we have $\pi_1^{\otimes t} \ll \pi_0^{\otimes t}$ for every finite t. Hence, intersecting over countably many times, $U_t \leq M_t$ for every $t \geq 0$, $\pi_1^\otimes$-almost surely. It follows that $\tau^{M^\infty} \leq \tau^{U^\infty}$, $\pi_1^\otimes$-almost surely. Since R is non-increasing,

$$\mathbb{E}_{\pi_1^\otimes}[R(\tau^{M^\infty})] \geq \mathbb{E}_{\pi_1^\otimes}[R(\tau^{U^\infty})] \,.$$

It remains to represent M^∞ as a canonical wealth process. Since M_t is $\sigma(X^{1:t})$-measurable and $\mathcal{X}^t$ is standard Borel, the Doob–Dynkin lemma gives Borel maps $m_t : \mathcal{X}^t \to \mathbb{R}_+$ such that $M_t = m_t(X^{1:t})$, $\pi_0^\otimes$-almost surely. We take $m_0 \equiv 1$. For $t \geq 1$ and $x^{1:t-1} \in \mathcal{X}^{t-1}$, define

$$b_t(x^{1:t-1}) = \int_{\mathcal{X}} m_t(x^{1:t-1}, x) \, \mathrm{d}\pi_0(x) \,.$$

By the martingale property, $b_t(X^{1:t-1}) = m_{t-1}(X^{1:t-1})$, $\pi_0^{\otimes(t-1)}$-almost surely. Let

$$N_t = \{x^{1:t-1} \in \mathcal{X}^{t-1} : b_t(x^{1:t-1}) \neq m_{t-1}(x^{1:t-1})\} \,.$$

Then N_t is Borel and $\pi_0^{\otimes(t-1)}(N_t) = 0$. Replacing m_t by

$$\widetilde{m}_t(x^{1:t-1}, x) = m_t(x^{1:t-1}, x) \mathbf{1}_{N_t^c}(x^{1:t-1}) + m_{t-1}(x^{1:t-1}) \mathbf{1}_{N_t}(x^{1:t-1}) \,,$$

[3]The result in Kechris [1995] is stated for Polish spaces. We apply it to a Polish topology on $\mathcal{X}$ generating the given Borel σ-field. The resulting Borel σ-field on $\mathcal{P}_{\mathcal{X}}$ is the one generated by the evaluation maps.

we still have $M_t = \widetilde{m}_t(X^{1:t})$, $\pi_0^{\otimes}$-almost surely, and now

$$\int_{\mathcal{X}} \widetilde{m}_t(x^{1:t-1}, x)\,\mathrm{d}\pi_0(x) = m_{t-1}(x^{1:t-1})$$

for every $x^{1:t-1} \in \mathcal{X}^{t-1}$.

Define

$$e_t(x^{1:t-1}, x) = \begin{cases} \widetilde{m}_t(x^{1:t-1}, x)/m_{t-1}(x^{1:t-1}) & \text{if } m_{t-1}(x^{1:t-1}) > 0\,, \\ 1 & \text{if } m_{t-1}(x^{1:t-1}) = 0\,. \end{cases}$$

Then e_t is Borel, non-negative, and satisfies

$$\int_{\mathcal{X}} e_t(x^{1:t-1}, x)\,\mathrm{d}\pi_0(x) = 1$$

for every $x^{1:t-1} \in \mathcal{X}^{t-1}$. Hence, for each $x^{1:t-1} \in \mathcal{X}^{t-1}$, the formula

$$\nu_t(x^{1:t-1}, B) = \int_B e_t(x^{1:t-1}, x)\,\mathrm{d}\pi_0(x)$$

defines a probability measure $\nu_t(x^{1:t-1}, \cdot\,) \in \mathcal{P}^{\pi_0}$. Moreover, since $x^{1:t-1} \mapsto \nu_t(x^{1:t-1}, B)$ is Borel for every Borel set $B \subseteq \mathcal{X}$, the map $x^{1:t-1} \mapsto \nu_t(x^{1:t-1}, \cdot\,)$ is Borel from $\mathcal{X}^{t-1}$ to $\mathcal{P}^{\pi_0}$.

Now define a policy A^∞ by setting $A_t(x^{1:t-1}) = \nu_t(x^{1:t-1}, \cdot\,)$. By the defining property of ϕ^{π_0} in Lemma 2, for each $x^{1:t-1} \in \mathcal{X}^{t-1}$ the map $x \mapsto \phi^{\pi_0}(A_t(x^{1:t-1}), x)$ is a version of $e_t(x^{1:t-1}, \cdot\,)$, since $e_t(x^{1:t-1}, \cdot\,)$ is a version of $\mathrm{d}\nu_t(x^{1:t-1}, \cdot\,)/\mathrm{d}\pi_0$. Hence $\phi^{\pi_0}(A_t(x^{1:t-1}), x) = e_t(x^{1:t-1}, x)$ for π_0-almost every x, and therefore also for π_1-almost every x, since $\pi_1 \ll \pi_0$. By induction over t, we get

$$\widehat{W}_t^{A^\infty} = M_t$$

for every $t \geq 0$, $\pi_1^{\otimes}$-almost surely. Consequently, $\tau^{A^\infty} = \tau^{M^\infty}$, $\pi_1^{\otimes}$-almost surely, and so

$$\mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{A^\infty})] = \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{M^\infty})] \geq \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{U^\infty})]\,.$$

Taking the supremum over e-processes gives

$$\sup_{U^\infty \text{ e-process}} \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{U^\infty})] \leq \sup_{A^\infty} \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{A^\infty})]\,.$$

Combining this with (6) yields

$$\sup_{U^\infty \text{ e-process}} \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{U^\infty})] = \sup_{A^\infty} \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{A^\infty})]\,.$$

Finally, applying Theorem 1 to the canonical action space $(\mathcal{P}^{\pi_0}, \phi^{\pi_0})$ gives

$$\sup_{A^\infty} \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{A^\infty})] = \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^{\otimes}}[R(\tau^{a^\infty})]\,,$$

and the claim follows. □

B.3 Proof of Proposition 2

We provide here a proof of Proposition 2, which we restate here.

Proposition. *Consider testing $\pi_0^{\otimes}$ against $\pi_1^{\otimes}$, with $\pi_1 \ll \pi_0$. Then there exists a Borel set $\Gamma^\star \subseteq \mathcal{X}^T$, viewed as an event depending only on the first T observations, such that $\pi_0^{\otimes}(\Gamma^\star) \leq \alpha$ and*

$$\sup_{U^\infty \text{ e-process}} \pi_1^{\otimes}\big(\tau^{U^\infty} \leq T\big) = \pi_1^{\otimes}(\Gamma^\star) = \max\big\{\pi_1^{\otimes}(\Gamma) \,:\, \Gamma \subseteq \mathcal{X}^T \text{ Borel}\,,\, \pi_0^{\otimes}(\Gamma) \leq \alpha\big\}\,.$$

Fixing Borel versions of the conditional probabilities, the supremum is achieved by the $\pi_0^{\otimes}$-martingale $U_t^\star(x^{1:t}) = \pi_0^{\otimes}(\Gamma^\star \mid x^{1:t})/\pi_0^{\otimes}(\Gamma^\star)$ if $\pi_0^{\otimes}(\Gamma^\star) > 0$, and by the constant e-process $U_t \equiv 1$ otherwise.

Proof. First, we show that there exists $\Gamma^\star$ as required by the statement. Define the finite $\mathbb{R}^2$-valued measure μ on $\mathcal{X}^T$, given by

$$\mu(\Gamma) = \left(\pi_0^{\otimes T}(\Gamma), \pi_1^{\otimes T}(\Gamma)\right).$$

Its range is defined as

$$\mathcal{R}_\mu = \{\mu(\Gamma) \,:\, \Gamma \subseteq \mathcal{X}^T \text{ Borel}\}.$$

By Liapounoff's vector-measure theorem [Liapounoff, 1940, Theorem 2][4], $\mathcal{R}_\mu$ is closed in $\mathbb{R}^2$, and hence compact since it is bounded. In particular, the set

$$\mathcal{C} = \mathcal{R}_\mu \cap \left((-\infty, \alpha] \times \mathbb{R}\right)$$

is compact and non-empty (as $(0,0) \in \mathcal{C}$). Hence, there exists $(p^\star, q^\star) \in \mathcal{C}$ such that

$$q^\star = \max\{q \,:\, (p,q) \in \mathcal{C}\}.$$

Then, there is a Borel $\Gamma^\star \in \mathcal{X}^T$ such that $\mu(\Gamma^\star) = (p^\star, q^\star)$. In particular, $\pi_0^{\otimes T}(\Gamma^\star) \leq \alpha$ and

$$\pi_1^{\otimes T}(\Gamma^\star) = \max\{\pi_1^{\otimes T}(\Gamma) \,:\, \Gamma \subseteq \mathcal{X}^T \text{ Borel}\,,\, \pi_0^{\otimes T}(\Gamma) \leq \alpha\}.$$

We now show that this set gives the optimal hard-deadline value. Let U^∞ be any e-process under $\pi_0^\otimes$, and define

$$\Gamma_U = \left\{x^{1:T} \in \mathcal{X}^T \,:\, \max_{0 \leq t \leq T} U_t(x^{1:t}) \geq \alpha^{-1}\right\}.$$

Then Γ_U is Borel and $\{\tau^{U^\infty} \leq T\} = \Gamma_U$. For the bounded stopping time $\tau^{U^\infty} \wedge T$, the e-process property gives $\mathbb{E}_{\pi_0^\otimes}\left[U_{\tau^{U^\infty} \wedge T}\right] \leq 1$. On the event $\{\tau^{U^\infty} \leq T\}$ we have $U_{\tau^{U^\infty} \wedge T} \geq \alpha^{-1}$, and therefore

$$\pi_0^{\otimes T}(\Gamma_U) = \pi_0^\otimes(\tau^{U^\infty} \leq T) \leq \pi_0^{\otimes T}(U_{\tau^{U^\infty} \wedge T} \geq \alpha^{-1}) \leq \alpha\,,$$

by Markov's inequality. By the definition of $\Gamma^\star$, it follows that $\pi_1^\otimes(\tau^{U^\infty} \leq T) = \pi_1^{\otimes T}(\Gamma_U) \leq \pi_1^{\otimes T}(\Gamma^\star)$. Taking the supremum over all e-processes gives

$$\sup_{U^\infty \text{ e-process}} \pi_1^\otimes(\tau^{U^\infty} \leq T) \leq \pi_1^{\otimes T}(\Gamma^\star).$$

It remains to show that this upper bound is attained. Suppose first that $\pi_0^{\otimes T}(\Gamma^\star) > 0$. Fix Borel versions of the conditional probabilities

$$x^{1:t} \mapsto \pi_0^\otimes(\Gamma^\star \mid X^{1:t} = x^{1:t})$$

for $0 \leq t \leq T$, and define

$$U_t^\star(x^{1:t}) = \frac{\pi_0^\otimes(\Gamma^\star \mid X^{1:t} = x^{1:t})}{\pi_0^{\otimes T}(\Gamma^\star)}$$

for $0 \leq t \leq T$. For $t > T$, set $U_t^\star(x^{1:t}) = U_T^\star(x^{1:T})$. Then $U^{\star,\infty}$ is a non-negative $\pi_0^\otimes$-martingale with initial value one, and hence an e-process. Moreover, for $x^{1:T} \in \Gamma^\star$,

$$U_T^\star(x^{1:T}) = \frac{1}{\pi_0^{\otimes T}(\Gamma^\star)} \geq \alpha^{-1}\,,$$

because $\pi_0^{\otimes T}(\Gamma^\star) \leq \alpha$. Hence $\Gamma^\star \subseteq \{x^{1:T} : \tau^{U^{\star,\infty}}(x^{1:\infty}) \leq T\}$, and so

$$\pi_1^{\otimes T}(\Gamma^\star) \leq \pi_1^\otimes(\tau^{U^{\star,\infty}} \leq T).$$

Combining this with the upper bound already proved gives

$$\pi_1^\otimes(\tau^{U^{\star,\infty}} \leq T) = \pi_1^{\otimes T}(\Gamma^\star).$$

Finally, if $\pi_0^{\otimes T}(\Gamma^\star) = 0$, then $\pi_1^{\otimes T}(\Gamma^\star) = 0$ since $\pi_1 \ll \pi_0$. In this case the constant process $U_t^\star \equiv 1$ is an e-process and satisfies

$$\pi_1^\otimes(\tau^{U^{\star,\infty}} \leq T) = 0 = \pi_1^{\otimes T}(\Gamma^\star).$$

This proves the claim. □

[4] Paper in Russian, with French summary at the end.

B.4 Proof of Proposition 3

Here we prove Proposition 3, which we restate for convenience.

Proposition. *Let $(a_\star, \eta_\star)$ satisfy (4), with $\eta_\star \in (0,1)$ and $\phi(a_\star, \cdot) \le \Phi$, π_1-almost surely. Then*

$$(\alpha/\Phi)^{\eta_\star} \le \mathbb{E}_{\pi_1^\otimes}[e^{-\tau^{a_\star}/T}] \le \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}[e^{-\tau^{a^\infty}/T}] \le \alpha^{\eta_\star}\,.$$

We recall that (4), mentioned in the statement, reads

$$\sup_{a\in\mathcal{A}} \mathbb{E}_{\pi_1}\big[\phi(a,\cdot)^{\eta_\star}\big] = \mathbb{E}_{\pi_1}\big[\phi(a_\star,\cdot)^{\eta_\star}\big] = e^{1/T}\,.$$

Proof. Fix a Markov policy a^∞, let $\widetilde{W}_t^{a^\infty}$ be the wealth process starting from wealth one, and write

$$Z_t^{a^\infty} = \alpha \widetilde{W}_t^{a^\infty}\,,$$

so that rejection occurs when $Z_t^{a^\infty} \ge 1$. Define

$$M_t = e^{-(t\wedge\tau^{a^\infty})/T}\big(Z^{a^\infty}_{t\wedge\tau^{a^\infty}}\big)^{\eta^\star}\,.$$

We first show that $(M_t)_{t\ge 0}$ is a non-negative supermartingale under $\pi_1^\otimes$. On the event $\{\tau^{a^\infty} \le t\}$, we have $M_{t+1} = M_t$. On the event $\{\tau^{a^\infty} > t\}$, the next action is $\sigma(X^{1:t})$-measurable. Hence, using independence under $\pi_1^\otimes$ and the defining property of $(a^\star, \eta^\star)$,

$$\begin{aligned}
\mathbb{E}_{\pi_1^\otimes}\left[M_{t+1} \mid X^{1:t}\right] &= \mathbb{E}_{\pi_1^\otimes}\left[e^{-(t+1)/T}(Z_t^{a^\infty})^{\eta^\star}\phi(a_t, X_{t+1})^{\eta^\star} \mid X^{1:t}\right] \\
&\le e^{-(t+1)/T}\big(Z_t^{a^\infty}\big)^{\eta^\star} \sup_{a\in\mathcal{A}} \mathbb{E}_{\pi_1}\left[\phi(a, X_1)^{\eta^\star}\right] \\
&= e^{-t/T}\big(Z_t^{a^\infty}\big)^{\eta^\star} = M_t\,.
\end{aligned}$$

Thus $(M_t)_{t\ge 0}$ is a non-negative supermartingale. Therefore, for every $n \ge 1$, $\mathbb{E}_{\pi_1^\otimes}[M_n] \le M_0 = \alpha^{\eta^\star}$. On the event $\{\tau^{a^\infty} \le n\}$, we have $M_n = e^{-\tau^{a^\infty}/T}(Z^{a^\infty}_{\tau^{a^\infty}})^{\eta^\star}$ and $Z^{a^\infty}_{\tau^{a^\infty}} \ge 1$, so

$$e^{-\tau^{a^\infty}/T}\mathbf{1}_{\{\tau^{a^\infty}\le n\}} \le M_n\,.$$

It follows that $\mathbb{E}_{\pi_1^\otimes}[e^{-\tau^{a^\infty}/T}\mathbf{1}_{\{\tau^{a^\infty}\le n\}}] \le \alpha^{\eta^\star}$. Sending $n\to\infty$ and using monotone convergence gives

$$\mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\infty}/T}\right] \le \alpha^{\eta^\star}\,.$$

Taking the supremum over Markov policies yields

$$\sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\infty}/T}\right] \le \alpha^{\eta^\star}\,.$$

We now consider the constant policy that always plays $a^\star$. Let $\widetilde{W}_t^{a^\star}$ be the corresponding wealth process starting from wealth one, and write $Z_t^{a^\star} = \alpha\widetilde{W}_t^{a^\star}$. Define $N_t = e^{-t/T}(Z_t^{a^\star})^{\eta^\star}$. Then $(N_t)_{t\ge 0}$ is a non-negative martingale under $\pi_1^\otimes$, because

$$\mathbb{E}_{\pi_1^\otimes}\left[N_{t+1} \mid X^{1:t}\right] = e^{-(t+1)/T}\big(Z_t^{a^\star}\big)^{\eta^\star}\mathbb{E}_{\pi_1}\left[\phi(a^\star, X_1)^{\eta^\star}\right] = N_t\,.$$

On the event $\{\tau^{a^\star} < \infty\}$, we have $Z^{a^\star}_{\tau^{a^\star}-1} < 1$ and $Z^{a^\star}_{\tau^{a^\star}} \ge 1$. Since $\phi(a^\star,\cdot) \le \Phi$, π_1-almost surely, this implies $1 \le Z^{a^\star}_{\tau^{a^\star}} \le \Phi$, $\pi_1^\otimes$-almost surely on $\{\tau^{a^\star} < \infty\}$. On the event $\{\tau^{a^\star} > n\}$, instead, $Z_n^{a^\star} < 1$. Hence $0 \le Z^{a^\star}_{\tau^{a^\star}\wedge n} \le \Phi$, $\pi_1^\otimes$-almost surely.

Because $N_0 = \alpha^{\eta^\star}$, by optional stopping [Williams, 1991, Theorem 10.10] at the bounded stopping time $\tau^{a^\star}\wedge n$,

$$\alpha^{\eta^\star} = \mathbb{E}_{\pi_1^\otimes}[N_0] = \mathbb{E}_{\pi_1^\otimes}[N_{\tau^{a^\star}\wedge n}] = \mathbb{E}_{\pi_1^\otimes}\left[e^{-(\tau^{a^\star}\wedge n)/T}\big(Z^{a^\star}_{\tau^{a^\star}\wedge n}\big)^{\eta^\star}\right]\,.$$

Sending $n \to \infty$ and using dominated convergence yields

$$\alpha^{\eta^\star} = \mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\star}/T}\big(Z^{a^\star}_{\tau^{a^\star}}\big)^{\eta^\star}\mathbf{1}_{\{\tau^{a^\star}<\infty\}}\right].$$

Since $Z^{a^\star}_{\tau^{a^\star}} \le \Phi$ on $\{\tau^{a^\star} < \infty\}$, we get

$$\alpha^{\eta^\star} \le \Phi^{\eta^\star}\mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\star}/T}\right].$$

Therefore,

$$\left(\frac{\alpha}{\Phi}\right)^{\eta^\star} \le \mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\star}/T}\right].$$

Since the constant policy is Markov, the middle inequality is immediate. Combining the three bounds gives

$$(\alpha/\Phi)^{\eta^\star} \le \mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\star}/T}\right] \le \sup_{a^\infty \text{ Markov}} \mathbb{E}_{\pi_1^\otimes}\left[e^{-\tau^{a^\infty}/T}\right] \le \alpha^{\eta^\star},$$

as claimed. □

B.5 Proof of Proposition 4

Here we prove Proposition 4, restated below.

Proposition. *Let* $G = \{(1-\eta)\log\mathbb{E}_{\pi_0}[L_\eta] : \eta \in [0,1), \mathbb{E}_{\pi_0}[L_\eta] < \infty\}$. *If* $1/T \in (0, \sup G)$, *then there exist* $\eta_\star \in (0,1)$ *and* $\pi_\star \ll \pi_0$ *satisfying* (4). *Moreover,* $\mathrm{d}\pi_\star/\mathrm{d}\pi_0 = L_{\eta_\star}/\mathbb{E}_{\pi_0}[L_{\eta_\star}]$ *and* $\log(\sup_{\pi\ll\pi_0}\mathbb{E}_{\pi_1}[(\mathrm{d}\pi/\mathrm{d}\pi_0)^{\eta_\star}]) = \eta^\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = 1/T$.

We recall that (4), mentioned in the statement, reads

$$\sup_{a\in\mathcal{A}}\mathbb{E}_{\pi_1}\left[\phi(a,\,\cdot\,)^{\eta_\star}\right] = \mathbb{E}_{\pi_1}\left[\phi(a_\star,\,\cdot\,)^{\eta_\star}\right] = e^{1/T}.$$

Proof. Let $L = L_0 = \mathrm{d}\pi_1/\mathrm{d}\pi_0$. We first show that there exists $\eta_\star \in (0,1)$ such that

$$(1-\eta_\star)\log\mathbb{E}_{\pi_0}[L_{\eta_\star}] = 1/T.$$

Indeed, since $1/T \in (0, \sup G)$, there exists $\bar\eta \in (0,1)$ such that $\mathbb{E}_{\pi_0}[L_{\bar\eta}] < \infty$ and

$$(1-\bar\eta)\log\mathbb{E}_{\pi_0}[L_{\bar\eta}] > 1/T.$$

The map $\eta \mapsto (1-\eta)\log\mathbb{E}_{\pi_0}[L_\eta]$ is finite and continuous on $[0,\bar\eta]$. Finiteness follows by interpolation, since $L_\eta = L^{1/(1-\eta)}$ and $1/(1-\eta) \le 1/(1-\bar\eta)$ for $\eta \le \bar\eta$. Specifically, let $q = 1/(1-\eta)$ and $\bar q = 1/(1-\bar\eta)$. Then for $\eta \in [0,\bar\eta]$ there is a $\theta \in [0,1]$ such that $q = (1-\theta) + \theta\bar q$, and by Hölder's inequality, we have $\mathbb{E}_{\pi_0}[L_\eta] = \mathbb{E}_{\pi_0}[L^q] = \mathbb{E}_{\pi_0}[L^{1-\theta}(L^{\bar q})^\theta] \le \mathbb{E}_{\pi_0}[L]^{1-\theta}\mathbb{E}_{\pi_0}[L^{\bar q}]^\theta < \infty$. Continuity follows from the observation that $(1-\eta)\log\mathbb{E}_{\pi_0}[L_\eta] = \log\|L\|_{1/(1-\eta)}$ and the continuity of $q \mapsto \|L\|_q$ on finite intervals where the largest moment is finite (this can be shown for instance from dominated convergence). Since the same map is equal to zero at $\eta = 0$, the intermediate value theorem gives $\eta_\star \in (0,\bar\eta)$ such that

$$(1-\eta_\star)\log\mathbb{E}_{\pi_0}[L_{\eta_\star}] = 1/T.$$

We now identify the maximiser in (4). Fix $\eta \in (0,1)$ such that $\mathbb{E}_{\pi_0}[L_\eta] < \infty$, and write $p = 1/\eta$ and $q = 1/(1-\eta)$. Then $1/p + 1/q = 1$. Let $\pi \ll \pi_0$ and set $f = \mathrm{d}\pi/\mathrm{d}\pi_0$. By Hölder's inequality,

$$\mathbb{E}_{\pi_1}[f^\eta] = \mathbb{E}_{\pi_0}[Lf^\eta] \le \mathbb{E}_{\pi_0}[L^q]^{1/q}\mathbb{E}_{\pi_0}[f]^\eta = \mathbb{E}_{\pi_0}[L_\eta]^{1-\eta}.$$

Thus

$$\sup_{\pi\ll\pi_0}\mathbb{E}_{\pi_1}\left[(\mathrm{d}\pi/\mathrm{d}\pi_0)^\eta\right] \le \mathbb{E}_{\pi_0}[L_\eta]^{1-\eta}.$$

The upper bound is attained by the probability measure $\pi_\eta \ll \pi_0$ defined by

$$\frac{\mathrm{d}\pi_\eta}{\mathrm{d}\pi_0} = \frac{L_\eta}{\mathbb{E}_{\pi_0}[L_\eta]}.$$

Indeed, for this choice,

$$\mathbb{E}_{\pi_1}\left[\left(\frac{\mathrm{d}\pi_\eta}{\mathrm{d}\pi_0}\right)^\eta\right] = \mathbb{E}_{\pi_0}\left[L\left(\frac{L^q}{\mathbb{E}_{\pi_0}[L^q]}\right)^\eta\right] = \mathbb{E}_{\pi_0}[L^q]^{1-\eta} = \mathbb{E}_{\pi_0}[L_\eta]^{1-\eta}\,,$$

where we used $1 + q\eta = q$. Applying this with $\eta = \eta_\star$ and setting $\pi_\star = \pi_{\eta_\star}$ gives

$$\frac{\mathrm{d}\pi_\star}{\mathrm{d}\pi_0} = \frac{L_{\eta_\star}}{\mathbb{E}_{\pi_0}[L_{\eta_\star}]}\,.$$

Moreover,

$$\log\left(\sup_{\pi \ll \pi_0} \mathbb{E}_{\pi_1}\left[(\mathrm{d}\pi/\mathrm{d}\pi_0)^{\eta_\star}\right]\right) = (1-\eta_\star)\log \mathbb{E}_{\pi_0}[L_{\eta_\star}] = 1/T\,.$$

Finally, since

$$\eta_\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = (1-\eta_\star)\log \mathbb{E}_{\pi_0}[L_{\eta_\star}]\,,$$

we obtain

$$\log\left(\sup_{\pi \ll \pi_0} \mathbb{E}_{\pi_1}\left[(\mathrm{d}\pi/\mathrm{d}\pi_0)^{\eta_\star}\right]\right) = \eta_\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = 1/T\,.$$

Thus $\pi_\star$ attains the supremum in (4), and (4) holds. □

B.6 Proof of Proposition 5

We prove Proposition 5, which we now recall.

Proposition. *Fix $a \in \mathcal{A}$, assume that $\phi(a,\cdot) > 0$, π_1-almost surely, and that $\gamma(a) = \mathbb{E}_{\pi_1}[\log\phi(a,\cdot)] \in [-\infty,\infty]$ is well-defined. Then, $\pi_1^\otimes(\tau_w^a < \infty) = 1$ for every $w \in (0,\alpha^{-1})$ if and only if either $\gamma(a) > 0$, or $\gamma(a) = 0$ and $\pi_1(\log\phi(a,\cdot) > 0) > 0$. Moreover, when $\gamma(a) < 0$, if $\phi(a,\cdot) \le \Phi$, π_1-almost surely, and there exists $\kappa > 0$ such that $\mathbb{E}_{\pi_1}[\phi(a,\cdot)^\kappa] = 1$, then $(\alpha w/\Phi)^\kappa \le \pi_1^\otimes(\tau_w^a < \infty) \le (\alpha w)^\kappa$ for every $w \in (0,\alpha^{-1})$.*

Proof. Write $Y_i = \log\phi(a, X_i)$ and $S_t = \sum_{i=1}^t Y_i$. Under $\pi_1^\otimes$, the variables $(Y_i)_{i\ge1}$ are i.i.d. with well-defined mean $\gamma(a)$. For initial wealth $w \in (0,\alpha^{-1})$, the constant policy that always plays a has wealth

$$\widetilde{W}_t^{w,a} = w\exp(S_t)\,,$$

and therefore

$$\tau_w^a = \inf\{t \ge 1 : S_t \ge \log(\alpha^{-1}/w)\}\,.$$

Set $c_w = \log(\alpha^{-1}/w) > 0$.

Suppose first that $\gamma(a) > 0$. By the strong law of large numbers, $S_t \to +\infty$, $\pi_1^\otimes$-almost surely. Hence S_t hits the level c_w in finite time almost surely, and so $\pi_1^\otimes(\tau_w^a < \infty) = 1$. If instead $\gamma(a) = 0$ and $\pi_1(Y_1 > 0) > 0$, then $(S_t)_{t\ge0}$ is a centred non-degenerate random walk. Since its increments have mean zero, it cannot drift to $+\infty$ or $-\infty$. By Feller's oscillation theorem for one-dimensional random walks [Feller, 1971, Chapter XII, Section 2, Theorem 1], $\limsup_t S_t = +\infty$ almost surely. Hence S_t hits the level c_w in finite time almost surely, and again $\pi_1^\otimes(\tau_w^a < \infty) = 1$.

We now prove the converse. If $\gamma(a) = 0$ and $\pi_1(Y_1 > 0) = 0$, then $Y_1 = 0$, π_1-almost surely, so $S_t = 0$ for every $t \ge 0$ and the positive level c_w is never reached. If $\gamma(a) < 0$, then the strong law of large numbers gives $S_t \to -\infty$, $\pi_1^\otimes$-almost surely. In particular, $\sup_t S_t < \infty$ almost surely. We claim that $\pi_1^\otimes(\sup_t S_t < c_w) > 0$. Indeed, choose $\varepsilon > 0$ such that $\pi_1(Y_1 \le -\varepsilon) > 0$. Since an independent copy of the random walk has finite supremum almost surely, we may choose m large enough that this supremum is smaller than $c_w + m\varepsilon$ with positive probability. On the event that the first m increments are all at most $-\varepsilon$ and the future supremum is smaller than $c_w + m\varepsilon$, the level c_w is never hit. Thus $\pi_1^\otimes(\tau_w^a < \infty) < 1$. This proves the first part of the proposition.

We now assume that $\gamma(a) < 0$, that $\phi(a,\cdot) \le \Phi$, π_1-almost surely, and that there exists $\kappa > 0$ such that $\mathbb{E}_{\pi_1}[\phi(a,\cdot)^\kappa] = 1$. Then $(\widetilde{W}_t^{w,a})^\kappa$ is a non-negative $\pi_1^\otimes$-martingale with initial value w^κ. Optional stopping [Williams, 1991, Theorem 10.10] at the bounded stopping time $\tau_w^a \wedge n$ gives

$$w^\kappa = \mathbb{E}_{\pi_1^\otimes}\left[(\widetilde{W}^{w,a}_{\tau_w^a\wedge n})^\kappa\right] \ge \alpha^{-\kappa}\pi_1^\otimes(\tau_w^a \le n)\,.$$

Sending $n \to \infty$ yields

$$\pi_1^{\otimes}(\tau_w^a < \infty) \leq (\alpha w)^{\kappa}\,.$$

It remains to prove the lower bound. If $\Phi = \infty$, the bound is trivial, so assume $\Phi < \infty$. Fix $b \in (0, w)$ and define $\sigma_b = \inf\{t \geq 0 : \widetilde{W}_t^{w,a} \leq b\}$. Since $\gamma(a) < 0$, we have $\widetilde{W}_t^{w,a} \to 0$, $\pi_1^{\otimes}$-almost surely. Hence $\sigma_b < \infty$ almost surely for every $b \in (0, w)$, and $\sigma_b \to \infty$ almost surely as $b \downarrow 0$. Optional stopping at $\tau_w^a \wedge \sigma_b \wedge n$ gives

$$w^{\kappa} = \mathbb{E}_{\pi_1^{\otimes}}\left[(\widetilde{W}^{w,a}_{\tau_w^a \wedge \sigma_b \wedge n})^{\kappa}\right]\,.$$

Letting $n \to \infty$ and splitting according to whether $\tau_w^a < \sigma_b$ or $\sigma_b < \tau_w^a$, we obtain

$$w^{\kappa} \leq \left(\frac{\Phi}{\alpha}\right)^{\kappa} \pi_1^{\otimes}(\tau_w^a < \sigma_b) + b^{\kappa}\,.$$

Here we used that, on $\{\tau_w^a < \infty\}$, the pre-rejection wealth is below α^{-1}, and therefore the rejection wealth is at most $\Phi\alpha^{-1}$. Sending $b \downarrow 0$ gives

$$w^{\kappa} \leq \left(\frac{\Phi}{\alpha}\right)^{\kappa} \pi_1^{\otimes}(\tau_w^a < \infty)\,,$$

and hence

$$(\alpha w/\Phi)^{\kappa} \leq \pi_1^{\otimes}(\tau_w^a < \infty)\,.$$

Combining the two bounds proves the claim. □

C Details for the examples

C.1 Gaussian example

Let $\pi_0 = \mathcal{N}(0, \sigma^2)$ and $\pi_1 = \mathcal{N}(\mu, \sigma^2)$, with $\mu > 0$ and $\sigma > 0$. We consider sequential tests for $\pi_0^{\otimes}$ against $\pi_1^{\otimes}$. We let $\mathcal{A}$ and ϕ be as in Section 5.

Hard deadline. Fix an integer $T \geq 1$ and consider the reward $R = \mathbf{1}_{[0,T]}$. By the Neyman–Pearson lemma, the optimal fixed-horizon rejection event is obtained by spending the available null mass on the largest values of the likelihood ratio. Thus the event $\Gamma^\star$ from Proposition 2 can be chosen in the form $\Gamma^\star = \{\mathrm{d}\pi_1^{\otimes}/\mathrm{d}\pi_0^{\otimes} \geq C\}$, where C is chosen so that $\pi_0^{\otimes}(\mathrm{d}\pi_1^{\otimes T}/\mathrm{d}\pi_0^{\otimes T} \geq C) = \alpha$.

We let $S_t(x^{1:t}) = \sum_{k=1}^t x_k$ and notice that

$$\frac{\mathrm{d}\pi_1^{\otimes T}}{\mathrm{d}\pi_0^{\otimes T}}(x^{1:T}) = \exp\left(\frac{\mu}{\sigma^2} S_T(x^{1:T}) - \frac{T\mu^2}{2\sigma^2}\right).$$

Since $\mu > 0$, this expression is strictly increasing in $S_T(x^{1:T})$. In particular, we can write $\Gamma^\star = \{S_T \geq c\}$, where c satisfies $\pi_0^{\otimes}(S_T \geq c) = \alpha$. Since $S_T \sim \mathcal{N}(0, T\sigma^2)$ under $\pi_0^{\otimes}$, we have

$$c = \sigma\sqrt{T}\, F^{-1}(1-\alpha),$$

where F is the cumulative distribution function of the standard normal.

The optimal e-process from Proposition 2 can be written as

$$U_t^\star(x^{1:t}) = \frac{\pi_0^{\otimes}(S_T \geq c \mid S_t)}{\alpha}.$$

Since under $\pi_0^{\otimes}$ we have $\Delta_{T,t} = S_T - S_t \sim \mathcal{N}(0, \sigma^2(T-t))$, independently of S_t, we obtain

$$U_t^\star(x^{1:t}) = \frac{h_t(S_t(x^{1:t}))}{\alpha},$$

where, for $t < T$,

$$h_t(s) = 1 - F\left(\frac{c-s}{\sigma\sqrt{T-t}}\right),$$

with terminal convention $h_T(s) = \mathbf{1}_{[c,\infty)}(s)$.

Notice that, for every $t < T$ and every $s \in \mathbb{R}$, the Gaussian increment $S_T - S_t$ has full support under $\pi_0^{\otimes}$. Hence

$$0 < h_t(s) < 1.$$

Consequently, for every $t < T$,

$$U_t^\star(x^{1:t}) = \frac{h_t(S_t(x^{1:t}))}{\alpha} < \alpha^{-1}.$$

Thus the exact hard-deadline Doob test does not reject before the deadline. It rejects at time T precisely on the event $\Gamma^\star = \{S_T \geq c\}$, since

$$U_T^\star(x^{1:T}) = \alpha^{-1}\mathbf{1}_{\{S_T \geq c\}}.$$

Following the discussion after Proposition 2, we now describe the canonical predictable *truncated* strategy A^T whose wealth process is $U^{\star,T}$ (for $t \geq T$, any continuation of the policy and wealth process are equivalent for the hard deadline objective). For $t = 1, \ldots, T$ and $x^{1:t-1} \in \mathbb{R}^{t-1}$, the action $A_t(x^{1:t-1})$ is the measure

$$A_t(x^{1:t-1})(\mathrm{d}x) = \frac{h_t(S_{t-1}(x^{1:t-1}) + x)}{h_{t-1}(S_{t-1}(x^{1:t-1}))}\,\mathrm{d}\pi_0(x).$$

Remarkably, although this strategy is written as a function of the whole past, it corresponds to a Markovian strategy, since it can be made to depend on the current wealth only. Indeed, the wealth at round $t-1$ is precisely

$$U_{t-1}^\star(x^{1:t-1}) = \frac{h_{t-1}(S_{t-1}(x^{1:t-1}))}{\alpha}.$$

In particular, we can write $W_{t-1} = w_{t-1}(S_{t-1}(x^{1:t-1}))$, where $w_{t-1}(s) = h_{t-1}(s)/\alpha$. Since h_{t-1} is strictly increasing, we can invert this relation. Let $s_{t-1} = w_{t-1}^{-1}$. Then

$$S_{t-1}(x^{1:t-1}) = s_{t-1}(W_{t-1})\,.$$

Since A_t depends on $x^{1:t-1}$ only through S_{t-1}, the strategy depends only on current wealth. More precisely, the equivalent Markovian policy is a^T, where for $t = 1, \dots, T$ and wealth w the action $a_t(w)$ is the measure

$$a_t(w)(\mathrm{d}x) = \frac{h_t(s_{t-1}(w) + x)}{h_{t-1}(s_{t-1}(w))}\,\mathrm{d}\pi_0(x) = \frac{1}{\alpha w} h_t(s_{t-1}(w) + x)\,\mathrm{d}\pi_0(x)\,,$$

where we used $h_{t-1}(s_{t-1}(w)) = \alpha w$ by construction.

Finally, an important remark is that the value of μ does not appear in the construction, except through its sign, which here we have assumed to be positive. The magnitude of μ affects the power but not the rejection region or the associated Doob strategy. This contrasts with the GRO criterion, which would play the constant action π_1.

Exponentially decaying reward. Fix $T > 0$ and let $R(t) = e^{-t/T}$. We now compute the EDO action from Proposition 4. Plugging the closed form

$$D_\xi(\pi_1 \| \pi_0) = \frac{\xi \mu^2}{2\sigma^2}$$

of the Rényi divergence into the characteristic relation defining $\eta_\star$, we obtain

$$\frac{1}{1-\eta_\star} = 1 + \frac{2\sigma^2}{\mu^2 T}\,.$$

Hence

$$L_{\eta_\star}(x) = \left(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_0}(x)\right)^{1/(1-\eta_\star)} = \exp\left(\left(\frac{\mu}{\sigma^2} + \frac{2}{\mu T}\right)x - \left(\frac{\mu^2}{2\sigma^2} + \frac{1}{T}\right)\right)\,.$$

By Proposition 4, the optimal constant action $\pi_\star$ has Lebesgue density proportional to $(\mathrm{d}\pi_0/\mathrm{d}x)L_{\eta_\star}$. Therefore

$$\frac{\mathrm{d}\pi_\star}{\mathrm{d}x}(x) \propto \exp\left(-\frac{x^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} + \frac{2}{\mu T}\right)x\right) \propto \exp\left(-\frac{1}{2\sigma^2}\left(x - \mu - \frac{2\sigma^2}{\mu T}\right)^2\right)\,.$$

Thus

$$\pi_\star = \mathcal{N}\left(\mu + \frac{2\sigma^2}{\mu T},\, \sigma^2\right)\,.$$

Compared with GRO, which plays $\pi_1 = \mathcal{N}(\mu, \sigma^2)$, EDO is more aggressive, shifting further away from the null. As $T \to \infty$, this additional shift vanishes and $\pi_\star$ recovers the GRO action π_1. Finally, we remark that by Proposition 5, this increased aggressiveness preserves power one if and only if

$$T \geq \frac{1}{\mathrm{KL}(\pi_1 || \pi_0)} = \frac{2\sigma^2}{\mu^2}\,.$$

C.2 Bernoulli example

Let $\pi_0 = \mathrm{Ber}(p_0)$ and $\pi_1 = \mathrm{Ber}(p_1)$, with $0 < p_0 < p_1 < 1$. We consider sequential tests for $\pi_0^\otimes$ against $\pi_1^\otimes$. We let $\mathcal{A}$ and ϕ be as in Section 5.

Hard deadline. As in the Gaussian example we define $S_t(x^{1:t}) = \sum_{k=1}^t x_k$. Again, the likelihood ratio $\mathrm{d}\pi_1^{\otimes T}/\mathrm{d}\pi_0^{\otimes T}$ depends on $x^{1:T}$ only through S_T, since we have

$$\frac{\mathrm{d}\pi_1^{\otimes T}}{\mathrm{d}\pi_0^{\otimes T}}(x^{1:T}) = \left(\frac{p_1}{p_0}\right)^{S_T(x^{1:T})} \left(\frac{1-p_1}{1-p_0}\right)^{T - S_T(x^{1:T})}\,.$$

Since $p_1 > p_0$, this expression is increasing in $S_T(x^{1:T})$. However, unlike in the Gaussian case, S_T is discrete. Hence it might be that no level $\{S_T = c\}$ has exactly probability α under $\pi_0^{\otimes}$. For this reason, in general the Neyman–Pearson event $\Gamma^\star$ (see Proposition 2) is not S_T-measurable.

To get a concrete example of the issue, take $T = 3$, $p_0 = 1/2$, $p_1 = 3/4$, and $\alpha = 1/4$. Under $\pi_0^{\otimes 3}$, every triple of observations has mass $1/8$, while the likelihood ratio is increasing in S_3. The best S_3-measurable upper-tail event is $\{S_3 = 3\} = \{111\}$[5], since $\pi_0^{\otimes}(S_3 \geq 2) = 4/8 > \alpha$, and its power is

$$\pi_1^{\otimes}(S_3 = 3) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}\,.$$

However, the unrestricted Neyman–Pearson event can include one tied boundary sequence, for example

$$\Gamma^\star = \{111, 110\}\,.$$

Then $\pi_0^{\otimes}(\Gamma^\star) = 2/8 = \alpha$, while

$$\pi_1^{\otimes}(\Gamma^\star) = \frac{27}{64} + \frac{9}{64} = \frac{36}{64}\,.$$

Thus the exact deterministic Neyman–Pearson event need not be S_T-measurable: restricting to S_T-measurable upper tails can lose power in the discrete case.

In order to identify $\Gamma^\star$, it is enough to decide how many sequences to select from each level set of S_T. For each $k = 0, \ldots, T$, let m_k denote the number of sequences selected from $\{S_T = k\}$. Since both $\pi_0^{\otimes}$ and $\pi_1^{\otimes}$ assign the same mass to all sequences in a fixed level, the Neyman–Pearson problem reduces to the finite integer programme

$$\max_{m_0,\ldots,m_T} \sum_{k=0}^{T} m_k p_1^k (1-p_1)^{T-k}$$

subject to

$$\sum_{k=0}^{T} m_k p_0^k (1-p_0)^{T-k} \leq \alpha\,, \qquad 0 \leq m_k \leq \binom{T}{k}\,, \quad m_k \in \mathbb{N}\,.$$

After fixing a deterministic tie-break within each level, for instance lexicographic order with $1 > 0$, any maximiser $(m_0^\star, \ldots, m_T^\star)$ defines an exact deterministic Neyman–Pearson event $\Gamma^\star$ by selecting the first $m_k^\star$ sequences from each level $\{S_T = k\}$.

When $p_0 \leq 1/2$ and $p_1 \geq 1/2$, the integer programme has a simpler greedy structure. Indeed, the null mass of a single sequence with k ones is $p_0^k(1-p_0)^{T-k}$, which is non-increasing in k. Hence, after all levels with larger likelihood ratio have been included, if the remaining level budget is too small to add another sequence from the current boundary level, then it is also too small to add any sequence from a lower level. Thus one may include all sequences above a boundary level and then fill the remaining budget with as many sequences as possible from that boundary level. More precisely, choose k_α such that

$$\pi_0^{\otimes}(S_T > k_\alpha) \leq \alpha < \pi_0^{\otimes}(S_T \geq k_\alpha)\,,$$

and set

$$r_\alpha = \left\lfloor \frac{\alpha - \pi_0^{\otimes}(S_T > k_\alpha)}{p_0^{k_\alpha}(1-p_0)^{T-k_\alpha}} \right\rfloor\,.$$

After fixing the same deterministic tie-break within $\{S_T = k_\alpha\}$, an exact Neyman–Pearson event is

$$\Gamma^\star = \{S_T > k_\alpha\} \cup \mathcal{C}_{k_\alpha, r_\alpha}\,,$$

where $\mathcal{C}_{k_\alpha, r_\alpha}$ denotes the first r_α sequences in the level $\{S_T = k_\alpha\}$.

Once an optimal $\Gamma^\star$ is determined, the optimal e-process is again the Doob martingale from Proposition 2. More explicitly, set $h_0 = \pi_0^{\otimes}(\Gamma^\star)$ and, for $t = 0, \ldots, T$,

$$h_t(x^{1:t}) = \pi_0^{\otimes}(\Gamma^\star \mid x^{1:t}) = \sum_{y \in \{0,1\}^{T-t}} \mathbf{1}_{\Gamma^\star}(x^{1:t}, y)\, p_0^{S_{T-t}(y)} (1-p_0)^{T-t-S_{T-t}(y)}\,.$$

[5]Here, 111 is the triple of observations $x_1 = 1, x_2 = 1, x_3 = 1$.

Then

$$U_t^\star(x^{1:t}) = \frac{h_t(x^{1:t})}{h_0}\,.$$

Following the discussion after Proposition 2, this e-process is implemented by the predictable strategy which, whenever $h_{t-1}(x^{1:t-1}) > 0$, plays the action

$$A_t(x^{1:t-1})(\{x\}) = \frac{h_t(x^{1:t-1},x)}{h_{t-1}(x^{1:t-1})}\,\pi_0(\{x\})\,, \qquad x \in \{0,1\}\,.$$

When $h_{t-1}(x^{1:t-1}) = 0$, the action can be chosen arbitrarily. The recursion $h_{t-1}(x^{1:t-1}) = \sum_{x\in\{0,1\}} h_t(x^{1:t-1},x)\pi_0(\{x\})$ shows that $A_t(x^{1:t-1})$ is a probability measure, and the corresponding wealth telescopes to $U_t^\star$. Contrarily to the Gaussian case, in general this strategy is not Markovian in the wealth.

For large T, however, explicitly identifying $\Gamma^\star$ and computing the conditional quantities $h_t(x^{1:t})$ quickly becomes computationally costly, since the exact construction may require keeping track of a tie-broken subset of a boundary level. In such cases, a practical simplification is to use the S_T-measurable upper-tail event obtained by ignoring the partial boundary correction. Namely, let

$$\widetilde{k} = \min\{k : \pi_0^{\otimes}(S_T \geq k) \leq \alpha\}\,,$$

and set $\widetilde{\Gamma} = \{S_T \geq \widetilde{k}\}$. This event is optimal among S_T-measurable upper-tail events, and it coincides with the exact Neyman–Pearson event whenever the level constraint aligns with a binomial upper tail. Otherwise, the discrepancy is confined to the boundary level. Heuristically, this correction becomes less important for large T: when p_0 is bounded away from 0 and 1, the mass of any single boundary atom is exponentially small in T, so ignoring a partial boundary selection has only a small effect on the rejection probability.

For this simplified event, the Doob quantities depend only on the current number of ones. More precisely, for $s \in \{0,\dots,t\}$,

$$\widetilde{h}_t(s) = \pi_0^{\otimes}(S_T \geq \widetilde{k} \mid S_t = s) = \sum_{j=(\widetilde{k}-s)_+}^{T-t} \binom{T-t}{j} p_0^j (1-p_0)^{T-t-j}\,.$$

The corresponding predictable action is therefore

$$\widetilde{A}_t(x^{1:t-1})(\{x\}) = \frac{\widetilde{h}_t(S_{t-1}(x^{1:t-1})+x)}{\widetilde{h}_{t-1}(S_{t-1}(x^{1:t-1}))}\,\pi_0(\{x\})\,, \qquad x \in \{0,1\}\,.$$

This gives a simple S_T-based implementation of the hard-deadline Doob strategy, which is the approximation we use when the exact boundary selection is computationally impractical.

Exponentially decaying reward. Fix $T > 0$ and let $R(t) = e^{-t/T}$. We now compute the EDO action from Proposition 4. For Bernoulli distributions, the Rényi divergence has the closed form

$$D_\xi(\pi_1\|\pi_0) = \frac{1}{\xi-1}\log\left(p_1^\xi p_0^{1-\xi} + (1-p_1)^\xi (1-p_0)^{1-\xi}\right)\,.$$

Plugging this into the characteristic relation defining $\eta_\star$, namely $\eta_\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = 1/T$, gives the scalar equation

$$(1-\eta_\star)\log\left(p_1^{1/(1-\eta_\star)} p_0^{-\eta_\star/(1-\eta_\star)} + (1-p_1)^{1/(1-\eta_\star)}(1-p_0)^{-\eta_\star/(1-\eta_\star)}\right) = \frac{1}{T}\,.$$

In contrast to the Gaussian case, this equation does not simplify to a closed-form expression for $\eta_\star$, but it is only one-dimensional and can easily be solved numerically.

We now set

$$L_{\eta_\star}(x) = \left(\frac{\mathrm{d}\pi_1}{\mathrm{d}\pi_0}(x)\right)^{1/(1-\eta_\star)}\,.$$

Thus

$$L_{\eta_\star}(1) = \left(\frac{p_1}{p_0}\right)^{1/(1-\eta_\star)}\,, \qquad L_{\eta_\star}(0) = \left(\frac{1-p_1}{1-p_0}\right)^{1/(1-\eta_\star)}\,.$$

By Proposition 4, the EDO action $\pi_\star$ is the $L_{\eta_\star}$-tilt of π_0. Hence $\pi_\star = \mathrm{Ber}(p_\star)$, where

$$p_\star = \frac{p_0 \left(\frac{p_1}{p_0}\right)^{1/(1-\eta_\star)}}{p_0 \left(\frac{p_1}{p_0}\right)^{1/(1-\eta_\star)} + (1-p_0)\left(\frac{1-p_1}{1-p_0}\right)^{1/(1-\eta_\star)}}. \tag{7}$$

Equivalently,

$$\log \frac{p_\star}{1-p_\star} = \frac{1}{1-\eta_\star} \log \frac{p_1}{1-p_1} - \frac{\eta_\star}{1-\eta_\star} \log \frac{p_0}{1-p_0}.$$

Since $p_1 > p_0$ and $\eta_\star > 0$, this tilt moves past the alternative, so $p_\star > p_1$. Compared with GRO, which plays $\pi_1 = \mathrm{Ber}(p_1)$, EDO is therefore more aggressive. As $T \to \infty$, we have $\eta_\star \to 0$, and this additional tilt vanishes, so $\pi_\star$ recovers the GRO action π_1.

Finally, by Proposition 5, this increased aggressiveness preserves power one if and only if

$$T \geq \frac{1}{\mathrm{KL}(\pi_1 || \pi_0)} = \left(p_1 \log \frac{p_1}{p_0} + (1-p_1) \log \frac{1-p_1}{1-p_0} \right)^{-1}.$$

D Experimental details

D.1 Details for Figure 1

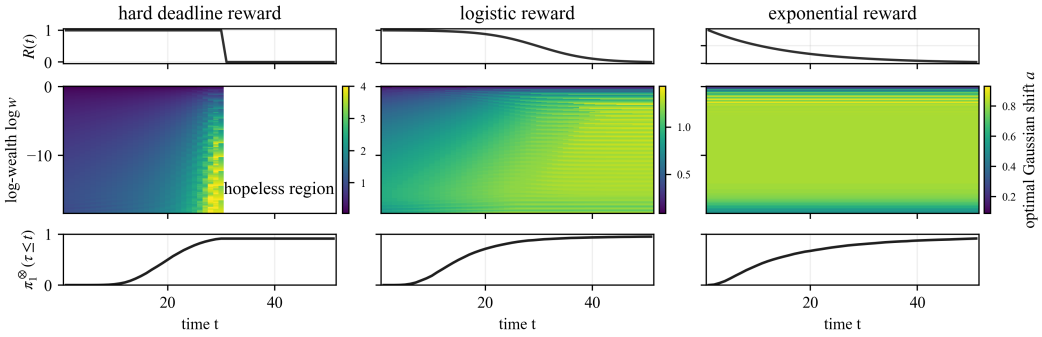


Figure 1 is obtained by solving the finite-horizon Bellman recursion numerically for a restricted Gaussian-shift action class. We take $\pi_0 = \mathcal{N}(0, 1)$, $\pi_1 = \mathcal{N}(0.6, 1)$ and $\alpha = 0.05$, and use actions $a \in [0, 4]$, corresponding to one-step e-variables $E_{[a]}(x) = \exp(ax - a^2/2)$. Thus the figure shows the Bellman optimum within this parametric class, not the unrestricted optimum over $\mathcal{P}^{\pi_0}$. In particular, for the hard-deadline reward, the exact unrestricted Gaussian Doob strategy discussed in Section 8 would not reject before the deadline.

The Bellman recursion is solved over the rescaled wealth $z = \alpha w \in [0, 1]$. We use a logarithmic grid for z on $[10^{-8}, 1]$ with 401 points, and a grid of 401 actions. Expectations under π_1 are evaluated by Gauss–Hermite quadrature with 41 nodes. To start the recursion, we have to truncate the rewards at a large value; in Fig. 1 the truncation is at 150. This barely changes the objectives since, for all rewards, the values above $t = 150$ are below 10^{-4}. The displayed panels show the early part of the policy, where the dependence on time and wealth is most visible.

The reward panels correspond to the hard-deadline reward $R(t) = \mathbf{1}_{t \le 30}$, the logistic reward $R(t) = (1 + \exp((t - 30)/5))^{-1}$, and the exponential reward $R(t) = e^{-t/15}$. The heatmaps show the Bellman-optimal Gaussian-shift action a as a function of time t and log-wealth $\log(\alpha w)$. The "hopeless region" marks states where the reward has value zero, and hence any valid rejection time has passed so that further betting is hopeless.

The three panels illustrate how the reward shape changes the optimal betting behaviour. Under the hard-deadline reward, the policy becomes increasingly aggressive as the deadline approaches, reflecting the fact that rejection after time 30 has no value. The logistic reward smooths this transition, leading to a more gradual change in actions around the same time scale. The exponential reward has no sharp deadline, and therefore produces a more stationary-looking policy away from the rejection boundary. The lower row shows the corresponding stopping-time CDFs under the alternative, estimated from 5000 paths using the same quadrature dynamics as in the Bellman computation.

D.2 Details for Figure 2

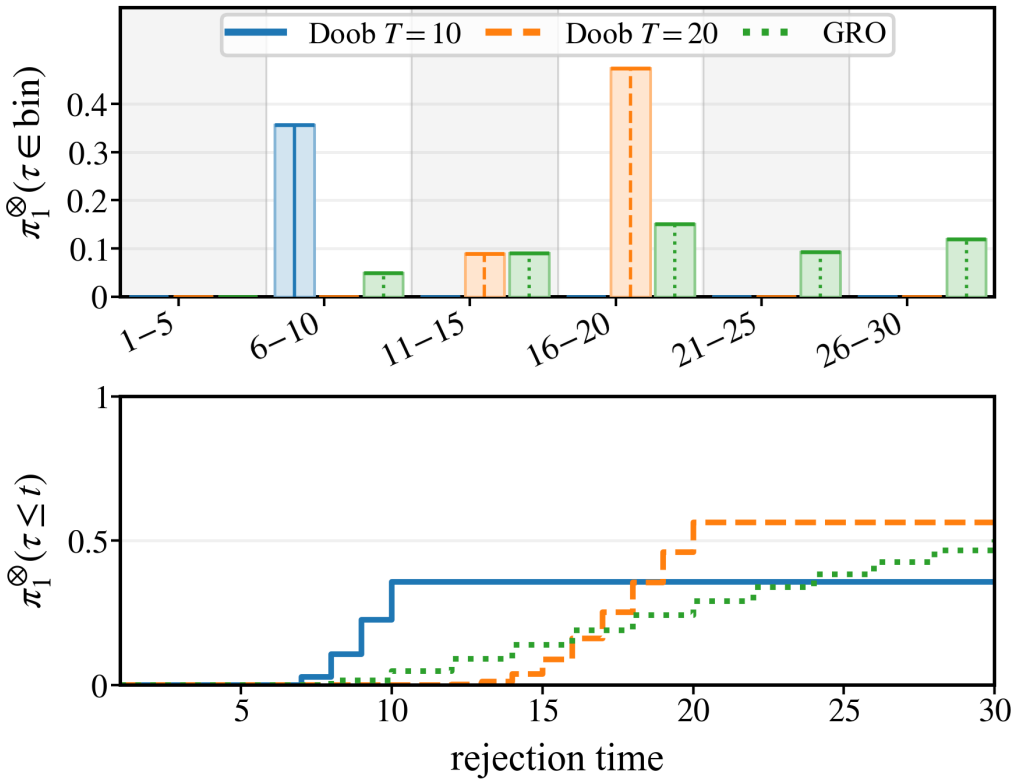


Figure 2 compares the hard-deadline Doob e-processes from Proposition 2 with GRO in the Bernoulli case. We take $\pi_0 = \mathrm{Ber}(0.4)$, $\pi_1 = \mathrm{Ber}(0.6)$ and $\alpha = 0.05$, and consider deadlines $T = 10$ and $T = 20$. For each deadline, the terminal Neyman–Pearson event is constructed exactly. Writing $S_T = \sum_{t=1}^{T} x_t$, we choose the smallest k such that $\pi_0^{\otimes T}(S_T \ge k) \le \alpha$. Since the Bernoulli model is atomic, the deterministic level-α event also includes a subset of the tied boundary level $S_T = k - 1$. We order boundary strings lexicographically with $1 > 0$, and include the first

$$r = \left\lfloor \frac{\alpha - \pi_0^{\otimes T}(S_T \ge k)}{p_0^{k-1}(1-p_0)^{T-k+1}} \right\rfloor .$$

This gives $k = 8$ and $r = 106$ for $T = 10$, and $k = 13$ and $r = 102809$ for $T = 20$.

For the resulting event $\Gamma^\star$, the Doob e-process is $U_t^\star = \pi_0^{\otimes T}(\Gamma^\star \mid x^{1:t})/\pi_0^{\otimes T}(\Gamma^\star)$. Its stopping-time distribution under $\pi_1^{\otimes}$ is computed exactly, by propagating the π_1-mass of alive prefixes and removing a prefix as soon as the conditional probability of $\Gamma^\star$ reaches $\pi_0^{\otimes T}(\Gamma^\star)/\alpha$. The GRO curve is computed in the same way, using the likelihood-ratio wealth, whose log at a state with s successes after t rounds is $s\log(p_1/p_0) + (t-s)\log((1-p_1)/(1-p_0))$. The top panel aggregates the probability mass of τ into time bins, while the bottom panel shows the corresponding CDF under the alternative.

D.3 Details for Figure 3

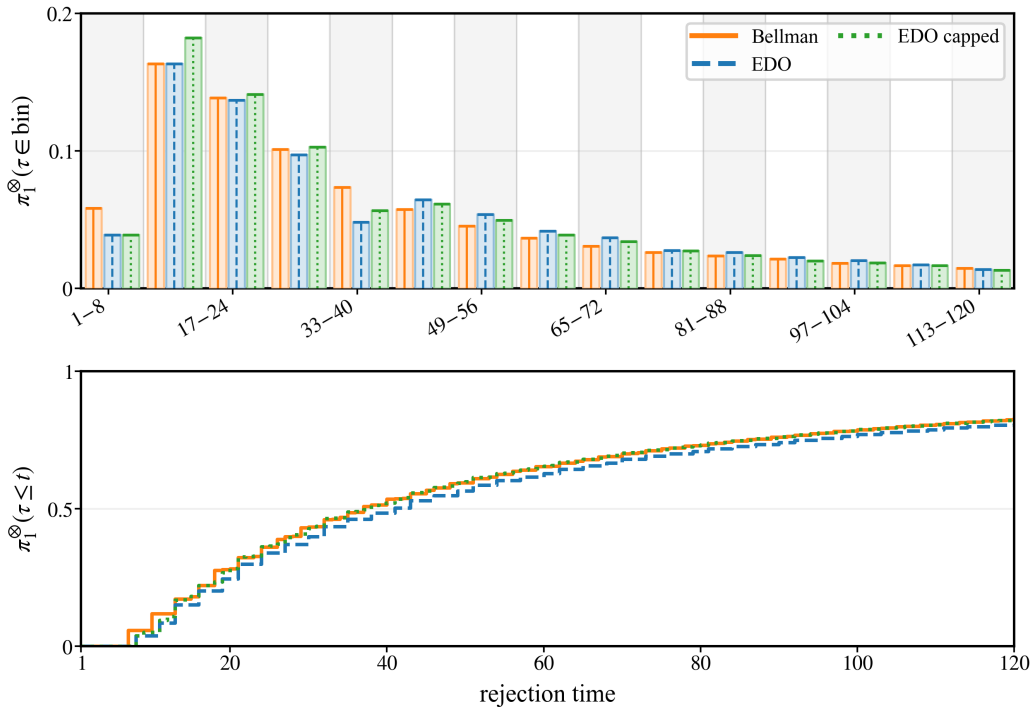


Figure 3 compares EDO with a numerical Bellman solution in the Bernoulli case. We take $\pi_0 = \mathrm{Ber}(1/2)$, $\pi_1 = \mathrm{Ber}(2/3)$ and $\alpha = 0.05$, and use the exponentially decaying reward $R(t) = e^{-t/T}$ with time scale $T = 30$. The EDO action is the constant action $\pi_\star = \mathrm{Ber}(p_\star)$ obtained from Proposition 4. In this case $\eta_\star$ is computed numerically by solving $\eta_\star D_{1/(1-\eta_\star)}(\pi_1 \| \pi_0) = 1/T$, and $p_\star$ is then obtained from the closed-form Bernoulli tilt (7).

The Bellman curve is computed by finite-horizon backward induction on the rescaled wealth $z = \alpha w$. We use horizon 180, plot rejection times up to 120, and discretise the state and action spaces using 401 wealth grid points and 401 actions. The wealth grid is logarithmic on $[10^{-10}, 1]$. The Bellman policy uses action-level overshoot capping, meaning that if an outcome would already cross the boundary, its payoff is capped at the value needed to reach $z = 1$. We compare it with the uncapped EDO action and with the same EDO action after applying this boundary-aware capping.

The stopping-time distributions are estimated under $\pi_1^{\otimes}$ using 300000 Monte Carlo paths, with common random numbers across the three strategies. The top panel aggregates the probability mass of τ into time bins, while the bottom panel shows the corresponding CDF under the alternative.

D.4 Details for Figure 4

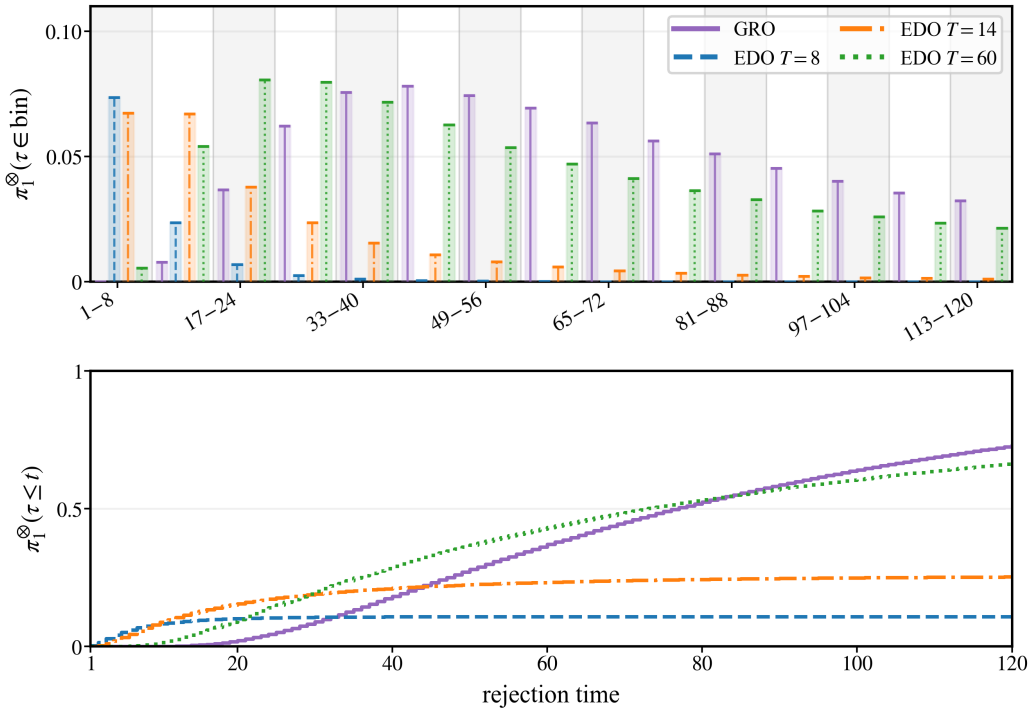


Figure 4 compares GRO with the EDO strategy in the Gaussian case (no capping). We take $\pi_0 = \mathcal{N}(0,1)$, $\pi_1 = \mathcal{N}(1/4,1)$ and $\alpha = 0.05$. GRO corresponds to the constant Gaussian-shift action $a = \mu = 1/4$. For each time scale $T \in \{8, 14, 60\}$, EDO uses the constant action from Appendix C.1, namely $a_\star(T) = \mu + 2/(\mu T)$. Thus smaller time scales lead to larger shifts away from the null, favouring earlier rejection. As T increases, $a_\star$ approaches the GRO action μ, which is asymptotically optimal for minimising the first-order rejection-time scale as $\alpha \to 0$. This is reflected in the CDFs: smaller time scales rise earlier but plateau below one, while the larger time scale more closely tracks GRO, which has the best large-time behaviour.

The stopping-time distributions are estimated under $\pi_1^{\otimes}$ using 300000 Monte Carlo paths, with common random numbers across all strategies. The horizon is 250, and rejection times are plotted up to 120. No capping is applied. The top panel aggregates the probability mass of τ into time bins, while the bottom panel shows the corresponding CDF under the alternative.

The power effect is visible through Proposition 5. For a constant Gaussian-shift action a, the log-growth under π_1 is $\gamma(a) = a\mu - a^2/2$. Hence power one holds exactly when $0 < a \le 2\mu$. For EDO this condition is equivalent to $T \ge 2/\mu^2$. In the present experiment, $2/\mu^2 = 32$, so the EDO curve with $T = 60$ has power one, whereas the smaller time scales $T = 8$ and $T = 14$ trade eventual power for earlier rejection.

D.5 Details for Figure 5

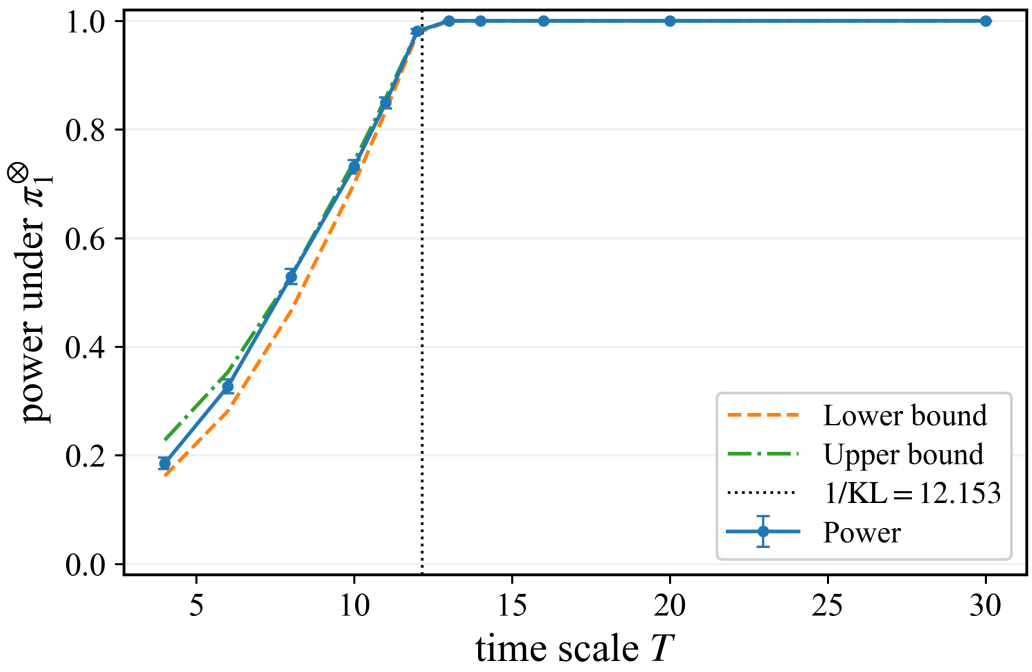


Figure 5 illustrates the power statement in Proposition 5 for the Bernoulli EDO constant strategy. We take $\pi_0 = \mathrm{Ber}(0.5)$, $\pi_1 = \mathrm{Ber}(0.7)$ and $\alpha = 0.05$. For each time scale T, we compute the EDO action by solving $\eta_\star D_{1/(1-\eta_\star)}(\pi_1\|\pi_0) = 1/T$. This determines the constant e-variable $E_\star = d\pi_\star/d\pi_0$, which in the Bernoulli case is specified by its two values $E_\star(0)$ and $E_\star(1)$.

The drift under the alternative is $\gamma = (1-p_1)\log E_\star(0) + p_1 \log E_\star(1)$. By Proposition 5, the strategy has power one when $\gamma \geq 0$. In this example the transition occurs at $T = \mathrm{KL}(\mathrm{Ber}(0.7)||\,\mathrm{Ber}(0.5))^{-1}$, where

$$\mathrm{KL}(\mathrm{Ber}(0.7)||\,\mathrm{Ber}(0.5)) = 0.7\log(1.4) + 0.3\log(0.6)\,.$$

This threshold is marked by the vertical dotted line.

For T below the threshold, the drift is negative. We then compute $\kappa > 0$ solving $\mathbb{E}_{\pi_1}[E_\star(X)^\kappa] = 1$, and plot the bounds from Proposition 5, namely $(\alpha/B_\star)^\kappa$ and α^κ, where $B_\star = \max\{E_\star(0), E_\star(1)\}$. The power itself is estimated under $\pi_1^{\otimes}$ by Monte Carlo, using 5000 paths for each value of T below the threshold. The error bars are Wilson confidence intervals for this Monte Carlo estimate, with the upper endpoint enlarged by the truncation bound used for paths stopped far below the rejection boundary. For T above the threshold, power is set to one by the proposition.

E Further discussion on the hard-deadline reward

We expand here on the hard-deadline construction of Proposition 2. In the simple-vs-simple case, the reward $R(t) = \mathbf{1}_{\{t \le T\}}$ Proposition 2 reduces the problem to a fixed-horizon Neyman–Pearson problem at time T. Namely, one first chooses a terminal event $\Gamma^\star \subseteq \mathcal{X}^T$ satisfying $\pi_0^{\otimes}(\Gamma^\star) \le \alpha$ and maximising $\pi_1^{\otimes}(\Gamma)$ among all terminal events $\Gamma \subseteq \mathcal{X}^T$ with $\pi_0^{\otimes}(\Gamma) \le \alpha$. The corresponding sequential test is then obtained by inducing this terminal event through the Doob martingale

$$U_t^\star(x^{1:t}) = \pi_0^{\otimes}(\Gamma^\star \mid x^{1:t}) / \pi_0^{\otimes}(\Gamma^\star)\,.$$

Thus the exact stopping rule is

$$\pi_0^{\otimes}(\Gamma^\star \mid x^{1:t}) \ge \pi_0^{\otimes}(\Gamma^\star)/\alpha\,.$$

This is the sense in which the hard-deadline optimum has a terminal character: the event being optimised lives at time T, and the sequential test rejects earlier only if the observations so far make this terminal event sufficiently likely under the null.

When $\pi_0^{\otimes T}$ is non-atomic, one can choose the Neyman–Pearson set $\Gamma^\star$ so that $\pi_0^{\otimes T}(\Gamma^\star) = \alpha$: since there are no atoms, the likelihood-ratio rejection region can be adjusted to have exactly the desired null probability. In this case $U_T^\star = \alpha^{-1}\mathbf{1}_{\Gamma^\star}$, so the induced Doob test rejects at the deadline exactly on the fixed-sample Neyman–Pearson event. Before the deadline, the stopping rule becomes $\pi_0^{\otimes}(\Gamma^\star \mid x^{1:t}) \ge 1$, and hence early rejection is possible only if the observations $x^{1:t}$ already imply that the final path will belong to $\Gamma^\star$, whatever the remaining observations may be.

This phenomenon becomes very explicit for Gaussian mean testing, where rejection before the deadline cannot happen, as discussed in Section 8. Let $\pi_0 = \mathcal{N}(0, \sigma^2)$ and $\pi_1 = \mathcal{N}(\mu, \sigma^2)$, with $\mu > 0$. Then the fixed-horizon Neyman–Pearson region is $\Gamma^\star = \{S_T \ge c\}$, where $S_T = \sum_{t=1}^T x_t$ and c is chosen so that $\pi_0^{\otimes}(S_T \ge c) = \alpha$. For every $t < T$, conditional on $x^{1:t}$, the remaining sum $S_T - S_t$ has full support on $\mathbb{R}$ under $\pi_0^{\otimes}$. Therefore $\pi_0^{\otimes}(S_T \ge c \mid x^{1:t}) < 1$ for every finite history $x^{1:t}$. Thus the exact Gaussian hard-deadline test never rejects before the deadline: it rejects at T precisely on $\Gamma^\star$, and hence coincides with the fixed-sample Neyman–Pearson test.

In Bernoulli testing, by contrast, the fixed-horizon problem is atomic. Since the likelihood ratio is increasing in S_T, the Neyman–Pearson ordering favours paths with more successes. If randomisation were allowed, this would give the usual upper-tail rule, with randomisation at a boundary value of S_T. For deterministic events, however, the atoms cannot be split arbitrarily, and the exact set $\Gamma^\star$ is obtained by a finite integer optimisation over paths, or equivalently over the numbers of strings selected at each value of S_T. Except for special values of α for which the desired level is attained exactly by a union of atoms, such a deterministic choice may satisfy $\pi_0^{\otimes}(\Gamma^\star) < \alpha$. This creates a

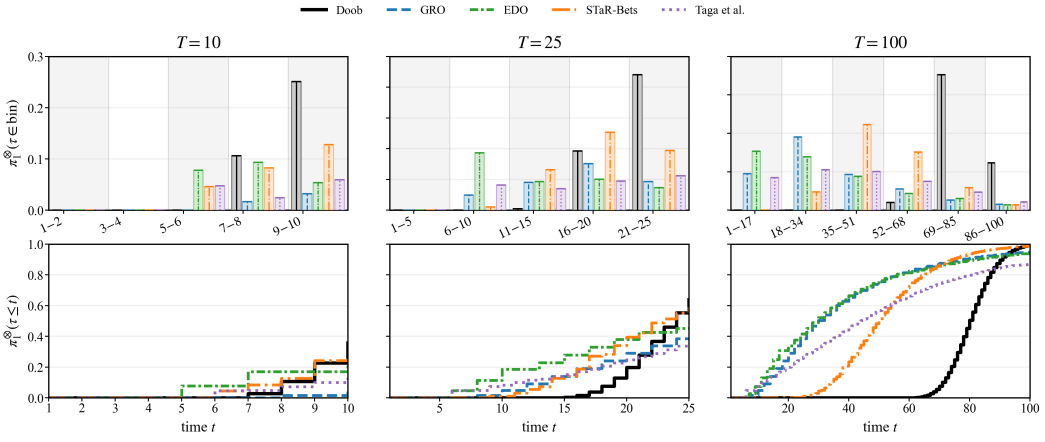


Figure 6: Rejection-time distributions for $\pi_0 = \mathrm{Ber}(0.4)$ and $\pi_1 = \mathrm{Ber}(0.6)$, with $\alpha = 0.05$, for deadlines $T = 10, 25, 100$. The top row shows coarse-binned rejection probabilities and the bottom row shows the corresponding cdfs. The Doob curve is the exact hard-deadline Doob/Neyman–Pearson benchmark, GRO and EDO are constant-bet strategies, and STaR-Bets and Taga et al. are simulated horizon-aware baselines.

minor distinction between scaling the Doob martingale by $\pi_0^{\otimes}(\Gamma^\star)$ and scaling it by α. In either case, since future observations are bounded, sufficiently favourable observations before time T can force membership in the terminal event, and hence allow earlier rejection. Nevertheless, the hard-deadline objective still assigns the same reward to all rejection times before T, and so does not itself favour rejection substantially earlier than the deadline. Consequently, even in Bernoulli examples the induced Doob test can remain conservative until late in the horizon. This is the phenomenon referred to in Section 6, and is one motivation for considering smoother rewards, such as the exponential-decay reward of Section 7.

Figure 6 compares the exact hard-deadline Doob construction with several betting rules in the Bernoulli example $\pi_0 = \mathrm{Ber}(0.4)$, $\pi_1 = \mathrm{Ber}(0.6)$ and $\alpha = 0.05$. The Doob curve is computed exactly from the deterministic Neyman–Pearson event with lexicographic tie-breaking, using the stopping rule $U_t^\star \geq \alpha^{-1}$, equivalently $\pi_0^{\otimes}(\Gamma^\star \mid x^{1:t}) \geq \pi_0^{\otimes}(\Gamma^\star)/\alpha$. As expected from its terminal-power objective, this curve is the most conservative among the horizon-aware hard-deadline procedures: rejection mass is pushed towards the deadline. At the same time, it attains the largest probability of rejection by the deadline, as guaranteed by Proposition 2. STaR-Bets from Voráček and Orabona [2025] performs particularly well in this example, achieving substantially earlier rejection while remaining close to the Doob benchmark near the deadline, and typically outperforming the method from Taga et al. [2026].

The comparison between GRO and EDO isolates the effect of changing the constant betting rule. GRO plays the likelihood-ratio bet, equivalently $\mathrm{Ber}(p_1)$, at every round. EDO instead plays the uncapped Bernoulli tilt $\mathrm{Ber}(p^\star)$ obtained from the exponential-decay criterion with time scale equal to the plotted horizon. For the shorter horizons, EDO outperforms GRO, betting more aggressively and achieving earlier rejection. For $T = 100$, the two curves are nearly indistinguishable, in line with the fact that EDO recovers GRO as the time scale grows.

For EDO, $p^\star$ is obtained as described in Appendix C. STaR-Bets is simulated under $\pi_1^{\otimes}$ using the finite-horizon target-recalculating rule of Voráček and Orabona [2025], with the empirical second-moment proxy regularised by the horizon-dependent bonus and clipped at $p_0(1-p_0)$ as in the paper's implementation. The Taga et al. curve is a simulated horizon-aware betting heuristic in the spirit of Taga et al. [2026]: we use the empirical Kelly stake as a baseline, mix it with the endpoint stake according to six linear or quadratic schedules with onsets $0.25T, 0.5T, 0.75T$, and reject using the average wealth across these schedules. Both simulated curves use 20000 Monte Carlo paths from $\pi_1^{\otimes}$ and the same random seed across horizons.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the paper's main contributions: the time-sensitive testing-by-betting formulation, the Markovian reduction, the simple-vs-simple equivalence with e-processes, and the analysis of hard-deadline and exponentially decaying rewards. These claims are formalised in Sections 3–7 and proved in Appendix B.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 9 discusses the main limitations of the present formulation.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper states the assumptions in the corresponding theorem and proposition statements. Complete proofs are provided in Appendix B, with external results cited where used.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The numerical results are illustrative and use synthetic Gaussian and Bernoulli settings. Appendix D describes the experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code is available at tobiaswegel/time-sensitive-anytime-valid.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper contains no training procedure or data splits, since the numerical examples are synthetic. Section 8, Appendix C, and Appendix D specify the Gaussian and Bernoulli settings, the reward choices, the Bellman approximation, the action and wealth grids, and the Monte Carlo simulation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The paper does not report error bars or statistical significance tests. The numerical figures are illustrative synthetic examples supporting theoretical results rather than empirical comparisons on real datasets.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: The experiments are small-scale and were run on a personal laptop in a few minutes, without any use of GPUs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research is theoretical and uses only synthetic numerical examples.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper is foundational statistical methodology and does not introduce a deployable system, model, dataset, or application-specific tool. No direct societal impact, either positive or negative, is anticipated beyond general methodological use in sequential testing.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release data or models with high misuse risk. It contains theoretical results and synthetic numerical examples only.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing datasets, codebases, models, or other external assets. Prior scholarly work used for theory and comparison is cited in the references.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce or release new assets such as datasets, code packages, trained models, or benchmarks.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing, user studies, data annotation, or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects, so IRB approval or equivalent review is not applicable.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: The use of LLMs is acknowledged in the LLM usage statement at the end of the paper.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.