UNIFORM RACE: PARAMETER-FREE APPROXIMATE REJECTION SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study approximate sampling: given N independent samples from a proposal distribution μ, the goal is to select one whose distribution is close to a target π specified only up to a normalizing constant. The seminal work of Block & Polyanskiy (2023) provides finite budget error bounds for approximate rejection sampling (RS) as a function of the algorithm's acceptance threshold M. The threshold M giving the smallest bound, however, depends on properties of (π, μ) that are typically unavailable from the observed sample. This raises a natural question: *Can one attain the best RS guarantee without taking M as input?* We answer this question affirmatively by proposing a parameter-free sampling algorithm called *uniform race* (UR), based on importance weights, which are ratios of target to proposal probabilities (or densities). It divides each observed weight by an independent uniform random variable to form a score and returns the candidate with the largest score. For every budget N, its total variation error satisfies the RS upper bound for every fixed threshold M simultaneously, thereby achieving the best such bound in hindsight. We also characterize its output distribution conditional on the largest score, identifying when it is exactly the target π. Uniform race has no larger total variation error than a natural budget-calibrated RS derived from Rohatgi et al. (2025) and sampling importance resampling (SIR). In particular, we exhibit instances where UR's error is exponentially smaller in N than that of either baseline. Furthermore, we establish conditions under which attaining this RS guarantee for every (π, μ) uniquely determines the selection probabilities as those of UR. Finally, test-time scaling experiments on LLM math-reasoning tasks corroborate the theoretical comparisons and demonstrate that UR remains competitive in ground-truth accuracy without requiring threshold selection.

1 INTRODUCTION

Inference-time computation often involves generating a finite number of samples from a proposal distribution and selecting one to approximate a target distribution. Inspired by this, we study the following approximate sampling problem: given N independent and identically distributed (i.i.d.) samples $Y_1, \ldots, Y_N$ drawn from a proposal distribution μ, select one whose marginal distribution is close to a target π that is specified only up to a normalizing constant. To describe how the target differs from the proposal, define the importance weight as $w := \frac{\mathrm{d}\pi}{\mathrm{d}\mu}$. For discrete distributions, this is simply $w(y) = \pi(y)/\mu(y)$, so $w(y)$ measures how much the proposal probability at y must be reweighted to obtain the target. We assume that the sampler does not evaluate this weight directly. Instead, at each sampled point Y_i for $i = 1, \ldots, N$, the sampler only has access to the *unnormalized* weight $w_u(Y_i) := Zw(Y_i)$, where $Z > 0$ is an *unknown* normalizing constant. Equivalently, $w(Y_i) = w_u(Y_i)/Z$, but neither Z nor $w(Y_i)$ is available to the sampler. We assume that the proposal covers the target, written $\pi \ll \mu$: every set with zero proposal probability also has zero target probability. Throughout the main text, $0 < w(Y) < \infty$ almost surely under μ; we consider zero weights separately in Appendix B.5.

We refer to the sampled points as *candidates*. Let $\widehat{Y}$ be the returned candidate and $P_N := \mathcal{L}(\widehat{Y})$ its marginal distribution, where $\mathcal{L}(X)$ denotes the law of X. Our goal is to make P_N close to π, measured by total variation (TV) distance,

$$D_{\mathrm{TV}}(P_N, \pi) := \sup_D |P_N(D) - \pi(D)|, \tag{1}$$

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

where D ranges over measurable sets. We note that the distribution P_N averages over both the candidates and the sampler's additional randomness.

A useful baseline for this objective is classical rejection sampling (RS), which produces an exact sample from π if a valid acceptance threshold M is available. For the threshold to be valid, it must satisfy $w(y) \leq M$ over all possible proposals. Given such a threshold, RS repeatedly proposes $Y \sim \mu$ and accepts with probability $w(Y)/M$ until the first acceptance. The returned sample follows π since the probability of proposing y and accepting it is proportional to $\mu(y)w(y) = \pi(y)$ in the discrete case; the same calculation holds for densities. Notably, this does not require prior knowledge of the normalizing constant Z. If only the unnormalized weights $w_u = Zw$ are available, the same acceptance test can be implemented using the corresponding upper bound ZM: $w_u(Y)/ZM = w(Y)/M$. Hence RS need not compute Z, but it still relies on a valid upper bound, called an *envelope*, on the importance weight, and may continue sampling until an acceptance occurs.

This exposes two desiderata in our setting: the sampler should respect a prescribed proposal budget N, and it should not require prior knowledge of a problem-dependent envelope. One way to accommodate both is to choose an arbitrary threshold $M > 0$, accept each proposal with probability (in normalized units) $\min\{w(Y)/M, 1\}$, and return an observed candidate if all N proposals are rejected. With the observed unnormalized weights, this still requires supplying the corresponding threshold ZM on the unnormalized scale.

The threshold now controls two sources of error. First, if M falls below some importance weights, the cap at one changes the distribution of accepted samples: weights above M are effectively replaced by M. This is the *clipping* effect. To quantify it, for $M > 0$, define

$$\mathcal{E}_M(\pi, \mu) := \mathbb{E}_\mu[(w(Y) - M)_+], \quad A_M(\pi, \mu) := 1 - \mathcal{E}_M(\pi, \mu) = \mathbb{E}_\mu[\min\{w(Y), M\}]. \tag{2}$$

Here $(x)_+ := \max\{x, 0\}$; $\mathcal{E}_M$ is the excess weight removed by clipping at M, while A_M is the retained clipped mass. Denote $\mathcal{E}_M(\pi, \mu)$ by $\mathcal{E}_M$ when (π, μ) is fixed. For $M \geq 1$, $\mathcal{E}_M$ is the standard E_M-divergence (Liu et al., 2017); see also Block & Polyanskiy (2023, Example 4). We use the same clipping definition for $0 < M < 1$. Notice that $\mathcal{E}_M$ bounds the TV error of the accepted distribution; it is nonincreasing in M and tends to zero as $M \to \infty$ for every fixed (π, μ). Second, increasing M lowers the acceptance probability and makes it more likely that all N proposals are rejected.

This tradeoff is central to the approximate RS using finite budget. Building on the analysis from Block & Polyanskiy (2023), Huang et al. (2025, Lemma D.4) analyze the clipped-threshold rule above and show that, for a fixed threshold M,

$$D_{\mathrm{TV}}(P^{\mathrm{RS}}_{N,M}, \pi) \lesssim \underbrace{2\mathcal{E}_M(\pi, \mu)}_{\text{clipping}} + \underbrace{\exp(-N/M)}_{\text{budget exhaustion}}, \tag{3}$$

where $P^{\mathrm{RS}}_{N,M}$ denotes the marginal output distribution of this finite budget procedure. The two terms favor opposite choices of M: a larger threshold reduces clipping, but makes exhausting the proposal budget more likely. Consequently, the threshold giving the best guarantee hinges on the unknown target–proposal pair, including importance weights that may never appear among the N observed candidates. This motivates the central question of this paper:

Can one attain the best fixed threshold RS guarantee without taking M as input?

Uniform race. To answer the question, we propose a parameter-free sampling algorithm, called *uniform race* (UR). It divides each observed weight by an independent uniform random variable on $(0, 1)$ and returns the candidate with the largest resulting score:

$$\widehat{I}_U := \arg\max_{i \in [N]} \frac{w_u(Y_i)}{U_i}, \qquad \widehat{Y}_U := Y_{\widehat{I}_U}, \qquad U_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, 1), \tag{4}$$

where $[N] := \{1, \ldots, N\}$ and the uniforms are independent of the proposals. Since Z is common to all candidates, it cancels from the ranking. By *parameter-free*, we mean that the sampler does not take an acceptance threshold as input, yet its guarantee adapts automatically to the best threshold in hindsight. It also does not require an estimate of the normalization constant Z; the proposal budget and any parameters that define the target remain inputs. Specifically, our main contributions are:

Main result (informal; Theorem 1). Let $P^U_N := \mathcal{L}(\widehat{Y}_U)$ be UR's marginal output distribution. For every fixed $N \geq 1$,

$$D_{\mathrm{TV}}(P^U_N, \pi) \leq \inf_{M > 0} \left\{ 2\mathcal{E}_M(\pi, \mu) + \exp(-N/M) \right\}. \tag{5}$$

Sampling rule	Sharp RS guarantee	Extra input	TV error as N grows
SIR	×	—	$\Theta(N^{-1})$
RS with $M = C_\infty^\pi$	✓	Exact envelope	$\Theta(3^{-N})$
Budget-calibrated RS	×	Calibration parameter δ	$\Omega(2^{-N/2})$
UR	✓	—	$\boldsymbol{\Theta(3^{-N})}$

Table 1: Comparison of sampling rules. TV orders are for $\mu = (1/2, 1/2)$, $\pi = (1/4, 3/4)$ and odd total budgets $N \geq 3$. The calibrated RS entry is a lower bound after optimizing δ before sampling.

The key difference from equation 3 is the *infimum over* M: UR satisfies the bound for every threshold simultaneously, so choosing a threshold is unnecessary. Theorem 1 gives a sharper version of this guarantee. Beyond this guarantee, we establish the following structural and comparative results.

- **Sharp guarantees under bounded weights.** Let $C_\infty^\pi := \operatorname{ess\,sup}_\mu w$ denote the smallest almost-sure upper bound on the importance weight. When it is finite, we show that UR's TV error is at most $(1 - 1/C_\infty^\pi)^N$, even though the sampler is not given this bound. This guarantee matches the smallest possible worst-case error over each bounded weight class; see Section 2 and Corollary 1.
- **Rationale behind the guarantee.** Using the normalized scores $S_i := w(Y_i)/U_i$ only for analysis, we characterize the output conditional on their maximum. In particular, when the weights are bounded, conditioning on $\max_i S_i > C_\infty^\pi$ gives exactly the target distribution π, although UR need not know this bound. Section 2 establishes the general identity and uses it to prove Theorem 1.
- **Comparison with budget-calibrated RS.** We next ask whether or not estimating the normalizer and choosing a threshold from the available budget can recover UR's accuracy. In particular, a direct budget-calibrated version of Rohatgi et al. (2025, Algorithm 5) uses pilot proposals for this estimate, sets a calibration parameter $\delta \in (0, 1)$, and then applies RS to fresh candidates. Theorem 2 shows that UR has no greater TV distance from π at the same budget, and even when UR uses only the fresh sampling portion of that budget. Moreover, on a fixed binary pair, the baseline-to-UR TV error ratio grows exponentially with N, even after optimizing prior to sampling. See Section 3.
- **Comparison with SIR.** Rather than setting an acceptance threshold, sampling importance resampling (SIR) (Rubin, 1987; Smith & Gelfand, 1992) selects candidate i with probability $w_u(Y_i)/\sum_j w_u(Y_j)$. We show that UR has no greater TV error for every (π, μ) and fixed N, and this also extends to every convex f-divergence. The gap can be substantial: on a fixed binary pair, UR's error decreases exponentially in N, whereas SIR's decreases as $1/N$. See Section 4.
- **Uniqueness of selection probabilities.** We also ask whether the universal RS guarantee allows other selection probabilities. Among rules that exploit only relative observed weights, apply the same rule to every target–proposal pair, and treat candidate order symmetrically, the bounded weight guarantee forces each candidate's selection probability to equal UR's. See Section 5.

Table 1 highlights the main distinction: UR alone combines the sharp RS guarantee with no envelope or additional calibration input, while matching the TV error of RS supplied with the exact envelope on the displayed binary pair. Overall, UR removes threshold calibration while ensuring the best fixed threshold RS guarantee in hindsight. Numerical experiments on LLM tasks corroborate these theoretical comparisons.

2 RS GUARANTEES WITHOUT A THRESHOLD

We now state the sharper guarantee underlying equation 5. With $\mathcal{E}_M$ and A_M defined in equation 2, the quantity A_M/M is the acceptance probability of one proposal under the fixed-threshold RS rule. Thus $(1 - A_M/M)^N$ is the probability that all N proposals are rejected.

Theorem 1 (RS without threshold selection)**.** *Under the assumptions of Section 1, for every fixed* $N \geq 1$*,*

$$D_{\mathrm{TV}}(P_N^U, \pi) \leq \inf_{M>0} \left\{ \mathcal{E}_M + \left(1 - \frac{A_M}{M}\right)^N \right\}. \tag{6}$$

To prove Theorem 1, we compare uniform race (UR) with rejection sampling (RS) at a fixed threshold and then optimize that threshold only in the analysis. To this end, fix $N \geq 1$ under the assumptions

of Section 1, and use the same candidates and uniforms for both procedures. Recall the normalized scores $S_i := \frac{w(Y_i)}{U_i}$ and their maximum $S_{(N)} := \max_{i\in[N]} S_i$, and define the threshold-crossing event $H_M := \{S_{(N)} \geq M\}$. By construction, $S_i \geq M$ exactly when $U_i \leq \min\{w(Y_i)/M, 1\}$. Hence H_M is exactly the event that RS accepts at least one proposal when the same uniforms are used; on H_M, RS returns the first accepted candidate. UR, by contrast, always returns the candidate with the largest score. We denote their conditional distributions on H_M by

$$Q_M^{\mathrm{RS}}(\mathrm{d}y) := \frac{\min\{w(y), M\}}{A_M}\mu(\mathrm{d}y), \qquad Q_{N,M}^U := \mathcal{L}(\widehat{Y}_U \mid H_M). \tag{7}$$

Proof of Theorem 1. The proof has two key ingredients. First, fix $M > 0$ and let $\delta_M := \Pr(H_M^c)$. Lemma 1 yields $\delta_M = (1 - A_M/M)^N$ and $D_{\mathrm{TV}}(Q_M^{\mathrm{RS}}, \pi) \leq \mathcal{E}_M$. Second, Lemma 2 then shows that selecting the largest accepted score has no greater conditional error than returning the first accepted candidate. When $\delta_M > 0$, write $F_{N,M}^U := \mathcal{L}(\widehat{Y}_U \mid H_M^c)$. By the law of total probability, $P_N^U = (1 - \delta_M)Q_{N,M}^U + \delta_M F_{N,M}^U$. With this decomposition, we obtain:

$$\begin{aligned} D_{\mathrm{TV}}(P_N^U, \pi) &\overset{\text{(i)}}{\leq} (1 - \delta_M) D_{\mathrm{TV}}(Q_{N,M}^U, \pi) + \delta_M D_{\mathrm{TV}}(F_{N,M}^U, \pi) \\ &\overset{\text{(ii)}}{\leq} (1 - \delta_M) D_{\mathrm{TV}}(Q_M^{\mathrm{RS}}, \pi) + \delta_M \\ &\overset{\text{(iii)}}{\leq} \mathcal{E}_M + \left(1 - \frac{A_M}{M}\right)^N, \end{aligned} \tag{8}$$

where (i) follows from the convexity of total variation; (ii) follows from Lemma 2 and $D_{\mathrm{TV}} \leq 1$; and (iii) follows from Lemma 1 and the definition of δ_M. If $\delta_M = 0$, then $P_N^U = Q_{N,M}^U$, and the same bound follows directly from Lemmas 1 and 2. Since P_N^U does not depend on the analysis threshold M, taking the infimum over $M > 0$ completes the proof. We note that the lemmas are stated in Section 2.2, with full proofs in Appendix B. □

2.1 Output distribution at a fixed largest score

We first characterize the output conditional on the largest score.

Proposition 1 (Output distribution at a fixed largest score)**.** *For every fixed $N \geq 1$, except on a set of score values having probability zero under the distribution of $S_{(N)}$,*

$$\mathcal{L}(\widehat{Y}_U \mid S_{(N)} = s) = \pi(\cdot \mid w < s), \tag{9}$$

whenever $\pi(w < s) > 0$. Here $\{w < s\} := \{y : w(y) < s\}$. In particular, if $C_\infty^\pi < \infty$, then

$$\mathcal{L}(\widehat{Y}_U \mid S_{(N)} = s) = \pi \tag{10}$$

for every $s > C_\infty^\pi$, except possibly on the same probability zero set.

Proof. See Appendix B.2. The proposition has a simple interpretation. For a fixed score s, states with $w(y) \geq s$ are excluded, while the remaining states follow π after renormalization. If $C_\infty^\pi < \infty$, then conditional on $S_{(N)} > C_\infty^\pi$, the returned candidate has distribution π. Finally, we note that the budget affects only the distribution of the largest score, not this conditional target rule. □

2.2 Acceptance probability and conditional error

We now return to the two lemmas used to prove Theorem 1. The first describes RS at a fixed threshold, and the second compares its conditional output distribution with that of UR.

Lemma 1. *For every $M > 0$, $0 < A_M/M \leq 1$ and*

$$\Pr(H_M^c) = \left(1 - \frac{A_M}{M}\right)^N, \qquad D_{\mathrm{TV}}(Q_M^{\mathrm{RS}}, \pi) \leq \mathcal{E}_M. \tag{11}$$

Proof. One proposal is accepted with probability $\mathbb{E}_\mu[\min\{w(Y)/M, 1\}] = A_M/M$, so independence gives the probability of no acceptance. Conditioning a proposal on acceptance yields Q_M^{RS}; Appendix B verifies that the first accepted candidate has this same distribution. To bound its error, observe that $A_M Q_M^{\mathrm{RS}}(\mathrm{d}y) = \min\{w(y), M\}\mu(\mathrm{d}y) \leq \pi(\mathrm{d}y)$. The residual measure has mass $1 - A_M$, so π is a mixture containing Q_M^{RS} with weight A_M. Consequently $D_{\mathrm{TV}}(Q_M^{\mathrm{RS}}, \pi) \leq 1 - A_M = \mathcal{E}_M$; when $A_M = 1$, the two distributions agree. □

Lemma 2. *For every $N \geq 1$ and $M > 0$, $D_{\mathrm{TV}}(Q^U_{N,M}, \pi) \leq D_{\mathrm{TV}}(Q^{\mathrm{RS}}_M, \pi)$.*

Proof sketch. Using the joint distribution from Proposition 1, we obtain two monotonicity properties. First, the density of $Q^U_{N,M}$ relative to π is nonincreasing in the weight. Second, the density ratio $dQ^U_{N,M}/dQ^{\mathrm{RS}}_M$ is nondecreasing in the weight. Together, these monotonicity properties show that selecting the largest accepted score does not increase the total variation error relative to the target. The formal proof is deferred to Appendix B, where Lemma 3 converts these two properties into the desired result. □

2.3 Bounded weights and minimax error

At $M = C^\pi_\infty < \infty$, both conditional distributions are exactly π, so the inequality in Lemma 2 is tight. Note that there is then no clipping, so $Q^{\mathrm{RS}}_{C^\pi_\infty} = \pi$. Averaging equation 10 over $H_{C^\pi_\infty}$ and applying Lemma 1 yield:

$$\mathcal{L}(\widehat{Y}_U \mid H_{C^\pi_\infty}) = \pi, \qquad \Pr(H^c_{C^\pi_\infty}) = \left(1 - \frac{1}{C^\pi_\infty}\right)^N. \tag{12}$$

Hence RS with $M = C^\pi_\infty$ has the same conditional output distribution π on $H_{C^\pi_\infty}$. Note that this does not imply equality of the full output distributions, since their conditional laws on $H^c_{C^\pi_\infty}$ may differ. Let $\delta_N := \Pr(H^c_{C^\pi_\infty}) = (1 - \frac{1}{C^\pi_\infty})^N$. When $\delta_N > 0$, let $F^U_{N,C^\pi_\infty} := \mathcal{L}(\widehat{Y}_U \mid H^c_{C^\pi_\infty})$. It follows that

$$\begin{aligned} P^U_N &= (1 - \delta_N)\pi + \delta_N F^U_{N,C^\pi_\infty}, \\ D_{\mathrm{TV}}(P^U_N, \pi) &= \delta_N D_{\mathrm{TV}}(F^U_{N,C^\pi_\infty}, \pi) \leq \delta_N. \end{aligned} \tag{13}$$

Here F^U_{N,C^π_∞} is simply UR's output conditioned on $H^c_{C^\pi_\infty}$; it is not an additional fallback step. If $\delta_N = 0$, then $P^U_N = \pi$ directly. In particular, Section 4 gives an example where UR and RS with $M = C^\pi_\infty$ also agree on the event of no acceptance, so their full laws coincide. Equation (13) gives exponential convergence whenever $C^\pi_\infty < \infty$. The next corollary places this bound alongside a second-moment bound and identifies the exact minimax error over the class $C^\pi_\infty \leq C$.

Corollary 1 (Error bounds and minimax optimality)**.** *Let $C^\pi := \mathbb{E}_\mu[w(Y)^2]$. For every $N \geq 1$,*

$$D_{\mathrm{TV}}(P^U_N, \pi) \leq \min\left\{1, \frac{2C^\pi}{eN}, \left(1 - \frac{1}{C^\pi_\infty}\right)^N\right\}. \tag{14}$$

If either C^π or C^π_∞ is infinite, its corresponding term is interpreted as the trivial bound 1. For every $C \geq 1$,

$$\inf_{\mathsf{A}} \sup_{(\pi,\mu):C^\pi_\infty \leq C} D_{\mathrm{TV}}(P^{\mathsf{A}}_N, \pi) = \left(1 - \frac{1}{C}\right)^N. \tag{15}$$

Here the infimum is over procedures A that return one of the N observed proposals. The supremum is over pairs satisfying the stated weight bound.

Appendix D gives both bounds: the lower bound follows from the fact that a selector cannot return a target state that does not appear among the N proposals.

3 Comparison with budget-calibrated RS

Section 2 compares UR with fixed threshold RS guarantees. We now ask whether an RS procedure that estimates the unknown normalizer from pilot samples and calibrates its threshold to the available proposal budget can match UR's actual sampling accuracy. A natural construction comes from the budget relation in Rohatgi et al. (2025, Algorithm 5); we call this specific reparameterization *budget-calibrated rejection sampling*.

The algorithm starts by estimating Z by averaging the unnormalized weights of n pilot proposals and then applies RS to fresh candidates. For a maximum proposal budget $N = 2n + 1 \geq 3$ and calibration parameter $\delta \in (0, 1)$, invert the relation $n = 4M \log(4/\delta)$ to set

$$M_{N,\delta} := \frac{n}{4\log(4/\delta)}, \qquad \widehat{Z} := \frac{1}{n}\sum_{i=1}^{n} w_u(Y_i). \tag{16}$$

Here δ is a calibration input; Appendix C.1 records the sufficient condition under which it upper-bounds the TV error. The procedure then tests $Y_{n+1}, \ldots, Y_{2n}$ in order, accepting Y_i with probability $\min\{w_u(Y_i)/(M_{N,\delta}\widehat{Z}), 1\}$. Budget-calibrated RS then returns the first accepted candidate, or the fresh proposal Y_{2n+1} if all n proposals are rejected. Let $P^R_{N,\delta}$ denote its output distribution. For a general integer budget $N \geq 3$, take $n = \lfloor (N-1)/2 \rfloor$ and leave at most one proposal unused.

3.1 Exponential separation for fixed target–proposal pair

We first show that the gap can grow exponentially with the budget on a single (π, μ), even after optimizing the calibration.

Proposition 2. *Let $\mu = (1/2, 1/2)$ and $\pi = (1/4, 3/4)$. For every odd total proposal budget $N \geq 3$,*

$$\inf_{\delta \in (0,1)} D_{\mathrm{TV}}(P^R_{N,\delta}, \pi) \geq \frac{1}{12} 2^{-(N-1)/2}, \qquad D_{\mathrm{TV}}(P^U_N, \pi) = \frac{3}{4} 3^{-N}. \tag{17}$$

The calibration may depend on (π, μ) and budget, but is chosen before observing the pilot.

Why calibration leaves error. Write $n = (N-1)/2$, $M = M_{N,\delta}$, and $\overline{W}_n := n^{-1} \sum_{i=1}^n w(Y_i)$, so the normalized rejection threshold is $T = M\overline{W}_n$. The normalized importance weights are $w(0) = 1/2$ and $w(1) = 3/2$. The lower bound reflects an unavoidable tradeoff: If $M \leq 2$, then on a rare pilot event the threshold falls below the larger weight $3/2$, so clipping compresses the preference for state 1. If $M > 2$, then clipping disappears on a constant-probability pilot event, but the acceptance probability becomes too small, so complete rejection exposes the fresh proposal. Therefore, either choice underweights state 1 and leaves error of order 2^{-n}.

We first make the clipping mechanism explicit. At threshold T, the acceptance probabilities of states 0 and 1 are $a_0(T) = \min\{1/(2T), 1\}$, and $a_1(T) = \min\{3/(2T), 1\}$. Since clipping can only reduce the ratio between the larger and smaller acceptance probabilities, $a_1(T)/a_0(T) \leq 3$. As μ is uniform, the accepted proposal therefore satisfies $Q^{\mathrm{RS}}_T(1) = \frac{a_1(T)}{a_0(T)+a_1(T)} \leq \frac{3}{4} = \pi(1)$. Also, note that the fresh proposal used after complete rejection assigns state 1 probability $1/2$. Hence every pilot-conditioned output underweights state 1, so averaging over the pilot cannot cancel the error.

When $M \leq 2$, consider the event that all n pilot weights equal $1/2$, which has probability 2^{-n}. This gives $T = M/2 \leq 1$. State 1 is always accepted, whereas state 0 is accepted with probability $a_0(T) = \min\{1, 1/(2T)\} \geq 1/2$. The probability of state 1 conditional on acceptance is therefore $1/(1 + a_0(T)) \leq 2/3$. The fallback assigns state 1 probability $1/2$, so the entire conditional output assigns state 1 probability at most $2/3$. Hence $D_{\mathrm{TV}}(P^R_{N,\delta}, \pi) \geq 2^{-n}(\frac{3}{4} - \frac{2}{3}) = \frac{2^{-n}}{12}$. For $M > 2$, the opposite failure occurs: on a constant probability pilot event there is no clipping, but each proposal is rejected with probability at least $1/2$, so complete rejection leaves an $\Omega(2^{-n})$ error through the fresh proposal. See Appendix C.3 for the full proof.

For UR, on the other hand, a score stays below $3/2$ exactly when $Y_i = 0$ and $U_i > 1/3$, with probability $(1/2)(2/3) = 1/3$. All N scores stay below with probability 3^{-N}, on which UR returns 0; on the complement its law is π by equation 12. Its exact TV error is therefore $(3/4)3^{-N}$.

Beyond a difference in exponential rates. Despite an error ratio of at least $\frac{1}{3}(9/2)^{(N-1)/2}$, both methods have $\Theta(\log(1/\xi))$ sample complexity for TV tolerance $\xi \downarrow 0$ with optimal tuning; see Appendix C.3. We now fix N and ask whether calibration can recover the bounded-weight guarantee of Corollary 1 within a pair-independent factor.

Corollary 2 (No uniform constant-factor guarantee)**.** *Fix an odd $N \geq 13$. There is no finite constant K_N, independent of (π, μ), satisfying*

$$\inf_{\delta \in (0,1)} D_{\mathrm{TV}}(P^R_{N,\delta}, \pi) \leq K_N \left(1 - \frac{1}{C^\pi_\infty}\right)^N$$

for all strictly positive target–proposal pairs on finite state spaces.

Proof. Proposition 4 in Appendix C.4 gives a binary family (π_η, μ_η) with lower weight proposal probability η and $C^{\pi_\eta}_\infty = (1 - \eta/2)^{-1}$. Therefore, $1 - 1/C^{\pi_\eta}_\infty = \eta/2$, and the optimized error $\Theta_N(\eta^2)$ divided by the benchmark $(\eta/2)^N$ is $\Theta_N(\eta^{2-N}) \to \infty$ as $\eta \downarrow 0$ at fixed N. □

3.2 General comparison

The comparison is not limited to these binary examples: UR has no greater TV error than budget-calibrated RS for every (π, μ).

Theorem 2 (Marginal error relative to budget-calibrated RS)**.** *For every pair satisfying the positive-weight assumptions of Section 1, every $n \geq 1$, and every $\delta \in (0, 1)$,*

$$D_{\mathrm{TV}}(P^U_{2n+1}, \pi) \leq D_{\mathrm{TV}}(P^U_{n+1}, \pi) \leq D_{\mathrm{TV}}(P^R_{2n+1,\delta}, \pi).$$

More generally,

$$D_f(P^U_{2n+1} \| \pi) \leq D_f(P^U_{n+1} \| \pi) \leq D_f(P^R_{2n+1,\delta} \| \pi) \tag{18}$$

for every convex $f : (0, \infty) \to \mathbb{R}$ with $f(1) = 0$. Here $D_f(P\|\pi) := \int f(\mathrm{d}P/\mathrm{d}\pi)\mathrm{d}\pi$ denotes f-divergence (Sason & Verdú, 2016).

Proof. See Appendix C.2. □

Consequently, UR needs only $n + 1$ proposals, no more than the baseline uses in total on any run. Giving UR the full $2n + 1$ budget cannot increase its divergence from π.

4 Comparison with sampling importance resampling (SIR)

A standard threshold-free alternative is SIR, which selects candidate i with probability $w_u(Y_i)/\sum_j w_u(Y_j)$. Let $\widehat{Y}_{\mathrm{SIR}}$ denote the returned candidate and $P^{\mathrm{SIR}}_N := \mathcal{L}(\widehat{Y}_{\mathrm{SIR}})$ its marginal output distribution. We first illustrate the difference on a binary example with $N = 2$, and then establish the general comparison at the same proposal budget.

4.1 Example where UR and SIR differ

Consider the binary family where $\mu = (\frac{1}{2}, \frac{1}{2})$ and $\pi = (\varepsilon, 1 - \varepsilon)$. Here $0 < \varepsilon < \frac{1}{2}$. At $\varepsilon = 1/4$, this is the fixed pair in Proposition 2. State 1 has the larger target probability, and the two importance weights are $(w_L, w_H) = (2\varepsilon, 2(1 - \varepsilon))$, where $C^\pi_\infty = 2(1 - \varepsilon)$. Fix $N = 2$ and write $B := (Y_1, Y_2)$. Each of $00, 01, 10, 11$ has probability $\frac{1}{4}$. Every rule returning an observed proposal must return state 0 on 00 and state 1 on 11. The methods can therefore differ only when both states are present.

On either mixed batch, SIR chooses state 1 with probability $w_H/(w_L + w_H) = 1 - \varepsilon$. For UR, let U_L, U_H be the independent uniform variables attached to the two candidates. The candidate in state 0 is selected exactly when $\frac{w_L}{U_L} > \frac{w_H}{U_H}$, equivalent to $U_L < \frac{\varepsilon}{1-\varepsilon} U_H$. Conditioning on $U_H = u$ yields:

$$\Pr(\widehat{Y}_U = 0 \mid B = 01) = \int_0^1 \Pr\left(U_L < \frac{\varepsilon}{1 - \varepsilon} u\right) \mathrm{d}u = \frac{\varepsilon}{2(1 - \varepsilon)}. \tag{19}$$

Hence UR chooses state 1 with probability $1 - \frac{\varepsilon}{2(1-\varepsilon)}$ on either mixed batch. RS with $M = w_H = C^\pi_\infty$ accepts state 0 with probability $\frac{\varepsilon}{1-\varepsilon}$ and state 1 with probability one. It returns the first accepted proposal, or an observed proposal if both are rejected. We note that this reference exploits the true unnormalized envelope $Z w_H$. Averaging over the four equally likely batches in turn gives

$$\Pr(\widehat{Y}_U = 1) = \frac{3}{4} - \frac{\varepsilon}{4(1 - \varepsilon)}, \qquad \Pr(\widehat{Y}_{\mathrm{SIR}} = 1) = \frac{3}{4} - \frac{\varepsilon}{2}. \tag{20}$$

For RS with $M = w_H$, averaging the probabilities of returning state 1 over the two mixed batches gives $1 - \frac{\varepsilon}{2(1-\varepsilon)}$, matching UR's probability. Hence, the two output distributions agree after averaging.

Where the improvement comes from. On a mixed batch, SIR chooses state 1 with its target probability $1 - \varepsilon$. This does not make the marginal output exact, because the homogeneous batches 00 and 11 contribute unequal errors. By contrast, UR partially compensates for this imbalance by choosing state 1 more often when both states are available. Consequently,

$$\underbrace{\frac{1 - 2\varepsilon}{4}}_{\text{SIR error}} = \underbrace{\frac{\varepsilon(1 - 2\varepsilon)}{4(1 - \varepsilon)}}_{\text{reduction under UR}} + \underbrace{\frac{(1 - 2\varepsilon)^2}{4(1 - \varepsilon)}}_{\text{UR error}}. \tag{21}$$

When $\varepsilon = \frac{1}{4}$, the identity gives $\frac{1}{8} = \frac{1}{24} + \frac{1}{12}$.

Why RS with $M = w_H = C_\infty^\pi$ agrees with UR. The calculation above shows that UR and RS with the exact envelope have the same output distribution. From the threshold view, on H_{w_H} both conditional output distributions are π. Moreover, since any state 1 proposal has score above w_H, $H_{w_H}^c$ can occur only on batch 00, where both rules return state 0. Appendix F.2 extends this equality to general two-level weights.

4.2 Comparison for general distributions

Theorem 3 (Marginal error relative to SIR)**.** *Under the positive weight assumptions in Section 1, for every fixed $N \geq 1$,*

$$D_{\mathrm{TV}}(P_N^U, \pi) \leq D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi), \qquad D_f(P_N^U \| \pi) \leq D_f(P_N^{\mathrm{SIR}} \| \pi) \tag{22}$$

for every convex $f : (0, \infty) \to \mathbb{R}$ with $f(1) = 0$, allowing infinite divergences.

Why the comparison holds. Appendix E proves that, for every $t > 0$, $\Pr(w(\widehat{Y}_{\mathrm{SIR}}) > t) \leq \Pr(w(\widehat{Y}_U) > t) \leq \pi(w > t)$. Hence, relative to SIR, UR shifts probability toward larger importance weights without overshooting the target tail probabilities. We highlight that this alone would not imply a TV comparison. A second property is that the density of the UR output relative to π is nonincreasing in the weight. Its excess probability is therefore at smaller weights and its shortfall at larger weights, so the threshold comparison controls the set attaining its TV error. Appendix E gives the formal argument and extends it to every convex f-divergence.

Beyond two proposals. We now return to the binary example and let the proposal budget grow.

Proposition 3. *Fix $\varepsilon \in (0, 1/2)$, and let $\mu = (1/2, 1/2)$ and $\pi = (\varepsilon, 1 - \varepsilon)$. For every $N \geq 1$,*

$$D_{\mathrm{TV}}(P_N^U, \pi) = (1 - \varepsilon) \left(\frac{1 - 2\varepsilon}{2(1 - \varepsilon)} \right)^N, \qquad D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi) = \frac{2\varepsilon(1 - \varepsilon)(1 - 2\varepsilon)}{N} + O_\varepsilon(N^{-3/2}). \tag{23}$$

Proof. See Appendix F.1. □

The separation persists even when both states appear in nearly every batch. SIR retains the leading bias of $O(1/N)$ caused by random candidate counts and normalization, whereas UR's remaining error is exponentially small. Appendix F.3 shows that this phenomenon extends to every fixed nontrivial pair with bounded positive weights.

5 Uniqueness of the selection probabilities

Sections 3 and 4 compare UR's output law with budget-calibrated RS and SIR, respectively. These comparisons show advantages over particular alternatives, but leave a structural question: *can different selection probabilities satisfy the RS guarantee for every target–proposal pair?* Within the class of sampling algorithms restricted as below, we prove that the answer is no: the guarantee determines the probabilities on every fixed positive weight vector, and they must coincide with UR's.

What can the sampler use? Fix $N \geq 2$ and let the algorithm A return one of the N candidates. We restrict our attention to a class of selection rules satisfying two structural properties: First, its selection probabilities depend only on the observed weights up to a common positive factor, not on candidate states or other information about (π, μ). The same rule must be used for every pair. Second, changing the order of the candidates must only reorder their selection probabilities. For instance, weights $(1, 2)$ and $(10, 20)$ must give the same probabilities, while $(2, 1)$ must exchange them. We note that both UR and SIR satisfy these restrictions. For a fixed weight vector $w_{1:N} := (w_1, \ldots, w_N)$, let $p_i^{\mathsf{A}}(w_{1:N})$ be the probability of selecting candidate i over the sampler's random choices. The restrictions require, for every $c > 0$ and permutation σ,

$$p_i^{\mathsf{A}}(c w_{1:N}) = p_i^{\mathsf{A}}(w_{1:N}), \qquad p_{\sigma(i)}^{\mathsf{A}}(\sigma w_{1:N}) = p_i^{\mathsf{A}}(w_{1:N}), \tag{24}$$

where $(\sigma w_{1:N})_{\sigma(i)} = w_i$. Write $p_i^U(w_{1:N})$ for UR's probability on the same vector and P_N^{A} for the output distribution after averaging over the proposal draws. The following theorem requires the bounded weight guarantee from Corollary 1.

Theorem 4. *Fix $N \geq 2$ and suppose A satisfies the information restriction above and equation 24. If a finite constant L_N, independent of (π, μ), satisfies*

$$D_{\mathrm{TV}}(P_N^{\mathsf{A}}, \pi) \leq L_N \left(1 - \frac{1}{C_\infty^\pi} \right)^N \tag{25}$$

for every target and proposal on a finite state space with strictly positive probabilities, then

$$p_i^{\mathsf{A}}(w_{1:N}) = p_i^{U}(w_{1:N}) \qquad \textit{for every } w_{1:N} \in (0,\infty)^N \textit{ and } i \in [N].$$

Conversely, UR satisfies the bound with $L_N = 1$.

Proof. See Appendix G. We note that the requirement ranges over all (π, μ) at fixed N, so a different rule may have smaller error on a particular pair without satisfying this requirement. □

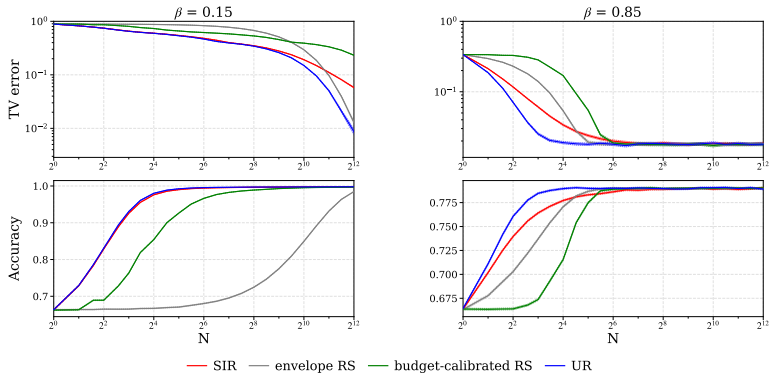


Figure 1: TV error (top) and ground-truth accuracy (bottom) w.r.t. sampling budget N on GSM8K Problem #258, for $\beta = 0.15$ (left) and $\beta = 0.85$ (right). The bands indicate 95% confidence intervals.

6 Test-time scaling experiments

We complement our theoretical findings with controlled LLM experiments on math reasoning tasks where the samplers approximate a reward-tilted target policy, also known as soft BoN (Aminian et al., 2026; Verdun et al., 2025). We evaluate both sampling fidelity, measured by TV distance, and ground-truth answer accuracy.

Experimental setup and evaluation. For each prompt x, we pre-generate a fixed response pool $\mathcal{A}_x^\star$ of $N^\star$ responses from `Llama-3.2-3B-Instruct` (Grattafiori et al., 2024). Let $\widehat{\mu}(y \mid x)$ denote the empirical distribution on $\mathcal{A}_x^\star$. For each sampling budget N, we draw N candidates i.i.d. from $\widehat{\mu}(\cdot \mid x)$ and apply each sampling procedure. Each response y is scored by the OASST reward model (Köpf et al., 2023), yielding $\widehat{r}(x, y)$. Motivated by the standard KL-regularized RLHF objective (Jaques et al., 2017; 2020; Rafailov et al., 2023), we define the reward-tilted target policy $\pi_\beta(y \mid x) \propto \widehat{\mu}(y \mid x) \exp(\frac{\widehat{r}(x,y)}{\beta}),\ y \in \mathcal{A}_x^\star$, so that the unnormalized importance weight for each sampler is $w_u(x, y) = \exp(\frac{\widehat{r}(x,y)}{\beta})$. We compare four methods: UR (OURS), ENVELOPE RS with threshold equal to the maximum weight in $\mathcal{A}_x^\star$, BUDGET-CALIBRATED RS, adapted from Rohatgi et al. (2025, Algorithm 5), and SIR. We evaluate on GSM8K (Cobbe et al., 2021), varying the sampling budget N and temperature parameter β. Estimates of the TV error $D_{\mathrm{TV}}(P_N, \pi_\beta)$ and ground-truth accuracy are based on 10^6 Monte Carlo trials, with ground-truth accuracy defined as the probability that the selected response has the correct final answer. We also evaluate on MATH500 (Hendrycks et al., 2021; Lightman et al., 2024), where we observe the same trends; additional results and implementation details are given in Appendix I.

Results and analysis. Figure 1 reports TV error (top) and ground-truth accuracy (bottom) for the same GSM8K problem under two values of β. Across both settings, UR achieves lower TV error than budget-calibrated RS and SIR, consistent with Theorems 2 and 3, while remaining comparable to envelope RS without requiring threshold tuning. Before reaching the Monte Carlo floor, UR exhibits substantially faster decay in TV error than budget-calibrated RS and SIR. The accuracy curves show how these distributional differences are reflected in ground-truth accuracy. For $\beta = 0.15$, the methods exhibit visible differences in TV error at larger sampling budgets, while their accuracies are already close to the common target accuracy. This indicates that the TV upper bound on the accuracy gap is loose in this regime. Specifically, accuracy probes only the probability of the correct answer event, so $\left|\mathrm{Acc}(P_N; x) - \mathrm{Acc}(\pi_\beta; x)\right| \le D_{\mathrm{TV}}(P_N, \pi_\beta)$. For $\beta = 0.85$, the accuracy gaps track the TV errors more closely, with UR approaching the common target accuracy at a smaller budget. Additional results across prompts and values of β, including MATH500, are given in Appendix I.3.

AI USE STATEMENT

In this work, we used generative AI tools to [help formulate mathematical claims, provide ingredients for proving mathematical claims, assist in the writing of proofs, design and provide feedback on research methodology and experiments, implement methods, and interpret experimental results]. We have not used generative AI tools for [generating synthetic datasets or qualitative and thematic data analysis], and [dataset cleaning and reformatting] were not applicable to this work. Additionally, we used generative AI tools to [suggest experimental parameters, create or edit software code, draft and edit parts of the manuscript, summarize and analyze existing literature, identify relevant literature, brainstorm research ideas, search for information, and improve the readability and organization of the paper]. We have reviewed all AI-assisted work. [All AI-assisted mathematical claims and proofs were independently verified and revised by at least two authors. All cited literature identified or discussed with AI assistance was manually checked against the original sources. AI-assisted code was verified and tested for correctness by two authors]. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Sergios Agapiou, Omiros Papaspiliopoulos, Daniel Sanz-Alonso, and Andrew M. Stuart. Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32(3):405–431, 2017. doi: 10.1214/17-STS611.

Gholamali Aminian, Idan Shenfeld, Amir R. Asadi, Ahmad Beirami, and Youssef Mroueh. Best-of-n through the smoothing lens: KL divergence and regret analysis. In *The Fourteenth International Conference on Learning Representations*, 2026.

Adam Block and Yury Polyanskiy. The sample complexity of approximate rejection sampling with applications to smoothed online learning. In *Proceedings of the Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 228–273, 2023. URL https://proceedings.mlr.press/v195/block23a.html.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

George Deligiannidis, Pierre E. Jacob, El Mahdi Khribch, and Guanyang Wang. On importance sampling and independent Metropolis–Hastings with an unbounded weight function. *arXiv preprint arXiv:2411.09514v3*, 2026. URL https://arxiv.org/abs/2411.09514. Version 3, July 2026. Originally posted in 2024.

Nick Duffield, Carsten Lund, and Mikkel Thorup. Priority sampling for estimation of arbitrary subset sums. *Journal of the ACM*, 54(6):32, 2007. doi: 10.1145/1314690.1314696.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*, 2021.

Audrey Huang, Adam Block, Qinghua Liu, Nan Jiang, Akshay Krishnamurthy, and Dylan J. Foster. Is best-of-N the best of them? coverage, scaling, and optimality in inference-time alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 25075–25126, 2025. URL https://proceedings.mlr.press/v267/huang25c.html.

Pierre E. Jacob, Christian P. Robert, and Murray H. Smith. Using parallel computation to improve independent Metropolis–Hastings based estimation. *Journal of Computational and Graphical Statistics*, 20(3):616–635, 2011. doi: 10.1198/jcgs.2011.10167.

Natasha Jaques, Shixiang Gu, Dzmitry Bahdanau, José Miguel Hernández-Lobato, Richard E Turner, and Douglas Eck. Sequence tutor: Conservative fine-tuning of sequence generation models with kl-control. In *International Conference on Machine Learning*, pp. 1645–1654. PMLR, 2017.

Natasha Jaques, Judy Hanwen Shen, Asma Ghandeharioun, Craig Ferguson, Agata Lapedriza, Noah Jones, Shixiang Gu, and Rosalind Picard. Human-centric dialog training via offline reinforcement learning. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pp. 3985–4003, 2020.

Andreas Köpf, Yannic Kilcher, Dimitri Von Rütte, Sotiris Anagnostidis, Zhi Rui Tam, Keith Stevens, Abdullah Barhoum, Duc Nguyen, Oliver Stanley, Richárd Nagyfi, et al. Openassistant conversations-democratizing large language model alignment. *Advances in neural information processing systems*, 36:47669–47681, 2023.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations*, volume 2024, pp. 39578–39601, 2024.

Jingbo Liu, Paul Cuff, and Sergio Verdú. E_γ-resolvability. *IEEE Transactions on Information Theory*, 63(5):2629–2658, 2017.

Jun S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996. doi: 10.1007/BF00162521.

Kerrie L. Mengersen and Richard L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24(1):101–121, 1996.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

Dhruv Rohatgi, Abhishek Shetty, Donya Saless, Yuchen Li, Ankur Moitra, Andrej Risteski, and Dylan J. Foster. Taming imperfect process verifiers: A sampling perspective on backtracking. *arXiv preprint arXiv:2510.03149*, 2025. URL https://arxiv.org/abs/2510.03149.

Donald B Rubin. Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm. *Journal of the American Statistical Association*, 82(398):542–543, 1987.

Igal Sason and Sergio Verdú. f-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, 2016.

Øivind Skare, Erik Bølviken, and Lars Holden. Improved sampling-importance resampling and reduced bias importance sampling. *Scandinavian Journal of Statistics*, 30(4):719–737, 2003.

Adrian FM Smith and Alan E Gelfand. Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician*, 46(2):84–88, 1992.

Claudio Mayrink Verdun, Alex Oesterling, Himabindu Lakkaraju, and Flavio P. Calmon. Soft best-of-n sampling for model alignment. In *2025 IEEE International Symposium on Information Theory (ISIT)*, volume 00, pp. 1–6. Institute of Electrical and Electronics Engineers (IEEE), June 2025. doi: 10.1109/isit63088.2025.11195541.

Guanyang Wang. Exact convergence analysis of the independent Metropolis–Hastings algorithms. *Bernoulli*, 28(3):2012–2033, 2022. URL https://arxiv.org/abs/2008.02455.

A Related work

Approximate rejection sampling. Our problem is closely related to the approximate rejection sampling (RS) framework of Block & Polyanskiy (2023): given N independent samples from a proposal distribution μ, the goal is to select one whose marginal distribution is close in total variation to a target distribution π. They characterize the minimax sample complexity of this problem under f-divergence constraints between target and proposal. Their upper bounds are obtained by truncating the target according to a likelihood ratio threshold and applying RS to the resulting truncated distribution. Our focus is complementary. Rather than optimizing sample complexity over a divergence class, we study a single parameter-free selection rule that uses only the observed unnormalized importance weights. In particular, UR requires neither the normalizing constant of the target nor an envelope bound or clipping threshold as an algorithmic input. The clipped RS formulation has also been used in recent inference-time sampling work (Huang et al., 2025).

Priority sampling. UR uses the same selection rule as a special case of classical priority sampling (Duffield et al., 2007). Given a fixed collection of weighted items, priority sampling assigns each item an independent priority w_i/U_i and retains the items with the largest priorities. When only one item is retained, this reduces exactly to the UR selection rule, $I = \arg\max_i \frac{w_i}{U_i}$. Classical priority sampling, however, is designed for a different statistical task: retain the top-k items with the largest priorities. Together with the $(k+1)$st largest priority as a threshold, the retained items are assigned adjusted weights that yield unbiased estimators of arbitrary subset sums (i.e., $\sum_{i\in A} w_i$ for subsets A of the population). Consequently, the classical analysis is concerned with inclusion, unbiased weight estimation, and the variance of subset-sum estimators. In our setting, the items themselves are random, $Y_i \overset{\text{iid}}{\sim} \mu$, and we instead study the marginal distribution of the selected candidate Y_I as an approximation to the target distribution π. Therefore, while UR is a special case of the classical priority selection rule, our objective and analysis concern the induced sampling distribution rather than subset-sum estimation.

Sampling importance resampling and importance sampling. Sampling importance resampling (SIR) also generates a candidate pool from μ and selects a candidate using importance weights. Conditional on $Y_{1:N}$, SIR selects index i with probability $\frac{w_u(Y_i)}{\sum_{j=1}^N w_u(Y_j)}$, and therefore coincides with sampling from the self-normalized empirical importance distribution. For finite N, the resulting marginal distribution is generally different from π, although it converges to the target as the candidate pool grows. Skare et al. (2003) propose adjusted resampling probabilities to reduce this finite sample bias and compare the resulting method with independent Metropolis–Hastings. Related work on self-normalized importance sampling studies the bias and stability of estimators based on the same normalized importance weights; see, e.g., Agapiou et al. (2017) and Deligiannidis et al. (2026). These works primarily concern estimation of expectations under π, whereas our primary object is the marginal distribution of a single selected candidate and its total variation distance from π.

Independent Metropolis–Hastings. Independent Metropolis–Hastings (IMH) also combines proposals from a fixed distribution μ with the importance ratio between π and μ. Unlike the one-shot selection rules considered here, IMH constructs a Markov chain with stationary distribution π. Its convergence properties and relationship with rejection and importance sampling have been studied extensively. Liu (1996) compares metropolized independent sampling with rejection sampling and importance sampling, while Mengersen & Tweedie (1996) establish classical convergence results for independence samplers. More recent work gives sharp convergence analyses (Wang, 2022) and studies IMH under unbounded weight functions (Deligiannidis et al., 2026). The independence of its proposals also permits parallel computation; Jacob et al. (2011) exploit batches of independent proposals to improve IMH-based estimation. Our setting differs in that no Markov chain is run: given a fixed budget of N independently generated and scored candidates, the algorithm returns one of them directly. This one-shot formulation is particularly natural for inference-time sampling, where candidate generation and scoring can be parallelized and the desired output is a single response rather than a sequence from a stationary Markov chain.

B Proof of the RS bound

Section 2 assembles Theorem 1 from the acceptance calculation and the comparison of conditional errors. We first prove the comparison in Lemma 2. The remaining subsections justify the conditioning,

give the exact clipping error, and extend the results to zero weights. Write $W := w(Y)$ for $Y \sim \mu$, with $0 < W < \infty$ almost surely until Appendix B.5.

B.1 Proof of Lemma 2

Fix $N \geq 1$ and $M > 0$. Let r_M and q_M be functions of the scalar weight such that

$$\frac{\mathrm{d}Q^U_{N,M}}{\mathrm{d}\pi}(y) = r_M(w(y)), \qquad \frac{\mathrm{d}Q^{\mathrm{RS}}_M}{\mathrm{d}\pi}(y) = q_M(w(y)).$$

We suppress the fixed N in r_M. Integrating the joint law equation 28 over $s \geq M$ gives, for $x > 0$,

$$r_M(x) = \frac{N}{\Pr(H_M)} \int_{\max\{x,M\}}^{\infty} \frac{G(s)^{N-1}}{s^2}\mathrm{d}s, \quad q_M(x) = \frac{\min\{x, M\}}{A_M x},$$

where $G(s) := \Pr(W/U \leq s)$ for an independent $U \sim \mathrm{Unif}(0, 1)$. Both functions are positive and finite, and their compositions with w integrate to one under π. The comparison lemma below requires two properties: r_M is nonincreasing, and r_M/q_M is nondecreasing. The first follows because the lower integration limit cannot decrease with x. For the second, both functions are constant on $x \leq M$. For $x > M$, their ratio is a positive constant times

$$x \int_x^{\infty} \frac{G(s)^{N-1}}{s^2}\mathrm{d}s = \int_0^1 G(x/u)^{N-1}\mathrm{d}u, \tag{26}$$

using $u = x/s$. The right side is nondecreasing in x because G is nondecreasing. The two expressions for the ratio agree at $x = M$, proving the second property. Lemma 3, with $P = \pi$, $Q = Q^{\mathrm{RS}}_M$, and $R = Q^U_{N,M}$, now proves Lemma 2.

Lemma 3 (A comparison of densities)**.** *Let $R \ll Q \ll P$ be probability measures whose densities have the form $\mathrm{d}Q/\mathrm{d}P = q(W) > 0$ and $\mathrm{d}R/\mathrm{d}P = r(W)$ for a measurable real function W. If r is nonincreasing and r/q is nondecreasing on the range of W, then*

$$D_{\mathrm{TV}}(P, R) \leq D_{\mathrm{TV}}(P, Q).$$

Proof. Let $D := \{y : r(W(y)) < 1\}$, the set where the density ratio of R to P is below one. Hence $D_{\mathrm{TV}}(P, R) = P(D) - R(D)$, and it is enough to show $R(D) \geq Q(D)$. Put $v := r/q$, so $\mathrm{d}R/\mathrm{d}Q = v(W)$ and $\int v(W)\mathrm{d}Q = 1$. If $y \in D$ and $z \notin D$, then $r(W(y)) < 1 \leq r(W(z))$. Since r is nonincreasing, $W(y) > W(z)$, and the monotonicity of v gives $v(W(y)) \geq v(W(z))$. Thus, when Q is multiplied by the density ratio $v(W)$ to obtain R, the factors on D are at least as large as those outside D. Because both measures have total mass one, this implies

$$\begin{aligned} R(D) - Q(D) &\overset{\text{(i)}}{=} R(D)Q(D^c) - Q(D)R(D^c) \\ &\overset{\text{(ii)}}{=} \int_D \int_{D^c} [v(W(y)) - v(W(z))]Q(\mathrm{d}z)Q(\mathrm{d}y) \\ &\overset{\text{(iii)}}{\geq} 0. \end{aligned}$$

Here (i) uses total mass one, (ii) substitutes $\mathrm{d}R = v(W)\mathrm{d}Q$, and (iii) uses the pointwise comparison above. The absolute integrand is bounded by $v(W(y)) + v(W(z))$, so it is integrable. It follows that

$$D_{\mathrm{TV}}(P, R) = P(D) - R(D) \leq P(D) - Q(D) \leq D_{\mathrm{TV}}(P, Q),$$

where the last inequality is the definition of TV. □

B.2 Proof of Proposition 1

We first consider the state and score of a single candidate, then account for the other candidates. Given $Y_i = y$, the change of variables $u = w(y)/s$ has absolute derivative $w(y)/s^2$ and requires $s > w(y)$. Since $w(y)\mu(\mathrm{d}y) = \pi(\mathrm{d}y)$, the joint distribution is

$$\Pr(Y_i \in \mathrm{d}y, S_i \in \mathrm{d}s) = \frac{\mathbb{1}\{w(y) < s\}}{s^2}\pi(\mathrm{d}y)\mathrm{d}s. \tag{27}$$

To be selected, this candidate must also have a larger score than the other $N - 1$ candidates. Let $G(s) := \Pr(S_i \leq s)$ be the distribution function of one score. Independence yields the factor $G(s)^{N-1}$. Crucially, this factor depends on s but not on y. The scores are atomless, so ties have

probability zero, and summing over the N possible selected indices yields:

$$\Pr(\widehat{Y}_U \in \mathrm{d}y, S_{(N)} \in \mathrm{d}s) = \frac{NG(s)^{N-1}}{s^2} \mathbb{1}\{w(y) < s\}\pi(\mathrm{d}y)\mathrm{d}s. \tag{28}$$

Integrating over y gives the density of $S_{(N)}$ as $\frac{NG(s)^{N-1}}{s^2}\pi(w < s)$. Therefore, for every measurable D and almost every such s,

$$\Pr(\widehat{Y}_U \in D \mid S_{(N)} = s) = \frac{\pi(D \cap \{w < s\})}{\pi(w < s)} = \pi(D \mid w < s),$$

proving equation 9; if $s > C_\infty^\pi$, $\pi(w < s) = 1$, so equation 10 holds. This completes the proof.

B.3 Conditioning on the largest score

The following identities are consequences of the joint law in Proposition 1. They also make precise the conditioning on a continuously distributed score.

Lemma 4 (Integrating the joint distribution)**.** *Integrating equation 28 over $s \geq M$ and dividing by $\Pr(H_M)$ gives r_M above. Integrating over all positive scores gives*

$$\frac{\mathrm{d}P_N^U}{\mathrm{d}\pi}(y) = N \int_{w(y)}^{\infty} \frac{G(s)^{N-1}}{s^2}\mathrm{d}s.$$

If $C_\infty^\pi < \infty$, then for every measurable state set D and Borel score set $B \subset (C_\infty^\pi, \infty)$,

$$\Pr(\widehat{Y}_U \in D, S_{(N)} \in B) = \pi(D)\Pr(S_{(N)} \in B). \tag{29}$$

Proof. The first two identities follow by integrating the nonnegative joint density. For $s > C_\infty^\pi$, the indicator $\mathbb{1}\{w(y) < s\}$ is one for π-almost every y. The joint density therefore factors into $\pi(\mathrm{d}y)$ and the density of $S_{(N)}$, which gives equation 29 after integration over $D \times B$. □

Conditioning on $S_{(N)} = s$ is not division by the probability of this event, which is zero. Instead, the joint density verifies the conditional probability formula

$$\Pr(\widehat{Y}_U \in D \mid S_{(N)} = s) = \frac{\pi(D \cap \{w < s\})}{\pi(w < s)}$$

where $\pi(w < s) > 0$. On the remaining score values we may define it to be $\pi(D)$. Those values carry no score probability, since the score density contains the factor $\pi(w < s)$. Equation (29) also shows that, conditional on $S_{(N)} > C_\infty^\pi$, the output state has law π independently of the largest score.

B.4 Accepted samples and the event of no acceptance

Put $a_M := A_M/M$ and let J_M be the first accepted index on H_M. For $J_M = i$, the first $i-1$ candidates must be rejected and candidate i must be accepted. Thus

$$\begin{aligned}\Pr(Y_{J_M} \in D, H_M) &\overset{\text{(i)}}{=} \sum_{i=1}^{N}(1-a_M)^{i-1}a_M Q_M^{\mathrm{RS}}(D) \\ &\overset{\text{(ii)}}{=} [1-(1-a_M)^N]Q_M^{\mathrm{RS}}(D),\end{aligned}$$

where (i) follows from the independence and the accepted law of one candidate; and (ii) follows from evaluating the geometric sum. Dividing by $\Pr(H_M) = 1-(1-a_M)^N$ proves the claimed conditional law. This calculation also explains why that law does not depend on N. The clipping bound in Lemma 1 can be sharpened to an exact identity:

$$\begin{aligned}D_{\mathrm{TV}}(\pi, Q_M^{\mathrm{RS}}) &\overset{\text{(i)}}{=} \int \left(w(y) - \frac{\min\{w(y), M\}}{A_M}\right)_+ \mu(\mathrm{d}y) \\ &\overset{\text{(ii)}}{=} \int \left(w(y) - \frac{M}{A_M}\right)_+ \mu(\mathrm{d}y) \\ &\overset{\text{(iii)}}{=} \mathcal{E}_{M/A_M} \leq \mathcal{E}_M,\end{aligned} \tag{30}$$

where (i) holds since TV is the integrated positive difference of the densities. For (ii), both positive parts vanish when $w(y) \le M$, since $A_M \le 1$, and their contents agree when $w(y) > M$. Step (iii) follows due to the definition of $\mathcal{E}_M$ evaluated at the threshold M/A_M. The last inequality uses $M/A_M \ge M$ and monotonicity in the threshold. Consequently, ensuring the factor $1 - \delta_M$ in equation 8 gives $(1 - \delta_M)\mathcal{E}_{M/A_M} + \delta_M$, where $\delta_M = (1 - A_M/M)^N$.

A direct acceptance interpretation for bounded weights. Suppose $C_\infty^\pi < \infty$. For candidate i, define

$$H_i := \left\{ U_i \le \frac{w(Y_i)}{C_\infty^\pi} \right\}, \quad V_i := \frac{C_\infty^\pi U_i}{w(Y_i)} \quad \text{on } H_i.$$

For every measurable D and $0 \le v \le 1$,

$$\Pr(Y_i \in D, H_i, V_i \le v) = \int_D \frac{v w(y)}{C_\infty^\pi} \mu(\mathrm{d}y) = \frac{v\pi(D)}{C_\infty^\pi}. \tag{31}$$

After division by $\Pr(H_i) = 1/C_\infty^\pi$, this factors as $\pi(D)v$. Hence, conditional on acceptance, Y_i has law π and is independent of $V_i \sim \mathrm{Unif}(0,1)$. Conditioning on which candidates are accepted preserves independence of these pairs, because each acceptance uses only its own candidate and uniform. Among the accepted candidates, UR selects the smallest V_i, equivalently the largest score C_∞^π/V_i. This choice uses variables independent of the accepted states, so the selected state still has law π. Together with the independent probability $1/C_\infty^\pi$ of each acceptance, this proves equation 12 directly.

B.5 Zero importance weights

Now allow $w \ge 0$ with $\mathbb{E}_\mu w = 1$. If all observed weights are zero, let both UR and SIR choose each of the N candidate indices with probability $1/N$. Otherwise neither rule selects a zero weight. Write $p_0 := \mu(w = 0) < 1$. The all zero event has probability p_0^N, and, if $p_0 > 0$, both conditional output laws on that event are $\mu(\cdot \mid w = 0)$. The joint density formulas still hold for positive scores, with an additional mass p_0^N at $S_{(N)} = 0$. For every $M > 0$, the conditional laws $Q_{N,M}^U$ and Q_M^{RS} put no mass on $\{w = 0\}$. Their density formulas and monotonicity properties on $\{w > 0\}$ are unchanged, so Lemma 2 still applies. The mixture argument in Theorem 1 then applies with the same bound on H_M^c. The bounded weight conclusion is unchanged, and Appendix D includes the extra zero score event in its moment argument. For the SIR comparison, the proof on a fixed batch in Appendix E applies after zero entries are removed from a batch with a positive weight. On a batch where all weights are zero, the two rules agree. On $\{w > 0\}$, the density of the UR law relative to π is still $h_N^U(w(y))$ from equation 51, but its integral is now $1 - p_0^N$. All remaining mass is on a set of target probability zero. Therefore the full probability gap is attained on

$$D := \{y : w(y) > 0, h_N^U(w(y)) < 1\}.$$

This is an upper weight set within $\{w > 0\}$, possibly the whole positive weight set. The selected-weight comparison, including the endpoint 0 where the positive masses agree, gives $P_N^U(D) \ge P_N^{\mathrm{SIR}}(D)$. Consequently,

$$D_{\mathrm{TV}}(P_N^U, \pi) = \pi(D) - P_N^U(D) \le \pi(D) - P_N^{\mathrm{SIR}}(D) \le D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi).$$

We retain the positive weight assumption for the general f-divergence statement, so no convention for singular mass is needed there.

B.6 Exponential relaxation: Derivation of equation 5

The output law $P_{N,M}^{\mathrm{RS}}$ uses the clipped acceptance rule and returns an observed candidate if all N proposals are rejected. Conditional on at least one acceptance, its law is Q_M^{RS}; on the complementary event, its TV error is at most one regardless of the specified fallback. Thus the same mixture argument as in equation 8 gives

$$D_{\mathrm{TV}}(P_{N,M}^{\mathrm{RS}}, \pi) \le \mathcal{E}_M + \left(1 - \frac{A_M}{M}\right)^N.$$

To obtain the simpler form used in the Introduction, note that $0 < A_M \le 1$, $0 \le \mathcal{E}_M < 1$, and $A_M = 1 - \mathcal{E}_M$. Then

$$\mathcal{E}_M + \left(1 - \frac{A_M}{M}\right)^N \overset{\text{(i)}}{\le} \mathcal{E}_M + \exp\left(-\frac{NA_M}{M}\right) \overset{\text{(ii)}}{\le} 2\mathcal{E}_M + A_M \exp(-N/M) \overset{\text{(iii)}}{\le} 2\mathcal{E}_M + \exp(-N/M).$$

Here (i) uses $1 - x \le \exp(-x)$. For (ii), convexity of the exponential gives

$$\exp(-A_M x) = \exp(\mathcal{E}_M \cdot 0 + A_M(-x)) \le \mathcal{E}_M + A_M \exp(-x), \qquad x := N/M.$$

(iii) uses $A_M \le 1$. This proves equation 3; applying it inside Theorem 1 proves equation 5. The term $\exp(-N/M)$ is not asserted to bound the rejection probability by itself when clipping occurs. The relaxation reallocates part of that probability bound to the clipping term.

C Comparison with budget-calibrated RS

We analyze the budget reparameterization in equation 16. Throughout this appendix, weights are positive and finite almost surely, and $N = 2n + 1 \ge 3$ is the maximum budget counting all pilot, rejection, and fallback proposals. For other integer budgets $N \ge 3$, use $n = \lfloor (N-1)/2 \rfloor$, leaving at most one proposal unused. Appendix C.1 records the conditional output law, and Appendix C.2 proves Theorem 2. Appendix C.3 proves Proposition 2 and verifies the N-dependent entries of Table 1. Appendix C.4 states and proves the complementary fixed budget local separation.

C.1 Baseline and its conditional law

Algorithm 5 of Rohatgi et al. (2025) exploits $n = 4M \log(4/\delta)$ pilot proposals, at most n fresh rejection proposals, and one fresh fallback upon complete rejection. Solving for M gives equation 16. The effective threshold is therefore $M_{N,\delta}\widehat{Z}$. Proposition F.2 of Rohatgi et al. (2025) applies when

$$M_{N,\delta} \ge 4C_\infty^\pi \quad \Longleftrightarrow \quad n \ge 16 C_\infty^\pi \log(4/\delta). \tag{32}$$

Under this sufficient condition, $D_{\mathrm{TV}}(P^R_{N,\delta}, \pi) \le \delta$. Therefore, the source provides an $O(C_\infty^\pi \log(1/\varepsilon))$ sufficient budget when $\delta = \varepsilon$ is an accuracy input. Notice that our comparisons concern the actual output law and do not assume this sufficient condition.

Condition on the pilot first. Write $M := M_{N,\delta}$ and define

$$\overline{W}_n := \frac{1}{n}\sum_{i=1}^n w(Y_i), \qquad T := M\overline{W}_n = \frac{M\widehat{Z}}{Z}, \qquad r_t := 1 - \frac{A_t}{t}.$$

The normalized threshold T is an analysis variable; the algorithm does not know Z. Conditional on the pilot, a fresh trial is accepted with probability A_T/T and its accepted law is Q_T^{RS}. All n trials fail with probability r_T^n, independently of the fresh fallback, whose law is μ. Therefore

$$P^R_{N,\delta} = \mathbb{E}\left[(1 - r_T^n)Q_T^{\mathrm{RS}} + r_T^n \mu\right]. \tag{33}$$

We note that the expectation is over the pilot. If $T \ge C_\infty^\pi$, the accepted law is exactly π; hence normalizer estimation affects the error through clipping and complete rejection.

C.2 Proof of Theorem 2

We first prove the TV comparison using only $k := n + 1$ UR proposals, then extend it to convex f-divergences. A final comparison between two UR budgets gives the same-budget assertion in Theorem 2. None of these arguments uses the source's sufficient condition equation 32.

Conditioning and randomizing the fresh-candidate order. Condition on the pilot and any independent threshold randomization, and denote the resulting fixed normalized threshold by T. Generate all $k = n + 1$ fresh candidates in advance, including the fallback, and uniformly permute them. Uniformly permuting these i.i.d. candidates preserves their joint law. Assign each candidate an

independent acceptance indicator with probability $\min\{w_i/T, 1\}$, including a virtual indicator for the last candidate. That indicator does not change the output: the last candidate is returned whenever all preceding candidates are rejected, regardless of its indicator. If an accepted index exists, the output is consequently the first accepted index in the permuted order; if no index is accepted, it is the last index. Conditional on the candidate list and indicators, the rule therefore selects uniformly among accepted indices, or uniformly among all indices when none is accepted. Conditional on T, the following lemma applies at the fixed threshold $\tau = T$.

Lemma 5 (A fixed batch comparison)**.** *Fix positive weights $w_{1:k}$ and $\tau > 0$. Retain index i independently with probability $\min\{w_i/\tau, 1\}$, select uniformly among retained indices, and use a uniform index if none is retained. Let $q_i(\tau)$ be its selection probabilities. For every $t > 0$,*

$$\sum_{i:w_i>t} q_i(\tau) \le \sum_{i:w_i>t} p_i^U(w_{1:k}).$$

The same inequality holds with $w_i \ge t$.

Proof. For $k = 1$, the rules coincide. Otherwise, put $w_{\max} := \max_i w_i$ and $I_t := \{i : w_i > t\}$. We show that moving the threshold to $w_{\max}$ cannot decrease the probability of selecting an index in I_t. At this threshold, the retained set is nonempty and the rule is UR's Bernoulli representation from Appendix E.3.

For $k \ge 2$, suppose $\tau \le w_{\max}$ and put $v_i := \min\{w_i, \tau\}$. The rule is UR on the clipped vector $v_{1:k}$. Couple this race with UR on $w_{1:k}$ using the same uniforms. The factor $w_i/v_i = \max\{1, w_i/\tau\}$ is nondecreasing in w_i. If the clipped race selects $i \in I_t$, then for every $j \notin I_t$,

$$\frac{w_i}{U_i} = \frac{w_i}{v_i}\frac{v_i}{U_i} > \frac{w_j}{v_j}\frac{v_j}{U_j} = \frac{w_j}{U_j}.$$

Ties have probability zero. Hence UR also selects within I_t, proving the comparison in this case.

Now consider the case where $\tau \ge w_{\max}$. Set $z_i := w_i/w_{\max}$, and write $\alpha := w_{\max}/\tau \in (0, 1]$. Integrating the reciprocal size of the retained set as in Appendix E.3 yields

$$q_i(w_{\max}/\alpha) = z_i \int_0^\alpha \prod_{j\ne i}(1 - z_j u)\mathrm{d}u + \frac{1}{k}\prod_{j=1}^{k}(1 - \alpha z_j).$$

The first term accounts for a nonempty retained set; and the second accounts for the fallback when the retained set is empty. Consequently,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha}\sum_{i\in I_t} q_i(w_{\max}/\alpha) &\overset{\text{(i)}}{=} \frac{1}{k}\sum_{i\in I_t}\sum_{j\notin I_t}\left[z_i\prod_{\ell\ne i}(1-\alpha z_\ell) - z_j\prod_{\ell\ne j}(1-\alpha z_\ell)\right] \\ &\overset{\text{(ii)}}{=} \frac{1}{k}\sum_{i\in I_t}\sum_{j\notin I_t}(z_i - z_j)\prod_{\ell\notin\{i,j\}}(1-\alpha z_\ell) \\ &\ge 0.\end{aligned}$$

Here (i) differentiates and groups terms across I_t and its complement. (ii) cancels the two terms containing $\alpha z_i z_j$. The last inequality uses $z_i \ge z_j$ for the indicated indices. Hence the probability increases up to $\alpha = 1$, where the rule is UR. Empty and full sets cause no difficulty, and the argument is unchanged for $I_t := \{i : w_i \ge t\}$. □

From weight ordering to divergence ordering. Average Lemma 5 over the fresh candidates and the pilot. Writing $\widehat{Y}_R$ for the baseline output and using k candidates for $\widehat{Y}_U$ gives

$$\Pr(w(\widehat{Y}_R) > t) \le \Pr(w(\widehat{Y}_U) > t),$$

with the same inequality for weights at least t. This establishes the required selected weight stochastic ordering. For the extension, notice first that both output laws have positive finite densities relative to π almost surely. For the baseline, positivity follows from the possibility of accepting its first fresh candidate, and finiteness follows from $P^R_{N,\delta}(D) \le k\mu(D)$. Write $h^R(y) := (\mathrm{d}P^R_{N,\delta}/\mathrm{d}\pi)(y)$. By Lemma 8, $\mathrm{d}P^U_k/\mathrm{d}\pi = h^U_k \circ w$, where h^U_k is nonincreasing. For $c > 0$, let $D_c := \{y : h^U_k(w(y)) > c\}$.

This is a lower weight set, so $P_k^U(D_c) \le P_{N,\delta}^R(D_c)$. For $X \sim \pi$,

$$\begin{aligned}\mathbb{E}[(h_k^U(w(X)) - c)_+] &\overset{\text{(i)}}{=} P_k^U(D_c) - c\pi(D_c) \\ &\overset{\text{(ii)}}{\le} P_{N,\delta}^R(D_c) - c\pi(D_c) \\ &\overset{\text{(iii)}}{\le} \mathbb{E}[(h^R(X) - c)_+].\end{aligned}$$

Here (i) integrates over the positive set; (ii) uses the probability ordering; and (iii) integrates the positive part over the whole space. At $c = 1$, this also recovers the TV comparison. Both density ratios have mean one, so affine terms have equal expectations. Positive sums of these hinge inequalities therefore compare every convex piecewise-linear function with finitely many breakpoints. For an arbitrary allowed convex f, subtract a supporting affine function at one, approximate the resulting nonnegative function increasingly by maxima of finitely many supporting affine functions and zero, and apply monotone convergence as in Appendix E. This proves the selection-stage inequality in equation 18, including infinite divergences. Nothing here required a sample-mean pilot: any positive finite threshold determined from an independent pilot and held fixed during the fresh stage is covered. The fresh-proposal fallback is essential to this argument; arbitrary other fallbacks and thresholds adapted to the fresh candidates are not covered.

Comparison under the same budget. It remains to compare two UR budgets. A decreasing upper bound would not establish that the actual error decreases.

Lemma 6 (UR's marginal divergences decrease with the budget)**.** *For every positive weight pair and integers* $1 \le k \le m$*,*

$$D_{\mathrm{TV}}(P_m^U, \pi) \le D_{\mathrm{TV}}(P_k^U, \pi).$$

More generally, $D_f(P_m^U \| \pi) \le D_f(P_k^U \| \pi)$ *for every convex* f *covered by Theorem 2.*

Proof. We first compare selected weights, then apply the same positive part argument. For a score value s with $\pi(w < s) > 0$,

$$\pi(w > t \mid w < s) = \begin{cases} 0, & s \le t, \\ 1 - \frac{\pi(w \le t)}{\pi(w < s)}, & s > t. \end{cases}$$

This probability is nondecreasing in s on the score values that can occur. Using the first k candidates among the same m candidate–uniform pairs couples their maxima so that $S_{(k)} \le S_{(m)}$. Proposition 1 then gives $P_k^U(w > t) \le P_m^U(w > t)$. The same reasoning applies to $w \ge t$, replacing $\pi(w \le t)$ by $\pi(w < t)$ in the second branch. For $c > 0$, the set $D_c := \{y : h_m^U(w(y)) > c\}$ is a lower weight set, and hence $P_m^U(D_c) \le P_k^U(D_c)$. Therefore, for $X \sim \pi$,

$$\begin{aligned}\mathbb{E}[(h_m^U(w(X)) - c)_+] &= P_m^U(D_c) - c\pi(D_c) \\ &\le P_k^U(D_c) - c\pi(D_c) \\ &\le \mathbb{E}[(h_k^U(w(X)) - c)_+].\end{aligned}$$

Setting $c = 1$ proves the TV assertion. Both density ratios have mean one, so the convex approximation argument above proves the general assertion, including infinite divergences. □

Combining the selection stage comparison with Lemma 6 therefore proves Theorem 2. For an even maximum budget, the baseline leaves one proposal unused and the same budget monotonicity applies.

Lastly, notice that the proof does not use the specific form of $M_{N,\delta}\widehat{Z}$: the same argument applies to any positive finite threshold determined from an independent pilot and held fixed during the rejection stage.

C.3 Proof of Proposition 2: exponential separation on a fixed pair

We prove Proposition 2 for the fixed pair $\mu = (1/2, 1/2)$ and $\pi = (1/4, 3/4)$. We then compare the budgets needed for a prescribed accuracy and verify the last column of Table 1. Throughout the fixed pair proof, $N = 2n + 1$ with $n \ge 1$.

Proof. First, notice that the normalized weights are $w(0) = \frac{\pi(0)}{\mu(0)} = 1/2$ and $w(1) = \frac{\pi(1)}{\mu(1)} = 3/2$. Fix a calibration and set $M := M_{N,\delta}$. It suffices to show that every multiplier $M > 0$ leaves TV error at least $2^{-n}/12$. We first record the output law of one accepted proposal at a fixed normalized threshold τ. The acceptance probabilities of states 0 and 1 are

$$a_0(\tau) = \min\left\{1, \frac{1}{2\tau}\right\}, \qquad a_1(\tau) = \min\left\{1, \frac{3}{2\tau}\right\}.$$

Since μ is uniform, conditional on acceptance, the probability of state 1 is

$$h(\tau) := Q_\tau^{\mathrm{RS}}(\{1\}) = \frac{a_1(\tau)}{a_0(\tau) + a_1(\tau)} = \begin{cases} 1/2, & 0 < \tau \le 1/2, \\ \tau/(\tau + 1/2), & 1/2 < \tau < 3/2, \\ 3/4, & \tau \ge 3/2. \end{cases}$$

Hence,

$$\frac{1}{2} \le h(\tau) \le \frac{3}{4} = \pi(1), \quad \text{for every } \tau > 0.$$

The fresh fallback is drawn from μ, so it assigns state 1 probability $1/2$ as well. Hence, for every realized pilot threshold, both the accepted law and the fallback underweight state 1 relative to its target probability $3/4$. In particular, the errors arising from different pilot outcomes all have the same sign and therefore cannot cancel when we average over the pilot.

Let r_T denote the rejection probability of one proposal at the random threshold T. Conditional on T, at least one of the n rejection stage proposals is accepted with probability $1 - r_T^n$, while all n proposals are rejected with probability r_T^n, in which case the fresh fallback is used. Therefore

$$\Pr(\widehat{Y}_R = 1 \mid T) = (1 - r_T^n)h(T) + \frac{1}{2}r_T^n.$$

Since this quantity is always at most $3/4$, the binary TV error has no absolute value cancellation, and equation 33 gives

$$\begin{aligned} D_{\mathrm{TV}}(P_{N,\delta}^R, \pi) &= \mathbb{E}\left[\frac{3}{4} - (1 - r_T^n)h(T) - \frac{1}{2}r_T^n\right] \\ &= \mathbb{E}\left[(1 - r_T^n)\left(\frac{3}{4} - h(T)\right) + \frac{1}{4}r_T^n\right]. \end{aligned} \tag{34}$$

We now lower-bound this expression for every $M > 0$. If $M \le 2$, consider the event that all n pilot weights equal the lower weight $1/2$. This event has probability 2^{-n} and gives

$$\overline{W}_n = \frac{1}{2}, \qquad T = M\overline{W}_n = \frac{M}{2} \le 1.$$

Since $h(\tau)$ is nondecreasing,

$$h(T) \le h(1) = \frac{1}{1 + 1/2} = \frac{2}{3}.$$

On this event, both the accepted law and the fallback assign state 1 probability at most $2/3$, so the conditional TV error is at least

$$\frac{3}{4} - \frac{2}{3} = \frac{1}{12}.$$

Since all conditional errors have the same sign, we obtain

$$D_{\mathrm{TV}}(P_{N,\delta}^R, \pi) \ge 2^{-n} \cdot \frac{1}{12}.$$

Now consider the case where $M > 2$. Since each pilot weight equals $1/2$ or $3/2$ with equal probability, $\overline{W}_n - 1$ is symmetric about zero. Hence $\Pr(\overline{W}_n \ge 1) \ge \frac{1}{2}$. On this event,

$$T = M\overline{W}_n \ge M > 2 > \frac{3}{2},$$

so there is no clipping and hence $h(T) = 3/4$. Moreover, since $\mathbb{E}_\mu[w(Y)] = 1$, the acceptance probability of one proposal is $1/T$, and therefore

$$r_T = 1 - \frac{1}{T} \ge \frac{1}{2}, \qquad r_T^n \ge 2^{-n}.$$

The clipping term in equation 34 vanishes, while the fallback term contributes at least $\frac{1}{4}r_T^n \geq \frac{1}{4}2^{-n}$. Since the event $\{\overline{W}_n \geq 1\}$ has probability at least $1/2$,

$$D_{\mathrm{TV}}(P^R_{N,\delta}, \pi) \geq \frac{1}{2} \cdot \frac{1}{4}2^{-n} = \frac{1}{8}2^{-n} \geq \frac{1}{12}2^{-n}.$$

Hence the same lower bound holds after taking the infimum over $\delta \in (0, 1)$.

The UR formula in Proposition 3 gives $D_{\mathrm{TV}}(P^U_k, \pi) = \frac{3}{4}3^{-k}$, which yields the claimed equality at $k = N = 2n + 1$. This completes the proof. □

The gap persists with fewer UR proposals. Every baseline run uses at least $n + 1$ proposals in total, including its pilot. Even with only this many proposals, UR satisfies

$$\frac{\inf_{\delta\in(0,1)} D_{\mathrm{TV}}(P^R_{2n+1,\delta}, \pi)}{D_{\mathrm{TV}}(P^U_{n+1}, \pi)} \geq \frac{1}{3}\left(\frac{3}{2}\right)^n. \tag{35}$$

With the same total budget $N = 2n + 1$ for both methods, the lower bound on the ratio is $\frac{1}{3}(9/2)^n$, as stated in Section 3. The different bases correspond to different UR budgets, not different target–proposal pairs.

Error ratio versus sample complexity. An exponentially growing error ratio does not imply different orders of optimized sample complexity. Let $\xi \in (0, 1)$ be a TV error tolerance, with $\xi \downarrow 0$. For UR, the exact formula gives a required budget of order $\log(1/\xi)$. For calibrated RS, the lower bound requires n to be at least of this order. Conversely, $C^\pi_\infty = 3/2$, and choosing $\delta = \xi$ in equation 32 gives TV error at most ξ whenever the integer n satisfies

$$n \geq 24 \log(4/\xi).$$

Hence both methods require $\Theta(\log(1/\xi))$ proposals when tuned for the prescribed accuracy. This statement allows the calibration to vary with the requested accuracy; it does not assert convergence for every fixed δ. Appendix C.4 addresses the separate question of matching the sharp guarantee uniformly over pairs at a fixed budget.

The budget-dependent entries of Table 1. The last column uses the fixed pair of Proposition 2. UR has exact error $(3/4)3^{-N}$, and known-envelope RS has the same law by Proposition 6, with a uniform observed-index fallback. Substituting $\varepsilon = 1/4$ into equation 23 gives

$$D_{\mathrm{TV}}(P^{\mathrm{SIR}}_N, \pi) = \frac{3}{16N} + O(N^{-3/2}).$$

For calibrated RS, equation 17 gives

$$\inf_{\delta\in(0,1)} D_{\mathrm{TV}}(P^R_{N,\delta}, \pi) \geq \frac{1}{12}2^{-(N-1)/2} = \frac{\sqrt{2}}{12}2^{-N/2}$$

for every odd $N \geq 3$. This proves the claimed $\Omega(2^{-N/2})$ lower bound; no matching $\Theta(2^{-N/2})$ rate is asserted. Appendix C.4 justifies the sharp guarantee column.

C.4 Local separation under optimal calibration

The fixed pair comparison in Section 3 distinguishes exponential rates as the budget grows. Here we fix the budget and vary the pair to ask whether calibrated RS can match the sharp bounded-weight guarantee within a factor independent of the pair. Keep the unnormalized weights of states 0 and 1 equal to $1/2$ and 1, respectively, and set

$$\mu_\eta := (\eta, 1 - \eta), \qquad \pi_\eta := \left(\frac{\eta}{2 - \eta}, \frac{2(1 - \eta)}{2 - \eta}\right), \qquad 0 < \eta < 1. \tag{36}$$

Hence η is the proposal probability of the lower weight state. The normalizer is $Z_\eta = 1 - \eta/2$, and $C^{\pi_\eta}_\infty = 1/Z_\eta$. In particular, the bounded-weight benchmark is $(1 - 1/C^{\pi_\eta}_\infty)^N = (\eta/2)^N$.

Proposition 4 (Local separation under optimal calibration)**.** *Fix an odd $N \geq 13$. For the family in equation 36, as $\eta \downarrow 0$,*

$$\inf_{\delta\in(0,1)} D_{\mathrm{TV}}(P^R_{N,\delta}, \pi_\eta) = \frac{N - 1}{4(N - 2)}\eta^2 + o_N(\eta^2), \tag{37}$$

whereas

$$D_{\mathrm{TV}}(P_N^U, \pi_\eta) = \frac{1-\eta}{1-\eta/2}\left(\frac{\eta}{2}\right)^N. \tag{38}$$

The infimum allows pair- and budget-dependent tuning before observing the pilot. The UR identity holds for every $N \geq 1$ and $0 < \eta < 1$.

Proof. Fix $n \geq 6$ and $N = 2n+1$, and let $\eta \downarrow 0$. For the lower bound, allow any deterministic multiplier $M > 0$ chosen before observing the pilot. Let J count the state 0 candidates among the n pilot proposals. Then $J \sim \mathrm{Bin}(n, \eta)$, and

$$\widehat{Z} = \frac{J/2 + n - J}{n} = 1 - \frac{J}{2n}, \qquad \tau_J := M\widehat{Z} = M\left(1 - \frac{J}{2n}\right).$$

Here τ_J is the raw threshold; its normalized counterpart is $T = \tau_J / Z_\eta$. Allowing all $M > 0$ enlarges the calibrated family. The upper-bound choice $M = 1$ is feasible in the original family because $\delta = 4\exp(-n/4) \in (0,1)$ gives $M_{N,\delta} = 1$.

The error conditional on the pilot. The accepted probability of state 0 is

$$q_\eta(\tau) = \begin{cases} \eta, & 0 < \tau \leq 1/2, \\ \frac{\eta/2}{\eta/2 + (1-\eta)\tau}, & 1/2 < \tau < 1, \\ \frac{\eta}{2-\eta}, & \tau \geq 1. \end{cases}$$

It is nonincreasing in τ and lies between $\pi_\eta(0)$ and $\mu_\eta(0) = \eta$. Write

$$d_\eta(\tau) := q_\eta(\tau) - \pi_\eta(0), \qquad g_\eta := \eta - \pi_\eta(0) = \frac{\eta(1-\eta)}{2-\eta}.$$

Let $r_\eta(\tau) := r_{\tau/Z_\eta}$ be the rejection probability of one fresh trial; the division by Z_η converts the raw threshold to the normalized units of Appendix C.1.

Both acceptance and fallback assign state 0 at least its target probability, so averaging over the pilot cannot cancel this excess. Consequently, equation 33 gives the exact TV error of the relaxed rule:

$$e_n(M,\eta) = \mathbb{E}\left[d_\eta(\tau_J) + r_\eta(\tau_J)^n\left(g_\eta - d_\eta(\tau_J)\right)\right] = \mathbb{E}\left[(1 - r_\eta(\tau_J)^n)d_\eta(\tau_J) + r_\eta(\tau_J)^n g_\eta\right]. \tag{39}$$

The expectation is over J. The integrand lies in $[d_\eta(\tau_J), g_\eta]$, a fact used for both bounds below.

Step 1: Upper bound from $M = 1$. For $J = 0$, the threshold is one, the accepted law is exact, and $r_\eta(1) = \eta/2$. The contribution is at most $g_\eta(\eta/2)^n = O_n(\eta^{n+1})$. For $J = 1$, put $v_n := 1 - 1/(2n) \in (1/2, 1)$, where we use $n \geq 2$. Direct expansion gives

$$d_\eta(v_n) = \frac{1}{2}\left(\frac{1}{v_n} - 1\right)\eta + O_n(\eta^2) = \frac{\eta}{2(2n-1)} + O_n(\eta^2).$$

Here $r_\eta(v_n) = \eta(1 - 1/(2v_n)) = O_n(\eta)$ and $\Pr(J=1) = n\eta + O_n(\eta^2)$. The contribution is consequently $n\eta^2/[2(2n-1)] + O_n(\eta^3)$. Finally, $\Pr(J \geq 2) = O_n(\eta^2)$ and each conditional error is at most $g_\eta = O(\eta)$. Hence

$$e_n(1,\eta) = \frac{n}{2(2n-1)}\eta^2 + O_n(\eta^3). \tag{40}$$

Step 2: Typical pilot forces $M_\eta \to 1$. Suppose $e_n(M_\eta, \eta) = O_n(\eta^2)$, and fix $\gamma \in (0, 1/2)$. If $M_\eta \leq 1 - \gamma$ along a subsequence, then the event $J = 0$ and the monotonicity of d_η give

$$e_n(M_\eta, \eta) \geq (1-\eta)^n d_\eta(1-\gamma), \qquad \frac{d_\eta(1-\gamma)}{\eta} \to \frac{\gamma}{2(1-\gamma)} > 0.$$

This contradicts the assumed second-order error. If instead $M_\eta \geq 1 + \gamma$ along a subsequence, then on $J = 0$ there is no clipping, but $r_\eta(M_\eta) = 1 - Z_\eta/M_\eta \geq \gamma/(1+\gamma)$. Hence

$$e_n(M_\eta, \eta) \geq (1-\eta)^n \left(\frac{\gamma}{1+\gamma}\right)^n g_\eta.$$

Since $g_\eta/\eta \to 1/2$, this again leaves first-order error. Therefore $M_\eta \to 1$.

Step 3: The event $J = 1$ forces the leading coefficient. For each η, choose $M_\eta > 0$ such that

$$e_n(M_\eta, \eta) \le \inf_{M>0} e_n(M, \eta) + \eta^3.$$

Step 1 makes this sequence second-order accurate, so Step 2 gives $M_\eta \to 1$. Hence $\tau_1 = M_\eta v_n \to v_n$, and

$$\frac{d_\eta(M_\eta v_n)}{\eta} \to \frac{1}{2(2n-1)}.$$

The nonnegative rejection term in equation 39 can only increase the error. Keeping just the $J = 1$ contribution therefore gives

$$\inf_{M>0} e_n(M, \eta) \ge \Pr(J = 1) d_\eta(M_\eta v_n) - \eta^3.$$

Dividing by η^2 and using $\Pr(J = 1)/\eta \to n$ proves

$$\liminf_{\eta \downarrow 0} \frac{\inf_{M>0} e_n(M, \eta)}{\eta^2} \ge \frac{n}{2(2n-1)}.$$

Together with the feasible upper bound in equation 40, this proves equation 37, since $n/[2(2n-1)] = (N-1)/[4(N-2)]$. In particular, the multiplier may depend on the pair and budget, but not on the realized pilot. □

UR's exact error. The largest normalized weight is $C_\infty^{\pi_\eta} = 1/Z_\eta$. A state 1 candidate crosses this score level almost surely. A state 0 candidate stays below it exactly when its uniform exceeds $1/2$, so a single candidate fails to cross with probability $\eta/2$. Independence gives $\Pr(H^c_{C_\infty^{\pi_\eta}}) = (\eta/2)^N$. On this event all candidates are in state 0, so UR returns 0; on the complement its output law is π_η by equation 12. The mixture identity in equation 13 therefore gives

$$D_{\mathrm{TV}}(P_N^U, \pi_\eta) = \pi_\eta(1) \left(\frac{\eta}{2}\right)^N = \frac{1-\eta}{1-\eta/2} \left(\frac{\eta}{2}\right)^N,$$

which proves equation 38. This identity holds for every $N \ge 1$ and $0 < \eta < 1$, not only in the local limit.

Comparison to others. At fixed N, exactly one state 0 candidate appears with probability $N\eta + O_N(\eta^2)$, and SIR selects it with probability $1/(2N-1)$. Batches containing at least two such candidates have total probability $O_N(\eta^2)$. Hence $P_N^{\mathrm{SIR}}(\{0\}) = N\eta/(2N-1) + O_N(\eta^2)$, and subtracting $\pi_\eta(0) = \eta/2 + O(\eta^2)$ yields

$$D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi_\eta) = \frac{\eta}{2(2N-1)} + O_N(\eta^2).$$

Known-envelope RS uses the exact raw envelope $Z_\eta C_\infty^{\pi_\eta} = 1$ and a uniform observed-index fallback. Its output law equals UR's on this two-weight family by Proposition 6. These calculations establish the local-error column. For the sharp guarantee column, UR satisfies the universal bounded-weight bound by Corollary 1, and known-envelope RS does so by its exact accepted law and complete-rejection probability. On this family the bound is $(1 - 1/C_\infty^{\pi_\eta})^N = (\eta/2)^N$, so the positive first- and second-order coefficients above rule it out for SIR and calibrated RS, even up to a pair-independent multiplicative constant at fixed odd $N \ge 13$. Even UR with only $n+1$ proposals has error $\Theta_n(\eta^{n+1})$ on this family. All local limits fix the budget; no remainder uniform in N is asserted.

D MOMENT BOUNDS AND MINIMAX RATES

The bounded weight upper bound in Corollary 1 follows from equation 12. We prove the second moment bound next, followed by the matching lower bound under a uniform weight bound. The final subsection gives a lower bound under a second moment constraint. These lower bounds use the same restriction: a sampler cannot return a state that never appears among its proposals.

D.1 DIRECT SECOND MOMENT UPPER BOUND

Assume $C^\pi = \mathbb{E}_\mu[w(Y)^2] < \infty$. By Proposition 1, conditioning on $S_{(N)} = s$ excludes exactly the target mass $\pi(w \ge s)$. We first average this conditional error, then bound the chance that the

largest score is too small. Let $X \sim \pi$ be independent of the entire race, used only for analysis. Recall $G(t) := \Pr(w(Y)/U \le t)$ for independent $Y \sim \mu$ and $U \sim \mathrm{Unif}(0,1)$. Then

$$\begin{aligned} D_{\mathrm{TV}}(P_N^U, \pi) &\overset{\text{(i)}}{\le} \mathbb{E}[\pi(w \ge S_{(N)})] \\ &\overset{\text{(ii)}}{=} \Pr(S_{(N)} \le w(X)) \\ &\overset{\text{(iii)}}{=} \mathbb{E}[G(w(X))^N], \end{aligned} \tag{41}$$

where (i) follows from convexity of TV and the conditional law; (ii) follows from using the independent target point $X \sim \pi$; and (iii) conditions on X and uses independence of the N scores. Under the zero weight convention of Appendix B.5, the conditional error at score zero is one, equal to $\pi(w \ge 0)$, so the same calculation applies. We now bound the chance that a single score exceeds $t > 0$:

$$\begin{aligned} 1 - G(t) &= \mathbb{E}_\mu \left[\min\left\{ \frac{w(Y)}{t}, 1 \right\} \right] \\ &\overset{\text{(i)}}{\ge} \mathbb{E}_\mu \left[\frac{w(Y)}{t + w(Y)} \right] \\ &\overset{\text{(ii)}}{=} \mathbb{E} \left[\frac{1}{t + w(X)} \right] \\ &\overset{\text{(iii)}}{\ge} \frac{1}{t + \mathbb{E}[w(X)]} = \frac{1}{t + C^\pi}, \end{aligned}$$

where (i) follows from the fact that $\min\{\frac{a}{t}, 1\} \ge \frac{a}{t+a}$ for $a \ge 0$; (ii) follows since $w \mathrm{d}\mu = \mathrm{d}\pi$; and (iii) follows from Jensen's inequality for the convex function $a \mapsto \frac{1}{t+a}$. This is where the second moment enters: $\mathbb{E}[w(X)] = \mathbb{E}_\mu[w(Y)^2] = C^\pi$. Independence and $1 - a \le e^{-a}$ give $G(t)^N \le e^{-N/(t+C^\pi)}$. Substituting this into equation 41,

$$D_{\mathrm{TV}}(P_N^U, \pi) \le \mathbb{E}\left[\exp\left(-\frac{N}{w(X) + C^\pi} \right) \right] \tag{42}$$

$$\le \frac{\mathbb{E}[w(X)] + C^\pi}{eN} = \frac{2C^\pi}{eN}, \tag{43}$$

where the last inequality follows from the fact that $e^{-x} \le 1/(ex)$ for $x > 0$. Together with the trivial bound 1, this proves the second moment term in Corollary 1.

Relation to the known SIR bound. The order C^π/N is not specific to UR. Agapiou et al. (2017, Theorem 2.1) bound the absolute bias of self-normalized importance sampling by $12C^\pi/N$ for bounded measurable functions (bounded by one in absolute value). For an event D, apply that bound to $2\mathbb{1}_D - 1$. The bias is then $2(P_N^{\mathrm{SIR}}(D) - \pi(D))$, giving $D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi) \le \min\{1, 6C^\pi/N\}$. Theorem 3 also transfers that rate to UR. The direct proof above gives the constant $2/e$ without using the SIR comparison, but does not establish optimality of that constant.

D.2 Exact bounded ratio minimax error

UR already supplies the upper bound in equation 15. For a lower bound, consider a binary pair with target mass one on state 1 and proposal probability $1/C$ on that state, where $C > 1$. Every observed-candidate sampler must return state 0 when all N proposals miss state 1. This is the same basic obstruction used in approximate sampling lower bounds (Block & Polyanskiy, 2023). To obtain the bound using strictly positive distributions, fix $C > 1$ and $0 < \eta < 1$ and take

$$\pi_\eta(1) = 1 - \eta, \quad \mu_\eta(1) = \frac{1 - \eta}{C}.$$

Then $w_\eta(1) = C$ and $w_\eta(0) = \eta/[1 - (1-\eta)/C] < C$. For any sampler A returning an observed candidate,

$$D_{\mathrm{TV}}(P_N^{\mathsf{A}}, \pi_\eta) \overset{\text{(i)}}{\ge} P_N^{\mathsf{A}}(\{0\}) - \eta$$

$$\overset{\text{(ii)}}{\geq} \left(1 - \frac{1-\eta}{C}\right)^N - \eta.$$

Here (i) evaluates the TV supremum on $\{0\}$; and (ii) uses the event that every proposal is state 0, on which the output must also be state 0. Letting $\eta \downarrow 0$ gives $(1 - 1/C)^N$, uniformly over the choice of sampler, and proves the minimax lower bound. The argument applies even to samplers with additional information about the pair. For $C = 1$, $w \leq 1$ and $\mathbb{E}_\mu w = 1$ imply $w = 1$ almost surely. Thus $\pi = \mu$ and returning the first proposal is exact.

D.3 Second moment minimax lower bound

A second moment constraint still allows a state to be much rarer under the proposal than under the target. We choose its proposal probability so that the chance of observing it is at most half its target probability.

Proposition 5 (A lower bound under a second moment constraint)**.** *Fix $R > 1$ and $N \geq 1$. For every procedure returning one of the N observed proposals, there is a finite strictly positive pair with $C^\pi \leq R$ and TV error at least*

$$\frac{1}{16} \min\left\{1, \frac{R-1}{N}\right\}.$$

For $R \geq 2$, this is at least $\frac{1}{32} \min\{1, R/N\}$.

Proof. Set

$$p := \min\left\{\frac{R-1}{8N}, \frac{1}{4}\right\}, \quad q := \frac{p}{2N}, \quad \pi(1) := p, \quad \mu(1) := q.$$

Both binary distributions are strictly positive. Since the output must be observed,

$$\Pr(\widehat{Y} = 1) \leq \Pr(\exists i : Y_i = 1) \leq Nq = p/2,$$

where the second inequality is the union bound. The TV error is therefore at least

$$p - \Pr(\widehat{Y} = 1) \geq p/2 \geq \frac{1}{16} \min\left\{1, \frac{R-1}{N}\right\}.$$

It remains to check the moment constraint. For a binary pair, direct substitution gives

$$C^\pi - 1 = \frac{(p-q)^2}{q(1-q)} \overset{\text{(i)}}{\leq} \frac{2Np}{1-q} \overset{\text{(ii)}}{\leq} \frac{16}{7}Np \overset{\text{(iii)}}{\leq} \frac{2}{7}(R-1) \leq R-1.$$

Here (i) uses $(p-q)^2 \leq p^2$ and $q = p/(2N)$; (ii) and (iii) use $q \leq 1/8$ and $Np \leq (R-1)/8$. Hence $C^\pi \leq R$. For $R \geq 2$, $R - 1 \geq R/2$ gives the second lower bound. □

For $R \geq 2$, the upper and lower bounds have the same dependence on R and N, up to constants. This is a minimax statement over a moment-constrained class, not a lower bound for every fixed pair.

E Proofs for the comparison to SIR

The proof of Theorem 3 uses two facts. UR returns weights above any threshold at least as often as SIR (Lemma 7), and UR's density relative to π is nonincreasing in the weight (Lemma 8). We first combine these facts to prove the theorem, then establish them directly from the selection rule.

E.1 Proof for convex divergences

Let h_N^U and h_N^{SIR} be the scalar functions in Lemma 8, so that

$$\frac{\mathrm{d}P_N^U}{\mathrm{d}\pi}(y) = h_N^U(w(y)), \quad \frac{\mathrm{d}P_N^{\mathrm{SIR}}}{\mathrm{d}\pi}(y) = h_N^{\mathrm{SIR}}(w(y)).$$

For $X \sim \pi$, both $h_N^U(w(X))$ and $h_N^{\mathrm{SIR}}(w(X))$ are positive, finite, and have mean one. We compare the amount by which each exceeds a level $c > 0$. Define $D_c := \{y : h_N^U(w(y)) > c\}$. Since h_N^U is nonincreasing, D_c has the form $\{w < t\}$ or $\{w \leq t\}$ for some threshold t, or is an empty

or full set. The threshold comparison in Lemma 7, with complements when necessary, gives $P_N^U(D_c) \le P_N^{\mathrm{SIR}}(D_c)$. Consequently,

$$\mathbb{E}[(h_N^U(w(X)) - c)_+] \overset{\text{(i)}}{=} P_N^U(D_c) - c\pi(D_c) \tag{44}$$

$$\overset{\text{(ii)}}{\le} P_N^{\mathrm{SIR}}(D_c) - c\pi(D_c) \tag{45}$$

$$\overset{\text{(iii)}}{\le} \int (h_N^{\mathrm{SIR}}(w(y)) - c)_+ \pi(\mathrm{d}y)$$

$$= \mathbb{E}[(h_N^{\mathrm{SIR}}(w(X)) - c)_+]. \tag{46}$$

Here (i) integrates the positive part over its positive set D_c. (ii) is the probability comparison on D_c. For (iii), the integral of $h_N^{\mathrm{SIR}}(w(y)) - c$ over D_c is at most the integral of its positive part over the whole space. At $c = 1$, the two expectations are exactly the TV errors, proving the desired result in TV distance. To pass to a convex f-divergence, first consider a convex piecewise linear function with finitely many breakpoints. It can be written as

$$g(t) = a + bt + \sum_{k=1}^{m} c_k (t - t_k)_+, \quad c_k \ge 0, \quad t_k > 0.$$

Each c_k is the increase in slope at a breakpoint. Because both density ratios have mean one, the affine terms have equal expectations. Applying equation 46 to each remaining term gives

$$\mathbb{E}[g(h_N^U(w(X)))] \le \mathbb{E}[g(h_N^{\mathrm{SIR}}(w(X)))].$$

Now let $f : (0, \infty) \to \mathbb{R}$ be convex with $f(1) = 0$. Choose the slope b of a supporting line at 1 and put $\widetilde{f}(t) := f(t) - b(t - 1) \ge 0$. Choose a countable dense set of points in $(0, \infty)$ and a supporting affine function of $\widetilde{f}$ at each point. Let g_m be the maximum of zero and the first m such functions. Then g_m is nonnegative and convex piecewise linear, and $g_m \uparrow \widetilde{f}$ pointwise. For completeness, supporting slopes are bounded on every compact subinterval of $(0, \infty)$ by secant slopes to nearby endpoints. At points of the dense set approaching t, continuity and this bound make the supporting lines approach $\widetilde{f}(t)$, proving the claimed limit. Monotone convergence yields

$$\mathbb{E}[\widetilde{f}(h_N^U(w(X)))] \le \mathbb{E}[\widetilde{f}(h_N^{\mathrm{SIR}}(w(X)))],$$

including infinite values. Adding back the affine part changes neither side because both means are one. This proves equation 22. The choices $f(t) = |t - 1|/2$, $t \log t$, $-\log t$, and $(t - 1)^2$ give TV, forward KL, reverse KL, and χ^2 divergence, respectively.

The target also bounds the selected weight probabilities. We additionally obtain the interpretation used in Section 4:

$$\Pr(w(\widehat{Y}_{\mathrm{SIR}}) > t) \le \Pr(w(\widehat{Y}_U) > t) \le \pi(w > t), \quad t > 0. \tag{47}$$

Lemma 7 gives the first inequality. For the second, take $A := \{y : w(y) > t\}$ and use that the density integrates to one:

$$P_N^U(A) - \pi(A) = P_N^U(A)\pi(A^c) - \pi(A)P_N^U(A^c)$$

$$= \int_A \int_{A^c} [h_N^U(w(y)) - h_N^U(w(z))]\pi(\mathrm{d}z)\pi(\mathrm{d}y)$$

$$\le 0.$$

The last inequality holds because $w(y) > w(z)$ on $A \times A^c$ and h_N^U is nonincreasing. The same proof works for $\{w \ge t\}$. No moment assumption beyond $\mathbb{E}_\mu w = 1$ is used.

E.2 Two supporting lemmas

Lemma 7 (Probability of returning a weight above a threshold)**.** *For every $t > 0$,*

$$\Pr(w(\widehat{Y}_{\mathrm{SIR}}) > t) \le \Pr(w(\widehat{Y}_U) > t). \tag{48}$$

The same inequality holds for weights at least t. Taking complements reverses the inequalities for weights below a threshold.

Proof. For $N = 1$, the claim is immediate. For $N \geq 2$, fix a positive vector $w_{1:N}$ and write $z_i := w_i/w_{\max}$, where $w_{\max} := \max_j w_j$. Let p_i^U and p_i^{SIR} be the selection probabilities on this fixed vector. We first compare candidates, then sum over those above t. A maximal weight candidate has score strictly above $w_{\max}$. Candidate i can therefore be selected only if $U_i < z_i$. Write $U_i = z_i u$. For $u \in (0, 1)$, its score is larger than candidate j's exactly when $U_j > z_j u$. Independence gives

$$p_i^U = z_i \int_0^1 \prod_{j \neq i} (1 - z_j u) \mathrm{d}u. \tag{49}$$

For distinct i, k with $w_i \geq w_k$, put $A_{ik}(u) := \prod_{j \notin \{i,k\}} (1 - z_j u)$. Subtracting the two formulas gives

$$\frac{p_i^U}{z_i} - \frac{p_k^U}{z_k} = (z_i - z_k) \int_0^1 u A_{ik}(u) \mathrm{d}u \geq 0.$$

Since $p_i^{\mathrm{SIR}} = z_i / \sum_j z_j$, this is equivalent to

$$p_i^U p_k^{\mathrm{SIR}} - p_i^{\mathrm{SIR}} p_k^U \geq 0 \quad \text{whenever } w_i \geq w_k. \tag{50}$$

Thus the change from SIR to UR favors larger weights in pairwise comparisons. Let $I_t := \{i : w_i > t\}$. Using that both probability vectors sum to one,

$$\sum_{i \in I_t} (p_i^U - p_i^{\mathrm{SIR}}) \overset{\text{(i)}}{=} \sum_{i \in I_t} \sum_{k \notin I_t} (p_i^U p_k^{\mathrm{SIR}} - p_i^{\mathrm{SIR}} p_k^U) \overset{\text{(ii)}}{\geq} 0.$$

In (i), the terms with both indices in I_t cancel. (ii) uses equation 50, since every weight in I_t is larger than every weight outside it. This includes empty and full sets I_t. Averaging over the proposals proves equation 48. Replacing I_t by $\{i : w_i \geq t\}$ proves the other endpoint convention. □

Lemma 8 (Density of the returned sample)**.** *Let $G(s) := \Pr(w(Y)/U \leq s)$ for independent $Y \sim \mu$ and $U \sim \mathrm{Unif}(0, 1)$. Then*

$$\frac{\mathrm{d}P_N^U}{\mathrm{d}\pi}(y) = h_N^U(w(y)), \quad h_N^U(x) := N \int_x^\infty \frac{G(s)^{N-1}}{s^2} \mathrm{d}s, \quad x > 0. \tag{51}$$

The function h_N^U is nonincreasing.

Proof. The density formula is the integral in Lemma 4. Its integrand is nonnegative, so increasing the lower limit x cannot increase the integral. For comparison, the SIR density is

$$\frac{\mathrm{d}P_N^{\mathrm{SIR}}}{\mathrm{d}\pi}(y) = h_N^{\mathrm{SIR}}(w(y)), \quad h_N^{\mathrm{SIR}}(x) := N\mathbb{E}\left[\frac{1}{x + \sum_{j=2}^N W_j}\right], \tag{52}$$

where $x > 0$ and $W_j = w(Y_j)$ for independent proposals. Indeed, a specified candidate at y contributes its SIR selection probability $w(y)/(w(y) + \sum_{j=2}^N W_j)$ times $\mu(\mathrm{d}y)$. There are N exchangeable candidate positions, and $w(y)\mu(\mathrm{d}y) = \pi(\mathrm{d}y)$, giving the formula. Both scalar functions are positive and at most N/x for $x > 0$, and their compositions with w integrate to one under π. At $N = 1$, both functions are $1/x$. □

E.3 Equivalent Bernoulli selection rule

UR has an equivalent rule on each fixed positive weight vector: retain candidates independently and select uniformly among those retained. Write $z_j := w_j / \max_k w_k$, draw independent $B_j \sim \mathrm{Bern}(z_j)$, and retain j when $B_j = 1$. If there are k retained indices, each is selected with probability $1/k$. At least one $z_j = 1$, so the retained set is nonempty. The denominator in z_j is the largest observed weight, not a bound on unobserved weights. Put $K_{-i} := \sum_{j \neq i} B_j$. By independence, the probability of selecting i is $z_i \mathbb{E}[(1 + K_{-i})^{-1}]$. Using $1/(k+1) = \int_0^1 x^k \mathrm{d}x$,

$$\mathbb{E}\left[\frac{1}{1 + K_{-i}}\right] \overset{\text{(i)}}{=} \int_0^1 \mathbb{E}[x^{K_{-i}}] \mathrm{d}x$$

$$\overset{\text{(ii)}}{=} \int_0^1 \prod_{j\neq i}(1 - z_j + z_j x)\mathrm{d}x$$

$$\overset{\text{(iii)}}{=} \int_0^1 \prod_{j\neq i}(1 - z_j t)\mathrm{d}t.$$

Here (i) interchanges a nonnegative integral and expectation. (ii) uses independence of the Bernoulli variables, and (iii) substitutes $t = 1 - x$. After multiplication by z_i, this is exactly equation 49. The two implementations therefore have the same selection probabilities on every fixed batch, though they need not select the same index when coupled using the same uniforms.

F BINARY EXAMPLES AND BIAS CALCULATIONS

Section 4 gives the direct comparison for two candidates. We first extend that binary calculation to every budget and identify the source of SIR's remaining error. We then prove the equality with RS with $M = C_\infty^\pi$ for two weight values, derive the general SIR bias and UR's corresponding adjustment, and conclude with a continuous example. Throughout the binary calculation,

$$\mu = \left(\frac{1}{2}, \frac{1}{2}\right), \qquad \pi = (\varepsilon, 1-\varepsilon), \qquad 0 < \varepsilon < \frac{1}{2},$$

with $w_L = 2\varepsilon$ and $w_H = 2(1-\varepsilon)$.

F.1 PROOF OF PROPOSITION 3: BINARY EXAMPLE FOR EVERY BUDGET

Let $K \sim \mathrm{Bin}\left(N, \frac{1}{2}\right)$ count the candidates in state 1, and put $X := K/N$. Thus X is the observed fraction of candidates in state 1. Define, for $x \in [0,1]$,

$$d(x) := \varepsilon + (1-2\varepsilon)x, \qquad g(x) := \frac{(1-\varepsilon)x}{d(x)}.$$

When $X = x$, the total observed weight is $2Nd(x)$, and SIR returns state 1 with probability $g(x)$. The exact output probabilities are

$$\Pr(\widehat{Y}_{\mathrm{SIR}} = 1) = \mathbb{E}[g(X)], \tag{53}$$

$$\Pr(\widehat{Y}_U = 1) = (1-\varepsilon)\left[1 - \left(\frac{1-2\varepsilon}{2(1-\varepsilon)}\right)^N\right]. \tag{54}$$

Although $g(\mathbb{E}[X]) = 1 - \varepsilon$, averaging $g(X)$ does not give this target probability. More precisely,

$$\pi(1) - \Pr(\widehat{Y}_{\mathrm{SIR}} = 1) = 4\varepsilon(1-\varepsilon)(1-2\varepsilon)\mathbb{E}\left[\frac{\left(X - \frac{1}{2}\right)^2}{d(X)}\right]. \tag{55}$$

On a binary state space, TV is the absolute error in the probability of returning state 1. Thus these formulae will give the claimed TV rates once we verify that both probabilities are below $1 - \varepsilon$. We first establish the output formulae by conditioning on K, then obtain the sign and size of SIR's error from the gap between g and its tangent at $\mathbb{E}[X] = 1/2$.

The two output distributions. It suffices to compute each selection probability conditional on K, since averaging over this binomial count gives the marginal output probability. For SIR, given $K = k$, the probability is

$$\frac{(1-\varepsilon)k}{(1-\varepsilon)k + \varepsilon(N-k)} = g\left(\frac{k}{N}\right).$$

Averaging over K proves equation 53. For UR and $k \geq 1$, fix a state 1 candidate and condition on its uniform variable being u. It beats another state 1 candidate when that candidate's uniform exceeds u, and a state 0 candidate when its uniform exceeds $\varepsilon u/(1-\varepsilon)$. The uniforms are independent, and ties have probability zero. Summing over the k possible selected state 1 candidates gives

$$\Pr(\widehat{Y}_U = 1 \mid K = k) = k\int_0^1 (1-u)^{k-1}\left(1 - \frac{\varepsilon}{1-\varepsilon}u\right)^{N-k}\mathrm{d}u.$$

The probability is zero when $k = 0$. Therefore,

$$\Pr(\widehat{Y}_U = 1) = \frac{1}{2^N} \int_0^1 \sum_{k=1}^{N} \binom{N}{k} k(1-u)^{k-1} \left(1 - \frac{\varepsilon}{1-\varepsilon} u\right)^{N-k} \mathrm{d}u$$

$$\overset{\text{(i)}}{=} \frac{N}{2} \int_0^1 \left(1 - \frac{u}{2(1-\varepsilon)}\right)^{N-1} \mathrm{d}u \tag{56}$$

$$\overset{\text{(ii)}}{=} (1-\varepsilon) \left[1 - \left(\frac{1-2\varepsilon}{2(1-\varepsilon)}\right)^{N}\right], \tag{57}$$

where (i) uses $k\binom{N}{k} = N\binom{N-1}{k-1}$ and the binomial theorem; and (ii) evaluates the integral, proving equation 54 directly from the selection rule.

The gap and its rate. To prove equation 55, it is enough to express $g(m) - g(X)$ as a centered linear term plus a nonnegative remainder, where $m := 1/2$. The linear term will disappear on taking expectations because $\mathbb{E}[X - m] = 0$. Since $g(m) = 1 - \varepsilon = \pi(1)$ and

$$g''(x) = -\frac{2\varepsilon(1-\varepsilon)(1-2\varepsilon)}{d(x)^3} < 0,$$

g lies below its tangent at m. The exact distance from that tangent is

$$g(m) + g'(m)(x-m) - g(x) = \frac{4\varepsilon(1-\varepsilon)(1-2\varepsilon)(x-m)^2}{d(x)}.$$

Taking expectations therefore proves equation 55. The formula is positive for every finite N, and attributes the gap to fluctuations in the candidate counts, not only to batches that miss a state. The bounds $\varepsilon \le d(X) \le 1 - \varepsilon$ and $\mathbb{E}[(X-m)^2] = 1/(4N)$ give

$$\frac{\varepsilon(1-2\varepsilon)}{N} \le \pi(1) - \Pr(\widehat{Y}_{\mathrm{SIR}} = 1) \le \frac{(1-\varepsilon)(1-2\varepsilon)}{N}.$$

For the leading coefficient, it suffices to keep the quadratic term in Taylor's formula and bound its expected remainder by $O_\varepsilon(N^{-3/2})$. Here $g''(m) = -16\varepsilon(1-\varepsilon)(1-2\varepsilon)$. For fixed $\varepsilon \in (0, 1/2)$, the third derivative of g is bounded on $[0,1]$, because $d(x) \ge \varepsilon$. Taylor's formula gives

$$\mathbb{E}[g(X)] = g(m) + g'(m)\mathbb{E}[X-m] + \frac{g''(m)}{2}\mathbb{E}[(X-m)^2] + O_\varepsilon\left(\mathbb{E}[|X-m|^3]\right)$$

$$= g(m) + \frac{g''(m)}{8N} + O_\varepsilon(N^{-3/2}).$$

Here the remainder follows from

$$\mathbb{E}[|X-m|^3] \le \left(\mathbb{E}[(X-m)^4]\right)^{3/4} = O(N^{-3/2}), \qquad \mathbb{E}[(X-m)^4] = \frac{3}{16N^2} - \frac{1}{8N^3}.$$

Consequently,

$$\pi(1) - \Pr(\widehat{Y}_{\mathrm{SIR}} = 1) = \frac{2\varepsilon(1-\varepsilon)(1-2\varepsilon)}{N} + O_\varepsilon(N^{-3/2}).$$

Both samplers assign state 1 less than its target probability. On a binary space, each TV error equals this gap, proving Proposition 3. The leading SIR coefficient is strictly positive for every fixed $0 < \varepsilon < 1/2$. The remainder constants may depend on ε; no uniform limit as ε approaches an endpoint is claimed.

F.2 Equality with RS at $M = C_\infty^\pi$ for two weight values

The equality with RS at $M = C_\infty^\pi$ extends beyond two states: the relevant restriction is that the weights take at most two values. We retain a specific fallback rule for RS: if all candidates are rejected, each of the N observed candidate indices is selected with probability $1/N$. This is uniform selection among indices, not among distinct states.

Proposition 6 (Two possible weight values)**.** *Suppose $C_\infty^\pi < \infty$ and $w(Y) \in \{a, C_\infty^\pi\}$ almost surely under μ, where $0 \le a < C_\infty^\pi$. Let RS use threshold $M = C_\infty^\pi$ and the uniform fallback above. If all sampled weights vanish, let UR also select each candidate index with probability $1/N$. Then, for*

every fixed $N \geq 1$,

$$P_N^U = P_{N,C_\infty^\pi}^{\mathrm{RS}}.$$

Proof. It suffices to compare the probability of selecting each candidate on a fixed list, after uniformly randomizing the order in which RS processes it. The randomization preserves the RS marginal law because the proposals are independent and identically distributed; UR already treats candidate order symmetrically. There are two cases to check: whether the list contains a weight equal to the global maximum C_∞^π.

First express the order-averaged RS rule in a useful form. Conditional on the list, assign independent acceptance indicators with probabilities w_i/C_∞^π, independently of the random order. Given a nonempty accepted set, its first index in that order is uniform over the set. Thus this RS rule selects uniformly among accepted indices, or uses the specified fallback if none is accepted.

If a weight C_∞^π is present, it is also the largest observed weight. The acceptance probabilities are exactly those in UR's Bernoulli representation from Appendix E.3. At least one index is accepted with probability one. The retained set has the same distribution in the two Bernoulli representations, and both select uniformly from it, so their conditional selection probabilities agree. Zero-weight indices, if present, are never accepted or selected and can be removed when applying that representation.

If no weight C_∞^π is present, every observed weight is a. UR is uniform by symmetry, including the specified convention when $a = 0$. Order-averaged RS is also symmetric over these indices, both on acceptance and on its fallback, so every index has probability $1/N$. These cases cover every list under the two-value assumption. Averaging the equal conditional probabilities over the candidates proves the claimed equality of marginal laws. □

The two-value restriction makes a batch without a maximal-weight candidate have equal weights throughout. With more weight values, such a batch can still contain unequal weights, so this second case of the proof no longer applies.

F.3 Leading SIR bias

The binary calculation attributes SIR's error to random counts and normalization. To examine the same effect beyond state probabilities, let ψ be a bounded measurable function of the returned sample, and write $P(\psi) := \int \psi \mathrm{d}P$ for its expectation under P. For example, choosing $\psi = \mathbb{1}_D$ makes $P(\psi)$ the probability of an event D; taking all bounded ψ will later recover the TV error. Throughout this subsection and the next subsection, assume $0 < w(Y) \leq C < \infty$ μ-almost surely for a fixed C. No positive lower bound on w is required. Put $a := \pi(\psi)$, $b := \|\psi\|_\infty$, and

$$A_\psi := \mathbb{E}_\pi[(w - C^\pi)\psi].$$

No monotonicity of ψ is assumed. We prove

$$P_N^{\mathrm{SIR}}(\psi) - \pi(\psi) = -\frac{A_\psi}{N} + O_C(bN^{-3/2}). \tag{58}$$

The coefficient is signed and may vanish for a particular ψ. The TV expansion below instead considers all bounded measurable functions.

Reduction to two estimates. We will separate the bias into an exactly computable leading term and a remainder uniform over bounded ψ. Write $W_i := w(Y_i)$ and define

$$\overline{W}_N := \frac{1}{N}\sum_{i=1}^N W_i, \qquad Z_N := \frac{1}{N}\sum_{i=1}^N W_i(\psi(Y_i) - a), \qquad D_N := \overline{W}_N - 1.$$

Here Z_N is a centered weighted fluctuation, not an estimate of the normalizing constant Z. Conditional on the candidates, SIR's expected value of ψ is their weighted average. Subtracting a gives

$$\frac{\sum_{i=1}^N W_i\psi(Y_i)}{\sum_{i=1}^N W_i} - a = \frac{Z_N}{\overline{W}_N}.$$

Although $\mathbb{E}[Z_N] = 0$, dividing by the dependent random denominator need not preserve that zero mean. The identity

$$\frac{1}{1+D_N} = 1 - D_N + \frac{D_N^2}{1+D_N}$$

separates the leading interaction with the denominator from a remainder:

$$P_N^{\mathrm{SIR}}(\psi) - a = -\mathbb{E}[Z_N D_N] + \mathbb{E}\left[\frac{Z_N D_N^2}{\overline{W}_N}\right]. \tag{59}$$

Thus it suffices to establish

$$\mathbb{E}[Z_N D_N] = \frac{A_\psi}{N}, \qquad \mathbb{E}\left[\left|\frac{Z_N D_N^2}{\overline{W}_N}\right|\right] = O_C(bN^{-3/2}).$$

The second estimate controls the signed remainder as well, and its uniform dependence on b will justify taking the TV supremum.

The leading coefficient. Let $Y \sim \mu$ and $W := w(Y)$. Both $W(\psi(Y) - a)$ and $W - 1$ have mean zero, so cross terms from distinct samples vanish. Consequently,

$$\begin{aligned}\mathbb{E}[Z_N D_N] &\overset{\text{(i)}}{=} \frac{1}{N}\mathbb{E}_\mu[W(\psi(Y) - a)(W-1)] \qquad (60)\\ &\overset{\text{(ii)}}{=} \frac{1}{N}\mathbb{E}_\mu[W^2(\psi(Y) - a)]\\ &\overset{\text{(iii)}}{=} \frac{1}{N}\mathbb{E}_\pi[(w - C^\pi)\psi]\\ &= \frac{A_\psi}{N}. \qquad (61)\end{aligned}$$

Here (i) uses independence and centering; (ii) uses $\mathbb{E}_\mu[W(\psi(Y) - a)] = 0$; and (iii) uses $w\mathrm{d}\mu = \mathrm{d}\pi$, $\mathbb{E}_\pi[w] = C^\pi$, and $a = \pi(\psi)$. Consequently, the leading bias in equation 59 comes from the dependence between the centered weighted sum and the total observed weight.

The remainder without a lower weight bound. We cannot bound $1/\overline{W}_N$ uniformly, because weights may approach zero. The required estimate will instead follow from two centered moments and a lower-tail probability. Set $B_N := \{\overline{W}_N \geq 1/2\}$. We first reduce to those quantities:

$$\begin{aligned}\mathbb{E}\left[\left|\frac{Z_N D_N^2}{\overline{W}_N}\right|\right] &\overset{\text{(i)}}{\leq} 2\mathbb{E}[|Z_N| D_N^2] + 2b\Pr(B_N^c)\\ &\overset{\text{(ii)}}{\leq} 2\left(\mathbb{E}[Z_N^2]\right)^{1/2}\left(\mathbb{E}[D_N^4]\right)^{1/2} + 2b\Pr(B_N^c).\end{aligned}$$

Here (i) follows since $1/\overline{W}_N \leq 2$ on B_N. On B_N^c, use the whole ratio instead: $Z_N/\overline{W}_N$ is a weighted average of ψ minus a, so its absolute value is at most $2b$; also $|D_N| \leq 1$ there. (ii) follows from Cauchy–Schwarz inequality.

It remains to bound the three quantities on the right. Independence and centering give

$$\mathbb{E}[Z_N^2] = \frac{1}{N}\mathbb{E}_\mu[W^2(\psi(Y) - a)^2] \leq \frac{4Cb^2}{N}.$$

In the fourth moment of the centered sum, only four equal indices or two equal pairs survive, so

$$\begin{aligned}\mathbb{E}[D_N^4] &= \frac{N\mathbb{E}_\mu[(W-1)^4] + 3N(N-1)\left(\mathbb{E}_\mu[(W-1)^2]\right)^2}{N^4}\\ &= O_C(N^{-2}).\end{aligned}$$

Finally, since $0 < W \leq C$ and $\mathbb{E}_\mu[W] = 1$, Hoeffding's inequality yields

$$\Pr(B_N^c) \leq \exp\left(-\frac{N}{2C^2}\right).$$

Substituting these estimates into the displayed bound gives $O_C(bN^{-3/2})$. Together with the leading coefficient, this proves equation 58 without an inverse moment assumption.

From expectation bias to TV error. The remainder is uniform over $\|\psi\|_\infty \le 1$, which is essential when taking the supremum that defines TV. Hence

$$\begin{aligned} D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi) &= \frac{1}{2} \sup_{\|\psi\|_\infty \le 1} |P_N^{\mathrm{SIR}}(\psi) - \pi(\psi)| \\ &\overset{\text{(i)}}{=} \frac{1}{2N} \sup_{\|\psi\|_\infty \le 1} |A_\psi| + O_C(N^{-3/2}) \\ &\overset{\text{(ii)}}{=} \frac{1}{2N} \mathbb{E}_\pi[|w - C^\pi|] + O_C(N^{-3/2}). \end{aligned} \tag{62}$$

Here (i) uses the uniform remainder in equation 58. For (ii), choose $\psi(y) = \mathrm{sign}(w(y) - C^\pi)$ to attain the supremum. The coefficient vanishes exactly when $w = C^\pi$ π-almost surely. Because the weights are positive, π and μ have the same null sets; hence w is then constant μ-almost surely. The identity $\mathbb{E}_\mu[w] = 1$ forces that constant to be one, so the coefficient vanishes exactly when $\pi = \mu$. The leading term in equation 58 is the classical asymptotic bias term for self-normalized importance sampling; see, for example, Deligiannidis et al. (2026, Theorem 2.1). The proof here additionally supplies the displayed remainder uniformly over bounded measurable functions ψ under a finite upper weight bound, without requiring a positive lower weight bound or an inverse moment assumption.

Combining equation 62 with Corollary 1, every fixed nontrivial pair with bounded positive weights therefore has SIR error of order $1/N$, whereas UR's TV error decays exponentially.

F.4 How UR cancels the leading SIR bias

With the definitions and assumptions above, we prove

$$P_N^U(\psi) - P_N^{\mathrm{SIR}}(\psi) = \frac{A_\psi}{N} + o(N^{-1}). \tag{63}$$

The marginal coefficient also follows from equation 58 and the exponential UR error bound. Here we derive it from the choices within a fixed batch, to explain how UR changes the SIR probabilities before averaging over proposals. Let p_i^U and p_i^{SIR} denote the selection probabilities conditional on the realized candidates, and define

$$\Delta_N(\psi) := \sum_{i=1}^N (p_i^U - p_i^{\mathrm{SIR}})\psi(Y_i).$$

Its expectation is the left side of equation 63. Write $W_\Sigma := \sum_{i=1}^N W_i$ and $Q_\Sigma := \sum_{i=1}^N W_i^2$. The event B_N from the preceding subsection is equivalently $\{W_\Sigma \ge N/2\}$. We first show how a batchwise expansion implies equation 63, then reduce that expansion to a candidate-wise estimate and prove the remaining integral approximation.

Proof from a batchwise expansion. We will establish, on B_N and for sufficiently large N depending only on C,

$$\Delta_N(\psi) = \frac{\sum_i W_i^2 \psi(Y_i)}{W_\Sigma^2} - \frac{Q_\Sigma \sum_i W_i \psi(Y_i)}{W_\Sigma^3} + O_C(bN^{-2}). \tag{64}$$

To see why this is sufficient, split the expected correction according to B_N:

$$N\mathbb{E}[\Delta_N(\psi)] = \mathbb{E}[N\Delta_N(\psi)\mathbb{1}_{B_N}] + \mathbb{E}[N\Delta_N(\psi)\mathbb{1}_{B_N^c}].$$

It is enough for the first term to converge to A_ψ and the second to vanish. Assuming equation 64, the strong law gives

$$\begin{aligned} N\Delta_N(\psi)\mathbb{1}_{B_N} &= \left[\frac{N^{-1}\sum_i W_i^2 \psi(Y_i)}{(W_\Sigma/N)^2} - \frac{(Q_\Sigma/N)(N^{-1}\sum_i W_i \psi(Y_i))}{(W_\Sigma/N)^3} \right] \mathbb{1}_{B_N} + O_C(b/N) \\ &\to \mathbb{E}_\pi[w\psi] - C^\pi \pi(\psi) \\ &= A_\psi \qquad \text{almost surely.} \end{aligned}$$

Here $W_\Sigma/N \to 1$, $Q_\Sigma/N \to C^\pi$, and the two weighted sample averages converge to $\mathbb{E}_\pi[w\psi]$ and $\pi(\psi)$; in particular, B_N occurs eventually almost surely. Each of the two leading terms, restricted to B_N, is at most $2Cb$ in absolute value. In particular, $Q_\Sigma \le CW_\Sigma$, $|\sum_i W_i\psi(Y_i)| \le bW_\Sigma$, and $W_\Sigma \ge N/2$. The remainder is uniform over batches in B_N, so dominated convergence gives $\mathbb{E}[N\Delta_N(\psi)\mathbb{1}_{B_N}] \to A_\psi$. On the complement, $|\Delta_N(\psi)| \le 2b$ and the preceding Hoeffding bound give

$$N\mathbb{E}[|\Delta_N(\psi)|\mathbb{1}_{B_N^c}] \le 2bN \exp\left(-\frac{N}{2C^2}\right) \to 0.$$

Hence the batchwise expansion implies equation 63. We now prove that expansion.

A sufficient candidate-wise estimate. Because $\Delta_N(\psi)$ is a sum over N candidates, it suffices to prove the following expansion uniformly in i on B_N:

$$p_i^U - p_i^{\mathrm{SIR}} = \frac{W_i}{W_\Sigma^2}\left(W_i - \frac{Q_\Sigma}{W_\Sigma}\right) + O_C(N^{-3}). \tag{65}$$

Multiplying by $\psi(Y_i)$ and summing gives the two leading terms in equation 64, with total remainder $O_C(bN^{-2})$. The reference value $Q_\Sigma/W_\Sigma = \sum_i p_i^{\mathrm{SIR}} W_i$ is the SIR-weighted average of the observed weights. The leading correction therefore increases the probabilities of candidates above this average and decreases those below it. This describes the leading term, not the exact sign of every finite-budget difference. Its total mass is zero:

$$\sum_i \frac{W_i}{W_\Sigma^2}\left(W_i - \frac{Q_\Sigma}{W_\Sigma}\right) = \frac{Q_\Sigma}{W_\Sigma^2} - \frac{Q_\Sigma}{W_\Sigma^2} = 0.$$

We next derive this estimate from the exact selection probabilities.

Reduction to the selection integral. Fix a positive batch and set

$$z_j := \frac{W_j}{\max_k W_k}, \qquad \lambda_i := \sum_{j\ne i} z_j, \qquad s_{2,i} := \sum_{j\ne i} z_j^2.$$

The exact selection law equation 49 gives

$$\frac{p_i^U}{z_i} = \int_0^1 \prod_{j\ne i}(1 - z_j t)\mathrm{d}t.$$

We will prove, uniformly for $\lambda_i \ge 1$,

$$\frac{p_i^U}{z_i} = \frac{1}{\lambda_i} - \frac{s_{2,i}}{\lambda_i^3} + O(\lambda_i^{-3}). \tag{66}$$

The remainder constant is independent of the batch. The second term cannot be absorbed into the remainder: its numerator $s_{2,i}$ can grow with λ_i, although $s_{2,i} \le \lambda_i$.

First use this integral approximation to obtain equation 65. Since $0 < z_i \le 1$,

$$p_i^{\mathrm{SIR}} = \frac{z_i}{\lambda_i + z_i} = \frac{z_i}{\lambda_i} - \frac{z_i^2}{\lambda_i^2} + O\left(\frac{z_i}{\lambda_i^3}\right).$$

Multiplying equation 66 by z_i and subtracting this expansion yields

$$p_i^U - p_i^{\mathrm{SIR}} = \frac{z_i}{\lambda_i^2}\left(z_i - \frac{s_{2,i}}{\lambda_i}\right) + O\left(\frac{z_i}{\lambda_i^3}\right). \tag{67}$$

On B_N, $W_i \le C$ and $W_\Sigma \ge N/2$ imply

$$\lambda_i \ge \frac{N/2 - C}{C} \ge \frac{N}{4C} \qquad \text{when } N \ge 4C.$$

Thus $\lambda_i \ge 1$ and the remainder is uniformly $O_C(N^{-3})$. In the original weights, the leading expression is

$$\frac{W_i^2}{(W_\Sigma - W_i)^2} - \frac{W_i(Q_\Sigma - W_i^2)}{(W_\Sigma - W_i)^3}.$$

Because $W_i \le C$, $Q_\Sigma \le CW_\Sigma$, and $W_\Sigma - W_i \ge N/4$, replacing $W_\Sigma - W_i$ by W_Σ in the denominators changes this expression by $O_C(N^{-3})$, uniformly in i. The term with numerator W_i^3 is

also $O_C(N^{-3})$. This proves equation 65, and hence equation 63, once the integral approximation is verified.

Proof of the integral approximation. The product in the selection integral is bounded by $e^{-\lambda_i t}$, so its mass concentrates near $t = 0$ as λ_i grows. The approximation in equation 66 will follow by replacing that product with $e^{-\lambda_i t}(1 - s_{2,i}t^2/2)$, since

$$\int_0^\infty e^{-\lambda_i t}\left(1 - \frac{s_{2,i}t^2}{2}\right)\mathrm{d}t = \frac{1}{\lambda_i} - \frac{s_{2,i}}{\lambda_i^3}.$$

To justify the replacement, it suffices to bound the difference on $[0, 1/2]$ and both tails starting at $1/2$ by $O(\lambda_i^{-3})$.

For the local difference, we claim that, uniformly on $0 \le t \le 1/2$,

$$\prod_{j\neq i}(1 - z_j t) = e^{-\lambda_i t}\left[1 - \frac{s_{2,i}t^2}{2} + O(\lambda_i t^3 + \lambda_i^2 t^4)\right].$$

Its integrated error is at most a constant times

$$\int_0^{1/2} e^{-\lambda_i t}(\lambda_i t^3 + \lambda_i^2 t^4)\mathrm{d}t \le \lambda_i \frac{3!}{\lambda_i^4} + \lambda_i^2 \frac{4!}{\lambda_i^5} = O(\lambda_i^{-3}),$$

using $\int_0^\infty t^k e^{-\lambda_i t}\mathrm{d}t = k!/\lambda_i^{k+1}$. The original product's tail is at most $e^{-\lambda_i/2}/\lambda_i$. The absolute tail of the approximating function is at most

$$\int_{1/2}^\infty e^{-\lambda_i t}\left(1 + \frac{\lambda_i t^2}{2}\right)\mathrm{d}t,$$

since $s_{2,i} \le \lambda_i$. Both tails are uniformly $O(\lambda_i^{-3})$ for $\lambda_i \ge 1$. Thus only the claimed local approximation remains to be checked.

The logarithmic series gives

$$\prod_{j\neq i}(1 - z_j t) = e^{-\lambda_i t}\exp\left(-\frac{s_{2,i}t^2}{2} - R_i(t)\right), \qquad 0 \le R_i(t) \le \frac{2}{3}\lambda_i t^3.$$

Specifically, $\sum_{k\ge 3} u^k/k \le 2u^3/3$ for $0 \le u \le 1/2$, and $\sum_{j\neq i} z_j^3 \le \lambda_i$. For $u, v \ge 0$,

$$|e^{-u-v} - 1 + u| \le |e^{-u} - 1 + u| + 1 - e^{-v} \le \frac{u^2}{2} + v.$$

Apply this with $u = s_{2,i}t^2/2$ and $v = R_i(t)$. Since $s_{2,i} \le \lambda_i$, this gives the claimed local remainder. This verifies equation 66 and completes the candidate-wise, batchwise, and marginal arguments above. The exponential UR target error follows from equation 12, not merely from cancellation of the displayed N^{-1} coefficients.

F.5 Continuous example

Consider $\mu = \mathrm{Unif}(0, 1)$ and $\pi(\mathrm{d}y) = 2y\mathrm{d}y$ on $(0, 1)$, so $w(y) = 2y$ and the target mean is $2/3$. The two output means are

$$\mathbb{E}[\widehat{Y}_U] = \frac{2}{3} - \frac{2^{1-N}}{3(N+1)}, \qquad \mathbb{E}[\widehat{Y}_{\mathrm{SIR}}] = \frac{2}{3} - \frac{1}{9N} + O(N^{-3/2}). \tag{68}$$

These formulae concern means, not TV distances. For UR, we first express the mean error through the largest score; this shows which part of its distribution must be computed. For SIR, it suffices to evaluate A_ψ from Appendix F.3 at $\psi(y) = y$.

UR: Reduce the mean error to scores below 2. The conditional law equation 9 has mean $s/3$ for $0 < s < 2$ and $2/3$ for $s \ge 2$. The first value is the mean of the target restricted to $(0, s/2)$; in the second case, no target mass is excluded. Consequently, the law of total expectation gives

$$\frac{2}{3} - \mathbb{E}[\widehat{Y}_U] = \mathbb{E}\left[\left(\frac{2}{3} - \frac{S_{(N)}}{3}\right)\mathbb{1}\{S_{(N)} < 2\}\right].$$

It remains to compute the largest-score density on $(0, 2)$ and evaluate this expectation. For independent $Y \sim \mu$ and $U \sim \mathrm{Unif}(0, 1)$, set $S := 2Y/U$. Since $\Pr(S \le s \mid Y = y) = (1 - 2y/s)_+$ for $s > 0$,

$$G(s) = \int_0^{\min\{1, s/2\}} (1 - 2y/s)\mathrm{d}y = \begin{cases} s/4, & 0 < s < 2, \\ 1 - 1/s, & s \ge 2. \end{cases}$$

Independence makes the largest-score CDF equal to $G(s)^N$. Its derivative on $(0, 2)$ is $Ns^{N-1}/4^N$. Substituting into the preceding expectation gives

$$\begin{aligned} \frac{2}{3} - \mathbb{E}[\widehat{Y}_U] &\overset{\text{(i)}}{=} \int_0^2 \left(\frac{2}{3} - \frac{s}{3}\right) \frac{Ns^{N-1}}{4^N} \mathrm{d}s \\ &\overset{\text{(ii)}}{=} \frac{2^{1-N}}{3(N+1)}. \end{aligned} \tag{69}$$

Step (i) uses the score density in the preceding reduction, and (ii) evaluates the polynomial integral.

SIR: Evaluate the general bias coefficient. For $\psi(y) = y$,

$$C^\pi = \frac{4}{3}, \qquad A_\psi = \int_0^1 (2y - 4/3)y(2y)\mathrm{d}y = 1 - \frac{8}{9} = \frac{1}{9}.$$

Substituting into equation 58 proves the second formula in equation 68. Although $w(y)$ approaches zero near the origin, it is positive almost surely and bounded above, so the assumptions of that bias calculation apply.

G Proof of Theorem 4

Fix $N \ge 2$ and maintain the structural restrictions on A mentioned in Section 5. The proof keeps relative weights fixed while varying their proposal probabilities. The resulting marginal output probabilities are polynomials in these probabilities; the required error rate forces their low-degree coefficients to agree with UR's, thereby determining the selection probabilities on each batch. We first illustrate the argument for two and three candidates, then provide the general proof.

Why accuracy determines a choice. Take $N = 2$ and keep the relative weights of states 0 and 1 fixed at $1/2$ and 1. We change only how often state 0 is proposed: let $\mu_\eta(0) = \eta$ and $\mu_\eta(1) = 1 - \eta$, where $0 < \eta < 1$. Since $\pi(y) \propto w_u(y)\mu(y)$,

$$\pi_\eta(0) = \frac{\eta/2}{\eta/2 + 1 - \eta} = \frac{\eta}{2 - \eta}, \qquad C_\infty^{\pi_\eta} = \frac{1}{1 - \eta/2} = \frac{2}{2 - \eta}.$$

Let q denote the probability that A selects state 0 when the batch contains both states. The sampler sees the same relative weights $(1/2, 1)$ on every such batch, regardless of η, and symmetry makes this probability independent of the order of the two candidates. Hence the same q must be used for every η. Decomposing over the three batch types,

$$P_2^{\mathsf{A}}(\{0\}) = \Pr(00) + \Pr(\{01, 10\}) \Pr(\text{return } 0 \mid \{01, 10\}) = \eta^2 + 2\eta(1 - \eta)q.$$

For small η, one can readily see that $P_2^{\mathsf{A}}(\{0\}) = 2q\eta + O(\eta^2)$ and $\pi_\eta(0) = \frac{\eta}{2} + O(\eta^2)$. Hence any mismatch between $2q$ and $1/2$ creates a first-order error in η. However, the assumed guarantee allows only second-order error, since $1 - \frac{1}{C_\infty^{\pi_\eta}} = \frac{\eta}{2}$. Since the state space is binary, $|P_2^{\mathsf{A}}(\{0\}) - \pi_\eta(0)| = D_{\mathrm{TV}}(P_2^{\mathsf{A}}, \pi_\eta)$ and equation 25 becomes:

$$\left|\eta^2 + 2\eta(1 - \eta)q - \frac{\eta}{2 - \eta}\right| = |P_2^{\mathsf{A}}(\{0\}) - \pi_\eta(0)| \le \frac{L_2\eta^2}{4} \iff \left|\eta + 2(1 - \eta)q - \frac{1}{2 - \eta}\right| \le \frac{L_2\eta}{4}.$$

Letting $\eta \downarrow 0$, this forces $2q = 1/2$, so $q = 1/4$. This is exactly UR's probability, since $\Pr(U_L < U_H/2) = 1/4$ for independent uniforms on $(0, 1)$. The general proof below applies a similar argument with several rare states, determining the probabilities on every possible batch.

G.1 THREE CANDIDATES SHOW HOW ACCURACY DETERMINES A BATCHWISE CHOICE

The preceding binary example determines a choice from a first-order term. With three candidates, a batch containing two rare states first appears in a second-order term. This is why one must examine more than the probability of seeing a single rare candidate.

Take distinct relative weights $a, b \in (0, 1)$ and a reference weight 1. Give these three states proposal probabilities $\eta_1, \eta_2, 1-\eta_1-\eta_2$, respectively, where $\eta_1, \eta_2 > 0$ and $\eta_1+\eta_2 < 1$. Write $\boldsymbol{\eta} := (\eta_1, \eta_2)$. The common normalizer is $Z_{\boldsymbol{\eta}} := 1-(1-a)\eta_1-(1-b)\eta_2$. The largest normalized importance weight is therefore

$$C_\infty^{\pi_{\boldsymbol{\eta}}} = \frac{1}{Z_{\boldsymbol{\eta}}},$$

and hence

$$1 - \frac{1}{C_\infty^{\pi_{\boldsymbol{\eta}}}} = (1-a)\eta_1 + (1-b)\eta_2.$$

The assumed bound equation 25, specialized to $N = 3$, gives

$$D_{\mathrm{TV}}(P_3^{\mathsf{A}}, \pi_{\boldsymbol{\eta}}) \le L_3 \left((1-a)\eta_1 + (1-b)\eta_2\right)^3.$$

Therefore, any first- or second-order mismatch between the sampler and the target would violate the cubic error bound.

First, consider the separate binary family containing only relative weights a and 1. Let ℓ be the probability that A selects the sole a candidate from $(a, 1, 1)$. The sampler returns the lower weight state with probability $3\eta_1(1-\eta_1)^2\ell + O(\eta_1^2) = 3\eta_1\ell + O(\eta_1^2)$, while the corresponding target probability is

$$\frac{a\eta_1}{1-(1-a)\eta_1} = a\eta_1 + O(\eta_1^2).$$

The assumed error is $O(\eta_1^3)$, so their linear coefficients agree and $\ell = a/3$.

Now return to the ternary state family and let x be the probability of selecting the a candidate from $(a, b, 1)$. The target probability of the state with relative weight a expands as:

$$\frac{a\eta_1}{Z_{\boldsymbol{\eta}}} = a\eta_1 + a(1-a)\eta_1^2 + a(1-b)\eta_1\eta_2 + O((\eta_1+\eta_2)^3).$$

Hence its mixed $\eta_1\eta_2$ coefficient is $a(1-b)$. We now compute the same coefficient for the sampler. The contribution of the batch composition $(a, 1, 1)$ is

$$3\eta_1(1-\eta_1-\eta_2)^2\frac{a}{3} = a\eta_1 - 2a\eta_1^2 - 2a\eta_1\eta_2 + O((\eta_1+\eta_2)^3),$$

so its mixed coefficient is $-2a$. The contribution of $(a, b, 1)$ is

$$6\eta_1\eta_2(1-\eta_1-\eta_2)x = 6x\eta_1\eta_2 + O((\eta_1+\eta_2)^3),$$

so its mixed coefficient is $6x$. Batches without an a candidate contribute zero. Every remaining composition with an a candidate contains either at least two a candidates or three lower weight candidates, so none contributes to the $\eta_1\eta_2$ coefficient. Therefore the sampler's mixed coefficient is $-2a + 6x$.

The cubic error bound forces this coefficient to match the target's, giving

$$-2a + 6x = a(1-b).$$

Solving for x gives

$$x = \frac{a(3-b)}{6} = a\int_0^1 (1-bu)(1-u)\mathrm{d}u, \tag{70}$$

which is precisely UR's probability from equation 49. The contribution from $(a, 1, 1)$ was determined first, leaving the mixed coefficient to determine the choice on $(a, b, 1)$. The general proof repeats this procedure in increasing order of the number of lower weight candidates.

G.2 General proof in three steps

Step 1. Construct a family with fixed relative weights. We construct a family of target–proposal pairs that keeps the relative weights fixed while making the lower weight states rare.

To this end, we choose $1 \le d \le N-1$ distinct relative weights $a_1, \dots, a_d \in (0,1)$ and let $a_0 := 1$. State 0 is the reference state with the largest relative weight, while states $1, \dots, d$ have smaller fixed weights. For $j = 1, \dots, d$, choose $\eta_j > 0$ with $\sum_{j=1}^{d} \eta_j < 1$, and set

$$\eta_0 := 1 - \sum_{j=1}^{d} \eta_j.$$

We consider the setting where the a_j's remain fixed throughout the argument, whereas the η_j's tend to zero.

Now write $\boldsymbol{\eta} := (\eta_1, \dots, \eta_d)$ and define the normalizing constant

$$Z_{\boldsymbol{\eta}} := \sum_{j=0}^{d} a_j \eta_j = 1 - \sum_{j=1}^{d} (1 - a_j)\eta_j.$$

The proposal, target, and normalized weights are, for $j = 0, \dots, d$,

$$\mu_{\boldsymbol{\eta}}(j) := \eta_j, \qquad \pi_{\boldsymbol{\eta}}(j) := \frac{a_j \eta_j}{Z_{\boldsymbol{\eta}}}, \qquad w_{\boldsymbol{\eta}}(j) := \frac{a_j}{Z_{\boldsymbol{\eta}}}. \tag{71}$$

Hence the fixed values a_j are unnormalized importance weights, in the same sense as $w_u = Zw$ in Section 1. Note that varying $\boldsymbol{\eta}$ changes their frequencies and their common normalization, but not their ratios. The largest normalized weight is $C_\infty^{\pi_{\boldsymbol{\eta}}} = 1/Z_{\boldsymbol{\eta}}$, so

$$1 - \frac{1}{C_\infty^{\pi_{\boldsymbol{\eta}}}} = \sum_{j=1}^{d} (1 - a_j)\eta_j. \tag{72}$$

As the smaller weights become rare, this quantity tends to zero and the required error becomes small.

Step 2: Express marginal output differences as a polynomial. We compare A with UR rather than directly with the target. The target probabilities involve the changing normalizer $Z_{\boldsymbol{\eta}}^{-1}$, whereas both samplers' output probabilities are polynomials in $\boldsymbol{\eta}$. To this end, let $\boldsymbol{n} := (n_1, \dots, n_d)$ be a count vector which records how many candidates from each lower weight state appear in the batch. Here $|\boldsymbol{n}| := \sum_{j=1}^{d} n_j \le N$, so the remaining $N - |\boldsymbol{n}|$ candidates are in state 0. For example, if $d = 2$, then $\boldsymbol{n} = (1, 2)$ means one candidate from state 1, two from state 2, and the remaining $N - 3$ candidates from state 0. Let $q_j^{\mathsf{A}}(\boldsymbol{n})$ be the probability of returning state j from this batch, and define $q_j^{U}(\boldsymbol{n})$ similarly. Unlike p_i^{A}, which refers to one candidate, q_j^{A} sums the probabilities of all candidates in state j. Neither the ordering nor the common normalization affects these probabilities, so they are independent of $\boldsymbol{\eta}$. Write $\Delta_j(\boldsymbol{n}) := q_j^{\mathsf{A}}(\boldsymbol{n}) - q_j^{U}(\boldsymbol{n})$.

Lemma 9 (Output probability differences are polynomials)**.** *For the pair in equation 71 and $j = 1, \dots, d$, let $D_j(\boldsymbol{\eta}) := P_N^{\mathsf{A}}(\{j\}) - P_N^{U}(\{j\})$. Then*

$$D_j(\boldsymbol{\eta}) = \sum_{|\boldsymbol{n}| \le N} \frac{N!}{(N - |\boldsymbol{n}|)! \prod_{h=1}^{d} n_h!} \eta_0^{N - |\boldsymbol{n}|} \prod_{h=1}^{d} \eta_h^{n_h} \Delta_j(\boldsymbol{n}). \tag{73}$$

This is a polynomial in $\eta_1, \dots, \eta_d$ of total degree at most N.

Proof. For each count vector, the multinomial factor times the powers of η_h is the probability of that batch composition. Multiplying by its conditional output difference and summing is the law of total probability. Since $\eta_0 = 1 - \sum_{h=1}^{d} \eta_h$, each term has degree at most N. No continuity of the selection functions is needed: their values on each fixed relative weight vector remain unchanged while the probabilities of those vectors vary. □

Step 3: Use the error guarantee to recover the batchwise probabilities. We now use the assumed bound equation 25 to determine the choices within each batch. Fix $\boldsymbol{v} \in (0, \infty)^d$ and let $\boldsymbol{\eta} = \theta \boldsymbol{v}$ for

sufficiently small $\theta > 0$. By equation 72, the triangle inequality and the two error bounds give

$$\begin{aligned}|D_j(\theta \boldsymbol{v})| &\overset{\text{(i)}}{\leq} D_{\mathrm{TV}}(P_N^{\mathsf{A}}, \pi_{\theta \boldsymbol{v}}) + D_{\mathrm{TV}}(P_N^{U}, \pi_{\theta \boldsymbol{v}}) \\ &\overset{\text{(ii)}}{\leq} (L_N + 1) \left(\theta \sum_{h=1}^{d} (1 - a_h) v_h \right)^N \\ &= O(\theta^N) \\ &= o(\theta^{N-1}),\end{aligned}$$

where (i) compares both samplers with the same target on the event $\{j\}$; and (ii) uses equation 25 and Corollary 1. The constant L_N cannot change with θ, since it is independent of the target and proposal. The following lemma explains why this small error removes every lower-degree term.

Lemma 10 (Small marginal differences remove the lower-degree terms)**.** *Suppose that, for every fixed $\boldsymbol{v} \in (0, \infty)^d$,*

$$D_j(\theta \boldsymbol{v}) = o(\theta^{N-1}) \qquad (\theta \downarrow 0).$$

Then every coefficient of D_j of total degree below N is zero.

Proof. Setting $\boldsymbol{\eta} = \theta \boldsymbol{v}$ scales all rare-state probabilities down together while keeping their proportions fixed. For this fixed $\boldsymbol{v}$, write

$$D_j(\theta \boldsymbol{v}) = c_0(\boldsymbol{v}) + c_1(\boldsymbol{v})\theta + \cdots + c_N(\boldsymbol{v})\theta^N.$$

If the first nonzero term had degree $m < N$, dividing by θ^m would give a nonzero limit. The assumed bound instead gives a limit of zero, including when $m = N - 1$. Hence $c_m(\boldsymbol{v}) = 0$ for every $m < N$.

This conclusion holds for every positive $\boldsymbol{v}$, not just one choice of proportions. For each m, $c_m(\boldsymbol{v})$ is exactly the degree-m part of D_j evaluated at $\boldsymbol{v}$. It is a polynomial that vanishes throughout the positive orthant, and therefore all its coefficients vanish. In particular, fix all but one coordinate in an open box and use that a univariate polynomial vanishing on an interval is zero; repeating this for the other coordinates proves the claim. Lastly, notice that varying the proportions is essential: a polynomial such as $\eta_1 - \eta_2$ vanishes when $\eta_1 = \eta_2$ without having zero coefficients. □

Lemma 11 (The coefficients determine the selection probabilities)**.** *If all coefficients of D_j of total degree below N vanish for $j = 1, \ldots, d$, then $\Delta_j(\boldsymbol{n}) = 0$ whenever $|\boldsymbol{n}| < N$. Allowing the relative weights to vary identifies each candidate's selection probability on every positive input vector.*

Proof. Fix $j \geq 1$ and proceed in increasing order of $m := |\boldsymbol{n}|$. For $m = 0$, state j is absent, so both rules assign it probability zero. Now fix $1 \leq m \leq N - 1$ and suppose all compositions with fewer than m lower weight candidates already have $\Delta_j = 0$. Their terms in equation 73 vanish entirely. Compositions with more than m lower weight candidates cannot contribute to degree m, since they contain more than m factors of the rare-state probabilities. For a composition with exactly m such candidates,

$$\eta_0^{N-m} = \left(1 - \sum_{h=1}^{d} \eta_h \right)^{N-m}$$

has constant term one, and all its other terms raise the total degree above m. Therefore the degree-m part is exactly

$$\sum_{|\boldsymbol{n}|=m} \frac{N!}{(N-m)! \prod_{h=1}^{d} n_h!} \Delta_j(\boldsymbol{n}) \prod_{h=1}^{d} \eta_h^{n_h}.$$

Each composition produces a distinct monomial with a positive multinomial coefficient. Since the displayed polynomial is identically zero, every $\Delta_j(\boldsymbol{n})$ with $|\boldsymbol{n}| = m$ is zero. Induction therefore proves the assertion for $j \geq 1$; the probabilities of state 0 then agree since each rule's probabilities sum to one.

Finally, divide an arbitrary positive weight vector by its maximum. If all entries become one, symmetry gives probability $1/N$ to each candidate. Otherwise, its distinct smaller values are some

$a_1, \ldots, a_d$, with multiplicities $n_1, \ldots, n_d$. At least one maximal entry remains, so $|\boldsymbol{n}| \le N-1$. The preceding induction identifies the total selection probability of each equal-weight group. Within a group, exchanging two candidates leaves the weight vector unchanged, so symmetry assigns them equal probabilities. In conclusion, the group probability determines each individual probability. □

Lemma 10 removes the lower-degree coefficients, and Lemma 11 determines every batchwise probability. Conversely, Corollary 1 supplies the UR bound with $L_N = 1$. This completes the proof of Theorem 4.

A weaker local condition. For fixed N, the same conclusion holds under the weaker requirement

$$\sup_{(\pi,\mu):C_\infty^\pi \le 1+\delta} D_{\mathrm{TV}}(P_N^{\mathsf{A}}, \pi) = o(\delta^{N-1}) \qquad (\delta \downarrow 0),$$

where the supremum ranges over finite target–proposal pairs with strictly positive probabilities. For each fixed $\boldsymbol{v} \in (0,\infty)^d$, equation 72 shows that $C_\infty^{\pi_{\theta \boldsymbol{v}}} - 1$ is proportional to θ to first order. The local requirement and UR's bound therefore give $D_j(\theta \boldsymbol{v}) = o(\theta^{N-1})$. Lemmas 10 and 11 then apply unchanged.

H NONUNIFORM DIVISORS CAN ALSO CONVERGE EXPONENTIALLY

Uniformity of the auxiliary variables is not necessary for exponential convergence. Consider the more general score

$$S_i^F := \frac{w(Y_i)}{T_i},$$

where the T_i are i.i.d. positive random variables independent of the proposals. As a simple example, let

$$T \sim \frac{1}{2}\,\mathrm{Unif}(0,1) + \frac{1}{2}\,\mathrm{Unif}(2,3).$$

Its CDF satisfies

$$F(t) = \frac{t}{2}, \qquad 0 \le t \le 1.$$

Suppose $C_\infty^\pi < \infty$. A candidate with state y has score at least C_∞^π exactly when

$$T_i \le \frac{w(y)}{C_\infty^\pi}.$$

Since $w(y)/C_\infty^\pi \le 1$,

$$\Pr\left(T_i \le \frac{w(y)}{C_\infty^\pi} \middle| Y_i = y\right) = \frac{w(y)}{2C_\infty^\pi}.$$

Let $\widehat{I}_F := \arg\max_{i\in[N]} S_i^F$ and $P_N^F := \mathcal{L}(Y_{\widehat{I}_F})$. Suppose $C_\infty^\pi < \infty$ and define

$$H_i := \left\{T_i \le \frac{w(Y_i)}{C_\infty^\pi}\right\}, \qquad V_i := \frac{C_\infty^\pi T_i}{w(Y_i)} \quad \text{on } H_i.$$

For every measurable set D and $0 \le v \le 1$,

$$\Pr(Y_i \in D, H_i, V_i \le v) = \int_D F\left(\frac{v w(y)}{C_\infty^\pi}\right)\mu(dy) = \frac{v}{2C_\infty^\pi}\pi(D).$$

Hence, one can observe that conditional on H_i, $Y_i \sim \pi$ and $V_i \sim \mathrm{Unif}(0,1)$ are independent. Conditioning on the set of crossing indices preserves independence across candidates, since each crossing event depends only on its own candidate and divisor. Among the candidates satisfying H_i, the rule selects the smallest V_i, since $S_i^F = C_\infty^\pi / V_i$ on H_i. Therefore, conditional on at least one threshold crossing, the selected state has law π.

Each candidate crosses independently with probability $1/(2C_\infty^\pi)$. It follows that

$$D_{\mathrm{TV}}(P_N^F, \pi) \le \left(1 - \frac{1}{2C_\infty^\pi}\right)^N.$$

In conclusion, we observe that uniform divisors are not necessary for exponential convergence. Notice that this bound has a slower exponential rate than UR's bounded weight guarantee though, since only half of the divisor mass lies in the linear part of the CDF.

I Experimental Details

This section provides additional details on the experimental setup, evaluation procedure, and results.

Datasets. Throughout the experiments, we evaluate on two mathematical reasoning benchmarks.

GSM8K (Cobbe et al., 2021) is a dataset created by OpenAI, consisting of 8.5K grade-level math word problems. For each evaluated prompt, we pre-generated a response pool $\mathcal{A}_x^\star$ of $N^\star := 4{,}096$ responses from the reference policy, with a maximum generation length of 500 tokens.

MATH500 (Hendrycks et al., 2021; Lightman et al., 2024) is a 500-question evaluation subset of the MATH benchmark, composed of competition-level mathematics problems. Compared to GSM8K, MATH500 requires more complex multi-step mathematical reasoning, providing a setting where we can evaluate our method with more challenging problems. For each problem (i.e., evaluated prompt), we pre-generated a response pool $\mathcal{A}_x^\star$ of 4,096 responses, with a maximum generation length of 1,024 tokens.

Reference policy and response generation. We use `meta-llama/Llama-3.2-3B-Instruct` (Grattafiori et al., 2024) as the reference policy. Specifically, for each evaluated prompt x, we pre-generate a fixed response pool

$$\mathcal{A}_x^\star := \{a_1, \ldots, a_{N^\star}\}, \qquad N^\star = 4{,}096.$$

We use a generation temperature of 0.3 and maximum generation lengths of 500 tokens for GSM8K and 1,024 tokens for MATH500 respectively. The response generation and reward model settings are summarized in Table 2. All response generation and reward model evaluation were run on two NVIDIA RTX 6000 Pro Blackwell GPUs.

Component	Setting	Value
Response generation	Model	`meta-llama/Llama-3.2-3B-Instruct`
	Max tokens	500 (GSM8K); 1,024 (MATH500)
	Generation temperature	0.3
	Pre-generated responses/prompt	4,096
Reward model	Model	`OpenAssistant/reward-model-deberta-v3-large-v2`

Table 2: Response generation and reward model settings.

Finite empirical proposal. The fixed response pool $\mathcal{A}_x^\star$ defines the empirical distribution $\widehat{\mu}(y \mid x) := \frac{1}{N^\star}\sum_{i=1}^{N^\star} \mathbb{1}\{a_i = y\}$. Since $\mathcal{A}_x^\star$ may contain repeated responses, let $\mathrm{supp}(\widehat{\mu}(\cdot \mid x))$ denote the set of distinct responses represented in $\mathcal{A}_x^\star$. For brevity, we write $\mathrm{supp}(\widehat{\mu})$ when the prompt x is fixed. This fixed empirical proposal allows us to draw repeated independent candidate sets *without additional LLM generations.* The pool size $N^\star$ is fixed throughout the experiment, whereas N denotes the sampling budget available to each sampling procedure and is varied to produce the curves.

Reward model and target policy. Each distinct response $y \in \mathrm{supp}(\widehat{\mu})$ is scored by `OpenAssistant/reward-model-deberta-v3-large-v2` (Köpf et al., 2023), yielding a reward $\widehat{r}(x, y)$. Motivated by the standard KL-regularized RLHF objective (Jaques et al., 2017; 2020; Rafailov et al., 2023), we define the reward-tilted target policy

$$\pi_\beta(y \mid x) = \frac{\widehat{\mu}(y \mid x)\exp(\widehat{r}(x, y)/\beta)}{\sum_{y' \in \mathrm{supp}(\widehat{\mu})} \widehat{\mu}(y' \mid x)\exp(\widehat{r}(x, y')/\beta)}, \quad y \in \mathrm{supp}(\widehat{\mu}).$$

Since $\mathrm{supp}(\widehat{\mu})$ is finite, the normalizing constant can be computed exactly for evaluation. The sampling procedures themselves receive only the unnormalized importance weights

$$w_u(x, y) = \exp\left(\widehat{r}(x, y)/\beta\right).$$

Hence, $(\pi_\beta, \widehat{\mu})$ forms a finite target–proposal pair to which the theoretical results in Sections 2–5 apply directly.

I.1 Sampling Algorithms

We consider the four sampling algorithms mentioned in Section 6. In particular, all four samplers receive the same N i.i.d. candidates $Y_1, \ldots, Y_N \overset{\text{iid}}{\sim} \widehat{\mu}(\cdot \mid x)$ and their unnormalized weights $w_u(x, Y_i)$. Each sampler returns exactly one of the observed candidates. If $N = 1$, all four methods return the sole candidate and hence coincide with $\widehat{\mu}$. As the implementations of UR and SIR are described clearly in previous sections, we provide the detailed implementations of the remaining two algorithms.

Envelope RS. This baseline implements the approximate rejection sampler of Section 1. For convenience, we parameterize its threshold directly on the unnormalized weight scale. Given a threshold $\tau > 0$, candidates are examined in order and candidate i is accepted with probability

$$\min\left\{\frac{w_u(x, Y_i)}{\tau}, 1\right\}.$$

The first accepted candidate is returned; if all N candidates are rejected, one of the observed candidates is returned uniformly at random. Equivalently, $\tau = ZM$ for the normalized threshold M used in the theoretical formulation.

For each (x, β), we set the threshold using the fixed response pool $\mathcal{A}_x^\star$ and use the same threshold for every N. Specifically, we set $\tau := \max_{y \in \mathrm{supp}(\widehat{\mu})} w_u(x, y)$. Hence, envelope RS serves as a proxy for the best fixed-threshold RS in hindsight.

Budget-calibrated RS. This baseline is adapted from Rohatgi et al. (2025, Algorithm 5) as described in Section 3. In particular, given a sampling budget N, set $n = \left\lfloor \frac{N-1}{2} \right\rfloor$. For $N \in \{1, 2\}$, for which $n = 0$, we define the baseline to return a single fresh candidate $Y_1 \sim \widehat{\mu}(\cdot \mid x)$. The algorithm first uses n candidates as pilot samples to estimate the normalizing constant, $\widehat{Z} := \frac{1}{n}\sum_{i=1}^{n} w_u(x, Y_i)$, and the threshold multiplier is set to $M_{N,\delta} := \frac{n}{4\log(4/\delta)}$, $\delta \in (0, 1)$. The algorithm then runs standard RS on the next n candidates with acceptance probability

$$\min\left\{\frac{w_u(x, Y_i)}{M_{N,\delta}\widehat{Z}}, 1\right\},$$

and the fresh fallback candidate Y_{2n+1} is returned if all are rejected. Hence, at most $2n + 1 \le N$ candidates are used. In addition, we select δ from $\{0.001, 0.01, 0.05, 0.1, 0.25, 0.5, 0.9\}$ separately for each (x, β, N). For the TV plots, we choose the value minimizing the estimated TV error; for the accuracy plots, we choose the value maximizing ground-truth accuracy. All candidate values of δ are evaluated on the same Monte Carlo runs.

I.2 Evaluation Metrics and Estimation

We report two complementary metrics: (1) the TV error $D_{\mathrm{TV}}(P_N, \pi_\beta)$ and (2) ground-truth accuracy.

TV error. Let $P_N(\cdot \mid x)$ denote the marginal distribution of the response returned by a sampler with budget N. Since all four algorithms are designed to approximate the same target distribution π_β, we measure sampling fidelity by $D_{\mathrm{TV}}(P_N, \pi_\beta)$. We note that this is also the quantity controlled directly by our theoretical results.

Ground-truth accuracy. Ground-truth accuracy is not the objective optimized by the samplers. Rather, it measures whether the sampled response gives the correct final answer to the underlying math problem. For a fixed prompt x, define the target accuracy

$$\mathrm{Acc}(\pi_\beta; x) := \sum_y \pi_\beta(y \mid x)\mathbb{1}\{y \text{ has the correct final answer}\}.$$

Since correctness corresponds to an event, $\left|\mathrm{Acc}(P_N; x) - \mathrm{Acc}(\pi_\beta; x)\right| \le D_{\mathrm{TV}}(P_N, \pi_\beta)$. Hence, as the output distribution approaches π_β, its ground-truth accuracy approaches the corresponding target accuracy. We report accuracy to assess whether improved sampling fidelity also appears in downstream task performance. Importantly, the target accuracy may be either higher or lower than the reference policy accuracy, so accuracy need not increase monotonically with N.

I.2.1 Estimating the TV Error on a Finite Pool

Computing the TV error requires knowledge of the target distribution π_β and the output law P_N. Note that the target distribution π_β can be computed exactly on the finite empirical support $\mathrm{supp}(\widehat{\mu})$, but the output law P_N of a sampling procedure should be estimated through repeated trials.

To estimate P_N, for each prompt x, temperature parameter β, and sampling budget N, each trial independently draws $Y_1, \dots, Y_N \overset{\mathrm{iid}}{\sim} \widehat{\mu}(\cdot \mid x)$ and applies the corresponding sampling procedure. We repeat this experiment m times, using the trial counts specified below and record the returned responses $\widehat{Y}^{(1)}, \dots, \widehat{Y}^{(m)}$. A fresh candidate set is drawn in every trial because P_N is the marginal law of the returned response, averaging over both candidate generation and the sampler's internal randomness. Reusing a single candidate set would instead estimate the output law conditional on that set. With this, we estimate P_N by its empirical frequencies,

$$\widehat{P}_N(y \mid x) := \frac{1}{m} \sum_{t=1}^{m} \mathbb{1}\{\widehat{Y}^{(t)} = y\}, \quad y \in \mathrm{supp}(\widehat{\mu}),$$

and report the plug-in TV estimate

$$\widehat{D}_{\mathrm{TV}}(P_N, \pi_\beta) := \frac{1}{2} \sum_{y \in \mathrm{supp}(\widehat{\mu})} \left|\widehat{P}_N(y \mid x) - \pi_\beta(y \mid x)\right|.$$

Monte Carlo floor. We highlight that even when a sampler is exact, the plug-in TV estimate above is generally positive because $\widehat{P}_N$ is computed from finitely many Monte Carlo trials. Consequently, once the true sampling error becomes sufficiently small, the observed TV error is limited by the statistical resolution of the estimator.

To quantify this effect without relying on an asymptotic approximation, we compute an exact sampler baseline. For each setting, we draw m independent samples directly from π_β, form their empirical distribution using the same estimator, and compute its TV distance from π_β. This provides the Monte Carlo floor at the same number of trials used for the sampling algorithms. We show this exact sampler floor as a dotted line in the corresponding appendix figures. Once a sampler reaches this level, differences of comparable magnitude cannot be reliably distinguished from Monte Carlo estimation error.

I.3 Further experimental results

This section provides additional empirical results supporting Section 6. We first summarize the theoretical comparisons relevant to the experiments and then examine their empirical behavior across prompts and values of β.

Theoretical comparisons and empirical reference points. Since $(\pi_\beta, \widehat{\mu})$ forms a valid finite target–proposal pair, the theoretical results apply directly to the corresponding population output distributions:

- **UR vs. envelope RS.** (Theorem 1) UR attains the best fixed-threshold RS guarantee in hindsight without requiring threshold information. Envelope RS serves as an empirical proxy for this benchmark, and we therefore expect the two methods to exhibit comparable empirical performance.
- **UR vs. budget-calibrated RS** (Theorem 2): for every prompt, β, sampling budget N, and δ, $D_{\mathrm{TV}}(P_N^U, \pi_\beta) \le D_{\mathrm{TV}}(P_{N,\delta}^R, \pi_\beta)$.
- **UR vs. SIR** (Theorem 3): for every prompt, β, and sampling budget N, $D_{\mathrm{TV}}(P_N^U, \pi_\beta) \le D_{\mathrm{TV}}(P_N^{\mathrm{SIR}}, \pi_\beta)$.

- **Monte Carlo resolution.** Once the true TV error becomes sufficiently small, the estimated curves approach the scale of the exact-sampler Monte Carlo baseline. Differences at this scale cannot be reliably distinguished from Monte Carlo estimation error.

Number of trials. All appendix results in this section use $m = 10^5$ independent runs. For Figure 1, we use $m = 10^6$ to reduce Monte Carlo estimation error.

Confidence intervals for TV error. Since $\widehat{D}_{\mathrm{TV}}$ is computed from the empirical output distribution $\widehat{P}_N$ based on m independent runs, we estimate its Monte Carlo variability using a bootstrap. For each $b = 1, \ldots, B$ with $B = 200$, we draw m responses independently from $\widehat{P}_N$, then form the corresponding empirical distribution $\widehat{P}_N^{\star(b)}$, and compute

$$\widehat{D}_b^{\star} := \frac{1}{2} \sum_{y \in \mathrm{supp}(\widehat{\mu})} \left| \widehat{P}_N^{\star(b)}(y \mid x) - \pi_\beta(y \mid x) \right|.$$

We take $\widehat{\mathrm{se}}$ to be the sample standard deviation of $\{\widehat{D}_b^{\star}\}_{b=1}^{B}$ and report the approximate 95% interval $\widehat{D}_{\mathrm{TV}} \pm z_{0.975}\widehat{\mathrm{se}}$. The interval reflects Monte Carlo variability of the estimator. The exact-sampler baseline separately shows the nonzero plug-in error induced by finite m when the true TV error is zero.

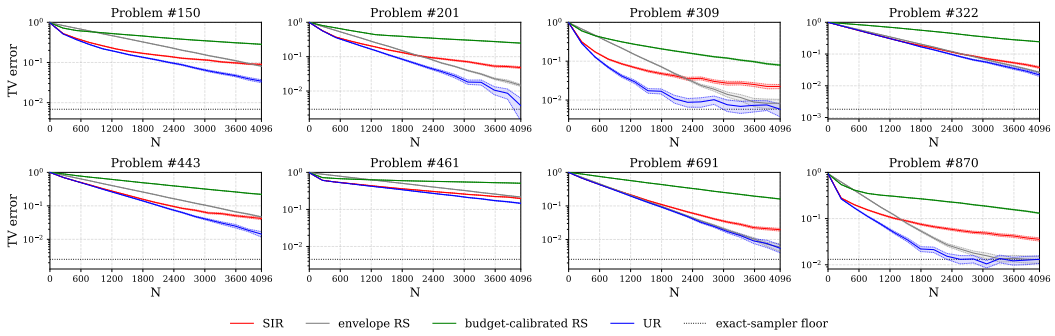
Confidence intervals for accuracy. For accuracy, each of the m independent runs yields a Bernoulli outcome. We report $\widehat{a} \pm t_{m-1,0.975}\widehat{\mathrm{se}}$, and $\widehat{\mathrm{se}} = \frac{s}{\sqrt{m}}$, where $\widehat{a}$ and s denote the sample mean and sample standard deviation of the outcomes, respectively. At $m = 10^5$, $t_{m-1,0.975} \approx 1.96$.

GSM8K: variation across prompts. Figure 2a reports the TV error at $\beta = 0.1$ for eight randomly selected GSM8K problems. The difficulty of approximating the target varies substantially across prompts. For some prompts, the methods reach the Monte Carlo floor within a small sampling budget, leaving little resolution for distinguishing them at larger N. For others, substantial separation remains throughout the tested range. Across the displayed prompts, UR achieves estimated TV error below or comparable to SIR and budget-calibrated RS, up to Monte Carlo resolution, consistent with Theorems 2 and 3. UR remains competitive with envelope RS without requiring threshold information.

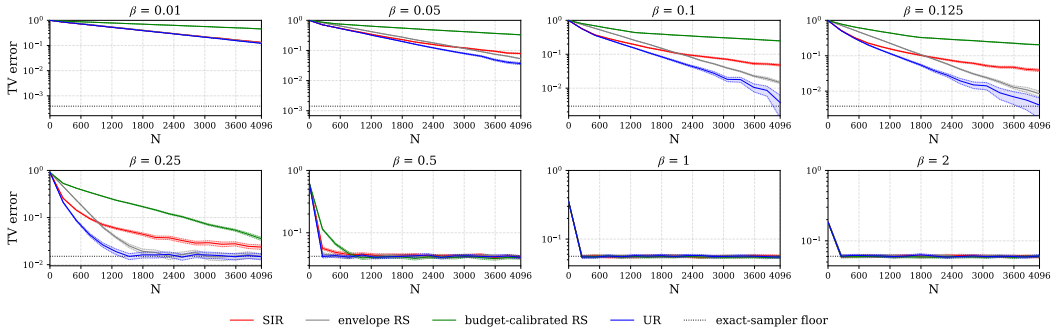
GSM8K: variation across β. Figure 2b fixes Problem #201 and varies $\beta \in \{0.01, 0.05, 0.1, 0.125, 0.25, 0.5, 1, 2\}$. Changing β changes the concentration of the reward-tilted target and therefore the finite-budget difficulty of the sampling problem. At smaller β, the target is more strongly tilted toward high-reward responses, whereas larger β makes π_β closer to the reference proposal $\widehat{\mu}$. Accordingly, the sampling difficulty varies substantially with β. The same qualitative comparison persists across the displayed values: UR remains competitive with other baselines without requiring threshold information.

GSM8K: Ground-truth accuracy. Figure 3 reports ground-truth accuracy across eight prompts at $\beta = 0.1$ and, for Problem #258, across $\beta \in \{0.01, 0.05, 0.1, 0.125, 0.25, 0.5, 1, 2\}$. Distributional closeness to π_β translates directly into closeness in ground-truth accuracy, since correctness corresponds to an event. However, the target accuracy may be either above or below the reference policy accuracy. Accordingly, as a sampler approaches π_β, its accuracy may either increase or decrease toward the corresponding target accuracy. For this reason, we use TV error as the primary metric for comparing the sampling procedures and accuracy as a complementary metric.

MATH500. Figures 4 and 5 report the MATH500 results across eight problems at $\beta = 0.1$ and across values of β for selected problems. Across the displayed settings, the qualitative behavior observed on GSM8K also appears on MATH500, indicating that the empirical comparisons carry over to these more challenging mathematical reasoning problems.

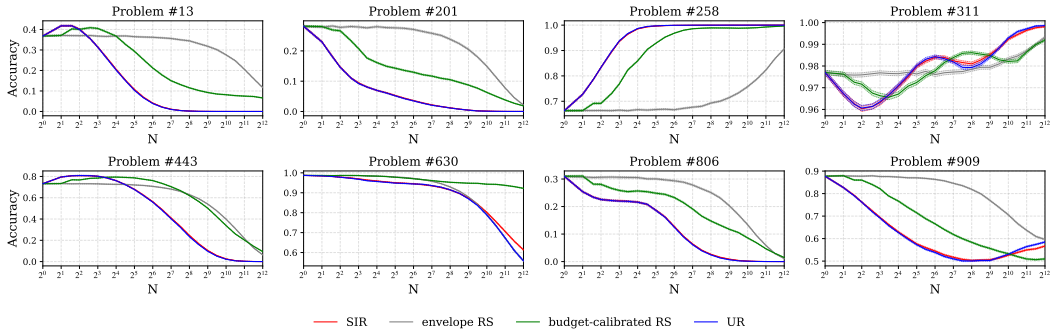


(a) $\beta = 0.1$ for eight randomly selected problems.

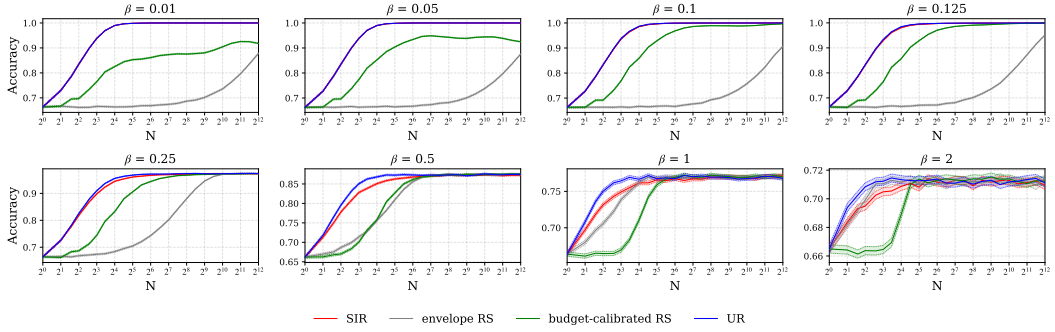


(b) Problem #201 across eight values of β from 0.01 to 2.

Figure 2: TV error $\widehat{D}_{\mathrm{TV}}(P_N, \pi_\beta)$ w.r.t sampling budget N on GSM8K. The dotted line indicates the exact-sampler Monte Carlo floor.

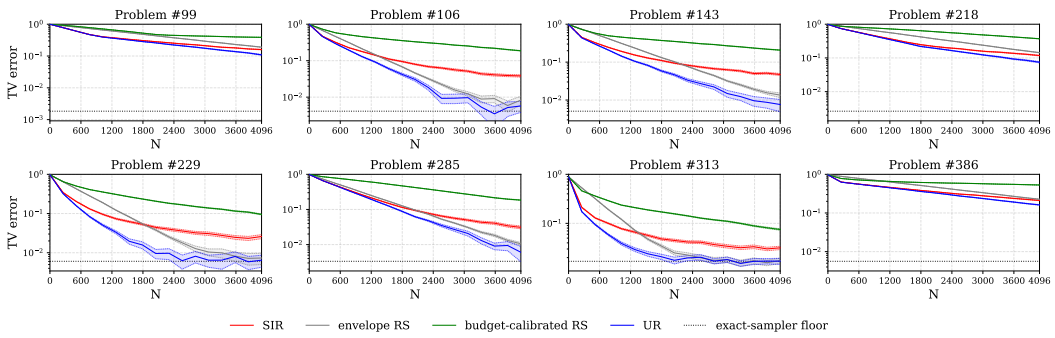


(a) $\beta = 0.1$ for eight randomly selected problems.

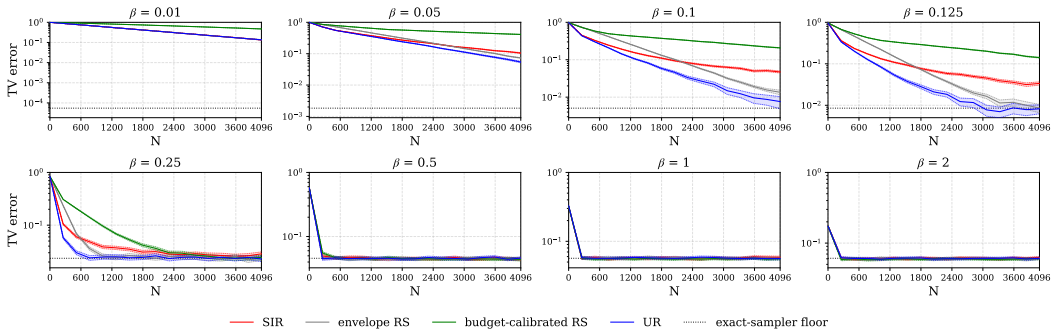


(b) Problem #258 for β from 0.01 to 2.

Figure 3: Ground-truth accuracy w.r.t. sampling budget N on GSM8K.

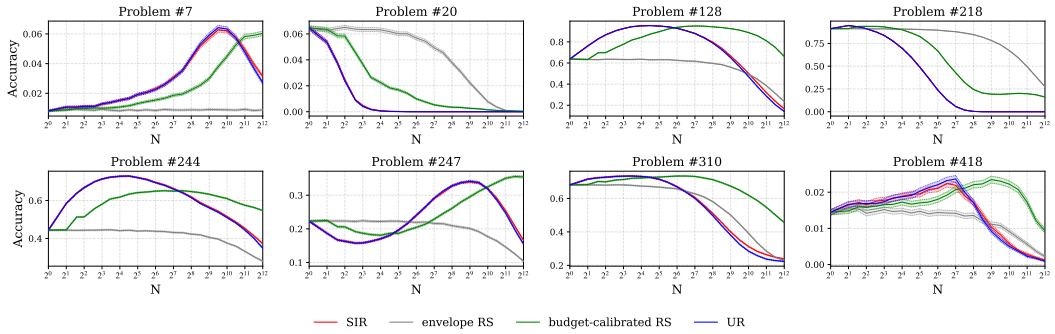


(a) $\beta = 0.1$ for eight randomly selected problems.

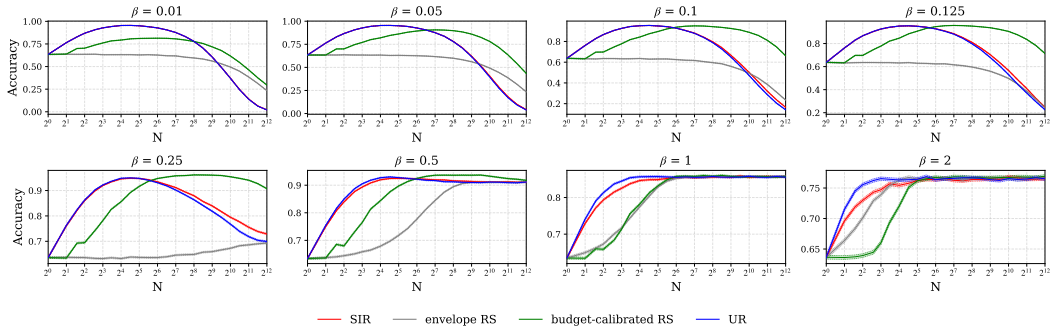


(b) Problem #143 for β from 0.01 to 2.

Figure 4: TV error $\widehat{D}_{\mathrm{TV}}(P_N, \pi_\beta)$ w.r.t sampling budget N on MATH500. The dotted line indicates the exact-sampler Monte Carlo floor.



(a) $\beta = 0.1$ for eight randomly selected problems.



(b) Problem #128 for β from 0.01 to 2.

Figure 5: Accuracy w.r.t sampling budget N on MATH500.