Counterexamples to a best-of-n win-rate bound at every budget

Dev D. Goyal

26 October 2026; revised 1 November 2026

Abstract

Beirami et al. conjectured an upper bound on the win rate of best-of-n sampling in terms of its Kullback–Leibler divergence from the reference distribution [2, Conjecture 7.2]. This note shows that the bound fails at every sampling budget. With two independent draws, the reference distribution $(2/5, 1/5, 1/5, 1/5)$ on four responses yields a selected distribution with win rate $163/250$, which lies strictly above the conjectured curve. An exact rational certificate proves the positive gap between the two expressions, which involve real logarithms. For every integer budget $n \geq 2$ there are also finite counterexamples whose largest reference probability is arbitrarily small. A Lean formalization proves the finite counterexample, its inverse-curve consequence and the theorem for every budget.

1 The proposed bound

Fix a prompt and responses $1, \dots, k$ with strictly increasing rewards. Let $p_i > 0$ be the reference probabilities, with $\sum_i p_i = 1$. Put

$$F_0 = 0, \qquad F_i = \sum_{j=1}^{i} p_j.$$

Best-of-n sampling returns the response with the highest reward among n independent draws. The number of draws is the sampling budget. The selected probabilities are

$$q_i = F_i^n - F_{i-1}^n. \tag{1}$$

Indeed, the probability that all draws have index at most i is F_i^n. A comparison with an independent reference draw gives one point for a strict win and one half for equality. Define

$$W = \sum_i q_i \left(F_{i-1} + \frac{p_i}{2} \right), \qquad D = \sum_i q_i \log \frac{q_i}{p_i}. \tag{2}$$

All logarithms are natural. Thus D is the Kullback–Leibler divergence of the selected distribution from the reference distribution.

Beirami et al. [2], Conjecture 7.2, propose

$$W \leq \ell^{-1}(D), \qquad \ell(w) = \log \frac{w}{1-w} + \frac{1}{w} - 2, \quad \frac{1}{2} \leq w < 1. \tag{3}$$

Their equation (28) uses the same half-credit convention. Their Lemma 5.2 gives (2). Differentiation gives

$$\ell'(w) = \frac{2w - 1}{w^2(1-w)}.$$

Hence ℓ is strictly increasing on $[1/2, 1)$. Also $\ell(1/2) = 0$ and $\ell(w) \to \infty$ as $w \uparrow 1$. Its inverse is therefore defined on $[0, \infty)$.

Theorem 1 below gives an exact counterexample to (3) with $n = 2$. Theorem 2 gives counterexamples at every budget $n \geq 2$, with every reference probability below any prescribed positive cap. The proved bounds of [2] are unaffected; Section 5 discusses prior work. The Lean formalization described in Section 4 proves both theorems.

2 An exact counterexample

Theorem 1. *Take $n = 2$ and $p = (2/5, 1/5, 1/5, 1/5)$. Then*

$$q = (4/25, 5/25, 7/25, 9/25), \qquad W = \frac{163}{250},$$

and

$$\ell(W) - D > \frac{1}{2445} > 0. \tag{4}$$

In particular, (3) is false.

Proof. Equation (1) gives the table

i	1	2	3	4
p_i	2/5	1/5	1/5	1/5
F_i	2/5	3/5	4/5	1
q_i	4/25	5/25	7/25	9/25
$F_{i-1} + p_i/2$	1/5	1/2	7/10	9/10.

The two probability vectors have positive entries and sum to one. Substitution gives

$$D = \frac{4}{25}\log\frac{2}{5} + \frac{7}{25}\log\frac{7}{5} + \frac{9}{25}\log\frac{9}{5}, \qquad \ell(W) = \log\frac{163}{87} - \frac{76}{163}.$$

Define the positive rational number

$$R = \frac{(163/87)^{25}(5/2)^4}{(7/5)^7(9/5)^9} = \frac{163^{25}5^{20}}{87^{25}2^4 7^7 9^9}.$$

Then

$$\ell(W) - D = \frac{\log R}{25} - \frac{76}{163}. \tag{5}$$

The exact integer comparison

$$120^{1400}87^{25}2^4 7^7 9^9 < 119^{1400}163^{25}5^{20} \tag{6}$$

gives $(120/119)^{1400} < R$. Repeated integer multiplication verifies this arithmetic certificate. The check does not require numerical evaluation of a logarithm. For $x > 0$, the elementary inequality $\log x \geq 1 - x^{-1}$ gives

$$\log R > 1400\log(120/119) \geq \frac{1400}{120} = \frac{35}{3}.$$

Combining this with (5) yields

$$\ell(W) - D > \frac{7}{15} - \frac{76}{163} = \frac{1}{2445}.$$

To show that the inverse is defined at D, first prove $D \geq 0$. Apply $\log x \geq 1 - x^{-1}$ to each q_i/p_i. This gives $q_i \log(q_i/p_i) \geq q_i - p_i$. Summation proves the claim. Thus $u = \ell^{-1}(D)$ exists. Since $D < \ell(W)$, strict monotonicity gives $u < W$, as required. □

The certificate (6) is checked by the following integer expression in Python:

```
assert 120**1400 * 87**25 * 2**4 * 7**7 * 9**9 < \
       119**1400 * 163**25 * 5**20
```

The gap is approximately 0.002375504922. This approximation is not used in the proof.

3 Every budget and arbitrarily small reference probabilities

Theorem 2. *For every integer $n \geq 2$ and every $\delta > 0$, there is a finite positive reference distribution with $\max_i p_i \leq \delta$ such that best-of-n has*

$$\frac{1}{2} \leq W < 1, \qquad D < \ell(W).$$

Proof. Set

$$a = \min\left\{\delta, \frac{1}{8(n+1)}\right\}, \qquad T = \frac{a^n}{48(n+1)}.$$

Choose an integer $m > (1-a)/T$, and put $h = (1-a)/m$. Use the ordered probabilities

$$p_0 = a, \qquad p_j = h \quad (1 \leq j \leq m).$$

All probabilities are positive and sum to one. Also $h < T \leq a \leq \delta$, since $0 < a < 1$ and $n \geq 2$. Thus each reference probability satisfies the prescribed cap.

The first step bounds the sums that define the divergence and win rate. For $x > 0$, define

$$f(x) = nx^{n-1}, \qquad K(x) = x^n\left(\log n + (n-1)\log x - \frac{n-1}{n}\right), \qquad J(x) = \frac{n}{n+1}x^{n+1}.$$

Their derivatives are $K' = f \log f$ and $J' = xf$. For $0 < u < v$, put $q = v^n - u^n$ and $A = q/(v-u) > 0$. The inequality $\log z \geq 1 - z^{-1}$ implies

$$f(x)\log\frac{f(x)}{A} - f(x) + A \geq 0.$$

Therefore $K(x) - x^n \log A - x^n + Ax$ is increasing. Evaluation at u, v gives

$$q\log\frac{q}{v-u} \leq K(v) - K(u). \tag{7}$$

Also $vx^n - J(x)$ has derivative $f(x)(v-x) \geq 0$ on $[u, v]$. Hence

$$J(v) - J(u) \leq vq = q\frac{u+v}{2} + \frac{v-u}{2}q. \tag{8}$$

Both bounds follow from calculus on a finite interval.

Apply (7) and (8) to the m intervals with endpoints $a + jh$. The differences of the primitives telescope. The first response has selected probability a^n and comparison score $a/2$. Put

$$D_0 = \log n - \frac{n-1}{n}, \qquad w_0 = \frac{n}{n+1}, \qquad c = \frac{n-1}{2(n+1)}.$$

The resulting bounds are

$$D \leq D_0(1-a^n), \qquad W \geq w_0 - ca^{n+1} - \frac{h}{2}. \tag{9}$$

For the second bound, the sum of the selected probabilities after the first response is $1 - a^n \leq 1$. Every comparison score is less than one. The selected probabilities are positive and sum to one. Therefore $W < 1$.

It remains to check the win-rate domain and prove a strict gap. For $n \geq 2$,

$$w_0 \geq \frac{2}{3}, \quad 0 < c < \frac{1}{2}, \quad a \leq \frac{1}{24}, \quad a^{n+1} \leq a, \quad h \leq \frac{1}{144}.$$

Equation (9) therefore gives

$$W \geq \frac{2}{3} - \frac{1}{48} - \frac{1}{288} > \frac{1}{2}.$$

The inequality $\log(1+x) > 2x/(x+2)$ for $x > 0$ gives

$$D_0 > \frac{(n-1)^2}{n(n+1)} \geq \frac{1}{6}.$$

The last comparison is equivalent to $(n-2)(5n-3) \geq 0$. Also $\ell(w_0) = D_0$ and

$$L := \ell'(w_0) = \frac{(n-1)(n+1)^2}{n^2}, \qquad 0 < L < n+1.$$

On $(0,1)$,

$$\ell''(w) = \frac{4w^2 - 5w + 2}{w^3(1-w)^2} > 0, \qquad 4w^2 - 5w + 2 = 4(w - 5/8)^2 + 7/16.$$

Thus the convex tangent inequality and (9) imply

$$\ell(W) - D \geq a^n D_0 - Lca^{n+1} - \frac{Lh}{2}.$$

The parameter choices give $Lca \leq 1/16$ and $Lh/2 \leq a^n/96$. Consequently

$$\ell(W) - D > a^n \left(\frac{1}{6} - \frac{1}{16} - \frac{1}{96} \right) = \frac{3a^n}{32} > 0.$$

This proves the theorem. □

4 Scope and formal verification

The result refutes the universal conjecture but does not determine the sharp replacement curve. The proved bounds $D \leq \log n - (n-1)/n$ and $W \leq n/(n+1)$ in [2] remain valid. The example uses distinct rewards. Equality in a comparison occurs when the selected response and the independent reference draw are the same response.

The Lean namespace `MathAndTheory.BestOfN` defines the cumulative probabilities and both quantities in (2). The theorem `selected_equals_independent_sampling` proves that (1) agrees with direct enumeration of all 16 independent sample pairs. Theorems `gap_positive` and `explicit_counterexample` prove (4). The theorem `inverse_curve_exists` proves the inverse-point conclusion without an existence assumption. The theorem `universal_curve_false` states the resulting disproof.

The namespace `MathAndTheory.BestOfNExtension` proves Theorem 2. The theorem `every_budget_small_atoms` includes every quantifier and conclusion in the paper's statement. It proves positivity, normalization, the probability cap, the win-rate domain, and the strict gap. The proof establishes the cumulative probabilities, the two interval bounds, the telescoping estimates, and each parameter bound. It uses the same distribution and integer condition as the proof above. The constants are not claimed to be optimal.

5 Prior work

The cited source states the proposed curve as a conjecture. Related work on soft best-of-n sampling [1] studies different divergence and regret bounds. No claim of first priority is made.

6 Research process and AI disclosure

This work is fully autonomous agentic research. Autonomous AI research agents produced the research, proofs, formalization and drafts. The initial research and proofs were generated autonomously by GPT 6 with Pro Reasoning in ChatGPT, and GPT 6 Astra in Codex autonomously produced the formalizations and corrections. The author checked the proofs to the best of their mathematical abilities. This check is not expert validation or peer review.

Opus 5.5 by Anthropic used with Claude Code revised this manuscript editorially on 1 November 2026: it rewrote the abstract and this disclosure, added a summary of the results to Section 1 and completed the references. It did not change the mathematics.

Dev D. Goyal is responsible for publication. The work has not been externally peer reviewed. Section 4 states the formal scope and the limits of the result, and Section 5 states its relation to prior work.

References

[1] Gholamali Aminian, Idan Shenfeld, Amir R. Asadi, Ahmad Beirami, and Youssef Mroueh. Best-of-N through the smoothing lens: KL divergence and regret analysis, 2025. URL https://arxiv.org/abs/2507.05913v1.

[2] Ahmad Beirami, Alekh Agarwal, Jonathan Berant, Alexander D'Amour, Jacob Eisenstein, Chirag Nagpal, and Ananda Theertha Suresh. Theoretical guarantees on the best-of-n alignment policy. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 3580–3602, 2025. URL https://proceedings.mlr.press/v267/beirami25a.html. Conjecture 7.2; arXiv:2401.01879v3, dated 29 May 2025.