DELETION OPTIMALITY CHARACTERIZES SHARED GUMBEL SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Shared randomness coordinates categorical samples across different distributions, supporting correlated sampling and counterfactual generation. What makes shared Gumbel sampling a distinguished choice among exact sampling rules? We show that a simple deletion principle determines its entire joint behavior. On a finite alphabet with at least three labels, any exact shared-randomness sampler that maximizes agreement between every distribution and its one-label deletions has the same finite-dimensional output laws as shared Gumbel sampling. Thus, optimizing agreement on these local pairs uniquely determines the coupling across all distributions. As a consequence, no exact shared-randomness sampler can match or improve Gumbel agreement on every distribution pair and strictly improve it on any pair. The characterization covers arbitrary measurable samplers. Its proof derives a common race representation from deletion optimality and yields quantitative certificates relating comparison reversals to deletion deficits. On countably infinite alphabets, we construct distinct deletion-optimal couplings within a common tail class and show that continuity under prefix truncation restores uniqueness. These results establish a local characterization of shared Gumbel sampling and identify the boundary between finite rigidity and countable freedom.

1 INTRODUCTION

Shared randomness lets separate sampling decisions move together. In distributed correlated sampling, each participant sees its own distribution and uses a common random seed to coordinate its output with the others. In counterfactual generation, reusing the noise links an observed output to its counterpart after an intervention. Both tasks require a rule that samples each distribution correctly and specifies how outputs across distributions are related. Exact marginal probabilities leave this joint behavior undetermined, making the choice of coupling a substantive part of the task.

Shared Gumbel sampling provides a simple coupling, used in counterfactual language generation (Ravfogel et al., 2025) and communication-free sampling (Daliri et al., 2025). Its equivalent exponential race assigns one independent clock to each label and rescales that clock by the label's probability. Recent work shows that Gumbel sampling improves pairwise agreement over weighted MinHash and asks whether a further Pareto improvement over Gumbel is possible (Daliri et al., 2025). This question concerns a single rule serving all distributions: can another exact sampler agree at least as often on every pair and more often on some pair?

We answer this question through a local characterization. Delete one label from a distribution and renormalize the remaining probabilities. Shared Gumbel sampling preserves the winner whenever the deleted label did not win, attaining the largest possible agreement on this pair. We prove that attaining this local optimum for every deletion already determines the entire joint sampling law on any finite alphabet with at least three labels. Every such sampler has the same finite-dimensional output laws as shared Gumbel sampling. Any different joint law therefore loses agreement on a deletion pair, establishing Gumbel's Pareto optimality.

The proof derives a common race representation from agreement constraints on an arbitrary measurable sampler. Deletion optimality first forces the winner to survive every nonwinner deletion. Three-label distributions then connect binary comparisons. If a label wins at lower odds but loses at higher odds, these connections propagate the reversal along a geometric chain. At its endpoints, exact marginal probabilities make that reversal arbitrarily unlikely. Nested binary comparisons yield

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.

random thresholds; consistency across triples makes these thresholds ratios of common scores. Finally, exact categorical marginals identify the exponential-ratio law.

The same proof also supplies finite certificates. Two endpoint ranking events cover every reverse comparison along a chosen chain. Their probabilities give a sharper bound than the elementary endpoint union, and a pointwise inequality adds the observed deletion deficits without assuming that defective rankings retain their ideal law. Finite linear programs illustrate the distinction between a valid certificate and the residual freedom left by a bounded collection of tests.

The finite-alphabet assumption is essential. On a countable alphabet, the tail fixes an absolute scale that cannot be removed by finitely many deletions. We exploit this scale to build exact, deletion-optimal samplers whose clocks refresh as a label's mass changes. The construction works within a single tail class, so merely tying together different classes does not restore uniqueness. Continuity under prefix truncation provides the missing connection and recovers the Gumbel law. Figure 1 contrasts the finite and countable mechanisms.

2 Related work

Correlated sampling. Shared randomness coordinates samples without revealing each participant's distribution. Bavarian et al. establish optimal worst-case guarantees (Bavarian et al., 2020), and Angel and Spinka study simultaneous pairwise agreement (Angel & Spinka, 2019). Multimaximal couplings also organize contextuality-by-default (Dzhafarov & Kujala, 2016). Daliri et al. compare Gumbel sampling with weighted MinHash (Broder, 1997; Manasse et al., 2010; Ioffe, 2010) and ask whether Gumbel admits a Pareto improvement (Daliri et al., 2025). List-level coupling extends matching to multiple proposals (Rowan et al., 2026). Our deletion-optimality condition identifies the entire finite-dimensional law of one sampler serving every input.

Choice representations. Yellott characterizes double-exponential noise within an independent Thurstone model (Yellott Jr, 1977); Bundesen characterizes independent races through proportional hazards (Bundesen, 1993). Gumbel perturbations support discrete sampling (Maddison et al., 2014; Huijben et al., 2022) and continuous relaxations (Maddison et al., 2016; Jang et al., 2016). We derive the race representation from deletion agreement. The sampler's implementation is unrestricted at the outset; independent exponential ratios emerge only at the final step, as the law identified by exact marginals.

Counterfactual coupling. Oberst and Sontag establish counterfactual stability for Gumbel-max structural causal models (Oberst & Sontag, 2019). Ravfogel et al. construct shared-Gumbel language-model counterfactuals (Ravfogel et al., 2025), while Lorberbom et al. learn generalized mechanisms for alternative coupling objectives (Lorberbom et al., 2021). These works fix the marginals and choose the dependence by a mechanism or objective; we identify the dependence by maximal deletion agreement across the simplex.

Speculative generation. Draft verification accelerates language-model sampling (Leviathan et al., 2023; Chen et al., 2023; Xia et al., 2024). Tree proposals, multiple heads, parallel lookahead and staged drafts broaden this execution pattern (Miao et al., 2024; Cai et al., 2024; Fu et al., 2024; Spector & Re, 2023). Daliri et al. use communication-free coupling to make speculative decoding drafter-invariant (Daliri et al., 2025).

3 Problem setup

Let $A = \{1, \dots, n\}$ and let $\Delta(A)$ denote its probability simplex. An exact shared-randomness sampler is a measurable map

$$F : \Delta(A) \times \Omega \longrightarrow A, \qquad \mathbb{P}(F(p) = i) = p_i \quad (p \in \Delta(A)). \tag{1}$$

All inputs use the same seed in Ω. Write $a_F(p, q) = \mathbb{P}(F(p) = F(q))$. The shared exponential race is

$$G(p) = \operatorname*{argmin}_{i : p_i > 0} \frac{E_i}{p_i}, \qquad E_1, \dots, E_n \overset{\text{iid}}{\sim} \operatorname{Exp}(1). \tag{2}$$

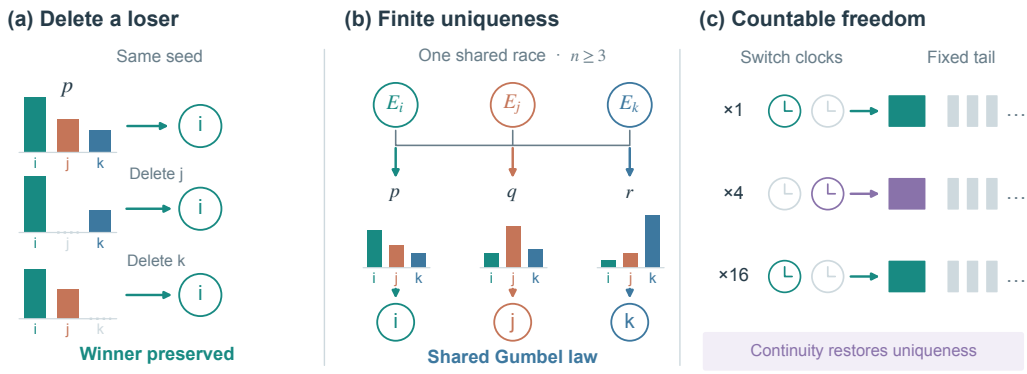


Figure 1: One deletion rule, two regimes (schematic). Finite alphabets with $n \geq 3$ share one exponential field. Countable alphabets admit alternating head clocks under rescaling, while the positive tail stays fixed.

Equivalently, maximize $\log p_i + Z_i$ with shared Gumbels $Z_i = -\log E_i$. Ties occur with probability zero for each fixed input. We call the collection of outputs over all inputs a *grand coupling*; equality of grand-coupling laws means equality on every finite list of inputs.

For p with support K, $|K| \geq 2$, and $i \in K$, delete label i and renormalize:

$$p_i^{-i} = 0, \qquad p_j^{-i} = \frac{p_j}{1 - p_i} \ (j \neq i), \qquad a_F(p, p^{-i}) \leq 1 - p_i. \tag{3}$$

The bound follows from $d_{\mathrm{TV}}(p, p^{-i}) = p_i$. The race attains it: survivor clock times are multiplied by a common positive factor, so the output changes exactly when i was the winner. A sampler is *deletion-maximal* (deletion-optimal) when every inequality in (3) is an equality. Define the nonnegative shortfalls, computed from outputs of the one sampler F, whose shared randomness couples every deletion edge:

$$d_i(p) = (1 - p_i) - a_F(p, p^{-i}), \qquad \delta(p) = \sum_{i \in K} d_i(p). \tag{4}$$

4 Deletion Rigidity on Finite Alphabets

Theorem 1 (Deletion rigidity). *Let $n \geq 3$ and let F satisfy* (1)*. The following are equivalent:*

1. *F is deletion-maximal.*
2. *For every p with support K of size at least two, $\sum_{i \in K} a_F(p, p^{-i}) = |K| - 1$.*
3. *For every finite list $p^1, \ldots, p^M \in \Delta(A)$, $(F(p^1), \ldots, F(p^M)) \overset{d}{=} (G(p^1), \ldots, G(p^M))$.*

The second condition packages the deletion tests at one menu into a scalar. More generally, any strictly positive deletion weights $\mu_i(p)$ give an equivalent equality

$$\sum_{i \in K} \mu_i(p) a_F(p, p^{-i}) = \sum_{i \in K} \mu_i(p)(1 - p_i), \tag{5}$$

because every summand has a nonnegative shortfall. Uniform deletion has ceiling $1 - 1/|K|$; deleting a label sampled from p has ceiling $1 - \sum_i p_i^2$. The latter expresses the same condition through Gini impurity.

Corollary 2 (A deletion witness and Pareto optimality). *If $n \geq 3$ and a finite-dimensional law of F differs from that of G, some supported deletion satisfies $a_F(p, p^{-i}) < a_G(p, p^{-i})$. On every finite alphabet, an exact sampler satisfying $a_F(p, q) \geq a_G(p, q)$ for all p, q has the Gumbel grand-coupling law.*

For $n \geq 3$, pairwise domination forces deletion maximality, so the theorem applies. For $n = 2$, Gumbel is the nested maximal Bernoulli coupling; domination forces its nested thresholds directly (Appendix A.4). Deletion tests alone carry no information in that case, since every deletion leaves a singleton. In both cases the conclusion is equality of grand-coupling laws.

4.1 From deletion agreement to binary nesting

Fix p with support K, $|K| \geq 2$, and put $J_i = \mathbf{1}\{F(p) = F(p^{-i})\}$. If $F(p) = k$, the deletion of k cannot agree, and at most $|K| - 1$ other deletions can agree. Thus

$$\delta(p) = \mathbb{E}\left[(|K| - 1) - \sum_{i \in K} J_i\right]. \tag{6}$$

The integrand counts failed nonwinner deletions. Under deletion maximality it vanishes almost surely: deleting any nonwinner preserves the winner. We call this *heredity*. At a ternary input, the binary faces consequently form a transitive tournament, with the full output as its top-ranked label.

For distinct labels i, j and odds $x > 0$, define

$$C_{ij}(x) = \mathbf{1}\left\{F\left(\frac{xe_i + e_j}{1 + x}\right) = i\right\}, \qquad \mathbb{P}(C_{ij}(x) = 1) = \frac{x}{1 + x}. \tag{7}$$

Exact marginals specify these probabilities but do not yet make the events nested. The essential step is to derive nesting using a third label.

Lemma 3 (Extreme-label amplification)**.** *Fix $0 < r < s$ and $\lambda = s/r$. For any $m \geq 1$ and $y > 0$, consider the normalized ternary weights*

$$q_t \propto (\lambda^t y, \lambda^t y/r, 1), \qquad q'_t \propto (\lambda^{t+1} y, \lambda^t y/r, 1), \quad 0 \leq t < m. \tag{8}$$

If F is exact and deletion-maximal on these ternaries, the reverse event $E = \{C_{ij}(r) = 1, C_{ij}(s) = 0\}$ satisfies

$$\mathbb{P}(E) \leq \frac{y}{1 + y} + \frac{1}{1 + \lambda^m y}. \tag{9}$$

Proof. Every low menu q_t has the same $i : j$ face with odds r; every high menu q'_t has the same face with odds s. The pair q_t, q'_t shares its $j : k$ face, while q'_t, q_{t+1} shares its $i : k$ face. On E, transitivity yields

$$C_{ik}(\lambda^{t+1} y) \leq C_{jk}(\lambda^t y/r) \leq C_{ik}(\lambda^t y). \tag{10}$$

Figure 2 displays these shared faces. Iterating shows that E is contained in $\{C_{ik}(y) = 1\} \cup \{C_{ik}(\lambda^m y) = 0\}$. Their exact marginal probabilities give (9). □

Choosing $y = \lambda^{-m/2}$ gives $\mathbb{P}(E) \leq 2/(1+\lambda^{m/2})$. Under the theorem's premise these finite bounds hold for every depth, so $\mathbb{P}(E) = 0$. This yields nesting for an arbitrary measurable sampler from finite comparisons alone.

4.2 From nested comparisons to the full output law

On a common probability-one event for rational odds, define

$$R_{ij} = \inf\{x \in \mathbb{Q}_{>0} : C_{ij}(x) = 1\}. \tag{11}$$

Nesting and exactness give $\mathbb{P}(R_{ij} \leq x) = x/(1 + x)$. Rational squeezing then identifies $C_{ij}(x) = \mathbf{1}\{R_{ij} \leq x\}$ almost surely for each fixed real $x > 0$. The thresholds are positive and finite, and reversing labels gives $R_{ji} = R_{ij}^{-1}$.

Ternary transitivity makes the thresholds multiplicatively consistent:

$$R_{ik} = R_{ij} R_{jk} \quad \text{almost surely.} \tag{12}$$

Indeed, if the left side were larger, positive rational odds just above R_{ij} and R_{jk} would make i beat j, j beat k, and k beat i. A strict inequality in the other direction gives the reverse cycle. Both contradict transitivity on a rational three-label menu.

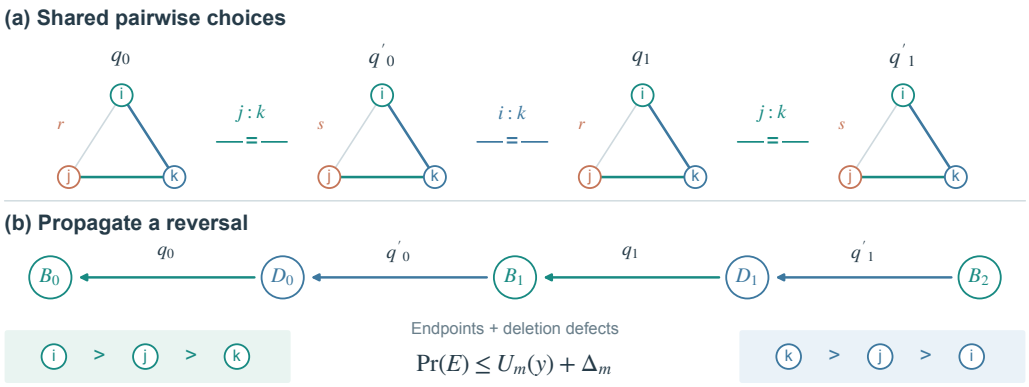


Figure 2: Shared binary faces connect four ternary menus. Arrows show implications on the fixed reverse event, with $B_t = C_{ik}(\lambda^t y)$ and $D_t = C_{jk}(\lambda^t y/r)$. Endpoint rankings cover reversal; $\Delta_m = \sum_{t<m}(\delta(q_t) + \delta(q'_t))$ sums the deletion deficits.

Set $S_1 = 1$ and $S_i = R_{i1}$ for $i \neq 1$. Equation (12) gives $R_{ij} = S_i/S_j$. Heredity now extends binary choices to every support by induction. If $F(p) = k$, delete a label other than k and a specified competitor j. The smaller menu still chooses k, so $S_k/p_k \le S_j/p_j$. Repeating for every competitor proves

$$F(p) = \operatorname*{argmin}_{i:p_i>0} \frac{S_i}{p_i} \quad \text{almost surely for each fixed } p. \tag{13}$$

The continuous pair-ratio laws eliminate ties. This constructs a shared score race without assuming independent scores.

It remains to identify its dependence. For positive $r_1, \dots, r_{n-1}$, choose $p_j = r_j p_n$ and $p_n = (1 + \sum_{j<n} r_j)^{-1}$. Exactness and (13) imply

$$\mathbb{P}\left(\frac{S_j}{S_n} \ge r_j \text{ for all } j < n\right) = \frac{1}{1 + \sum_{j<n} r_j} = \int_0^\infty e^{-t} \prod_{j<n} e^{-r_j t}\, dt. \tag{14}$$

The final integral is the survival function of $(E_j/E_n)_{j<n}$ for independent unit exponentials. Upper orthants determine the ratio-vector law, hence every finite collection of outputs in (13). This proves Theorem 1. The construction uses simultaneous probability-one events for countably many rational comparisons and each finite target family. Details are in Appendix A.

5 Finite certificates and deletion defects

The amplification proof identifies the process using arbitrarily long finite chains. A sharper finite certificate shows how much information each chain contains and how deletion errors accumulate. For a hereditary ternary menu, exact winner and binary-face marginals determine the six ranking probabilities:

$$\mathbb{P}(\Pi_p = i \succ j \succ k) = \frac{p_i p_j}{1 - p_i}. \tag{15}$$

One way to see this is to subtract the event that j wins the full menu from the event that j wins its binary contest with k. The remaining ranking is exactly $i \succ j \succ k$. Equation (15) is the three-label Plackett–Luce law.

Theorem 4 (Two-atom certificate with defects)**.** *For the chain* (8)*, every exact sampler satisfies*

$$\mathbb{P}(E) \le U_m(y) + \sum_{t=0}^{m-1} \big(\delta(q_t) + \delta(q'_t)\big), \tag{16}$$

$$U_m(y) = \frac{y^2/r}{(1 + y + y/r)(1 + y/r)} + \frac{1}{(s+1)\big(1 + \lambda^m y(1 + 1/s)\big)}. \tag{17}$$

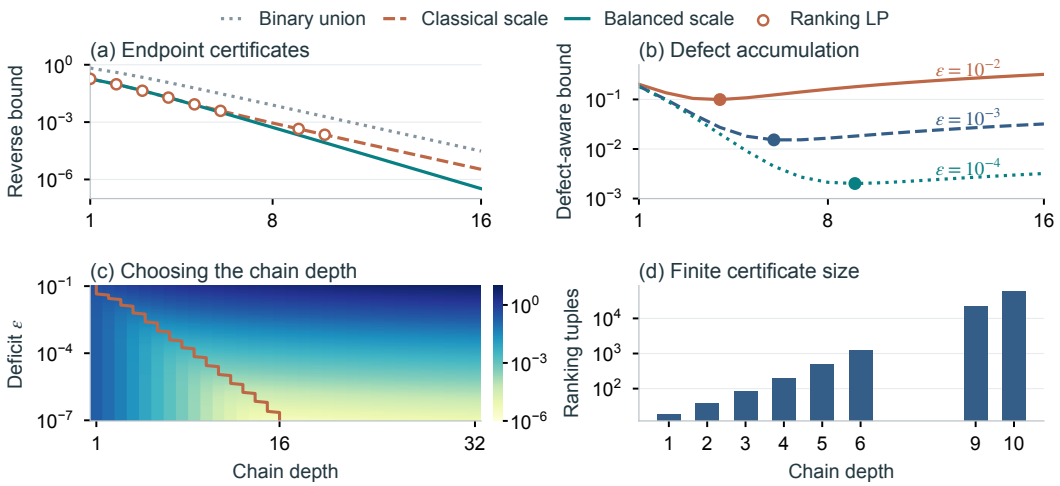


Figure 3: Finite certificates at $r = 1/2$, $s = 2$: endpoint covers and ranking-LP optima, defect-aware bounds, optimal-depth landscape and compatible ranking counts. Dots and the orange path identify minimizing displayed depths.

If each displayed ternary has $\delta \le \varepsilon$, put $c_{r,s} = (rs/[2(s+1)^2])^{1/3}$ and $y = c_{r,s}\lambda^{-m/3}$. Then

$$\mathbb{P}(E) \le C_{r,s}\lambda^{-2m/3} + 2m\varepsilon, \qquad C_{r,s} = 3c_{r,s}^2/r. \tag{18}$$

To understand the two atoms, first take zero deficits. On E, every low ranking places i above j, and every high ranking places j above i. If q_0 does not rank $i \succ j \succ k$, its shared $j : k$ face makes q_0' rank $k \succ j \succ i$. The shared $i : k$ face then forces q_1 to rank $k \succ i \succ j$, and the next shared $j : k$ face forces q_1' to rank $k \succ j \succ i$. The same propagation reaches the last high menu. Thus

$$E \subseteq \{\Pi_{q_0} = i \succ j \succ k\} \cup \{\Pi_{q'_{m-1}} = k \succ j \succ i\}. \tag{19}$$

The two atom probabilities, evaluated by (15), are exactly the two terms of $U_m(y)$. They retain the runner-up information discarded by the binary endpoint union.

For nonzero deficits, we account for each failed nonwinner deletion in a seedwise inequality. Write $B_t = C_{ik}(\lambda^t y)$, $D_t = C_{jk}(\lambda^t y/r)$, $H = 1 - C_{ij}(s)$, $X_t = F(q_t)$, $X_t' = F(q_t')$, and let $h(q)$ be the integrand of (6). Then

$$\mathbf{1}_E \le D_0 - \mathbf{1}\{X_0 = j\} + H - \mathbf{1}\{X'_{m-1} = j\} + \sum_{t=0}^{m-1} \big(h(q_t) + h(q_t')\big). \tag{20}$$

The proof is a telescoping chain of local inequalities, given in Appendix B. Exact one-dimensional marginals make the expected endpoint differences equal to the two terms in (17), even when heredity fails. Taking expectations proves (16). Finally, $U_m(y) \le y^2/r + s/((s+1)^2\lambda^m y)$; balancing these terms proves (18).

For the Bernoulli outputs at odds $r < s$, exact marginals give a useful interpretation:

$$a_G(P_{ij}(r), P_{ij}(s)) - a_F(P_{ij}(r), P_{ij}(s)) = 2\mathbb{P}(E), \tag{21}$$

where $P_{ij}(x) = (xe_i + e_j)/(1+x)$. Choosing $m = \lceil 3\log(1/\varepsilon)/(2\log\lambda)\rceil$ for $0 < \varepsilon < 1$ yields an $O(\varepsilon\log(1/\varepsilon))$ agreement-loss bound at fixed r, s. A randomized version chooses one of the $2m$ ternaries uniformly and then one of its three deletions uniformly, independently of the sampler's seed. If its mean agreement is $\bar{a}$, then

$$\mathbb{P}(E) \le U_m(y) + 6m\left(\frac{2}{3} - \bar{a}\right). \tag{22}$$

This converts mean test agreement into a certificate for a specified binary pair. At zero deficit, tolerance η requires $O(\log(1/\eta))$ chain depth at fixed r, s, with a linear number of certificate terms. Positive deficits introduce an optimal-depth trade-off (Figure 3b,c). The analytic test needs no enumeration; finite programs instead ask whether the chosen constraints admit a matching primal witness.

Table 1: Finite witnesses at $r = 1/2$, $s = 2$, $y = 2^{-m}$. The ranking-LP optima and exact two-atom covers differ by less than 10^{-14} at every listed depth. Tuples count compatible rankings.

Depth	Tuples	Exact cover	LP reverse probability
1	18	$11/60$	1.833×10^{-1}
2	38	$2/21$	9.524×10^{-2}
3	84	$94/2145$	4.382×10^{-2}
6	1248	$278/71489$	3.889×10^{-3}
9	21932	$134662/305342985$	4.410×10^{-4}
10	57358	$177154/809769987$	2.188×10^{-4}

5.1 Numerical certificates

Finite-chain linear programs put probability mass on compatible deterministic assignments, impose exact marginals and selected deletion constraints, and maximize reverse mass. At zero deficit, assignments can be compressed to face-compatible tuples of ternary rankings with the marginals (15). A feasible primal describes the freedom left by those selected tests; an upper bound valid on every assignment applies to every universal sampler.

Table 1 reports certificate instances at the classical scale $y = 2^{-m}$. The ranking-LP optima match the analytic two-atom values to the displayed precision. At depth one, equality has a direct proof: condition on the shared $j : k$ face and apply the Fréchet bound in each fiber. The two fibers contribute $1/10$ and $1/12$, giving $11/60$. Increasing depth reduces the remaining reverse mass.

The fiberwise Fréchet calculation can sharpen the cover at other odds. For $r = 2/3$, $s = 5$, $y = 5/6$, and $m = 1$, it gives the exact value $1/6$, below $961/5661$. For positive deficits, the full-assignment LPs at $\varepsilon = 0.02$ attain $11/60 + 0.04$ at $m = 1$ and $2/21 + 0.08$ at $m = 2$. At these depths the finite relaxation attains the defect term of (16) exactly. Appendix C gives its constraints and additional exact checks.

The certificate thus separates two resources. Increasing chain depth adds shared-face constraints and reduces endpoint reverse mass. Each observed deletion deficit contributes an explicit additive cost. Figure 3c shows the depth minimizing their sum, so the chain depth can be matched to the measured deficit. The ranking programs complement this bound by constructing joint assignments that attain it at the listed classical scales.

For arbitrary finite target families, an existence result complements the explicit binary bound.

Proposition 5 (Finite rational target certificate)**.** *Let $n \geq 3$, let P be a finite family of rational distributions, and let $\eta > 0$. There are a finite deletion-closed set $W \subset \Delta(A) \cap \mathbb{Q}^A$ containing P and a rational $\varepsilon > 0$ such that every joint law on W with exact marginals and $\delta(p) \leq \varepsilon$ for all nonsingleton $p \in W$ has target law within η in total variation of the shared Gumbel law on P.*

Compactness proves the result: a sequence with vanishing deficits but a fixed target discrepancy would converge to a deletion-maximal process on the rational simplex, contradicting rigidity. Exhaustive rational LP search therefore finds a certificate for each rational target and tolerance (Appendix C).

6 The countable scale boundary

Now let the alphabet be $\mathbb{N}$. The exponential race remains exact because its total rate is one. Finite deletions still preserve every nonwinner. What changes is the relation between finite-support and infinite-support inputs, and between different mass scales within the same tail class.

Let $T_M p$ be the normalized restriction of p to $\{1, \ldots, M\}$, for prefixes of positive mass. Call F *truncation-continuous* when $F(T_M p) \to F(p)$ in probability for each p.

Theorem 6 (Countable characterization)**.** *An exact sampler on $\Delta(\mathbb{N})$ has the Gumbel grand-coupling law if and only if it is deletion-maximal and truncation-continuous.*

For sufficiency, apply finite rigidity to the union of the truncated supports, enlarged to at least three labels, and pass to the limiting output vector. For necessity, a prefix race agrees with the full race exactly when the full winner lies in the prefix. Hence its disagreement probability is the omitted mass, which tends to zero. These probabilities and deletion agreements depend only on two-dimensional laws (Appendix D).

6.1 The freedom between tail classes

Finite deletion cannot connect every countable distribution to every other one. For example, the Borel statistic $\rho(p) = \limsup_{n\to\infty} p_n^{1/n}$ is invariant under a finite deletion and renormalization. Use one exponential field when $\rho(p) < 2/5$ and an independent field otherwise. This sampler remains exact and deletion-maximal because no tested edge crosses the partition. For $p_n = 2^{-n}$ and $q_n = 2 \cdot 3^{-n}$, the two outputs use independent fields, whereas the ordinary shared race couples them with much higher agreement:

$$a_F(p, q) = \frac{2}{5}, \qquad a_G(p, q) = \sum_{n\geq 1} \frac{2}{3^n + 1} \approx 0.8081. \tag{23}$$

Every finite prefix has $\rho = 0$, so the truncated outputs at p use the first field rather than the independent field of $F(p)$ and converge to an independent copy of the full marginal. Their limiting agreement with $F(p)$ is $\sum_n p_n^2 = 1/3$, not one. This construction shows a gap between different tail classes. A stronger obstruction persists after fixing a single class: within-class mass changes can select different clocks even though every finite deletion keeps the surviving clocks unchanged.

6.2 Clocks can change within a single tail class

To see why a continuity condition is needed, fix a positive summable sequence $a = (a_n)$. Consider the Borel class of distributions that are eventually proportional to a. Each such distribution has a unique representation $p_n = c(p)u_n$ with $u_n = a_n$ eventually. Deleting a coordinate changes $c(p)$ but leaves every surviving u_n unchanged. Choose a fixed scale ratio $\lambda > 1$ and set

$$\chi(u) = \lfloor \log_\lambda u \rfloor \bmod 2, \qquad F(p) = \operatorname*{argmin}_{n:u_n>0} \frac{\xi_{n,\chi(u_n)}}{u_n}. \tag{24}$$

The clock pairs $(\xi_{n,0}, \xi_{n,1})$ are independent across labels and have unit-exponential marginals. Within a pair, the two clocks may be independent or dependent. For each fixed menu, the selected label clocks are independent exponentials, so the sampler is exact. Deletion preserves all surviving times, so it is deletion-maximal. Multiplying one u_n by λ flips its phase and can change its clock. Extending the rule by an independent ordinary race off this deletion-invariant class gives a measurable sampler on the full countable simplex.

For independent phases, compare two menus whose rates for label i are a and λa, whose rate for label j is b, and whose remaining rates sum to R. The reverse-winner event is that i wins the first menu and j wins the second. Its probability is

$$\frac{b}{b + \lambda a + R} - \frac{b}{b + (\lambda + 1)a + R} > 0. \tag{25}$$

Conditioning on the j clock shows why: the first i clock has rung, its refreshed clock has not, and no tail clock has rung. Appendix D evaluates the resulting exponential integral. Increasing a winning label's rate in an ordinary shared race can never cause this reversal. With $b = a/r$ and fixed tail rate R, (25) tends as $a \to \infty$ to

$$V_{\mathrm{ind}}(r, \lambda) = \frac{r}{(1 + \lambda r)(1 + (\lambda + 1)r)}. \tag{26}$$

At $r = 1/2$ and $\lambda = 4$ this is $1/21$. For the finite rates $(a, b, R) = (1/2, 1, 2)$, the probability is instead $1/55$. Figure 4c–e gives the dependence on the tail-to-head rate ratio.

6.3 Adjacent-scale dependence and clock reuse

The phase rule in (24) can put antithetic clocks on every adjacent scale edge: $\xi_{n,0} = -\log U_n$ and $\xi_{n,1} = -\log(1 - U_n)$, with independent uniform U_n across labels. For $b = a/r$ and vanishing

relative tail rate R/a, the reverse probability becomes

$$V(r,\lambda) = \frac{1}{r}\int_0^\infty e^{-t/r}\min\{1-e^{-t}, e^{-\lambda t}\}\,dt. \tag{27}$$

The minimum is the Fréchet intersection bound between one clock falling below t and the other exceeding λt, attained by these antithetic clocks. At $r = 1/2$, $\lambda = 4$, if $x + x^4 = 1$, it equals

$$V(1/2, 4) = \frac{1 - 2x^2 + x^3}{3} \approx 0.1102. \tag{28}$$

Across an odd number k of scale steps the phases are opposite, giving $V(r, \lambda^k)$. Across an even number of steps the same clock is reused and the reverse probability is exactly zero. Figure 4 displays both the inter-class gap and this within-class parity structure. All these samplers are exact.

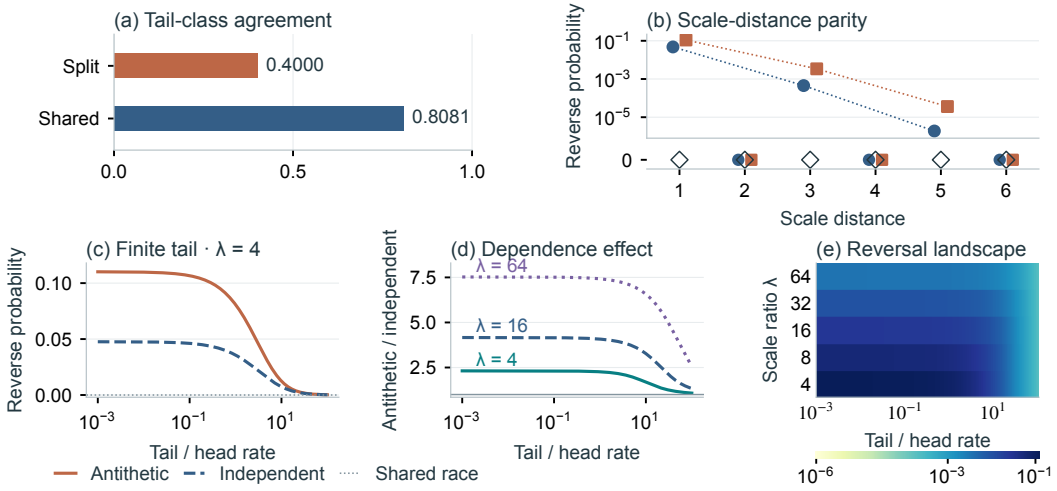


Figure 4: Countable nonuniqueness: tail-class agreement, scale parity, finite-tail reversals and their dependence on coupling and scale ratio. Curves evaluate the clock construction analytically; even-distance reversals are exactly zero.

The obstruction to repeating the finite proof can now be stated directly. With only two retained labels, multiplying their unnormalized weights by a common factor leaves the distribution unchanged. All low ternaries in (8) therefore refer to one binary comparison, and all high ternaries refer to another. With a fixed positive tail, the putative binary face still contains that tail. Scaling the two head weights changes its normalized distribution. The local inequalities then concern different reverse events at different scales; they cannot telescope into a bound for one fixed event. Truncation continuity reconnects those scale-dependent menus to their finite restrictions.

7 Limitations

The explicit error bound controls a specified binary comparison. For an arbitrary finite rational target family, the certificate follows from compactness and exhaustive search; obtaining a useful bound on its size, or an explicit modulus relating deletion deficits to the full target law, remains open.

8 Conclusion

Local deletion optimality determines the shared Gumbel joint sampling law on finite alphabets with at least three labels. This characterization establishes its Pareto optimality and derives its race structure directly from agreement constraints. Quantitative certificates connect imperfect deletion agreement to binary comparison reversals. On countable alphabets, scale-dependent clocks allow distinct deletion-optimal couplings within one tail class, while prefix-truncation continuity restores uniqueness. These results identify when local agreement fixes a global sampling rule.

REFERENCES

David S Ahn, Federico Echenique, and Kota Saito. On path independent stochastic choice. *Theoretical Economics*, 13(1):61–85, 2018.

Omer Angel and Yinon Spinka. Pairwise optimal coupling of multiple random variables. *arXiv preprint arXiv:1903.00632*, 2019.

August A Balkema and Laurens De Haan. Residual life time at great age. *The Annals of probability*, pp. 792–804, 1974.

Mohammad Bavarian, Badih Ghazi, Elad Haramaty, Pritish Kamath, Ronald L Rivest, and Madhu Sudan. Optimality of correlated sampling strategies. *Theory of Computing*, 16(1), 2020.

Andrei Z Broder. On the resemblance and containment of documents. In *Proceedings. Compression and Complexity of SEQUENCES 1997 (Cat. No. 97TB100171)*, pp. 21–29. IEEE, 1997.

Claus Bundesen. The relationship between independent race models and luce's choice axiom. *Journal of Mathematical Psychology*, 37(3):446–471, 1993.

Tianle Cai, Yuhong Li, Zhengyang Geng, Hongwu Peng, Jason D Lee, Deming Chen, and Tri Dao. Medusa: Simple llm inference acceleration framework with multiple decoding heads. *arXiv preprint arXiv:2401.10774*, 2024.

Christopher P Chambers, Yusufcan Masatlioglu, and Christopher Turansick. Correlated choice. *Theoretical Economics*, 19(3):1087–1117, 2024.

Charlie Chen, Sebastian Borgeaud, Geoffrey Irving, Jean-Baptiste Lespiau, Laurent Sifre, and John Jumper. Accelerating large language model decoding with speculative sampling. *arXiv preprint arXiv:2302.01318*, 2023.

Majid Daliri, Christopher Musco, and Ananda Theertha Suresh. Coupling without communication and drafter-invariant speculative decoding. In *2025 IEEE International Symposium on Information Theory (ISIT)*, pp. 1–6. IEEE, 2025.

Peter Donnelly and Paul Joyce. Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Processes and their Applications*, 31(1):89–103, 1989.

Ehtibar N Dzhafarov and Janne V Kujala. Contextuality-by-default 2.0: Systems with binary random variables. In *International Symposium on Quantum Interaction*, pp. 16–32. Springer, 2016.

Yichao Fu, Peter Bailis, Ion Stoica, and Hao Zhang. Break the sequential dependency of llm inference using lookahead decoding. *arXiv preprint arXiv:2402.02057*, 2024.

Emil Julius Gumbel. *Statistics of extremes*. Courier Corporation, 2004.

Iris AM Huijben, Wouter Kool, Max B Paulus, and Ruud JG Van Sloun. A review of the gumbel-max trick and its extensions for discrete stochasticity in machine learning. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):1353–1371, 2022.

Sergey Ioffe. Improved consistent sampling, weighted minhash and l1 sketching. In *2010 IEEE international conference on data mining*, pp. 246–255. IEEE, 2010.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

Yaniv Leviathan, Matan Kalman, and Yossi Matias. Fast inference from transformers via speculative decoding. In *International conference on machine learning*, pp. 19274–19286. PMLR, 2023.

Torgny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.

Guy Lorberbom, Daniel D Johnson, Chris J Maddison, Daniel Tarlow, and Tamir Hazan. Learning generalized gumbel-max causal mechanisms. *Advances in Neural Information Processing Systems*, 34:26792–26803, 2021.

R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2005.

Chris J Maddison, Daniel Tarlow, and Tom Minka. A* sampling. *Advances in neural information processing systems*, 27, 2014.

Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016.

Mark Manasse, Frank McSherry, and Kunal Talwar. Consistent weighted sampling. Technical report, Technical Report MSR-TR-2010-73, Microsoft Research, 2010.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. 1972.

Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Zeyu Wang, Zhengxin Zhang, Rae Ying Yee Wong, Alan Zhu, Lijie Yang, Xiaoxiang Shi, et al. Specinfer: Accelerating large language model serving with tree-based speculative inference and verification. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*, pp. 932–949, 2024.

Michael Oberst and David Sontag. Counterfactual off-policy evaluation with gumbel-max structural causal models. In *International Conference on Machine Learning*, pp. 4881–4890. PMLR, 2019.

Hermann Þórisson. Coupling, stationarity, and regeneration. 2000.

James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pp. 119–131, 1975.

Shauli Ravfogel, Anej Svete, Vésteinn Snæbjarnarson, and Ryan Cotterell. Gumbel counterfactual generation from language models. In *International Conference on Learning Representations*, volume 2025, pp. 90486–90512, 2025.

Sidney I Resnick and Sidney I Resnick. *Extreme values, regular variation and point processes*, volume 4. Springer, 1987.

Joseph Rowan, Truong Buu Phan, and Ashish Khisti. List-level distribution coupling with applications to speculative decoding and lossy compression. *Advances in Neural Information Processing Systems*, 38:5576–5619, 2026.

Benjamin Spector and Chris Re. Accelerating llm inference with staged speculative decoding. *arXiv preprint arXiv:2308.04623*, 2023.

Louis L Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Heming Xia, Zhe Yang, Qingxiu Dong, Peiyi Wang, Yongqi Li, Tao Ge, Tianyu Liu, Wenjie Li, and Zhifang Sui. Unlocking efficiency in large language model inference: A comprehensive survey of speculative decoding. *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 7655–7671, 2024.

John I Yellott Jr. The relationship between luce's choice axiom, thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2): 109–144, 1977.

AI-USE DISCLOSURE

Generative AI was used in the underlying research workflow to develop proof arguments and counterexamples, write verification code, and interpret computational results. It was also used to select and organize evidence, draft and revise this manuscript, create figures from analytic expressions and recorded results, and inspect the resulting figures and PDF.

A PROBABILITY-ONE EVENTS AND FINITE-ALPHABET DETAILS

Sampler identities below concern a fixed finite input family unless a countable family is specified.

A.1 HEREDITY AND RANKING PROBABILITIES

For a fixed nonsingleton p, exactness implies $F(p) \in \operatorname{supp}(p)$ almost surely, and the same holds for its finitely many deletions. On this event, deleting the winner cannot agree. Consequently $h(p) = (|\operatorname{supp}(p)| - 1) - \sum_i J_i$ is a nonnegative integer. If $\delta(p) = 0$, then $h(p) = 0$ almost surely, proving heredity. In particular, on a ternary menu, its winner beats both competitors in the binary deletion faces. A tournament on three vertices with a vertex beating the other two is transitive. Its remaining edge decides the order of the two nonwinners. Seedwise, heredity is Sen's contraction consistency, which for single-valued choices is equivalent to path independence; the stochastic form of path independence characterizes Luce-type rules (Ahn et al., 2018). Correlated choice studies which joint distributions of several decision makers' choices arise from correlated random utilities (Chambers et al., 2024); our grand coupling is the analogous joint object, with menus in place of decision makers.

For a permutation (i, j, k) of the three labels, the event that the $j : k$ face selects j consists of three rankings: ijk, jik, and jki. The last two are exactly the event that the full menu selects j. Thus

$$\mathbb{P}(ijk) = \mathbb{P}(F(p^{-i}) = j) - \mathbb{P}(F(p) = j) = \frac{p_j}{1 - p_i} - p_j = \frac{p_i p_j}{1 - p_i}.$$

This identifies all six ranking probabilities without an independence assumption. These are the ranking probabilities of Luce's choice model (Luce, 2005), which conditional logit realizes with independent Gumbel utilities (McFadden, 1972). Among independent-noise models of comparative judgment (Thurstone, 1927), Yellott's theorem (Yellott Jr, 1977) singles out the same double-exponential law; here that law is derived from deletion constraints. Zero coordinates are handled by restricting to positive support; no ternary ranking is assigned to a binary input.

A.2 THRESHOLD CONSTRUCTION

Intersecting the amplification events over finitely many label pairs and countably many rational $r < s$ gives simultaneous rational nesting. Since $x/(1 + x)$ tends to zero and one at the endpoints, R_{ij} is strictly positive and finite almost surely.

For a fixed real $x > 0$, choose rational sequences $r_\ell \uparrow x$ and $s_\ell \downarrow x$. Applying the same pairwise nesting result to r_ℓ, x, s_ℓ yields

$$C_{ij}(r_\ell) \le C_{ij}(x) \le C_{ij}(s_\ell) \quad \text{for every } \ell$$

on a probability-one event. Exactness gives the continuous distribution function $x/(1 + x)$, so $\mathbb{P}(R_{ij} = x) = 0$; the squeeze identifies the comparison with its threshold event. The reciprocal identity follows because the same two-point distribution read with labels reversed has complementary outputs, and its odds are reciprocal.

Ternary transitivity is also available simultaneously for all rational positive weight triples. If $R_{ik} > R_{ij}R_{jk}$, choose rationals $a > R_{ij}$ and $b > R_{jk}$ with $ab < R_{ik}$. The weights $(ab, b, 1)$ produce the forbidden order $i \succ j \succ k \succ i$. If the inequality is reversed, choose $a < R_{ij}$ and $b < R_{jk}$ with $ab > R_{ik}$ to obtain the opposite cycle. These rational witnesses establish the cocycle on one countable probability-one event.

A.3 SUPPORT INDUCTION AND LAW IDENTIFICATION

The score representation is immediate for supports of size one or two. For a fixed larger support K, intersect the heredity event at p with the induction events at its finitely many deletions. If $F(p) = k$ and $j \neq k$, choose $i \in K \setminus \{k, j\}$. The output at p^{-i} is still k and the smaller race representation implies $S_k/p_k \le S_j/p_j$. The factors introduced by renormalization are common to all survivors. Each possible equality has probability zero because S_k/S_j has a continuous distribution. This proves the unique-minimum representation at p.

Upper orthants determine finite rectangles by inclusion–exclusion, hence the Borel law of the normalized scores. Each finite output vector is a measurable function of those ratios outside finitely many null tie hyperplanes. Equation (14) therefore identifies its law. Common random rescaling is invisible to outputs; independent unnormalized scores on the original seed space are not identified.

A.4 Binary alphabets and private randomness

Write a binary distribution as $(u, 1-u)$. If $u < v$, every maximal Bernoulli coupling has zero probability of choosing the first label at u and the second label at v. Global agreement domination over Gumbel forces maximal coupling for every such pair, since the Gumbel race attains the coupling inequality $\mathbb{P}(X = Y) \le 1 - d_{\mathrm{TV}}$ (Lindvall, 2002; Þórisson, 2000). Rational nesting and squeezing give one uniform threshold for every fixed finite list, proving the binary case of Corollary 2. Deletion optimality alone cannot do this: deleting either label leaves a deterministic singleton and achieves the deletion bound for every exact coupling, including nonnested ones.

Suppose two participants instead use different measurable algorithms and independent private randomness conditional on public randomness U. At a diagonal input (p, p), domination over the common Gumbel rule requires agreement one. If their conditional output distributions are $\alpha(U)$ and $\beta(U)$, the conditional agreement is $\langle \alpha, \beta \rangle \le 1$. Equality holds only if both distributions put mass one on the same label. For each fixed p, their outputs are therefore almost surely a common deterministic function of U. A measurable version can be selected, for example by the least label maximizing the conditional probability. For a finite collection of inputs, intersect the corresponding full-probability events. Pairwise agreement is unchanged, and the common-sampler theorem applies. This reduction uses diagonal domination, not the deletion-only premise.

B Pointwise proof of the robust certificate

We first establish (20) pointwise for every deterministic assignment on the gadget, then use exact marginal probabilities when taking expectations.

Write $L = C_{ij}(r)$, $H = 1 - C_{ij}(s)$, $B_t = C_{ik}(\lambda^t y)$, and $D_t = C_{jk}(\lambda^t y / r)$. The low output is X_t and the high output is X'_t. Directly counting agreement with the three deletion faces gives the following table.

Full output	$h(q_t)$	$h(q'_t)$
i	$2 - L - B_t$	$1 + H - B_{t+1}$
j	$1 + L - D_t$	$2 - H - D_t$
k	$B_t + D_t$	$B_{t+1} + D_t$

All entries are nonnegative. For arbitrary L, H, the endpoint expressions

$$h(q_0) + D_0 - \mathbf{1}\{X_0 = j\}, \qquad h(q'_{m-1}) + H - \mathbf{1}\{X'_{m-1} = j\}$$

are nonnegative, as inspection of the three respective rows shows. Thus the right-hand side of (20) is nonnegative off E.

On E we have $L = H = 1$. The first endpoint expression is at least D_0, and the second is at least $1 - D_{m-1}$. For every internal bridge $0 \le t < m - 1$, the same table yields

$$h(q'_t) \ge B_{t+1} - D_t, \qquad h(q_{t+1}) \ge D_{t+1} - B_{t+1}.$$

Adding the bridge inequalities and the two endpoint bounds telescopes to $D_0 + (D_{m-1} - D_0) + 1 - D_{m-1} = 1$. For $m = 1$ there are no internal bridges and the endpoint sum alone is at least one. This proves the pointwise inequality for all depths.

Now impose exact marginals. At the first low menu,

$$\mathbb{E}[D_0 - \mathbf{1}\{X_0 = j\}] = \frac{y/r}{1 + y/r} - \frac{y/r}{1 + y + y/r} = \frac{y^2/r}{(1 + y + y/r)(1 + y/r)}.$$

At the last high menu, whose $i : j$ odds are s,

$$\mathbb{E}[H - \mathbf{1}\{X'_{m-1} = j\}] = \frac{1}{s+1} - \frac{\lambda^m y/s}{1 + \lambda^m y(1 + 1/s)} = \frac{1}{(s+1)(1 + \lambda^m y(1 + 1/s))}.$$

Finally, $\mathbb{E}h(q) = \delta(q)$ by (6). Their sum proves (16). The defect counts every failed nonwinner deletion, including those occurring with a transitive binary tournament.

B.1 Rates and the randomized test

For $c = c_{r,s}$ and $y = c\lambda^{-m/3}$, the envelope satisfies

$$\frac{y^2}{r} + \frac{s}{(s+1)^2\lambda^m y} = \left(\frac{c^2}{r} + \frac{s}{(s+1)^2 c}\right)\lambda^{-2m/3} = \frac{3c^2}{r}\lambda^{-2m/3}.$$

At $r = 1/2$, $s = 2$, $c = 18^{-1/3}$ and $C = 6 \cdot 18^{-2/3} \approx 0.8736$. The classical scale $y = 2^{-m}$ gives the exact expression

$$U_m(2^{-m}) = \frac{2}{(2^m+2)(2^m+3)} + \frac{1}{3 + 9 \cdot 2^{m-1}}.$$

These expressions produce the curves in Figures 3 and 5. The balanced scale minimizes the simple envelope, not necessarily U_m at each finite depth. For instance, at depth one the rational choice $y = 1/3$ gives $8/45$, below both the classical value $11/60$ and the balanced-scale value.

For the Bernoulli outputs $C_{ij}(r)$ and $C_{ij}(s)$, let $e = \mathbb{P}(C_{ij}(r) = 1, C_{ij}(s) = 0)$. Marginal subtraction gives $\mathbb{P}(C_{ij}(r) = 0, C_{ij}(s) = 1) = s/(1+s) - r/(1+r) + e$. Agreement is therefore its maximal value minus $2e$, proving (21). With the depth chosen in the main text, the resulting explicit bound is

$$a_G - a_F \le 2C_{r,s}\varepsilon + \left(\frac{6}{\log\lambda}\log\frac{1}{\varepsilon} + 4\right)\varepsilon.$$

For the randomized test, each ternary's three deletion agreements sum to $2 - \delta(q)$. Averaging over $2m$ menus and three deletions gives $\bar{a} = 2/3 - \sum_q \delta(q)/(6m)$, yielding (22). This identity converts expected deletion agreement into a bound on the specified binary pair.

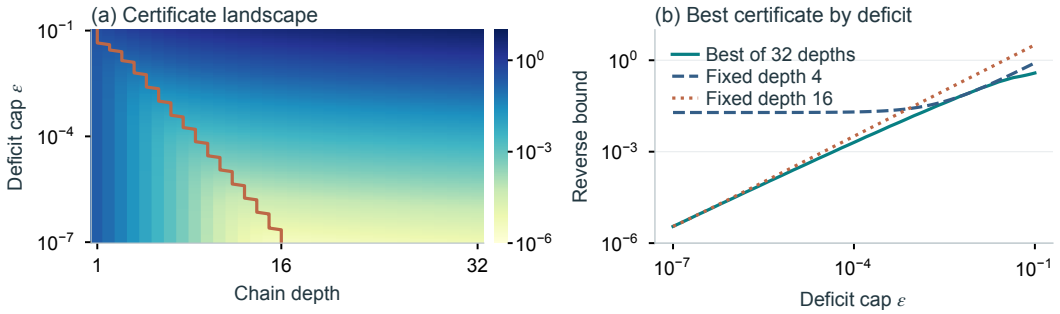


Figure 5: Analytic certificates at $r = 1/2$, $s = 2$. Left: $U_m(c_{r,s}4^{-m/3}) + 2m\varepsilon$ and the best-depth path. Right: the best of 32 depths versus two fixed depths.

C Finite coupling programs and exact checks

Let W be a finite set of distributions containing the chosen ternaries and all their distinct binary faces. A deterministic assignment z chooses one supported label z_p at each $p \in W$. Associate a nonnegative mass w_z to each assignment. The full-assignment linear program imposes

$$\sum_z w_z = 1, \qquad \sum_z w_z \mathbf{1}\{z_p = i\} = p_i \quad (p \in W,\ i \in \mathrm{supp}(p)),$$

$$\sum_z w_z \left(2 - \sum_{i \in \mathrm{supp}(p)} \mathbf{1}\{z_p = z_{p^{-i}}\}\right) \le \varepsilon \quad \text{for each selected ternary } p.$$

Its objective is $\sum_z w_z \mathbf{1}\{z_{P_{ij}(r)} = i, z_{P_{ij}(s)} = j\}$. Coincident faces are a single variable. At zero deficit, heredity restricts each menu to one of six rankings. The ranking program enumerates tuples of these rankings consistent on all shared faces and constrains each menu's six marginals to (15).

Table 2: Exhaustive pointwise-dual checks and positive-deficit LP values. The inequality holds on every deterministic assignment of each listed gadget.

Depth	Assignments	Dual violations	Deficit cap	LP reverse probability
1	288	0	0.02	0.2233
2	10368	0	0.02	0.1752
3	373248	0	—	—

The checks cover 1458 rational deletion identities and recover the six ranking masses by a rank-six rational constraint matrix. The dual enumeration tests every deterministic assignment, supplementing the all-depth algebraic proof.

At depth one, conditioning on the shared $j : k$ comparison separates the two-menu coupling into two separate two-marginal coupling (transport) problems (Villani et al., 2009). In each fiber, the maximum intersection of $\{i \succ j\}$ in the low ranking and $\{j \succ i\}$ in the high ranking is the minimum of their prescribed masses. At $(r, s, y) = (1/2, 2, 1/2)$ those minima are $1/10$ and $1/12$. At $(2/3, 5, 5/6)$ their sum is instead $1/6$, while the two selected endpoint atoms total $961/5661$. This gives an exact gap between the cover and the optimal finite coupling.

The finite program specifies a joint law on W. A universal sampler additionally satisfies face identifications over the full simplex.

C.1 Proof of the finite rational target theorem

The proof of Theorem 1 restricts to rational inputs: use rational odds and rational chain scales, and use rational upper orthants to identify the score-ratio law. Rational scales can approach the optimizer closely enough to make the finite amplification bounds vanish. Thus exactness and deletion maximality on the countable rational simplex identify all of its finite-dimensional laws.

Suppose Proposition 5 failed for a fixed P and $\eta > 0$. Enumerate the rational simplex and take an increasing finite deletion-closed exhaustion W_ℓ containing P. At each ℓ, failure supplies a coupling on W_ℓ with exact marginals, all deficits at most $1/\ell$, and target total-variation distance at least η. Extend it to the unmentioned rational inputs using independent exact marginals. These laws lie on the compact metrizable product $A^{\Delta(A)\cap\mathbb{Q}^A}$; probability measures on this space have a weakly convergent subsequence.

Each fixed marginal constraint and deletion-agreement constraint is a continuous cylinder condition. In the limit all marginals are exact and every deficit is zero. Since the target output space is finite, its total-variation distance is continuous in its finitely many atom probabilities and remains at least η. Rational-skeleton rigidity says that the limiting target law is exactly Gumbel, a contradiction.

For rational P and rational $\eta > 0$, enumerate finite rational witness sets and positive rational deficit caps. For each candidate, maximize the positive and negative deviation from the Gumbel target probability for every subset of the finite target output space. A finite set is accepted if all these LP optima are strictly below η. The Gumbel target probabilities are rational: normalized independent exponentials are uniform on a simplex, each fixed output assignment imposes homogeneous rational linear comparisons, and these define rational polytopes of rational volume. Thus all the programs have rational data and the comparisons can be exact. The existence proof guarantees termination of an exhaustive search.

D Countable races, measurability, and counterexamples

D.1 The infinite exponential race and truncation

For a countable positive-rate sequence (u_n) with finite sum $U > 0$, independent exponentials of rates u_n have minimum survival function

$$\mathbb{P}(\inf_n T_n > t) = \prod_n e^{-u_n t} = e^{-Ut}.$$

There are almost surely finitely many arrivals in each bounded time interval: the expected count is $\sum_n (1 - e^{-u_n t}) \le tU$. The infimum is therefore attained, and continuous pairwise times give a unique minimum. Integration shows that label n wins with probability u_n/U. This proves exactness for the countable races used here. More generally, the order of arrivals is a size-biased permutation of (u_n/U) (Donnelly & Joyce, 1989). In Gumbel form the race maximizes $\log u_n + Z_n$ with standard Gumbel perturbations $Z_n = -\log E_n$ (Gumbel, 2004), and the maximum is again Gumbel with location $\log U$, an instance of the max-stability of the Gumbel family (Resnick & Resnick, 1987).

For an ordinary shared race, let I be the full winner at p. Every prefix containing I has the same winner after normalization. Hence $G(T_M p) \to G(p)$ almost surely, and indeed

$$\mathbb{P}(G(T_M p) \neq G(p)) = \sum_{n>M} p_n.$$

For finitely many inputs, the union bound gives joint convergence. Apply finite rigidity to the union of truncated supports, enlarged to three labels if needed, then pass equality of the truncated laws to the limit. Conversely, Gumbel two-dimensional laws imply the displayed disagreement probability and all deletion equalities, proving Theorem 6.

D.2 A Borel tail class and its invariant coordinates

Fix $a_n > 0$ with $\sum_n a_n < \infty$, and define

$$\mathcal{O}_a = \{p : \text{for some } N \text{ and } c > 0,\ p_n = c a_n \text{ for every } n > N\}.$$

For fixed N, this is the intersection of the Borel equalities $p_n/a_n = p_{N+1}/a_{N+1} > 0$ for $n > N$. Taking a countable union gives a Borel set. The eventual ratio $c(p)$ is unique and is a measurable function on this class; set $u_n = p_n/c(p)$. Finite modifications of the head do not change the defining tail, and deleting a head or tail label only changes finitely many coordinates and the normalization factor. Thus $\mathcal{O}_a$ is invariant both under deletion and under reversal of a deletion when the restored coordinate is specified. No deletion edge crosses between this class and its complement.

The phase function is Borel on positive masses; zero rates are excluded. At fixed p, deterministic phase selection preserves unit-exponential marginals and independence across labels. The countable minimum, with a least-index tie convention, is measurable, unique, and exact by the race argument above. Deletion preserves surviving u_n and phases, so it changes only a deleted winner. An independent ordinary race off the deletion-invariant $\mathcal{O}_a$ completes the global sampler.

D.3 Independent-phase reversal

Let T_j have exponential rate b and let T_R be the minimum of the remaining clocks, with rate R. For opposite independent phases, let the two times for label i have rates a and λa. Conditioning on $T_j = t$ makes the reverse event require the first i time below t, the second above t, and $T_R > t$. The refreshed clock, the second i time, is memoryless: its excess over any threshold is again exponential with the same rate, the shape-zero case of the generalized Pareto excess law (Balkema & De Haan, 1974; Pickands III, 1975), whose excess distribution does not depend on the threshold. Independence gives

$$\int_0^\infty b e^{-(b+R)t}(1 - e^{-at})e^{-\lambda a t}\, dt = \frac{b}{b + \lambda a + R} - \frac{b}{b + (\lambda + 1)a + R}.$$

Taking $b = a/r$ and $a \to \infty$ gives (26). To realize these menus inside $\mathcal{O}_a$, change only finitely many head coordinates of the invariant sequence u and renormalize. The total tail rate remains fixed while the chosen head rates increase. Because scaling a head by λ flips phase, the reverse event has positive probability at every finite scale, in contrast to an ordinary shared race. Figure 6 evaluates these reverse probabilities at finite tail rates.

D.4 Antithetic phases and the closed form

Conditioning on the competing j clock and rescaling time by a, the large-scale reverse event for label i asks for $\xi_0 < t$ and $\xi_1 > \lambda t$. The individual event probabilities are $1 - e^{-t}$ and $e^{-\lambda t}$. With $\xi_0 = -\log U$ and $\xi_1 = -\log(1 - U)$, their intersection has probability equal to the smaller one.

This gives (27). The construction realizes the same countermonotone copula on every adjacent edge of the scale graph, since the phases alternate.

Let $x \in (0,1)$ solve $x + x^\lambda = 1$ and let $t_* = -\log x$. Splitting the integral at t_* gives

$$V(r,\lambda) = \frac{1}{r}\int_0^{t_*} e^{-t/r}(1-e^{-t})\,dt + \frac{1}{r}\int_{t_*}^{\infty} e^{-(1/r+\lambda)t}\,dt$$
$$= 1 - x^{1/r} - \frac{1-x^{1+1/r}}{1+r} + \frac{x^{1/r}(1-x)}{1+\lambda r}.$$

At $r = 1/2$, $\lambda = 4$, using $x^4 = 1 - x$ reduces this to $(1 - 2x^2 + x^3)/3$. At odd distance k, replace λ by λ^k. At even distance, $\chi(\lambda^k u) = \chi(u)$, so the same clock is used and a winning label cannot lose after its rate increases.

Table 3: Analytic large-scale reverse probabilities in the two-phase processes at $r = 1/2$, $\lambda = 4$. These are limits as the fixed tail becomes negligible relative to the head rates.

Scale distance k	Shared race	Independent phases	Antithetic phases
1	0	$1/21$	1.102×10^{-1}
2	0	0	0
3	0	$1/2211$	3.403×10^{-3}
4	0	0	0
5	0	$1/526851$	3.618×10^{-5}
6	0	0	0

D.5 Agreement in the inter-class example

For arbitrary positive p, q, condition on the clock E_i to obtain the shared-race agreement formula

$$a_G(p,q) = \sum_i \left(\sum_j \max\left\{ \frac{p_j}{p_i}, \frac{q_j}{q_i} \right\} \right)^{-1}.$$

For $p_i = 2^{-i}$ and $q_i = 2 \cdot 3^{-i}$, the inner sum equals $\sum_{j=1}^{i} 3^{i-j} + \sum_{j>i} 2^{i-j} = (3^i + 1)/2$, giving (23). Independent fields instead have agreement $\sum_i 2 \cdot 6^{-i} = 2/5$. Finally $\sum_i p_i^2 = \sum_i 4^{-i} = 1/3$ gives the limiting prefix agreement in the split-field construction. All series converge absolutely, so the conditioning and summation are justified by nonnegativity.

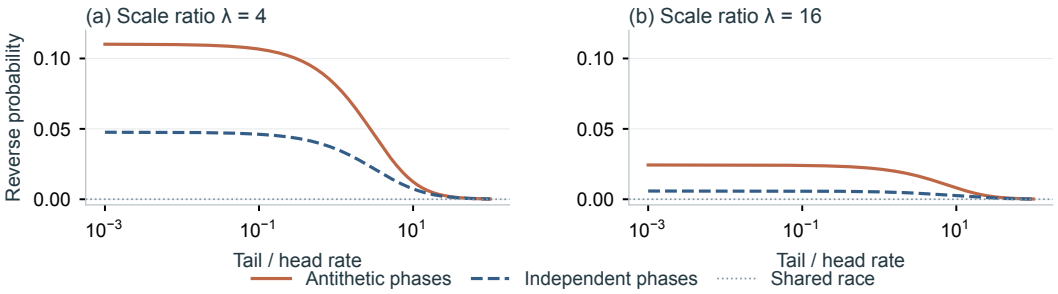


Figure 6: Exact finite-tail reverse probabilities at $r = 1/2$ and two scale ratios. Antithetic and independent phases admit reversals; the shared race remains at zero.