hexagon:2611.00732v1

Extremizers of the expected missing mass under an entropy constraint

Christopher D. Long
galizur@gmail.com

Abstract

For a discrete probability distribution $p = (p_i)$, the expected mass of symbols absent from t independent samples is $\Phi_t(p) = \sum_i p_i(1-p_i)^t$. For every positive integer t and finite entropy budget $h \geq 0$, we prove that every ℓ^1-local maximizer under $H(p) \leq h$ has at most two distinct positive coordinate values, and all but at most one positive coordinate are equal. This holds on finite and countably infinite alphabets and improves the four-size bound of Berend, Kontorovich, and Zagdanski (2017). On a countably infinite alphabet, we give an exact finite classification of all global maximizers: a unique one-light or uniform candidate is compared with an explicitly bounded family of one-heavy candidates, whose masses are uniquely specified by entropy equations. The one-light or uniform candidate is uniquely optimal for $t = 1, 2$. For fixed $h > 0$, the optimal value $B_t(h)$ satisfies

$$\frac{h}{B_t(h)} = \log t + \log\log t + 2 + O_h\left(\frac{\log\log t}{\log t}\right).$$

Every optimizer is eventually one-heavy, with $m_t \sim ht$ equal light atoms of size $q_t \sim 1/(t\log t)$. The two-size proof combines a three-coordinate entropy-curvature estimate with a finite coefficient comparison. Crossing theorems describe the winning intervals of the candidate distributions, and the classification gives exact extrema for singleton counts, discovery increments, and expected coverage.

1 Introduction

Let I be a nonempty finite or countably infinite alphabet, and write

$$H(p) = -\sum_{i\in I} p_i \log p_i, \qquad \Phi_t(p) = \sum_{i\in I} p_i(1-p_i)^t,$$

$$\mathcal{P}_h(I) = \left\{ p_i \geq 0 : \ \sum_{i\in I} p_i = 1, \ H(p) \leq h \right\}.$$

Here $0\log 0 = 0$ and logarithms are natural. For integer $t \geq 1$, Φ_t is the expected mass of symbols absent from t independent samples. An *atom size* is a distinct positive coordinate value.

Without an entropy constraint, Berend and Kontorovich [1] showed that finite-alphabet local maximizers have all but possibly one coordinate equal, with the remaining coordinate no smaller. For the entropy-constrained problem, Berend, Kontorovich, and Zagdanski [2, Theorem 1] obtained a four-size bound for finite-alphabet local maximizers and asked whether four sizes can occur. They also characterized the maximizing shape for $t = 1$ [2, Proposition 3]. The following theorem replaces four by two.

Theorem 1.1. *For every integer $t \geq 1$ and $0 \leq h < \infty$, every local maximizer of Φ_t on $\mathcal{P}_h(I)$, in the ℓ^1 topology, has at most two atom sizes and has finite support. The maximum is attained. If I is countably infinite, every local maximizer satisfies $H(p) = h$.*

On countably infinite alphabets, attainment and finite support of global maximizers were established in [2, Theorem 2], and entropy saturation in [2, Proposition 1 and Corollary 1]. The finite-subvector argument in their proof of Theorem 2 also yields a four-size bound, hence finite support, for ℓ^1-local maximizers: fixing the complement makes each finite subvector locally optimal with its own mass and entropy budget, to which their four-size argument applies. We include the compactness and atom-splitting arguments for completeness.

Theorem 5.2 strengthens the two-size conclusion: every nonuniform local maximizer has one exceptional atom and an equimass block. For a countably infinite alphabet, Theorem 5.4 reduces the global problem to a finite candidate list with an explicit multiplicity cutoff. Theorem 6.1 determines the next terms beyond the first-order asymptotic $h/\log t$ from [2, Theorems 3–4], and also determines the asymptotic sizes and multiplicities of the maximizing atoms.

The proof of the two-size theorem occupies the next three sections. Section 5 gives the multiplicity restriction and finite classification, and Section 6 gives the asymptotics and occupancy consequences. The appendices contain the entropy-curvature estimate, crossing comparisons, the finite-alphabet classification, and rationally certified examples.

2 The three-coordinate reduction

Fix three distinct positive coordinates $x < y < z$, keeping all other coordinates unchanged, and put

$$\begin{gathered} S = x + y + z, \quad E = xy + xz + yz, \quad P = xyz, \quad D(s) = (s+x)(s+y)(s+z), \\ I_0 = \int_0^\infty \frac{\mathrm{d}s}{D(s)}, \qquad I_1 = \int_0^\infty \frac{s\,\mathrm{d}s}{D(s)}, \qquad m = I_1/I_0, \\ \mathcal{K} = \frac{1}{I_0}\int_0^\infty \frac{(s-m)^2}{D(s)^2}\,\mathrm{d}s. \end{gathered} \tag{2.1}$$

The map $(x, y, z) \mapsto (S, E, P)$ has nonzero Vandermonde Jacobian at a triple of distinct roots. Thus, with S fixed, (E, P) are smooth local coordinates; sufficiently small changes preserve positivity and distinctness. The three-coordinate entropy $\mathcal{H} = -x\log x - y\log y - z\log z$ satisfies

$$\mathcal{H}_E = I_1, \qquad \mathcal{H}_P = I_0. \tag{2.2}$$

Indeed the roots of $W(u) = u^3 - Su^2 + Eu - P$ satisfy $u_E = -u/W'(u)$ and $u_P = 1/W'(u)$. Use

$$\log u = \int_0^\infty \left(\frac{1}{s+1} - \frac{1}{s+u}\right)\mathrm{d}s$$

and the partial fractions, with each sum over $u \in \{x, y, z\}$,

$$\sum_u \frac{1}{(s+u)W'(u)} = \frac{1}{D(s)}, \qquad \sum_u \frac{u}{(s+u)W'(u)} = -\frac{s}{D(s)}.$$

Differentiation under the integral gives

$$\mathcal{H}_{EE} = -\int_0^\infty \frac{s^2}{D(s)^2}\,\mathrm{d}s, \quad \mathcal{H}_{EP} = -\int_0^\infty \frac{s}{D(s)^2}\,\mathrm{d}s, \quad \mathcal{H}_{PP} = -\int_0^\infty \frac{\mathrm{d}s}{D(s)^2}.$$

Since $\mathcal{H}_P = I_0 > 0$, the entropy level set through the triple is a smooth curve $P = P(E)$, and implicit differentiation yields

$$P' = -m, \qquad P'' = \mathcal{K}. \tag{2.3}$$

We shall use the following estimate, proved in Appendix A.

Lemma 2.1. *For every positive triple,*

$$0 < m \le S/3, \qquad \mathcal{K} \ge \frac{16}{15(S+m)}. \tag{2.4}$$

Equality in $m \le S/3$, or in the lower bound for $\mathcal{K}$, holds if and only if $x = y = z$.

Let $f_t(u) = u(1-u)^t$ and $F_t(E,P) = f_t(x) + f_t(y) + f_t(z)$. For $\phi(E) = F_t(E, P(E))$, the chain rule and (2.3) give

$$\begin{aligned} \phi' &= \mathcal{D}_t := (F_t)_E - m(F_t)_P, \\ \phi'' &= \mathcal{C}_t := (F_t)_{EE} - 2m(F_t)_{EP} + m^2(F_t)_{PP} + \mathcal{K}(F_t)_P. \end{aligned} \tag{2.5}$$

At a local maximum containing the selected triple, both directions along this entropy level curve are feasible, whether or not the entropy constraint is active. Hence $\mathcal{D}_t = 0$ and $\mathcal{C}_t \le 0$. At fixed E, decreasing P decreases $\mathcal{H}$, because $\mathcal{H}_P > 0$, and is also feasible. Local maximality therefore gives $(F_t)_P \ge 0$.

If $t > 1$ and $(F_t)_P = 0$, stationarity gives $(F_t)_E = 0$ as well. Since (E, P) are local coordinates on the fixed-sum plane, this forces $f_t'(x) = f_t'(y) = f_t'(z)$. Rolle's theorem would then give two distinct zeros of f_t'' between x and z, whereas

$$f_t''(u) = t(1-u)^{t-2}((t+1)u - 2)$$

has only one zero in $(0,1)$. Thus $R_t := (F_t)_P > 0$. We will prove

$$\mathcal{D}_t = 0, \quad R_t > 0, \quad S \le 1 \quad \Longrightarrow \quad \mathcal{C}_t > 0, \tag{2.6}$$

contradicting $\phi'' \le 0$. The case $t = 1$ is immediate: $F_1 = S - S^2 + 2E$ gives $\mathcal{D}_1 = 2$, contradicting stationarity.

3 The finite coefficient comparison

For three nonnegative variables $r = (r_1, r_2, r_3)$, let

$$G_r(w) = \prod_{i=1}^{3} (1 - r_i w)^{-1} = \sum_{j \ge 0} h_j(r) w^j, \qquad h_j = 0 \quad (j < 0).$$

Thus h_j is the complete homogeneous polynomial of degree j. We use

$$h_j(r)^2 - h_{j-1}(r) h_{j+1}(r) = h_j(r_1 r_2, r_1 r_3, r_2 r_3) \ge 0 \qquad (j \ge 0). \tag{3.1}$$

For distinct positive variables, this identity follows by substituting $h_j(r) = \sum_i r_i^{j+2} / \prod_{k \ne i}(r_i - r_k)$; continuity gives the general case. If $r \ne 0$, then every $h_j(r) > 0$ for $j \ge 0$, and h_j / h_{j-1} is nonincreasing for $j \ge 1$, strictly decreasing when all three variables are positive.

Lemma 3.1. *If $r \in [0,1]^3$ and $\sigma = 3 - \sum_i r_i > 0$, then*

$$b_j := h_j(r) - h_{j-1}(r) \le C_\sigma := \max\{1, \sigma^{-1}\} \qquad (j \ge 0). \tag{3.2}$$

Proof. The case $j = 0$ is immediate. Fix $N \ge 1$. If $b_N \le 0$, there is nothing to prove, so suppose $b_N > 0$. This excludes $r = 0$. By (3.1),

$$\frac{h_j}{h_{j-1}} \ge \frac{h_N}{h_{N-1}} > 1 \qquad (1 \le j \le N),$$

so $b_j > 0$ for $0 \le j \le N$.

The regime $\sigma \ge 2$ is incompatible with $b_N > 0$. Indeed, every monomial in h_N occurs at least once in $(r_1 + r_2 + r_3)h_{N-1}$, so

$$h_N \le (r_1 + r_2 + r_3)h_{N-1} \le h_{N-1} \qquad \text{if } \sigma \ge 2.$$

We may therefore assume $0 < \sigma < 2$.

For $N \ge 2$, differentiation of the generating function gives

$$(\partial_{r_i} - \partial_{r_k})b_N = (r_i - r_k)[w^{N-2}]\frac{(1-w)G_r(w)}{(1-r_i w)(1-r_k w)}.$$

The coefficient on the right is

$$\sum_{\nu=0}^{N-2} h_\nu(r_i, r_k) b_{N-2-\nu}, \qquad h_\nu(r_i, r_k) = \sum_{a+b=\nu} r_i^a r_k^b,$$

which is positive whenever $b_N > 0$. Thus the derivative is nonnegative if $r_i \ge r_k$. For $N = 1$, $b_1 = \sum_i r_i - 1$ is constant at fixed sum.

Choose two coordinates in $(0,1)$ with $r_i \ge r_k$, and transfer mass from the smaller to the larger:

$$r_i(s) = r_i + s, \qquad r_k(s) = r_k - s, \qquad 0 \le s \le \min\{1 - r_i, r_k\}.$$

The ordering is preserved since $r_i(s) - r_k(s) = r_i - r_k + 2s \ge 0$. Writing $g(s) = b_N(r(s))$, we have $g'(s) \ge 0$ whenever $g(s) > 0$. If g had a first zero s_*, it would be nondecreasing on $[0, s_*)$, and continuity would give $g(s_*) \ge g(0) > 0$, a contradiction. Hence positivity persists and b_N does not decrease throughout the transfer.

Keep boundary coordinates fixed and repeat. Each transfer puts at least one additional coordinate at 0 or 1, so at most two transfers leave at most one coordinate in $(0,1)$. At the fixed sum, the resulting vector is, up to permutation,

$$(1, 1, 1-\sigma) \quad (0 < \sigma \le 1), \qquad (1, 2-\sigma, 0) \quad (1 \le \sigma < 2).$$

The corresponding coefficients of $(1-w)G_r(w)$ are

$$\sum_{k=0}^{N} (1-\sigma)^k \le \sigma^{-1}, \qquad (2-\sigma)^N \le 1,$$

respectively. Since the transfers do not decrease the original positive coefficient, these endpoint bounds prove (3.2). The formulas also hold at $\sigma = 1$ by continuity. □

Lemma 3.2. *Let $\beta > q_1 > q_2 > q_3 > 0$, $\alpha > 0$, and $n \geq 0$ be an integer. Write $G = G_q$, $k_j = h_j(q) - \alpha h_{j-1}(q)$, and suppose $k_n > 0$, $k_{n+1} \geq \beta k_n$. Then*

$$\begin{aligned} J &:= [w^{n-1}](1-\alpha w)(1-\beta w)^2 G(w)^2 \\ &< \max\left\{\frac{1}{\beta}, \frac{1}{3\beta - \sum_i q_i}\right\} k_n, \end{aligned} \tag{3.3}$$

where negative-index coefficients are zero.

Proof. For $n = 0$, $J = 0$, $k_0 = 1$, and the right-hand side is positive since $3\beta - \sum_i q_i > 0$. Let $n \geq 1$, abbreviate $h_j = h_j(q)$, and put

$$B_j = h_j - \beta h_{j-1}, \qquad a_j = k_j - \beta k_{j-1} = B_j - \alpha B_{j-1}.$$

By (3.1),

$$\frac{h_{n+1}}{h_n} - \frac{k_{n+1}}{k_n} = \frac{\alpha(h_n^2 - h_{n-1}h_{n+1})}{h_n k_n} > 0.$$

Thus $h_{n+1}/h_n > \beta$, and the decreasing h-ratios give $B_j > 0$ for $0 \leq j \leq n+1$.

The partial-fraction representation is

$$B_j = \sum_{i=1}^{3} c_i q_i^j, \qquad c_i = \frac{q_i(q_i - \beta)}{\prod_{k \neq i}(q_i - q_k)},$$

with $c_1 < 0$, $c_2 > 0$, $c_3 < 0$. Define the exponential interpolation

$$g(s) = c_1(q_1/q_2)^s + c_2 + c_3(q_3/q_2)^s, \qquad g(j) = B_j/q_2^j.$$

We have $g'' < 0$. Since $g(0) > 0$ and $g(n+1) > 0$, concavity makes g positive on $[0, n+1]$. Therefore $(\log g)'' = (g''g - (g')^2)/g^2 < 0$, and the ratios B_j/B_{j-1} strictly decrease for $1 \leq j \leq n+1$. The hypothesis gives

$$B_{n+1} - \alpha B_n = a_{n+1} = k_{n+1} - \beta k_n \geq 0.$$

Consequently $B_j/B_{j-1} > B_{n+1}/B_n \geq \alpha$ for $1 \leq j \leq n$, so $a_j > 0$ for $0 \leq j \leq n$, with $a_0 = 1$.

Let $C = \max\{1, \beta/(3\beta - \sum_i q_i)\}$. Applying Lemma 3.1 to $r_i = q_i/\beta$ gives $B_j \leq C\beta^j$ for every $j \geq 0$. Hence

$$J = \sum_{j=0}^{n-1} a_j B_{n-1-j} \leq C \sum_{j=0}^{n-1} a_j \beta^{n-1-j} = C k_{n-1} < \frac{C}{\beta} k_n.$$

The equality telescopes using $a_j = k_j - \beta k_{j-1}$ and $k_{-1} = 0$; the last inequality is $a_n > 0$. □

4 Completion of the two-size proof

Proof of Theorem 1.1. By the three-coordinate reduction, it remains to prove (2.6) for $t \geq 2$. Set $n = t - 2$, $\alpha = t/(t+1)$, $\beta = 1 + m$, and $q = (1-x, 1-y, 1-z)$. In the notation of Lemma 3.2,

$$\begin{aligned} R_t &= (t+1)k_n, \\ \mathcal{D}_t &= (t+1)(k_{n+1} - \beta k_n), \\ \mathcal{C}_t &= (t+1)(\mathcal{K} k_n - J). \end{aligned} \tag{4.1}$$

To verify these identities, denote the second divided difference by

$$\psi[x,y,z] = \sum_{u\in\{x,y,z\}} \frac{\psi(u)}{W'(u)}.$$

Root differentiation gives $(F_t)_P = f_t'[x,y,z]$ and $(F_t)_E = -(uf_t'(u))[x,y,z]$. For integer $j \geq 0$, the second divided difference of $(1-u)^j$ at x,y,z is $h_{j-2}(q)$, by partial fractions. Since

$$f_t'(u) = (t+1)(1-u)^t - t(1-u)^{t-1},$$

we obtain $(F_t)_P = (t+1)k_n$ and $(F_t)_E = (t+1)(k_{n+1}-k_n)$, which give the first two identities in (4.1). This is where integrality of t enters: the powers are polynomials, and $n = t-2$ is a nonnegative integer indexing the coefficient comparison.

For the third identity, observe that

$$G_q(w)^{-1} = 1 - (3-S)w + (3-2S+E)w^2 - (1-S+E-P)w^3.$$

With m frozen at the point of differentiation,

$$(\partial_E - m\partial_P)G_q = -w^2(1-\beta w)G_q^2.$$

Along the entropy curve, $\frac{d}{dE}G_q = (\partial_E - m\partial_P)G_q$ and $\beta' = m' = -\mathcal{K}$. Differentiating

$$\frac{\mathcal{D}_t}{t+1} = [w^{n+1}](1-\alpha w)(1-\beta w)G_q$$

therefore gives

$$\begin{aligned}\frac{\mathcal{C}_t}{t+1} &= [w^{n+1}](1-\alpha w)\big(\mathcal{K}wG_q - w^2(1-\beta w)^2G_q^2\big)\\ &= \mathcal{K}k_n - J.\end{aligned}$$

Here $\mathcal{C}_t = d\mathcal{D}_t/dE$ by (2.5); the $\mathcal{K}$ term accounts for the variation of m along the curve.

At a stationary triple, $0 < x < y < z < 1$ gives $\beta > q_1 > q_2 > q_3 > 0$. Also $R_t > 0$ and $\mathcal{D}_t = 0$ give $k_n > 0$ and $k_{n+1} = \beta k_n$. Thus all hypotheses of Lemma 3.2 hold. Since $S \leq 1$, it yields

$$\frac{J}{k_n} < \max\left\{\frac{1}{1+m}, \frac{1}{S+3m}\right\} \leq \frac{1}{S+m}.$$

Lemma 2.1 and (4.1) now imply

$$\frac{\mathcal{C}_t}{R_t} = \mathcal{K} - \frac{J}{k_n} > \frac{1}{15(S+m)} > 0. \tag{4.2}$$

This proves (2.6), excluding every local maximum with three distinct positive coordinate values. The same calculation also gives the quantitative bound $\mathcal{C}_t > R_t/(20S)$, since $m \leq S/3$.

These perturbations change only three coordinates and tend to zero in ℓ^1, so the argument applies equally to countably infinite alphabets. Each fixed positive size can occur only finitely often in a probability vector. Hence at most two positive sizes imply finite support.

For completeness, we give the compactness argument of [2, Lemmas 1–2 and Theorem 2]. On a countably infinite alphabet, identify I with $\mathbb{N}$. Both H and Φ_t are symmetric and are unaffected by insertion or deletion of zero coordinates. It therefore suffices to take a maximizing sequence whose coordinates are nonincreasing. For such a probability vector, $p_i \leq 1/i$, and for $N \geq 1$,

$$h \geq H(p) \geq \sum_{i>N} p_i \log\frac{1}{p_i} \geq \log(N+1)\sum_{i>N} p_i.$$

Choose a coordinatewise convergent subsequence $p^{(k)} \to p$ by a diagonal argument. The uniform tail bound prevents loss of mass: passing to the limit in finite partial sums and then letting $N \to \infty$ gives $\sum_i p_i = 1$. The same tail bound, together with convergence on each finite initial segment, gives $\|p^{(k)} - p\|_1 \to 0$. Since $-u \log u \geq 0$ on $[0,1]$, Fatou's lemma gives

$$H(p) \leq \liminf_{k\to\infty} H(p^{(k)}) \leq h.$$

Finally, $f_t \in C^1[0,1]$ and $f_t(0) = 0$, so

$$|\Phi_t(p) - \Phi_t(q)| \leq L_t \|p - q\|_1, \qquad L_t = \max_{0\leq u\leq 1} |f_t'(u)| < \infty.$$

Thus the maximum is attained. On a finite alphabet, attainment follows directly from compactness of the feasible set and continuity of the objective.

For entropy saturation, use the atom-splitting argument of [2, Proposition 1]. Let I be countably infinite and let p be a local maximizer with $H(p) < h$. Its support is finite, so an unused coordinate is available. Replace a positive atom a by $a-\varepsilon$ and put ε in that unused coordinate, where $0 < \varepsilon < a$. The new vector p^ε satisfies

$$\|p^\varepsilon - p\|_1 = 2\varepsilon, \qquad H(p^\varepsilon) - H(p) = a\log a - (a-\varepsilon)\log(a-\varepsilon) - \varepsilon \log \varepsilon \longrightarrow 0.$$

The entropy increment is positive, and

$$f_t(a-\varepsilon) + f_t(\varepsilon) > (a-\varepsilon)(1-a)^t + \varepsilon(1-a)^t = f_t(a).$$

For sufficiently small ε, this is a feasible improvement arbitrarily close to p, contradicting local maximality. Hence every local maximizer satisfies $H(p) = h$. □

5 One exceptional atom and a finite classification

5.1 The multiplicity restriction

The following elementary criterion isolates the additional second-variation argument. It uses only perturbations of four coordinates.

Lemma 5.1 (Repeated-size criterion)**.** *Let $f \in C[0,1] \cap C^2(0,1)$, and let $J \subset (0,1)$ be an interval containing every u for which $f''(u) \leq 0$. Suppose that $g(u) = uf''(u)$ satisfies*

$$g(v) < \max\{g(x), g(y)\} \qquad (x < v < y, \quad x, y \in J). \tag{5.1}$$

At a finite-dimensional local maximum of $\sum_i f(p_i)$ under normalization and an upper Shannon-entropy constraint, at most one distinct positive coordinate value can occur more than once. The same conclusion holds on a countable alphabet whenever the objective is well-defined and the point is locally maximal under finite-coordinate perturbations.

Proof. If a value u occurs twice, replacing these copies by $u+\varepsilon$ and $u-\varepsilon$ preserves mass and strictly decreases entropy. Local maximality and Taylor expansion therefore imply $f''(u) \leq 0$.

Suppose two distinct values $x < y$ both occur at least twice, and vary just (x, x, y, y). On these four coordinates, the gradients of mass and entropy are independent. The fixed-mass, fixed-entropy set through the point is therefore a smooth manifold, whether or not the entropy inequality is active. Equality-constrained first-order conditions give constants λ, μ such that

$$f'(u) = \lambda + \mu(-\log u - 1), \qquad u \in \{x, y\}.$$

In particular,

$$-\mu = \frac{f'(y) - f'(x)}{\log(y/x)} = \frac{\int_x^y g(u)\,\dfrac{\mathrm{d}u}{u}}{\log(y/x)}. \tag{5.2}$$

A pair-difference direction within either repeated block is tangent to both constraints. The second-order condition on this manifold is

$$f''(u) + \frac{\mu}{u} \le 0, \qquad u \in \{x, y\}.$$

Thus $g(x), g(y) \le -\mu$. But $x, y \in J$, and (5.1) makes the weighted average in (5.2) strictly smaller than $\max\{g(x), g(y)\}$, a contradiction. All perturbations used here have finite support, which also proves the countable-alphabet assertion. □

Theorem 5.2 (One exceptional atom)**.** *Under the hypotheses of Theorem 1.1, every nonuniform local maximizer has, up to permutation and zero coordinates, the form*

$$p_m(z) = \left(z, \underbrace{q, \ldots, q}_{m \text{ copies}}\right), \qquad q = \frac{1-z}{m}, \quad m \in \mathbb{N}, \quad 0 < z < 1. \tag{5.3}$$

Thus either $z < q$ (one light atom) or $z > q$ (one heavy atom).

Proof. For $t = 1$, $uf_t''(u) = -2u$ is strictly decreasing. For $t > 1$, the set where $f_t'' \le 0$ is $J = (0, 2/(t+1)]$, and

$$\frac{\mathrm{d}}{\mathrm{d}u}\big(uf_t''(u)\big) = -t(1-u)^{t-3}\big(t(t+1)u^2 - 4tu + 2\big).$$

The quadratic roots are

$$r_\pm = \frac{2t \pm \sqrt{2t(t-1)}}{t(t+1)}.$$

They satisfy $0 < r_- < 2/(t+1) < r_+$. Hence $uf_t''(u)$ strictly decreases and then strictly increases on J, so it satisfies (5.1). Lemma 5.1 shows that two distinct repeated sizes are impossible. Theorem 1.1 supplies finite support and at most two sizes, so at least one of their multiplicities is one. □

A uniform law on $m+1$ points is the limiting case $z = q = 1/(m+1)$. A nonuniform binary law has both a one-light and a one-heavy description, which differ only by permutation. For support size at least three the two nonuniform descriptions are distinct.

5.2 The entropy-matched candidates

For the rest of this section the alphabet is countably infinite. Define

$$B_t(h) = \max_{p \in \mathcal{P}_h(\mathbb{N})} \Phi_t(p).$$

If $h = 0$, only point masses are feasible, and $B_t(0) = 0$. Fix $h > 0$ and put $k = \lfloor e^h \rfloor$. For $m \ge 1$, set

$$\begin{aligned} E_m(z) &= -z\log z - (1-z)\log\frac{1-z}{m}, \\ V_{m,s}(z) &= z(1-z)^s + (1-z)\left(1 - \frac{1-z}{m}\right)^s. \end{aligned} \tag{5.4}$$

The functions extend continuously to the entropy endpoints. Since

$$E'_m(z) = \log \frac{1-z}{mz}, \tag{5.5}$$

E_m increases from $\log m$ to $\log(m+1)$ on $[0, 1/(m+1)]$ and decreases from $\log(m+1)$ to 0 on $[1/(m+1), 1]$.

There is a unique $a \in [0, 1/(k+1))$ with $E_k(a) = h$. Define

$$p_{\mathrm{L}} = p_k(a), \qquad y = \frac{1-a}{k}, \qquad \ell(s) = V_{k,s}(a). \tag{5.6}$$

If $h = \log k$, then $a = 0$, and deleting its zero coordinate gives the uniform law on k points. Otherwise p_{L} is one-light and has $k+1$ positive coordinates.

For every integer $m \geq k$, there is a unique $z_m \in (1/(m+1), 1)$ with $E_m(z_m) = h$. Define

$$q_m = \frac{1-z_m}{m}, \qquad p_m^{\mathrm{H}} = p_m(z_m), \qquad b_m(s) = V_{m,s}(z_m). \tag{5.7}$$

The dependence on the fixed entropy h is suppressed in this notation. Theorem 5.2 and entropy saturation show that every global maximizer is among p_{L} and the p_m^{H}, $m \geq k$. When $0 < h < \log 2$, p_1^{H} is the binary duplicate of p_{L}. The real parameter s in $\ell(s)$ and $b_m(s)$ will also be useful for comparing these candidate values; the structural theorems continue to require an integer sample size.

5.3 A uniform cutoff for the heavy family

Lemma 5.3 (Heavy-tail cutoff)**.** *Fix $h > 0$ and a real exponent $s \geq 1$. Interpolate the heavy formulas by allowing $m > e^h - 1$ to be real and taking the unique root $E_m(z_m) = h$ with $z_m > 1/(m+1)$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}m} b_m(s) < 0 \qquad \textit{whenever } m \geq sh + 1. \tag{5.8}$$

Proof. This is an interpolation of the formulas, not a probability vector with nonintegral multiplicities. Write $z = z_m$, $q = (1-z)/m$, and $\sigma = \log(z/q) > 0$. The entropy equation is

$$h = -\log z + (1-z)\sigma. \tag{5.9}$$

Implicit differentiation gives

$$\frac{\mathrm{d}z}{\mathrm{d}m} = \frac{q}{\sigma}, \qquad \frac{\mathrm{d}q}{\mathrm{d}m} = -\frac{q(1+\sigma)}{m\sigma}. \tag{5.10}$$

With $f_s(u) = u(1-u)^s$, it follows that

$$\frac{\mathrm{d}b_m(s)}{\mathrm{d}m} = \frac{q}{\sigma}\Psi_s(z,q), \qquad \Psi_s(z,q) = f_s'(z) - f_s'(q) + sq\sigma(1-q)^{s-1}. \tag{5.11}$$

If $z \leq 1/s$, hold z fixed and put

$$A_z(u) = f_s'(u) - su\log(z/u)(1-u)^{s-1}.$$

Direct differentiation yields

$$A_z'(u) = s(1 + \log(z/u))(1-u)^{s-2}(su - 1).$$

This is strictly negative for $q < u < z$. Hence $\Psi_s(z,q) = A_z(z) - A_z(q) < 0$.

If $z > 1/s$, then $f_s'(z) < 0$, and

$$\Psi_s(z,q) = f_s'(z) - (1-q)^{s-1}\big(1 - q - sq(1+\sigma)\big).$$

By (5.9) and $\log z < z - 1$,

$$(1-z)(1+\sigma) = h + \log z + 1 - z < h.$$

Consequently $sq(1+\sigma) < sh/m$ and $q < 1/m$. If $m \geq sh + 1$, then

$$1 - q - sq(1+\sigma) > 1 - \frac{sh+1}{m} \geq 0.$$

Thus $\Psi_s(z,q) < 0$ in this case as well. □

Theorem 5.4 (Exact finite classification)**.** *For integer $t \geq 1$ and $h > 0$, put*

$$M = M(t,h) = \max\{k, \lceil th + 1 \rceil\}, \qquad k = \lfloor e^h \rfloor. \tag{5.12}$$

Then

$$B_t(h) = \max\{\ell(t), b_k(t), b_{k+1}(t), \ldots, b_M(t)\}. \tag{5.13}$$

Up to permutation and zero coordinates, the global maximizers are exactly the candidates in this list attaining the displayed maximum. In particular, every global maximizer has at most $M + 1$ positive coordinates.

Proof. The candidate list before truncation is exhaustive by Theorem 5.2 and entropy saturation. Every candidate is feasible. Lemma 5.3 gives $b_j(t) < b_M(t)$ for every integer $j > M$, which proves the formula and its equality statement. □

Thus the atom sizes are specified by unique monotone entropy roots, and the remaining comparison is finite. All ties are retained. Appendix B gives crossing intervals specifying when each candidate wins; Appendix C treats a fixed finite alphabet, where the entropy constraint can be slack.

5.4 Sample sizes one and two

Corollary 5.5. *For $t = 1$ and $t = 2$, the unique global maximizer on a countably infinite alphabet, up to permutation and zero coordinates, is p_{L}. Equivalently, $B_t(h) = \ell(t)$ for $h > 0$.*

The $t = 1$ case recovers the maximizing shape in [2, Proposition 3].

Proof. It suffices to exclude every distinct heavy candidate with $m \geq 2$. Write $z = rq$, $q = 1/(m+r)$, and $r > 1$. The fixed-mass, fixed-entropy second-order condition in a pair-difference direction among the repeated q-atoms is

$$q f_t''(q) + \mu \leq 0, \qquad \mu = \frac{f_t'(q) - f_t'(z)}{\log r}, \tag{5.14}$$

by the same equality-manifold argument as in Lemma 5.1. For $t = 1$, the left side is

$$2q\left(\frac{r-1}{\log r} - 1\right) > 0.$$

For $t = 2$, it is

$$\frac{N_m(r)}{(m+r)^2 \log r}, \qquad N_m(r) = (r-1)(4m+r-3) - (4m+4r-6)\log r.$$

Now

$$\begin{aligned} N_m(r) &= N_2(r) + 4(m-2)(r-1-\log r), \\ N_2(r) &= (r-1)(r+5) - (4r+2)\log r, \\ N_2'(r) &= 2\left(r - \frac{1}{r} - 2\log r\right) > 0 \qquad (r > 1). \end{aligned}$$

The last inequality follows by differentiating its parenthesis, whose derivative is $(r-1)^2/r^2$ and whose value at 1 is zero. Since $N_2(1) = 0$, we have $N_m(r) > 0$, again contradicting (5.14). The only possible heavy candidate with $m = 1$ is the binary duplicate of p_{L}. Attainment now proves the claim, including uniqueness. □

6 Sharp asymptotics and occupancy consequences

Throughout this section the entropy $h > 0$ is fixed and the alphabet is countably infinite. Berend, Kontorovich, and Zagdanski proved $B_t(h) \sim h/\log t$ by matching first-order upper and lower bounds [2, Theorems 3–4]. The classification yields the following refinement and the scales of all optimizing atoms.

Theorem 6.1 (Fixed-entropy asymptotics)**.** *As integer $t \to \infty$,*

$$\frac{h}{B_t(h)} = \log t + \log\log t + 2 + O_h\left(\frac{\log\log t}{\log t}\right). \tag{6.1}$$

Every global maximizer is eventually one-heavy. Writing any such maximizer as

$$p_t = \left(1 - L_t, \underbrace{q_t, \ldots, q_t}_{m_t \text{ copies}}\right), \qquad m_t q_t = L_t,$$

we have, uniformly over the choice of a maximizer,

$$L_t \sim \frac{h}{\log t}, \qquad q_t \sim \frac{1}{t\log t}, \qquad m_t \sim ht. \tag{6.2}$$

Proof. Put $T = \log t$. For a heavy vector with total light mass $L = mq$, the exact entropy and objective are

$$\begin{aligned} h &= -(1-L)\log(1-L) - L\log q, \\ \Phi_t(p) &= (1-L)L^t + L(1-q)^t. \end{aligned} \tag{6.3}$$

It is convenient to set

$$A(L) = -\frac{1-L}{L}\log(1-L) = 1 - \frac{L}{2} + O(L^2) \qquad (L \downarrow 0). \tag{6.4}$$

All constants below may depend on h, but not on the maximizing vector.

A lower bound with integer multiplicity. Choose $m = \lceil ht \rceil$ and solve the heavy entropy equation for L. For sufficiently large t, $h < \log(m+1)$, so this root exists. Its entropy can also be written as $h = H_2(L) + L \log m$, where $H_2(L) = -L \log L - (1-L)\log(1-L) \geq 0$. Thus $L \leq h/\log m = O_h(T^{-1})$. With $d = h/L$, the entropy equation becomes

$$d = \log(m/h) + \log d + A(L) = T + \log d + 1 + O_h(T^{-1}).$$

Since $d \geq \log m$ and $d - \log d$ is increasing for $d > 1$, this gives

$$d = T + \log T + 1 + O_h\left(\frac{\log T}{T}\right). \tag{6.5}$$

For example, substituting $T + \log T + 1$ in $d - \log d$ gives an error $O(\log T/T)$, and the derivative is bounded away from zero in this range. Moreover,

$$tq = \frac{th}{md} = \frac{1}{d} + O_h\left(\frac{1}{tT}\right), \qquad -t\log(1-q) = \frac{1}{d} + O_h\left(\frac{1}{tT}\right).$$

Since $B_t(h) \geq L(1-q)^t$, it follows that

$$\begin{aligned} \frac{h}{B_t(h)} &\leq d \exp\left(\frac{1}{d} + O_h\left(\frac{1}{tT}\right)\right) \\ &= T + \log T + 2 + O_h\left(\frac{\log T}{T}\right). \end{aligned} \tag{6.6}$$

In particular $B_t(h) \geq c_h/T$ for some $c_h > 0$ and all sufficiently large t.

A uniform upper bound. At fixed h, the light or uniform candidate has a fixed finite positive support, so its objective decays exponentially in t. The preceding lower bound and Theorem 5.4 therefore imply that every maximizer is eventually heavy.

Consider any such maximizer, with largest atom $z = 1 - L$. Since $H(p) \geq -\log z$, we have $z \geq e^{-h}$ and $L \leq \rho_h := 1 - e^{-h} < 1$. The heavy atom contributes at most ρ_h^t. Put $u = tq$ and $d = h/L$. From

$$L(1-q)^t = B_t(h) - (1-L)L^t \geq \frac{c_h}{2T}$$

for large t, and $(1-q)^t \leq e^{-u}$, we get $u \leq \log(2T/c_h) = O_h(\log T)$. The entropy identity gives $h \geq L(T - \log u)$, so $L = O_h(T^{-1})$. Therefore

$$d = T - \log u + 1 + O_h(T^{-1}). \tag{6.7}$$

Let $R = Le^{-u}$. We have $R \geq c_h/(2T)$ and $B_t(h) \leq R + \rho_h^t$. Consequently

$$\frac{h}{B_t(h)} \geq de^u - O_h(T^2\rho_h^t) = de^u - o_h(T^{-1}). \tag{6.8}$$

This explicitly controls the error from discarding the heavy atom.

The bounds (6.6) and (6.8) force $u < 1$ for all sufficiently large t. Indeed, if $u \geq 1$, then (6.7) and $u \leq \log(2T/c_h)$ imply

$$de^u \geq e(T - O_h(\log\log T)),$$

contradicting (6.6). For $u < 1$, equation (6.7) gives $d \geq T$ for large t. Using $e^u \geq 1 + u$ and putting $v = Tu$, we obtain

$$\begin{aligned} de^u &\geq d + Tu \\ &\geq T + \log T + 1 + v - \log v - O_h(T^{-1}) \\ &\geq T + \log T + 2 - O_h(T^{-1}), \end{aligned} \tag{6.9}$$

because $v - \log v \geq 1$. Combining this with (6.6) and (6.8) proves (6.1).

The optimizing scales. The same inequalities give

$$0 \le v - \log v - 1 = O_h\Big(\frac{\log T}{T}\Big).$$

The function on the left has its unique zero at $v = 1$ and tends to infinity at both ends of $(0, \infty)$. Hence $v \to 1$, uniformly over all maximizers. Thus $tq = u \sim 1/T$. Equation (6.7) now gives $d \sim T$, so $L = h/d \sim h/T$ and $m = L/q \sim ht$, as required. □

Corollary 6.2 (Expectation-level sample complexity)**.** *For fixed $h > 0$, define*

$$N_h(\varepsilon) = \min\{t \in \mathbb{N} : B_t(h) \le \varepsilon\}.$$

Then

$$N_h(\varepsilon) \sim \frac{\varepsilon}{e^2 h} \exp\Big(\frac{h}{\varepsilon}\Big) \qquad (\varepsilon \downarrow 0). \tag{6.10}$$

Proof. The function $B_t(h)$ is strictly decreasing in integer t: at an optimizer for $t+1$, positive entropy implies that every positive atom is below one, so $\Phi_t(p) > \Phi_{t+1}(p)$. Also $B_t(h) \to 0$ by Theorem 6.1. Put $A = h/\varepsilon$ and $S = A - \log A - 2$. For any fixed $\eta > 0$, apply (6.1) at the integers nearest $e^{S-\eta}$ and $e^{S+\eta}$. Their reciprocal objective values are $A - \eta + o(1)$ and $A + \eta + o(1)$, respectively. They therefore bracket $N_h(\varepsilon)$. It follows that $\log N_h(\varepsilon) = S + o(1)$, proving the assertion. □

Corollary 6.3 (Singletons, discovery, and coverage)**.** *Let K_n be the number of distinct symbols in n independent samples, let $K_{n,1}$ be the number occurring exactly once, and let C_n be the probability mass of the observed symbols. Then*

$$\begin{aligned} \max_{H(p)\le h} \mathbb{E}_p K_{n,1} &= nB_{n-1}(h) \quad (n \ge 2), \\ \max_{H(p)\le h} \mathbb{E}_p(K_{t+1} - K_t) &= B_t(h) \qquad (t \ge 1), \\ \min_{H(p)\le h} \mathbb{E}_p C_t &= 1 - B_t(h) \quad (t \ge 1). \end{aligned} \tag{6.11}$$

The optimizers are exactly those in Theorem 5.4, with the corresponding sample-size index.

Proof. Linearity of expectation gives

$$\mathbb{E}_p K_{n,1} = n\sum_i p_i(1-p_i)^{n-1}, \qquad \mathbb{E}_p(K_{t+1} - K_t) = \sum_i p_i(1-p_i)^t,$$

and $\mathbb{E}_p C_t = 1 - \Phi_t(p)$. These identities prove every assertion. □

In particular the singleton estimator $K_{n,1}/n$ has maximal expectation $B_{n-1}(h)$. The optimization in (6.11) concerns expectations, rather than estimation error or tail probabilities.

A The entropy-curvature estimate

Proof of Lemma 2.1. For a uniform triple $x = y = z = a$, direct integration gives

$$I_0 = \frac{1}{2a^2}, \qquad I_1 = \frac{1}{2a}, \qquad m = a, \qquad \mathcal{K} = \frac{4}{15a} = \frac{16}{15(S+m)}.$$

This proves the lemma, with equality, in the uniform case. We first establish a kernel estimate and its mixture form, and then apply them to a nonuniform triple.

For $a > 0$, set $\kappa_a(s) = 2a^2/(s+a)^3$, a probability density on $[0,\infty)$ of mean a, and $J_*(a) = \int_0^\infty (s-1)^2 \kappa_1(s)\kappa_a(s)\,\mathrm{d}s$. The kernel estimate is

$$J_*(a) \geq \frac{2(9-a)}{15(a+1)} = \frac{13}{15} - \frac{a}{3} + \frac{(a-1)^2}{3(a+1)}. \tag{A.1}$$

Here is an explicit positive remainder, with $b_v = 1 + v(a-1)$:

$$J_*(a) - \frac{2(9-a)}{15(a+1)} = \frac{2(a-1)^2}{5(a+1)} \int_0^1 \frac{v(1-v)^5(2b_v^3 + 12b_v^2 - 10b_v + 13)}{b_v^4}\,\mathrm{d}v.$$

The cubic is positive since it equals $2b_v^3 + 12(b_v - 5/12)^2 + 131/12$ and $b_v > 0$. For verification, direct integration gives, for $a \neq 1$,

$$J_*(a) = \frac{4a^2(a^2+10a+13)\log a - 2(a-1)(7a^3+33a^2+9a-1)}{(a-1)^5}.$$

For $M(a) = \frac{15}{2}(a-1)^5(a+1)(J_*(a) - 2(9-a)/(15(a+1)))$, with its smooth extension at $a = 1$, one has

$$M^{(j)}(1) = 0 \quad (0 \leq j \leq 5), \qquad M^{(6)}(a) = \frac{360(a-1)(2a^3+12a^2-10a+13)}{a^4}.$$

Taylor's integral remainder gives the displayed identity on both sides of 1; directly $J_*(1) = 8/15$.

For a probability law λ supported in a compact subset of $(0,\infty)$, put $m_\lambda = \int v\,\mathrm{d}\lambda(v)$ and $G(s) = \int \kappa_v(s)\,\mathrm{d}\lambda(v)$. Scaling (A.1) gives

$$\begin{aligned}\int_0^\infty (s-m_\lambda)^2 \kappa_{m_\lambda}(s)\kappa_v(s)\,\mathrm{d}s &= m_\lambda J_*(v/m_\lambda) \\ &\geq \frac{13m_\lambda}{15} - \frac{v}{3} + \frac{(v-m_\lambda)^2}{3(v+m_\lambda)}.\end{aligned}$$

Averaging and using $\int v\,\mathrm{d}\lambda(v) = m_\lambda$ yields

$$\int_0^\infty (s-m_\lambda)^2 \kappa_{m_\lambda}(s)G(s)\,\mathrm{d}s \geq \frac{8m_\lambda}{15} + \frac{1}{3}\int \frac{(v-m_\lambda)^2}{v+m_\lambda}\,\mathrm{d}\lambda(v).$$

Expand the nonnegative integral $\int_0^\infty (s-m_\lambda)^2(G - \kappa_{m_\lambda})^2\,\mathrm{d}s$ and use $\int_0^\infty (s-m_\lambda)^2\kappa_{m_\lambda}(s)^2\,\mathrm{d}s = 8m_\lambda/15$. It follows that

$$\int_0^\infty (s-m_\lambda)^2 G(s)^2\,\mathrm{d}s \geq \frac{8m_\lambda}{15} + \frac{2}{3}\int \frac{(v-m_\lambda)^2}{v+m_\lambda}\,\mathrm{d}\lambda(v). \tag{A.2}$$

Now let (x,y,z) be nonuniform. Let (U_1,U_2) have density 2 on $\{(u_1,u_2) : u_1,u_2 \geq 0,\ u_1+u_2 \leq 1\}$, set $U_3 = 1-U_1-U_2$, and put $V = xU_1 + yU_2 + zU_3$. Thus (U_1,U_2,U_3) is uniform on the probability simplex and $\mathbb{E}V = S/3$. Two elementary integrations over the simplex give

$$D(s)^{-1} = \mathbb{E}(s+V)^{-3}, \qquad I_0 = \tfrac{1}{2}\mathbb{E}V^{-2}, \quad I_1 = \tfrac{1}{2}\mathbb{E}V^{-1}.$$

Tilt its law by $\mathrm{d}\lambda(v) = v^{-2}\,\mathrm{d}\mathbb{P}_V(v)/(2I_0)$. Then

$$G(s) = \frac{1}{I_0 D(s)}, \quad \mathbb{E}_\lambda V = m, \quad b_2 := \mathbb{E}_\lambda V^2 = \frac{1}{2I_0}, \quad \mathbb{E}_\lambda V^3 = (S/3)b_2.$$

Consequently

$$\mathbb{E}_\lambda[(V-m)^2(V+m)] = (S/3-m)b_2.$$

Since V is positive and nonconstant, the left-hand side is strictly positive, so $0 < m < S/3$. In particular, the following denominators are positive. Cauchy–Schwarz gives

$$\mathbb{E}_\lambda \frac{(V-m)^2}{V+m} \geq \frac{\left(\mathbb{E}_\lambda(V-m)^2\right)^2}{\mathbb{E}_\lambda[(V-m)^2(V+m)]} = \frac{(b_2-m^2)^2}{(S/3-m)b_2}.$$

Substitute this bound into (A.2), where $m_\lambda = m$, and multiply by I_0 to obtain

$$\mathcal{K} \geq \frac{8}{15}I_1 + \frac{(1-2mI_1)^2}{S-3m}.$$

Also $D(s) \leq (s+S/3)^3$ by the arithmetic–geometric mean inequality, so $j := SI_1 \geq 3/2$. Put $r = m/S$, $d = 1-3r > 0$, and $v = j - 3/2 \geq 0$. The identities

$$\frac{8}{15}j + \frac{(1-2rj)^2}{d} - \frac{36-48r}{25} = \frac{(2rv-3d/5)^2}{d} + \frac{8}{15}dv \geq 0,$$

$$(1+r)\frac{36-48r}{25} - \frac{16}{15} = \frac{4(1-3r)(12r+7)}{75} > 0$$

give $\mathcal{K} > 16/(15(S+m))$ for a nonuniform triple. Together with the uniform calculation, this proves the lemma and its equality assertions. □

B Crossing structure of the candidate family

Fix $h > 0$ and retain the notation of Section 5. For real $s \geq 0$, $\ell(s)$ and $b_m(s)$ are exponential sums associated with fixed finite-support distributions. The following comparisons describe their crossings exactly. Their use in classifying the original optimization is restricted to integer sample sizes.

B.1 Two elementary identities

Lemma B.1. *For every discrete probability distribution,*

$$H(p) = \sum_{j=1}^{\infty} \frac{\Phi_j(p)}{j}, \tag{B.1}$$

with equality in $[0,\infty]$. In particular, if two distributions have the same finite entropy, the weighted series of the differences of their Φ_j values converges absolutely and sums to zero.

Proof. For $0 < u \leq 1$, $-\log u = \sum_{j\geq 1}(1-u)^j/j$. Multiplication by p_i and Tonelli's theorem prove (B.1). Absolute convergence of the difference follows by bounding it termwise by the sum of the two nonnegative series. □

Lemma B.2 (Exponential-sum zero counts)**.** *A nonzero sum of r exponentials with distinct positive bases has at most $r-1$ real zeros, counted with multiplicity. If*

$$F(s) = Aa^s - Bb^s - Cc^s + Dd^s, \qquad 0 < a < b < c < d, \quad A,B,C,D > 0,$$

then F has at most two real zeros, counted with multiplicity.

Proof. For the first assertion, divide by one exponential and differentiate. The derivative has $r-1$ exponential terms. Induction and Rolle's theorem, including repeated zeros, prove the claim.

For the second assertion, put $U(s) = F(s)/b^s$. Then

$$\frac{U'(s)}{(d/b)^s} = A\log(a/b)(a/d)^s - C\log(c/b)(c/d)^s + D\log(d/b).$$

The right-hand side has strictly positive derivative, so it has at most one zero, necessarily simple. Hence U' has at most one zero, counted with multiplicity. Rolle's theorem gives at most two for U and F. □

B.2 Heavy–heavy and light–heavy crossings

Theorem B.3 (Heavy–heavy single crossing)**.** *For each pair of integers $k \le i < j$, there is a unique positive crossing $\tau_{ij} > 1$. It is simple, and*

$$\begin{cases} b_j(s) < b_i(s), & 0 < s < \tau_{ij}, \\ b_j(s) > b_i(s), & s > \tau_{ij}. \end{cases} \tag{B.2}$$

Proof. Equations (5.10) give $z_i < z_j$ and $q_j < q_i$. Since $q_i < z_i$, the bases in

$$b_j(s) - b_i(s) = z_j(1-z_j)^s - z_i(1-z_i)^s - (1-z_i)(1-q_i)^s + (1-z_j)(1-q_j)^s$$

are strictly increasing in the displayed order, and their coefficient signs are $+,-,-,+$. Lemma B.2 gives at most two real zeros, counting multiplicity. One is $s=0$, since both distributions have mass one. The difference is positive for all sufficiently large s, because $q_j < q_i$. By Lemma B.1, it must be negative at some positive integer. Hence there is exactly one positive crossing, and the zeros at zero and at that crossing are both simple. The sign orientation follows, and the negative integer implies $\tau_{ij} > 1$. □

Theorem B.4 (Light–heavy crossings)**.** *Fix $m \ge k$ such that p_m^{H} is not the binary duplicate of p_{L}. There are endpoints $2 < \alpha_m < \beta_m \le \infty$ such that*

$$b_m(s) > \ell(s) \quad\Longleftrightarrow\quad \alpha_m < s < \beta_m \qquad (s > 0). \tag{B.3}$$

More precisely, if p_{L} is uniform or $q_m \le a$, then $\beta_m = \infty$ and α_m is the unique positive zero of $b_m - \ell$. If $0 < a < q_m$, then α_m and β_m are its exactly two positive zeros. Every finite positive crossing is simple; outside the closed interval the inequality in (B.3) *is strictly reversed.*

Proof. First note the ordering $q_m < y < z_m$. The first inequality follows from $q_m < 1/(m+1) \le 1/(k+1) < y$. To prove $z_m > y$, keep one y-atom of p_{L} and average its remaining masses over m coordinates, padding with zeros if necessary. This strictly increases entropy unless $k = m = 1$, the excluded binary duplicate. Thus $E_m(y) > h$. Since $y > 1/(m+1)$ and E_m is strictly decreasing on the heavy branch, $z_m > y$ follows. This averaging argument also applies to the uniform case $a = 0$.

Corollary 5.5 gives

$$b_m(1) < \ell(1), \qquad b_m(2) < \ell(2). \tag{B.4}$$

If $a = 0$, the difference is a three-term exponential sum with increasing bases and signs $+,-,+$. It is positive for large s. By Lemma B.2, the zero at $s=0$ and a single simple crossing $\alpha_m > 2$ exhaust its zeros.

Suppose $a > 0$. We have

$$b_m(s) - \ell(s) = z_m(1 - z_m)^s - (1 - a)(1 - y)^s + (1 - z_m)(1 - q_m)^s - a(1 - a)^s.$$

If $q_m < a$, the increasing-base signs are $+, -, -, +$, and the difference is positive for large s. The same two-zero bound gives a unique simple positive crossing $\alpha_m > 2$. If $q_m = a$, the last two terms with base $1 - a$ combine, with coefficient $1 - z_m - a = (m - 1)a > 0$. The strict inequality holds because the only $m = 1$ case is the excluded duplicate. This leaves a three-term sum and gives the same conclusion.

Finally, if $a < q_m$, the increasing-base signs are $+, -, +, -$, so there are at most three real zeros. One is $s = 0$, and the difference is negative for large s. It is also negative at $s = 1, 2$ by (B.4). Since its entropy series sums to zero, it must be positive at some integer $j \geq 3$. There is therefore one crossing between 2 and j, and another after j. These, together with zero, exhaust the permitted zeros and must all be simple. They give exactly the interval and signs stated in the theorem. □

B.3 Finite phase intervals

The crossing results give an alternative exact description of the finite maximum in Theorem 5.4. Every endpoint below is a specified root with a proved crossing orientation.

For a heavy multiplicity $m \geq k$ other than the binary duplicate, put

$$T_m = \tau_{m,m+1}, \qquad R_m = \max\{m + 1, \lceil hT_m + 1 \rceil\}, \tag{B.5}$$

and define the finite extrema

$$\begin{aligned} \underline{s}_m &= \max(\{\alpha_m\} \cup \{\tau_{j,m} : k \leq j < m\}), \\ \overline{s}_m &= \min(\{\beta_m\} \cup \{\tau_{m,j} : m < j \leq R_m\}). \end{aligned} \tag{B.6}$$

The second set contains $\tau_{m,m+1}$, so its minimum is finite even when $\beta_m = \infty$.

Corollary B.5 (Winning intervals)**.** *For integer $t \geq 1$, the heavy distribution p_m^{H} is globally optimal if and only if*

$$\underline{s}_m \leq t \leq \overline{s}_m. \tag{B.7}$$

If the lower endpoint exceeds the upper endpoint, this multiplicity never wins. The light or uniform candidate is globally optimal if and only if

$$t \notin \bigcup_{\substack{k \leq m \leq M(t,h) \\ p_m^{\mathrm{H}} \neq p_{\mathrm{L}}}} (\alpha_m, \beta_m), \tag{B.8}$$

where equality or inequality of distributions is understood up to permutation and zero coordinates. All endpoint ties are included.

Proof. Theorem B.3 gives the lower crossing constraints against smaller heavy multiplicities and the upper constraints against larger ones. Theorem B.4 gives the closed interval $[\alpha_m, \beta_m]$ for comparison with the baseline.

A candidate with multiplicity m loses to $m + 1$ when $t > T_m$. For $1 \leq t \leq T_m$, we have $R_m \geq ht + 1$, so Lemma 5.3 implies $b_j(t) < b_{R_m}(t)$ for every $j > R_m$. It is therefore enough to compare with the finitely many larger multiplicities in (B.6). This proves (B.7). The light criterion follows directly from Theorem 5.4 and the strict sign intervals in (B.3). □

At fixed entropy, heavy winning multiplicities cannot decrease as the integer sample size increases. Indeed, if $i < j$ and $b_j(t_1) \geq b_i(t_1)$, then $t_1 \geq \tau_{ij}$; for every $t_2 > t_1$ we have $b_j(t_2) > b_i(t_2)$. This assertion concerns comparisons as the sample size varies. The heavy sequence in m need not be unimodal at a fixed sample size, as Appendix D shows.

C A fixed finite alphabet

Let $B_{t,N}(h)$ denote the maximum when the alphabet has exactly N available symbols; zero coordinates are permitted. If $N = 1$ or $h = 0$, only point masses can maximize, and the maximum is zero. Suppose now that $N \geq 2$, $h > 0$, and $t \geq 1$ is an integer.

Form the saturated candidate list from Theorem 5.4, retaining only distributions with at most N positive coordinates. In addition, let z run through all roots in $(0, 1)$ of

$$f_t'(z) = f_t'\left(\frac{1-z}{N-1}\right) \tag{C.1}$$

for which $E_{N-1}(z) \leq h$, and add $p_{N-1}(z)$ to the list. Explicitly, (C.1) is the polynomial equation

$$\begin{aligned}(N-1)^t(1-z)^{t-1}(1-(t+1)z)\\ -(N-2+z)^{t-1}\big((t+1)z+N-t-2\big) = 0.\end{aligned} \tag{C.2}$$

This polynomial is nonzero and has degree at most t. The full-support uniform law is included as the root $z = 1/N$ whenever it is feasible.

Theorem C.1. *The maximum $B_{t,N}(h)$ is the largest objective value in the finite list just described. Up to permutation, its maximizing vectors are exactly the candidates attaining that value.*

Proof. Theorems 1.1 and 5.2 give the possible shapes. If a maximizer has $H(p) = h$, it is an entropy-matched candidate whose support fits in N. A heavy candidate with multiplicity above $M(t, h)$ cannot maximize even on this alphabet: the candidate at $M(t, h)$ has smaller support, the same entropy, and a strictly larger objective.

If a maximizer has $H(p) < h$, it has full support. Otherwise the atom-splitting perturbation in the proof of Theorem 1.1 is available on an unused coordinate and strictly improves the objective. On full support the entropy inequality is locally inactive, so ordinary stationarity under normalization gives equality of f_t' at every positive size. This is (C.1) for a nonuniform one-exceptional vector; a uniform vector satisfies it as well.

Equation (C.2) is obtained by clearing the factor $(N-1)^t$. It is not identically zero: the derivative difference in (C.1), extended to $z = 1$, is -2 for $t = 1$ and -1 for $t > 1$. Thus at most t roots are added. Every listed vector is feasible, and the list contains every maximizer; attainment proves the theorem. □

D Certified nonmonotonicity examples

The finite classification need not reduce to a unimodal search in the heavy multiplicity or to a single transition as entropy increases. The following strict inequalities are verified by the accompanying `missing_mass_extremizers_certificates.py`, which uses only integer and rational arithmetic from the Python standard library.

At $h = 19/8$ and $t = 8$, the heavy candidates satisfy

$$\begin{aligned}0.456467562315 < b_{10}(8) < 0.456467562316,\\ 0.455820123956 < b_{11}(8) < 0.455820123957,\\ 0.457052010607 < b_{12}(8) < 0.457052010608.\end{aligned} \tag{D.1}$$

In particular $b_{10}(8) > b_{11}(8) < b_{12}(8)$, so the heavy sequence is not unimodal. The cutoff is $M = 20$. Comparing the light candidate and all heavy candidates $10 \leq m \leq 20$ gives the unique winner $m = 14$, with

$$0.458686581808 < B_8(19/8) < 0.458686581809.$$

Its heavy atom lies in

$$\frac{34393713226485311588012958416 6}{10^{30}} < z_{14} < \frac{34393713226485311588012958416 7}{10^{30}},$$

and every other positive atom is $(1 - z_{14})/14$.

For $t = 3$, the type of the unique maximizer changes from one-light to one-heavy and back to one-light across the three entropy budgets below. In the two numerical columns, each displayed decimal d is a strict lower bound and $d + 10^{-12}$ is a strict upper bound.

h	$\ell(3)$	$\max_{m \geq k} b_m(3)$	M	Optimal type
11/10	0.296480675541	0.294054230107	5	One-light
6/5	0.327755692052	0.333702359978	5	One-heavy
7/5	0.424289708030	0.403758744660	6	One-light

In the middle row the winning heavy multiplicity is $m = 3$.

For completeness, the certificate method is as follows. The script stores a rational bracket (a, b) of width 10^{-30} for each entropy root and verifies the appropriate entropy signs at its endpoints. Logarithms are bounded after reduction to $x = 2^j y$, $1 \leq y < 2$, by setting $v = (y-1)/(y+1) \in [0, 1/3)$ and using

$$0 \leq \log y - 2 \sum_{r=0}^{R-1} \frac{v^{2r+1}}{2r+1} \leq \frac{2v^{2R+1}}{(2R+1)(1-v^2)}. \tag{D.2}$$

The same formula at $v = 1/3$ bounds $\log 2$. The script uses $R = 40$ and reverses the appropriate interval endpoints when multiplying by a negative j. For $z \in (a, b)$, positivity of the weights and bases gives

$$V_{m,t}(z) \geq a(1-b)^t + (1-b)\left(1 - \frac{1-a}{m}\right)^t,$$
$$V_{m,t}(z) \leq b(1-a)^t + (1-a)\left(1 - \frac{1-b}{m}\right)^t.$$

These are rational bounds for integer t. The stored data certify the entropy index k, the cutoff M, every entropy-root bracket, all strict objective comparisons, and the completeness of each candidate list.

AI disclosure

GPT-6 Astra and Claude Opus 5 were used for proof discovery, simplification, and editorial suggestions. The author is fully responsible for the correctness of the arguments.

References

[1] D. Berend and A. Kontorovich, *The missing mass problem*, Statistics & Probability Letters **82** (2012), no. 6, 1102–1110. doi:10.1016/j.spl.2012.02.014.

[2] D. Berend, A. Kontorovich, and G. Zagdanski, *The expected missing mass under an entropy constraint*, Entropy **19** (2017), no. 7, article 315, 13 pp. doi:10.3390/e19070315.