hexagon:2611.01227v1

RÉNYI ENTROPIES OF RECIPROCAL SUBSET SUMS

pengcheng chen

Abstract. Let $Y_n = \sum_{a \le n} X_a/a$ for independent fair Bernoulli variables X_a. We prove that the absolute difference between the Rényi entropies of any two distinct fixed nonnegative orders, including the endpoint orders, is comparable to the logarithm of the number of distinct reciprocal subset sums. The proof decodes many independent factors with the same nonuniform six-denominator law. For a uniformly chosen configuration, the logarithm of the size of its fibre concentrates about $n \log 2 - H_1(Y_n)$ within $O(n/\log n)$, a smaller scale than the entropy gaps. Both conclusions extend to a fixed number of disjoint labelled subsets. As an application of support estimates and modular collision counting, we give a scale for injectivity of reduction modulo primes, with a common exceptional set on the injective side.

1. Introduction and results

Erdős and Graham [4, p. 43] asked about the number of distinct subsums of the first n terms of the harmonic series and the largest subfamily with all subsums distinct. These are now catalogued as problems 320 and 321. Bleicher and Erdős [1, 2] obtained bounds involving iterated logarithms. Bettin, Grenié, Molteni and Sanna [3] proved a lower bound with a uniform stopping condition. Luo, Yang and Zhu [8] give a matching upper bound in a public manuscript associated with a proof claim accepted on the problem website. Kominers and Neu [6] state a finer asymptotic for the single-colour support in a working paper. We prove the upper bound needed here directly.

Our object is the probability distribution on the image of the subset-sum map. Fix an integer $m \ge 1$. Let $X_1, \dots, X_n$ be independent and uniform on $\{0, 1, \dots, m\}$. Label 0 denotes an unused denominator and the remaining labels denote ordered, disjoint sets. Write

$$Y_n = \left(\sum_{a \le n : X_a = i} \frac{1}{a} \right)_{i=1}^{m}, \qquad \mu_n(s) = \mathbb{P}(Y_n = s), \qquad \mathcal{V}_m(n) = \operatorname{supp} \mu_n.$$

For a finite set E of positive integers, Y_E, μ_E and $\mathcal{V}_m(E)$ have the analogous meanings, with independent uniform labels on E. Let $F_m(n; s)$ count the configurations with output s, so that

$$\mu_n(s) = (m+1)^{-n} F_m(n; s). \tag{1}$$

Date: Revised draft, 26 November 2026.

In particular $m = 1$ is the usual model of reciprocal subset sums, including the empty sum.

For a finite probability law μ, we use Rényi entropy [12] in natural logarithms,

$$H_\alpha(\mu) = \frac{1}{1-\alpha} \log \sum_{s \in \operatorname{supp} \mu} \mu(s)^\alpha \quad (0 < \alpha < \infty,\ \alpha \neq 1),$$

with the endpoint definitions

$$\begin{aligned} H_0(\mu) &= \log |\operatorname{supp} \mu|, \\ H_1(\mu) &= -\sum_s \mu(s) \log \mu(s), \\ H_\infty(\mu) &= -\log \max_s \mu(s). \end{aligned}$$

We abbreviate $H_\alpha(n) = H_\alpha(\mu_n)$ and $H_\alpha(E) = H_\alpha(\mu_E)$. These quantities depend on m, which is fixed throughout. All logarithms are natural. Write $\log_j t$ for the j-fold iterated natural logarithm wherever it is defined. In the product below, stop before the first factor less than 1. For sufficiently large real t, set

$$\mathcal{A}(t) = \prod_{j \geq 3 :\, \log_j t \geq 1} \log_j t, \qquad \mathcal{B}(t) = \frac{t}{\log t} \mathcal{A}(t).$$

Theorem 1.1 (Entropy separation)**.** *For every fixed $m \geq 1$, as $n \to \infty$,*

$$H_\alpha(n) \asymp_m \mathcal{B}(n) \qquad (0 \leq \alpha \leq \infty), \tag{2}$$

with comparison constants independent of α. For every fixed pair $0 \leq \alpha < \beta \leq \infty$,

$$H_\alpha(n) - H_\beta(n) \asymp_{m,\alpha,\beta} \mathcal{B}(n). \tag{3}$$

The common order in (2) can also be read off from the construction in [3, §2] and the support upper bound. Their proof of Theorem 1 counts a set of $\gg \mathcal{B}(n)$ denominators up to n whose reciprocals have distinct subset sums (see also [8, Theorem 1.3 and Lemma 4.1]). Indeed, the largest index of a nontrivial relation with coefficients in $\{-1, 0, 1\}$ would violate the defining condition of their set $\mathcal{U}$. Given all labels off $\mathcal{U} \cap [1, n]$, each coordinate sum determines which members of this set carry its label. Each conditional atom therefore has mass at most $(m+1)^{-|\mathcal{U} \cap [1,n]|}$, and averaging gives the same bound on $\max_s \mu_n(s)$. Since $\mathcal{V}_m(n) \subseteq \mathcal{V}_1(n)^m$, the single-colour support upper bound and entropy monotonicity now give (2). The lower bound for the difference in (3) requires an additional source of nonuniformity.

Theorem 1.2 (Typical fibres)**.** *For each fixed $m \geq 1$ there exist $C_m, c_m > 0$ such that for all sufficiently large n,*

$$\mathbb{P}\left(|\log F_m(n; Y_n) - n\log(m+1) + H_1(n)| > C_m \frac{n}{\log n}\right) \leq 3\exp\left(-c_m \frac{n}{(\log n)^3}\right).$$

Here the random output is sampled by a uniform random configuration, not by a uniform random point of $\mathcal{V}_m(n)$.

For $m = 1$, the signed harmonic sum with signs $2X_a - 1$ is $2Y_n - \sum_{a \leq n} 1/a$. Bettin, Molteni and Sanna [9, Theorem 2.1] estimate its mass in short real intervals. Conlon et al. [10, Theorem 1] determine the exponential growth rate of the fibre at each fixed positive rational target. Liu and Sawhney [11, Theorem 1.2] independently determine this rate at the target 1. In Theorem 1.2 the target itself is sampled with probability proportional to its fibre size.

These results distinguish four natural measures of fibre size. The logarithms of the average fibre over the support, the configuration-typical fibre, the configuration-weighted average fibre and the largest fibre are respectively

$$\begin{gathered} n\log(m+1) - H_0(n), \\ n\log(m+1) - H_1(n) + O_m(n/\log n), \\ n\log(m+1) - H_2(n), \\ n\log(m+1) - H_\infty(n). \end{gathered} \tag{4}$$

By Theorems 1.1 and 1.2, consecutive terms differ by positive quantities of order $\mathcal{B}(n)$, with high probability for comparisons involving the random fibre. Thus the distinction among these measures persists at the entire arithmetic correction scale.

For our modular application let N be a large integer and put $L = \log N$, $U = \log L$ and $Q_* = LU/\mathcal{A}(L)$. For a prime $p > q$ there is a well-defined map $\mathcal{V}_m(q) \to \mathbb{F}_p^m$ given by rational reduction in each coordinate.

Theorem 1.3 (Modular aliasing)**.** *There exist absolute constants $0 < a < b$ such that the following statements hold for all sufficiently large N.*

1. *There is a set of at most $N^{1/2}$ exceptional primes such that, for every nonexceptional prime p with $N/(\lfloor aQ_* \rfloor + 1) < p \leq N$ and every integer $1 \leq q \leq \lfloor aQ_* \rfloor$, the reduction map $\mathcal{V}_m(q) \to \mathbb{F}_p^m$ is injective.*
2. *For every integer $q \geq bQ_*$ and every prime $q < p \leq N$, this reduction map is not injective.*

The constants and exceptional set may be chosen independently of m.

In particular, take a prime $\sqrt{N} < p \leq N$ and set $q = \lfloor N/p \rfloor$, so that $q < p$. If $q \leq \lfloor aQ_* \rfloor$, part (1) gives injectivity outside its exceptional set, since $p > N/(\lfloor aQ_* \rfloor + 1)$. If $q \geq bQ_*$, part (2) gives a collision.

2. Exact arithmetic factorization

Bleicher and Erdős [2, proof of Theorem 2, pp. 603–604] distinguished the rational outputs of large-prime blocks by bounding their cleared numerators. Kominers and Neu [6, Proposition 3.1] use modular block coordinates to bound the support. With the cutoff below, the total output determines each rational block output and the remainder. We use this fact to keep the full configuration multiplicities.

Write $D_k = \operatorname{lcm}(1,\dots,k)$ and $h_k = \sum_{a\le k} a^{-1}$. We use the classical prime number theorem in the forms $\log D_k = \psi(k) \sim k$, $\vartheta(x) \ll x$ and $\pi(x) = \operatorname{Li}(x) + O_A(x/(\log x)^A)$ for any fixed $A > 0$, as in [7, p. 5 and Ch. 6].

Proposition 2.1 (Exact decoding)**.** *For sufficiently large n, set*

$$K = \left\lfloor \tfrac{1}{2}\log n \right\rfloor, \qquad y = \frac{n}{K+1}.$$

For primes $y < p \le n$ let $k_p = \lfloor n/p \rfloor$ and $B_p = \{p, 2p, \dots, k_p p\}$. These blocks are pairwise disjoint. Put $R = [1,n] \setminus \bigcup_{y<p\le n} B_p$, where $[1,n]$ denotes the integer interval. The map

$$(s_R, (s_p)_{y<p\le n}) \longmapsto s_R + \sum_{y<p\le n} \frac{s_p}{p} \tag{5}$$

is a bijection from $\mathcal{V}_m(R) \times \prod_{y<p\le n} \mathcal{V}_m(k_p)$ onto $\mathcal{V}_m(n)$. Under this bijection μ_n is the product law of the factors. Consequently, for all $0 \le \alpha \le \infty$,

$$H_\alpha(n) = \sum_{y<p\le n} H_\alpha(k_p) + H_\alpha(R), \qquad 0 \le H_\alpha(R) \le H_0(R) \ll_m n/\log n. \tag{6}$$

Proof. For large n, $y > \sqrt{n}$ and $D_K h_K < y$: indeed $\log(D_K h_K) \sim K$ and $\log y = \log n - O(\log\log n)$. The blocks are disjoint since a product of two primes exceeding y is greater than n. All combinations of factor outputs in (5) are attainable by assigning labels in disjoint blocks, so the map is surjective.

Fix a total output s and a block prime p. The vector ps has p-integral coordinates since $p^2 > n$. All summands from denominators outside B_p reduce to zero modulo p. Thus $ps \bmod p = s_p \bmod p$. As $p > K$, multiplication by D_K is invertible modulo p. Each coordinate of $D_K s_p$ is an integer between 0 and $D_K h_K < p$, so its residue determines it uniquely. Recovering every s_p and then subtracting their scaled contributions recovers s_R. For the unique decoded tuple the configuration count is

$$F_m(n;s) = F_m(R;s_R) \prod_{y<p\le n} F_m(k_p;s_p),$$

where $F_m(R;s_R)$ counts labellings of R. This follows by choosing each block labelling separately. Division by $(m+1)^n$ gives the product law, since $|R| + \sum_p k_p = n$. For $0 < \alpha < \infty$ with $\alpha \ne 1$, the power sums of this product

factor. Support sizes and maximum atom masses multiply, and Shannon entropies add, proving (6) also for $\alpha = 0, 1, \infty$.

Finally, all prime factors of denominators in R are at most y. Their least common multiple $D_R = \operatorname{lcm}\{a : a \in R\}$ (with $D_R = 1$ if R is empty) satisfies

$$\log D_R \le \vartheta(y) + \sum_{2 \le r \le (\log n)/(\log 2)} \vartheta(n^{1/r}) \ll y + \sqrt{n} \log n \ll n/\log n.$$

Each coordinate of Y_R lies in $D_R^{-1}\mathbb{Z} \cap [0, h_n]$, giving $H_0(R) \le m \log(1 + D_R h_n) \ll_m n/\log n$. Monotonicity of Rényi entropy in its order proves the remaining bound. □

3. The common scale

Lemma 3.1. *Uniformly in $0 \le \alpha \le \infty$, the function $g(n) = H_\alpha(n)$ satisfies*

$$g(n) = \frac{n}{\log n} \left(\int_1^{\log n} \frac{g(\lfloor v \rfloor)}{v^2}\, dv + O_m(1) \right). \tag{7}$$

Proof. The number of primes with $k_p = j$ for $1 \le j \le K$ is $\pi(n/j) - \pi(n/(j+1))$. The quantitative prime number theorem gives uniformly in this range

$$\frac{n}{j(j+1)\log n} \left(1 + O\left(\frac{\log\log n}{\log n} \right) \right) + O_A\left(\frac{n}{j(\log n)^A} \right).$$

Since $0 \le g(j) \le j \log(m+1)$, the relative-error contribution after multiplication by $g(j)$ and summation is $O_m(n(\log\log n)^2/(\log n)^2)$, and the second error totals $O_m(nK/(\log n)^A)$. Take $A = 3$. Proposition 2.1 now gives

$$g(n) = \frac{n}{\log n} \sum_{j \le K} \frac{g(j)}{j(j+1)} + O_m(n/\log n).$$

The sum is the integral up to $K+1$, and extending its endpoint to $\log n$ incurs $O_m(1)$ in the integral by the same linear bound on g. This proves (7). □

Lemma 3.2. *For fixed m, $H_0(n) \asymp_m \mathcal{B}(n)$. Furthermore, $\mathcal{A}(n) \to \infty$ and $\mathcal{A}(n) \le (\log_3 n)^2$ for large n.*

Proof. For $v \ge 1$, let $P(v)$ be the product of the successive iterated logarithms of v that are at least 1, starting at $\log v$. The empty product is 1. This function is nondecreasing and continuous since each new factor enters at value 1. Take $t_0 = e^e$ and use the product in the definition of $\mathcal{A}(t)$ for $t \ge t_0$, assigning value 1 to an empty product. Extend $\mathcal{B}(t)$ as the constant $\mathcal{B}(t_0)$ on $[1, t_0]$. The extension is positive and nondecreasing. For $x = \log_3 n \ge 1$ one has $\mathcal{A}(n) = xP(x)$. For $t \ge t_0$, $\mathcal{A}(t) = P(\log\log t)$, so the substitution $v = \log\log t$ gives

$$\int_1^{\log n} \frac{\mathcal{B}(t)}{t^2}\, dt = c_0 + \int_1^x P(v)\, dv, \qquad c_0 = \int_1^{t_0} \frac{\mathcal{B}(t)}{t^2}\, dt.$$

For $x \geq 2e$, monotonicity of all the factors except the first also gives

$$P(x/2) \leq P(x)\frac{\log(x/2)}{\log x}.$$

It follows that

$$xP(x) - \int_1^x P(v)\,dv \geq (x/2-1)P(x)\frac{\log 2}{\log x} \longrightarrow \infty.$$

Let c_m bound the absolute error inside the parentheses in (7), and take $g = H_0$. First choose a threshold so that the deficit above is at least $c_0 + c_m$ for every subsequent n. Then choose $C \geq 1$ with $g(j) \leq C\mathcal{B}(j)$ up to that threshold. Since $\lfloor v \rfloor < n$ throughout the integral, strong induction and monotonicity bound the parentheses by $C(c_0 + \int_1^x P(v)\,dv) + c_m \leq CxP(x)$. This gives $g(n) \leq C\mathcal{B}(n)$.

For the lower bound, the set E_n in [3, eq. (1)] equals $\mathcal{V}_1(n)$, including the empty sum. The third case of their Theorem 1 gives

$$\log|\mathcal{V}_1(n)| \geq (\log 2)\left(2 - \frac{3}{\log_k n}\right)\frac{n}{\log n}\prod_{j=3}^{k}\log_j n$$

for all integers $k \geq 4$ and n such that $\log_k n \geq 3/2$. For sufficiently large n choose the last k with $\log_k n \geq 3$, so $k \geq 4$. The coefficient in parentheses is at least 1. The next iterate is less than 3, the following one is less than $\log 3$, and any further iterate, if needed, is less than 1. The omitted factors therefore have product at most $3\log 3$. Hence $\log|\mathcal{V}_1(n)| \gg \mathcal{B}(n)$. Embedding the single-colour support in the first coordinate gives the lower bound for every m.

Finally $P(v) \leq v$: this follows from $P(v) = \log v\, P(\log v)$ for $v \geq e$, the initial case $1 \leq v < e$, and $(\log v)^2 \leq v$. Consequently $\mathcal{A}(n) = xP(x) \leq x^2$, while $\mathcal{A}(n) \geq x \to \infty$. □

4. Separation of the entropy orders

For a nonuniform law μ on its finite positive support the function $f(t) = \log\sum_s \mu(s)^t$ is strictly convex since its second derivative is the variance of $\log\mu(s)$ under weights proportional to $\mu(s)^t$. Since $f(1) = 0$, the negative secant slopes through $(1,0)$ give H_t, with the tangent value at 1. This proves strict decrease between all finite orders, starting at 0. Given any finite α, take a finite $\gamma > \alpha$ to get $H_\alpha > H_\gamma \geq H_\infty$ by passage to the limit.

Proof of Theorem 1.1. Fix an integer $T > 6$ sufficiently large that Proposition 2.1 holds at all lengths $v \geq T$, its cutoff K is at least 6, and the prime-counting bounds below are valid uniformly. Apply the proposition recursively to every complete interval factor of length at least T, leaving each smooth remainder as a terminal factor without further decomposition. The complete interval children of a node of length v have lengths at most

$(\log v)/2$, so the tree is finite. Write $\mathcal{I}$ for the set of internal nodes and n_w for the length at node w, and put

$$W = \sum_{w \in \mathcal{I}} \frac{n_w}{\log n_w}.$$

All the entropy quantities are sums over the resulting independent factors. The support entropies of the smooth remainders sum to at most $C_m W$. Each internal node has at most $\pi(n_w) \ll n_w/\log n_w$ complete interval children, and the terminal complete intervals have length less than T. Each terminal complete interval is counted among the children of its unique parent and contributes at most $T\log(m+1)$ to H_0. Therefore, for $n \geq T$,

$$H_0(n) \leq C_{m,T} W. \tag{8}$$

At every internal node of length v, the primes in $(v/7, v/6]$ give length-six children; all are in the decoding range since $K \geq 6$. The prime number theorem shows that there are $\gg v/\log v$ such primes, uniformly for $v \geq T$ after increasing T if necessary. Each of these children is terminal, and children from different parents are distinct factors of the product. Hence there are $\gg W$ length-six terminal factors.

Uniform labellings at length six induce a nonuniform output law for every $m \geq 1$. Its zero vector has only one configuration, whereas the vector $(1/2, 0, \ldots, 0)$ has at least two configurations, arising from $\{2\}$ and $\{3, 6\}$ in the first colour with the remaining labels unused. In particular, $H_\infty(6) > 0$ and for every fixed $\alpha < \beta$ the difference $\delta_{\alpha,\beta} = H_\alpha(6) - H_\beta(6)$ is strictly positive.

All remaining independent factors have nonnegative entropy and nonnegative entropy gaps. Additivity, followed by (8), gives

$$H_\infty(n) \gg_m W \gg_m H_0(n), \qquad H_\alpha(n) - H_\beta(n) \gg_{m,\alpha,\beta} W \gg_{m,\alpha,\beta} H_0(n).$$

Upper bounds follow immediately from monotonicity and nonnegativity. Lemma 3.2 proves (2) and (3). The bounds for each individual order are uniform in the order because $H_\infty \leq H_\alpha \leq H_0$. □

Bettin, Molteni and Sanna [9, proof of Proposition 2.7] also exploit collisions inside small blocks to bound a support size. Here the terminal block outputs can be recovered jointly from the full rational output, so that their entropy gaps add.

5. Information concentration and fibres

Proof of Theorem 1.2. Put $I_n = -\log \mu_n(Y_n)$. Its expectation is $H_1(n)$. By the exact factorization,

$$I_n = I_R + \sum_{y < p \leq n} I_p,$$

where the terms are independent and I_p is the information of a length-k_p block. Every atom of positive probability in that block has probability at

least $(m+1)^{-k_p}$. Thus $0 \le I_p \le k_p \log(m+1)$ and

$$\sum_{y<p\le n} (k_p \log(m+1))^2 \ll_m K \sum_{y<p\le n} k_p \ll_m n \log n.$$

Hoeffding's inequality [5, Theorem 2], with deviation $n/\log n$, gives

$$\mathbb{P}\left(\left| \sum_{y<p\le n} (I_p - \mathbb{E}I_p) \right| > n/\log n \right) \le 2\exp(-c_m n/(\log n)^3).$$

The smooth factor need not have uniformly bounded information at each atom on the desired scale. Instead use the general bound for a law of support size M,

$$\mathbb{P}(-\log \mu(Y) > \log M + t) \le e^{-t}, \qquad t \ge 0,$$

which follows by summing the masses of at most M atoms of probability below e^{-t}/M. Here $\log M = H_0(R) \ll_m n/\log n$. On the complementary event, $0 \le I_R \le \log M + t$ and $0 \le H_1(R) \le \log M$, so $|I_R - H_1(R)| \le \log M + t$. Take $t = n/\log n$ and combine this with the two-sided block bound. The total failure probability is at most $2\exp(-c_m n/(\log n)^3) + \exp(-n/\log n)$. Decreasing c_m if necessary and using (1) proves the assertion. □

For clarity, define $F_{\mathrm{avg}} = (m+1)^n/|\mathcal{V}_m(n)|$ and $F_{\max} = \max_s F_m(n;s)$. Equation (1) gives exactly

$$\log F_{\mathrm{avg}} = n\log(m+1) - H_0(n), \qquad \log F_{\max} = n\log(m+1) - H_\infty(n),$$

and also

$$\log \mathbb{E}[F_m(n;Y_n)] = n\log(m+1) + \log \sum_s \mu_n(s)^2 = n\log(m+1) - H_2(n).$$

Together with the typical-fibre theorem these are the four assertions in (4). In particular, since $n/\log n = o(\mathcal{B}(n))$, all the consecutive logarithmic gaps are of order $\mathcal{B}(n)$ with probability tending to 1 when a random fibre is involved.

6. Modular projection

Lemma 6.1. *For fixed $d > 0$, with $L = \log N$, $U = \log L$ and $Q_* = LU/\mathcal{A}(L)$,*

$$\mathcal{B}(dQ_*) \sim dL.$$

Proof. By Lemma 3.2, $\mathcal{A}(L) \le (\log_3 L)^2 = o(U)$ and $\mathcal{A}(L) \to \infty$. Hence $L = o(Q_*)$ and $Q_* = o(L^2)$. Away from transition points put $D(v) = v(\log P(v))'$. This is zero on $(1,e)$ and at most 1 on (e,e^e). For $v > e^e$ the product identity gives $D(v) = (1 + D(\log v))/\log v \le 2$ by induction. The transition points are locally finite and P is continuous there. Integrating on the successive smooth intervals thus gives $|\log P(v) - \log P(w)| \le 2|\log(v/w)|$ for $v, w \ge 1$. For $L/2 \le t \le L^2$ one has $\log\log t = \log U + O(1)$ uniformly. Applying this bound to $v = \log\log t$ and $w = \log U$ gives $\mathcal{A}(t)/\mathcal{A}(L) \to 1$ uniformly.

For each fixed $d > 0$, dQ_* is eventually in this interval and $\log(dQ_*) = U + \log(dU/\mathcal{A}(L)) \sim U$. Substitution gives the assertion. □

The collision count below uses the cleared-difference argument in the proof of [6, Lemma 3.2]. We apply it once to the largest short support, so the resulting exceptional set works for every smaller length. A comparison of support size with a prime also occurs in the proof of [2, Theorem 4], which studies the largest denominator needed to represent rationals with a given prime denominator.

Proof of Theorem 1.3. Take $a > 0$ sufficiently small and put $Q = \lfloor aQ_* \rfloor$. Lemmas 3.2 and 6.1 ensure that $2\log|\mathcal{V}_1(Q)| \le L/3$ for large N since $\mathcal{B}(Q) \le \mathcal{B}(aQ_*)$. Every prime $p > N/(Q+1)$ exceeds Q, and has logarithm at least $L - O(\log L)$. For distinct $s, t \in \mathcal{V}_1(Q)$, the integer $D_Q(s-t)$ is nonzero and of absolute value at most $D_Q h_Q$. Its logarithmic absolute value is $O(Q)$. Therefore it has $O(1 + Q/L)$ distinct prime divisors exceeding $N/(Q+1)$. By a union bound over all ordered pairs of distinct rational outputs, there are at most

$$|\mathcal{V}_1(Q)|^2 O(1 + Q/L) \le e^{L/3} O(1 + Q/L) \le N^{1/2}$$

primes in this range at which reduction is not injective. The exceptional set is fixed by N, Q and does not depend on the number of colours. Since $\mathcal{V}_1(q) \subseteq \mathcal{V}_1(Q)$ for $q \le Q$, the same set works for every such q. Each coordinate of a multicolour output belongs to the single-colour support, so coordinatewise injectivity implies injectivity of the vector reduction map.

Conversely, choose b so large that $\log|\mathcal{V}_1(\lfloor bQ_* \rfloor)| > L$ for large N, using $\lfloor bQ_* \rfloor \ge (b/2)Q_*$, Lemma 6.1 with $d = b/2$, the monotonicity of $\mathcal{B}$, and the lower bound in Lemma 3.2. If $q \ge bQ_*$ and $q < p \le N$, then $|\mathcal{V}_1(q)| > N \ge p$. Reduction is defined since $p > q$, but cannot be injective by the pigeonhole principle. Embedding the single-colour support in the first coordinate proves noninjectivity for every $m \ge 1$ as well. □

7. Further questions

The separation theorem determines a common scale, not precise leading constants. In particular, the existence of an ordinary limit of $H_\alpha(n)/H_0(n)$ for a nontrivial order α is not established here. The distinction between sampling by a random configuration and sampling uniformly from the rational support is important; concentration in the second sampling model is another question. Finally, modular noninjectivity is much weaker than complete coverage of a finite field. None of our results implies that a prescribed residue has a representation by a sparse reciprocal sum.

References

[1] M. N. Bleicher and P. Erdős, The number of distinct subsums of $\sum_{i=1}^{N} 1/i$, *Math. Comp.* **29** (1975), 29–42.

[2] M. N. Bleicher and P. Erdős, Denominators of Egyptian fractions. II, *Illinois J. Math.* **20** (1976), 598–613.
[3] S. Bettin, L. Grenié, G. Molteni and C. Sanna, A lower bound for the number of Egyptian fractions, *Math. Comp.*, published electronically 22 January 2026, doi:10.1090/mcom/4190, arXiv:2509.10030.
[4] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Monographies de L'Enseignement Mathématique, no. 28, Geneva, 1980.
[5] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30.
[6] S. D. Kominers and J. Neu, The asymptotic number of distinct reciprocal subset sums, working paper, July 2026, author manuscript.
[7] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I: Classical Theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, 2007.
[8] Y. Luo, R. Yang and K. Zhu, Distinct harmonic subset sums, public manuscript, 2026, author manuscript, problem 320 claim record.
[9] S. Bettin, G. Molteni and C. Sanna, Small values of signed harmonic sums, *C. R. Math. Acad. Sci. Paris* **356** (2018), 1062–1074.
[10] D. Conlon, J. Fox, X. He, D. Mubayi, H. T. Pham, A. Suk and J. Verstraëte, A question of Erdős and Graham on Egyptian fractions, *Discrete Anal.* (2025), Paper No. 28, 13 pp.
[11] Y. P. Liu and M. Sawhney, On further questions regarding unit fractions, *Int. Math. Res. Not. IMRN* (2026), no. 2, rnaf382.
[12] A. Rényi, On measures of entropy and information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, University of California Press, Berkeley, 1961, 547–561.