A proof of the Shepp–Olkin entropy monotonicity conjecture

Erwan Hillion[*]   Oliver Johnson[†]

Abstract

Consider tossing a collection of coins, each fair or biased towards heads, and take the distribution of the total number of heads that result. It is natural to suppose that this distribution should be 'more random' when each coin is fairer. In this paper, we prove a 40 year old conjecture of Shepp and Olkin, by showing that the Shannon entropy is monotonically increasing in this case, using a construction inspired by optimal transport theory. We discuss whether this result can be generalized to q-Rényi and q-Tsallis entropies, for a range of values of q.

1   Introduction and notation

We consider the entropy of Poisson–binomial random variables (sums of independent Bernoulli random variables). Given parameters $\mathbf{p} = (p_1, \ldots, p_n)$ (where $0 \le p_i \le 1$) we will write $f_{\mathbf{p}}$ for the probability mass function of the random variable $B_1 + \ldots + B_n$, where B_i are independent with $B_i \sim \text{Bernoulli}(p_i)$. We can write the Shannon entropy as a function of the parameters:

$$H(\mathbf{p}) = -\sum_{k=0}^{n} f_{\mathbf{p}}(k) \log f_{\mathbf{p}}(k). \tag{1.1}$$

Shepp and Olkin [16] made the following conjecture "on the basis of numerical calculations and verification in the special cases $n = 2, 3$":

Conjecture 1.1 (Shepp–Olkin monotonicity conjecture). *If all $p_i \le 1/2$ then $H(\mathbf{p})$ is a non-decreasing function of $\mathbf{p}$.*

---
[*]Aix Marseille Univ, France. E-mail: erwan.hillion@univ-amu.fr
[†]University of Bristol, United Kingdom. E-mail: maotj@bristol.ac.uk

The main contribution of the present paper is to prove that this conjecture is correct, and to give conditions for equality (see Theorem 3.6). The heart of our argument will be a representation of one probability distribution in terms of another, via so-called mixing coefficients α_k (see Definition 2.1) previously introduced in [7] in a form motivated by optimal transport. The key property of these coefficients is that their differences are decreasing functions (see Proposition 2.8). We analyse the resulting expressions using elementary tools; specifically the interplay between the product rule (3.1) for discrete differentiation and an integration by parts formula, Lemma 3.3. It may be surprising that only such simple tools are required to resolve this conjecture, however we remark that the interplay between product rules and integration by parts lies behind much of the power of the Bakry-Émery Γ_2-calculus for continuous random variables (as alluded to in the dedication of [1]). This may suggest that the Shepp–Olkin conjecture should be viewed in the context of an emerging discrete Bakry-Émery theory (see also [4, 9]).

Given the naturalness of Conjecture 1.1, and its simplicity of statement, it is perhaps surprising that it has remained open for over 40 years (although [16] was published as a book chapter in 1981, the conjecture was first formulated in the corresponding technical report [15] in July 1978). Indeed, we are not even aware of any published work that resolves special cases.

In the same paper [16], Shepp and Olkin also conjectured that the entropy $H(\mathbf{p})$ is concave in $\mathbf{p} \in [0,1]^n$, which was proved in [7] and [8], using constructions based on optimal transport which we will describe below. Similarly to the monotonicity conjecture, little published work had previously addressed the entropy concavity conjecture, though limited progress in special cases had been made in [18] and [6].

In their original paper [16], Shepp and Olkin did prove some related results. In particular, they showed [16, Theorem 1] that the entropy $H(\mathbf{p})$ is Schur concave in $\mathbf{p}$, and hence deduced a maximum entropy property for binomial random variables (see also contemporary work of Mateev [11], as well as later extensions by [5, 19]). Further, they showed [16, Theorem 2] that the entropy of $H(\mathbf{p})$ is concave in a single argument p_i (see (1.4) below) and [16, Theorem 4] that the entropy of the binomial distribution is concave. The Schur concavity property of Shepp and Olkin was generalized to a wider range of functionals and parametric families including negative binomials by Karlin and Rinott [10].

However, none of these results appear to be particularly relevant to the monotonicity conjecture, Conjecture 1.1. We can reformulate the conjecture to say that if $p_i \leq q_i \leq 1/2$ for each $i = 1, 2, \ldots n$ then $H(\mathbf{p}) \leq H(\mathbf{q})$. Clearly, by symmetry of the arguments, it is sufficient to verify this in the case where $p_i = q_i$ for $i = 1, 2, \ldots n - 1$, which in turn means that it is enough to check that $\frac{\partial}{\partial p_n} H(\mathbf{p}) \geq 0$. Without loss of generality we will assume throughout that all p_i are non-zero, so that $f(k)$ is non-zero for all k in $\{0, 1, \ldots, n\}$.

In this case, following standard calculations in [16], we can take $\mathbf{p}(t) = (p_1, \ldots, p_{n-1}, p_n(t))$ where $p_n(t) = p_n + t$, omit the subscript on $f_{\mathbf{p}(t)}$ for brevity and write $f(k) = p_n(t)g(k-1) + (1 - p_n(t))g(k)$ to deduce

$$\frac{\partial f}{\partial t}(k) = g(k-1) - g(k), \quad \text{for } k = 0, 1, \ldots, n, \tag{1.2}$$

where g is the probability mass function of $B_1 + \ldots + B_{n-1}$, which is supported on the set $\{0, \ldots, n-1\}$ and does not depend on t. Here and throughout we take $g(-1) = g(n) = 0$ if necessary.

As in [16, Theorem 2] we can use (1.2) to evaluate the first two derivatives of $H(\mathbf{p}(t))$ as a function of t. Direct substitution gives

$$\frac{\partial H}{\partial t} = \sum_{k=0}^{n} (g(k) - g(k-1)) \log f(k), \tag{1.3}$$

$$\frac{\partial^2 H}{\partial t^2} = -\sum_{k=0}^{n} \frac{(g(k) - g(k-1))^2}{f(k)}. \tag{1.4}$$

The negativity of each term in (1.4) tells us directly that (as proved in [16, Theorem 2]) the entropy $H(\mathbf{p}(t))$ is concave in t (of course, we also know this from the full Shepp–Olkin theorem proved in [7, 8]) so it is sufficient to prove that the derivative $\frac{\partial H}{\partial t}$ is non-negative in the case $p_n = 1/2$, since the derivative is therefore larger for any smaller values of p_n.

However, at this stage, further progress is elusive. Considering convolution with B_n means that we can express $f(k) = (g(k) + g(k-1))/2$. However, substituting this in (1.3) does not suggest an obvious way forward in general, though it is possible to use the resulting formula to resolve certain special cases. For example, careful cancellation in the case where $p_1 = p_2 = \ldots = p_{n-1} = 1/2$ and hence g is binomial allows us to deduce that, in this case, the entropy derivative (1.3) equals zero (see Example 2.7 below for an alternative view of this). However, this calculation does not give any particular insight into why the binomial case might be extreme in the sense of the conjecture.

Instead of expressing f as a linear combination of g, our key observation is that we can express g as a weighted linear combination of f, as described in the following section.

2 Entropy derivative and mixing coefficients

The following construction and notation were introduced in [7], based on the 'hyper-geometric thinning' construction of Yu [19, Definition 2.2]. The key is to observe that in general we can write

$$g(k) = \alpha_{k+1} f(k+1) + (1 - \alpha_k) f(k), \tag{2.1}$$

for certain 'mixing coefficients' $(\alpha_k)_{k=0,1,\ldots,n}$. The general construction for α_k in the case of Shepp–Olkin paths is given in [7, Proposition 5.1], but in the specific case where only p_n varies, in the case $p_n = 1/2$, we can simply define the following values:

Definition 2.1. *For $k = 0, \ldots, n$, define*

$$\alpha_k := \frac{g(k-1)}{2f(k)} = \frac{g(k-1)}{g(k-1) + g(k)}. \tag{2.2}$$

By direct calculation using (2.2) we can see that in this case

$$\alpha_{k+1} - \alpha_k = \frac{g(k)^2 - g(k-1)g(k+1)}{(g(k-1) + g(k))(g(k) + g(k+1))},$$

so the fact that $g(k)^2 > g(k-1)g(k+1)$ (see [12]) means that the mixing coefficients satisfy

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \ldots < \alpha_n = 1. \tag{2.3}$$

In [7, Proposition 5.2] this result was stated in the form $\alpha_{k-1} \leq \alpha_k$, but the strict inequalities will help us to resolve the case of equality in Conjecture 1.1. It will often be useful for us to observe that Definition 2.1 implies that for $k = 0, 1, \ldots, n$

$$g(k) = 2\alpha_{k+1} f(k+1) = 2(1 - \alpha_k) f(k), \tag{2.4}$$

and that for $k = 0, 1, \ldots, n-1$

$$\alpha_{k+1} g(k+1) = (1 - \alpha_{k+1}) g(k). \tag{2.5}$$

Remark 2.2. Summing (2.4), we can directly calculate that

$$\sum_{k=0}^{n} f(k) \alpha_k = \frac{1}{2}, \tag{2.6}$$

which will play an important role in our proof of Conjecture 1.1. Further, it is interesting to note by rearranging (2.2) that $\alpha_k \leq 1/2$ if and only if $g(k-1) \leq g(k)$, which by the unimodality of g (see for example [12]) means that $k \leq \mathrm{mode}(g)$. This may suggest that the Shepp-Olkin conjecture can be understood as relating to the skewness of the random variables g. Direct calculation shows that the centred third moment of $B_1 + \ldots + B_{n-1}$ is $\sum_{i=1}^{n-1} p_i(1-p_i)(1-2p_i) \geq 0$, but it is not immediately clear how this positive skew will affect the entropy of f.

Remark 2.3. In [7] we used these mixing coefficients $(\alpha_k)_{k=0,1,\ldots,n}$ to formulate a discrete analogue of the Benamou–Brenier formula [2] from optimal transport theory, which gave an understanding of certain interpolation paths of discrete probability measures (including Shepp–Olkin paths) as geodesics in a metric space. We do not require this interpretation here, but simply study the properties of α_k in their own right.

We now define a function which will form the basis of our proof of Conjecture 1.1:

Definition 2.4. *Define*

$$\psi(\alpha) := \alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) - (2 - 2 \log 2)(\alpha - 1/2),$$

where we take $0 \log 0 = 0$ to ensure that ψ is continuous at 0 and at 1.

Remark 2.5.

1. Note that we can express

$$\psi(\alpha) = -\sum_{r=1}^{\infty} \frac{(2\alpha - 1)^{2r+1}}{2r(2r+1)}, \tag{2.7}$$

   as a power series with only odd terms with all coefficients negative. Further, comparison with $\sum_{r=1}^{\infty} \frac{1}{2r(2r+1)} = 1 - \log 2 < \infty$ shows that this series converges absolutely for all $\alpha \in [0, 1]$.

2. Hence the ψ is non-increasing and antisymmetric about $1/2$ (with $\psi(\alpha) = -\psi(1-\alpha)$) and with $\psi(0) = 1 - \log 2 > 0$, $\psi(1/2) = 0$ and $\psi(1) = -(1 - \log 2) < 0$.

We can express the derivative of entropy in terms of these functions and the mixing coefficients, as follows:

Proposition 2.6.

$$\frac{\partial H}{\partial t} = 2 \sum_{k=0}^{n} f(k) \psi(\alpha_k). \tag{2.8}$$

Hence by (2.7), the entropy derivative (1.3) is positive if each of the odd centred moments $\sum_{k=0}^{n} f(k)(\alpha_k - 1/2)^{2r+1} \leq 0$, for $r = 1, 2, \ldots$.

Proof. Using the fact that $g(-1) = g(n) = 0$ and adding cancelling terms into the sum (1.3), we can use $g(k)/(2f(k)) = 1 - \alpha_k$ and $g(k-1)/(2f(k)) = \alpha_k$ (see (2.4)) to obtain

that

$$\frac{\partial H}{\partial t} = \sum_{k=0}^{n-1} g(k) \log f(k) - \sum_{k=1}^{n} g(k-1) \log f(k)$$

$$= \sum_{k=0}^{n-1} g(k) \log \left(\frac{f(k)}{g(k)/2} \right) - \sum_{k=1}^{n} g(k-1) \log \left(\frac{f(k)}{g(k-1)/2} \right)$$

$$= 2 \sum_{k=0}^{n-1} f(k) \left(-(1-\alpha_k) \log(1-\alpha_k) \right) + 2 \sum_{k=1}^{n} f(k) \left(\alpha_k \log \alpha_k \right)$$

$$= 2 \sum_{k=0}^{n} f(k) \left(\alpha_k \log \alpha_k - (1-\alpha_k) \log(1-\alpha_k) \right) \tag{2.9}$$

since $\alpha_0 = 0$ and $\alpha_n = 1$ so that $\alpha_0 \log \alpha_0 = (1-\alpha_n) \log(1-\alpha_n) = 0$.

Since by (2.6) the $\sum_{k=0}^{n} f(k)\alpha_k = 1/2$, subtracting off the linear term makes no difference to the sum and we can rewrite (2.9) as $2 \sum_{k=0}^{n} f(k)\psi(\alpha_k)$, as required.

We can exchange the order of summation in

$$2 \sum_{k=0}^{n} f(k)\psi(\alpha_k) = -2 \sum_{k=0}^{n} f(k) \sum_{r=1}^{\infty} \frac{(2\alpha_k - 1)^{2r+1}}{2r(2r+1)}$$

$$= -\sum_{r=1}^{\infty} \frac{2^{2r+1}}{r(2r+1)} \sum_{k=0}^{n} f(k)(\alpha_k - 1/2)^{2r+1},$$

because of Fubini's theorem, since as mentioned above the power series for ψ converges absolutely. Hence if each odd centered moment is negative then the entropy derivative (1.3) is positive. □

Notice that, using Remark 2.5 and Proposition 2.6, we can immediately deduce that the entropy derivative must lie between $-2(1 - \log 2)$ and $2(1 - \log 2)$.

Example 2.7. In the case where $p_1 = p_2 = \ldots = p_n$, since g is Binomial$(n-1, 1/2)$, we know that $g(k)/g(k-1) = (n-k)/k$ for $k = 1, \ldots, n$, so that

$$\alpha_k = \frac{1}{1 + g(k)/g(k-1)} = \frac{1}{1 + (n-k)/k} = \frac{k}{n}$$

Using this, and the fact that $\alpha_0 = 0$, since f is Binomial$(n, 1/2)$ we know that $f(k) = f(n-k)$ and $\alpha_k - 1/2 = -(\alpha_{n-k} - 1/2)$ so each odd centred moment satisfies

$$\sum_{k=0}^{n} f(k)(\alpha_k - 1/2)^{2r+1} = -\sum_{k=0}^{n} f(n-k)(\alpha_{n-k} - 1/2)^{2r+1}$$

$$= -\sum_{l=0}^{n} f(l)(\alpha_l - 1/2)^{2r+1}$$

by relabelling, and hence equals zero.

We shall argue that the binomial example, Example 2.7, represents the extreme case using the following property, which will be key for us:

Proposition 2.8. If all the $p_i \le 1/2$ then

$$\alpha_{k+1} - \alpha_k \quad \text{is decreasing in } k = 0, \ldots, n-1. \tag{2.10}$$

Proof. See Appendix A. □

Note that comparing averages, and taking the values of $\alpha_0 = 0$ and $\alpha_n = 1$ from (2.3), Proposition 2.8 implies that

$$\frac{\alpha_k}{k} = \frac{\sum_{\ell=0}^{k-1}(\alpha_{\ell+1} - \alpha_\ell)}{k} \geq \frac{\sum_{\ell=0}^{n-1}(\alpha_{\ell+1} - \alpha_\ell)}{n} = \frac{\alpha_n}{n} = \frac{1}{n},$$

or that if all the $p_i \leq 1/2$ then $\alpha_k \geq k/n$ for $0 \leq k \leq n$, showing that the binomial distribution of Example 2.7 is the extreme case in this sense.

3 Proof of Shepp–Olkin monotonicity conjecture

We are now in a position to complete our proof of Conjecture 1.1.

Definition 3.1. *First we introduce some further notation:*

1. *We denote $\beta_k = \alpha_k - 1/2$. In particular $\alpha_{k+1} + \alpha_{k-p} - 1 = \beta_{k+1} + \beta_{k-p}$. As proven in (2.3) and Proposition 2.8 respectively, the sequence $(\beta_k)_k$ is increasing and $(\beta_{k+1} - \beta_k)_k$ is decreasing.*

2. *We define the family $(A_p(k))_k$ by $A_0(k) = 1$ and, for $p \geq 1$, $A_p(k) = \prod_{j=0}^{p-1} \alpha_{k-j}$. Here if $p \geq k+1$, we take $A_p(k) = 0$ (this reflects that the product includes the term $\alpha_0 = 0$).*

3. *We define the family $(B_p(k))_k$ by $B_0(k) = 1$ and, for $1 \leq p \leq k+1$, write $B_p(k) = \prod_{j=0}^{p-1}(\beta_{k+1} - \beta_{k-j})$. For other values of k, $B_p(k)$ is not defined.*

4. *The notation ∇ stands for the left-derivative operator: $\nabla v(k) = v(k) - v(k-1)$. This operator satisfies a product rule of the form:*

$$\nabla(vw)(k) = v(k)\nabla w(k) + w(k-1)\nabla v(k). \tag{3.1}$$

5. *For $n \geq 1$ and $p \geq 1$ we define the polynomial (symmetric in its inputs)*

$$Q_{n,p}(X_1, \ldots, X_p) = \sum X_1^{i_1} \ldots X_p^{i_p}, \tag{3.2}$$

   where the sum is taken over all the p-tuples $0 \leq i_1, \ldots, i_p \leq n$ such that $i_1 + \ldots + i_p = n$. We also set $Q_{0,p} = 1$ and $Q_{-1,p} = 0$ for all $p \geq 1$. Clearly, $Q_{n,p}(X_1, \ldots, X_p)$ is non-negative if $X_1, \ldots, X_p$ are non-negative.

We now state three technical lemmas that we will require in the proof; each of these are proved in Appendix B. First, the fact that $B_p(k)$ is decreasing in k:

Lemma 3.2. *Given $p \geq 0$, we have $\nabla B_p(k) = B_p(k) - B_p(k-1) \leq 0$ for $k \geq p$.*

We next give an integration by parts formula. Note that although we restrict the range of summation for technical reasons, the values $A_p(k)$ and $A_{p+1}(k)$ are zero outside the respective ranges:

Lemma 3.3. *For any function $v(k)$ that is well-defined on $p \leq k \leq n-1$ and any $p \geq 0$ we have*

$$\sum_{k=p}^{n-1} g(k)A_p(k)v(k)(\beta_{k+1} + \beta_{k-p}) = \sum_{k=p+1}^{n-1} g(k)A_{p+1}(k)\nabla v(k). \tag{3.3}$$

Finally, a result concerning differences of the Q polynomials:

Lemma 3.4. *For $n \geq 0$ and $p \geq 1$, we have*

$$Q_{n,p}(X_2, \ldots, X_{p+1}) - Q_{n,p}(X_1, \ldots, X_p) = (X_{p+1} - X_1)Q_{n-1,p+1}(X_1, \ldots, X_{p+1}). \tag{3.4}$$

We now state and prove the key Proposition:

Proposition 3.5. *Fix $r \geq 1$. The sequence $(S_{r,p})_{1 \leq p \leq r+1}$ defined by*

$$S_{r,p} = \sum_{k=p-1}^{n-1} g(k)A_{p-1}(k)B_p(k)Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2)\,(\beta_{k+1} + \beta_{k-p+1}),$$

is increasing in p. Here note that $S_{r,r+1} = 0$ since $Q_{-1,p+1} = 0$.

(Note that by restricting to this range of summation the $B_p(k)$ is well-defined).

Proof. We first take $v(k) = B_p(k)Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2)$ in the integration by parts formula (3.3) from Lemma 3.3 to write

$$S_{r,p} = \sum_{k=p}^{n-1} g(k)A_p(k)\nabla\left[B_p(k)Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2)\right],$$

where here and throughout the proof, the ∇ refers to a difference in the k parameter. Now, using the product rule (3.1) with $v(k) = B_p(k)$ and $w(k) = Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2)$ yields:

$$
\begin{aligned}
S_{r,p} &= \sum_{k=p}^{n-1} g(k)A_p(k)B_p(k)\nabla\left[Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2)\right] \\
&\quad + \sum_{k=p}^{n-1} g(k)A_p(k)Q_{r-p,p+1}(\beta_k^2, \ldots, \beta_{k-p}^2)\nabla B_p(k)
\end{aligned}
$$

In order to transform the first sum, we use equation (3.4) from Lemma 3.4:

$$
\begin{aligned}
&\nabla\left[Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2)\right] \\
&= Q_{r-p,p+1}(\beta_{k+1}^2, \ldots, \beta_{k-p+1}^2) - Q_{r-p,p+1}(\beta_k^2, \ldots, \beta_{k-p}^2) \\
&= (\beta_{k+1}^2 - \beta_{k-p}^2)Q_{r-p-1,p+2}(\beta_{k+1}^2, \ldots, \beta_{k-p}^2) \\
&= (\beta_{k+1} - \beta_{k-p})Q_{r-p-1,p+2}(\beta_{k+1}^2, \ldots, \beta_{k-p}^2)(\beta_{k+1} + \beta_{k-p}).
\end{aligned}
$$

Further Lemma 3.2 gives that $\nabla B_p(k) \leq 0$ for $k \geq p$ and since the arguments are positive $Q_{r-p,p+1}(\beta_k^2, \ldots, \beta_{k-p}^2) \geq 0$, so the second sum is ≤ 0. We finally conclude that

$$
\begin{aligned}
S_{r,p} &\leq \sum_{k=p}^{n-1} g(k)A_p(k)B_{p+1}(k)Q_{r-p-1,p+2}(\beta_{k+1}^2, \ldots, \beta_{k-p}^2)\,(\beta_{k+1} + \beta_{k-p}) \\
&= S_{r,p+1},
\end{aligned}
$$

as we have $B_p(k)(\beta_{k+1} - \beta_{k-p}) = B_{p+1}(k)$. □

We are now able to prove the following theorem, which confirms Conjecture 1.1:

Theorem 3.6. *If all $p_i \leq 1/2$ then $H(\mathbf{p})$ is a non-decreasing function of $\mathbf{p}$. Equality holds if and only if each p_i equals 0 or $\frac{1}{2}$.*

Proof. As described in Proposition 2.6, it is sufficient for us to prove that for every $r \geq 1$ we have

$$\sum_{k=0}^{n} f(k)(\alpha_k - 1/2)^{2r+1} = \sum_{k=0}^{n} f(k)\beta_k^{2r+1} \leq 0. \tag{3.5}$$

Using (2.4) we know that $\alpha_k = g(k-1)/(2f(k))$ and $1 - \alpha_k = g(k)/(2f(k))$, so that (subtracting these two expressions)

$$\beta_k = \alpha_k - 1/2 = \frac{(g(k-1) - g(k))}{4f(k)}.$$

This means that, using a standard factorization of $\left(\beta_{k+1}^{2r} - \beta_k^{2r}\right)$, since $g(-1) = g(n) = 0$ we can write

$$
\begin{aligned}
4 \sum_{k=0}^{n} f(k) \beta_k^{2r+1} \\
= \sum_{k=0}^{n} (g(k-1) - g(k)) \beta_k^{2r} \\
= \sum_{k=0}^{n-1} g(k) \left(\beta_{k+1}^{2r} - \beta_k^{2r}\right) \\
= \sum_{k=0}^{n-1} g(k) \left(\beta_{k+1} - \beta_k\right) \left(\beta_{k+1} + \beta_k\right) Q_{r-1,2}(\beta_{k+1}^2, \beta_k^2) \qquad (3.6) \\
= S_{r,1}.
\end{aligned}
$$

However, Proposition 3.5 gives

$$
S_{r,1} \leq S_{r,r+1} = 0,
$$

and we are done. Note that an examination of (3.6) allows us to deduce conditions under which equality holds for the cubic case ($r = 1$). In this case we can rewrite (3.6) using the integration by parts formula (3.3) as

$$
\begin{aligned}
\sum_{k=0}^{n-1} g(k) \left(\beta_{k+1} - \beta_k\right) \left(\beta_{k+1} + \beta_k\right) \\
= \sum_{k=0}^{n-1} g(k) \alpha_k \left((\beta_{k+1} - \beta_k) - (\beta_k - \beta_{k-1})\right) \qquad (3.7)
\end{aligned}
$$

Here $g(k)$ and α_k are positive for $k \geq 1$, and Proposition 2.8 tells us that the second bracket is negative, and so the centered third moment equals zero if and only $\beta_{k+1} - \beta_k$ is constant in k, which means that $\alpha_k = k/n$. However, (2.6) tells us that this implies that

$$
\frac{1}{2} = \sum_{k=0}^{n} f(k) \alpha_k = \frac{1}{n} \sum_{k=0}^{n} f(k) k = \frac{1}{n} \sum_{i=1}^{n} p_i,
$$

so that equality can hold if and only if $p_i \equiv 1/2$. $\qquad \square$

4 Monotonicity of Rényi and Tsallis entropies

As in [8, Section 4] where a similar discussion considered the question of concavity of entropies, we briefly discuss whether Theorem 3.6 may extend to prove that q-Rényi and q-Tsallis entropies are always increasing functions of $\mathbf{p}$ for $p_i \leq 1/2$. We make the following definitions, each of which reduce to the Shannon entropy (1.1) as $q \to 1$.

Definition 4.1. *For $f_{\mathbf{p}}$ as defined above, for $0 \leq q \leq \infty$ define*

1. q-Rényi entropy [14]: $\quad H_{R,q}(\mathbf{p}) = \dfrac{1}{1-q} \log \left(\sum_{x=0}^{n} f_{\mathbf{p}}(x)^q\right),$ \qquad (4.1)

2. q-Tsallis entropy [17]: $\quad H_{T,q}(\mathbf{p}) = \dfrac{1}{q-1} \left(1 - \sum_{x=0}^{n} f_{\mathbf{p}}(x)^q\right).$ \qquad (4.2)

Note that, unlike the concavity case of [8, Section 4], since they are both monotone functions of $\sum_{x=0}^{n} f_{\mathbf{p}}(x)^q$, both $H_{R,q}(\mathbf{p})$ and $H_{T,q}(\mathbf{p})$ will be increasing in $\mathbf{p}$ in the same

cases. We can provide analogues of (1.3) and (1.4) by (for $q \neq 1$)

$$\frac{\partial H_{T,q}}{\partial t} = -\frac{q}{q-1} \sum_{k=0}^{n} \left(g(k-1) - g(k)\right) f(k)^{q-1}, \tag{4.3}$$

$$\frac{\partial^2 H_{T,q}}{\partial t^2} = -q \sum_{k=0}^{n} \left(g(k-1) - g(k)\right)^2 f(k)^{q-2}. \tag{4.4}$$

Again, the second term is negative, and therefore $H_{T,q}(\mathbf{p})$ will be increasing for all $p_n \leq 1/2$ if it is increasing in the case $p_n = 1/2$. Clearly for $q = 0$ (4.3) shows that the entropy is constant (indeed we know that in this case $H_{R,q} = \log(n+1)$ and $H_{T,q} = n$).

Curiously, we can simplify (4.3) in the case of collision entropy ($q = 2$) by substituting for f as a linear combination of g (which is the argument that did not work for $q = 1$).

Lemma 4.2. *For $q = 2$, if $p_n \leq 1/2$*

$$\frac{\partial H_{T,q}}{\partial t} = (1 - 2p_n) \sum_{k=0}^{n} \left(g(k-1) - g(k)\right)^2 \geq 0. \tag{4.5}$$

Proof. In (4.3) we obtain

$$\begin{aligned}
\frac{\partial H_{T,q}}{\partial t} &= -2 \sum_{k=0}^{n} \left(g(k-1) - g(k)\right) f(k) \\
&= -2 \sum_{k=0}^{n} \left(g(k-1) - g(k)\right) \left((1 - p_n)g(k) + p_n g(k-1)\right) \\
&= 2 \sum_{k=0}^{n} (1 - p_n - p_n) g(k)^2 - (1 - p_n - p_n) g(k) g(k-1),
\end{aligned}$$

which is equal to the term stated in (4.5) by relabelling. Note that (curiously) this property will hold for any g, including the mass function of any $B_1 + \ldots + B_{n-1}$ (not necessarily with $p_i < 1/2$). $\qquad\square$

It may be natural to conjecture that Tsallis (and hence Rényi) entropy is increasing for all q. However, the following example shows that this property in fact can fail for $q > 2$ (note that Rényi entropy is not concave in the same range – see [8, Lemma 4.3]).

Example 4.3. Consider $n = 2$ with $p_1 = 1/2 - \epsilon$ and $p_2 = 1/2$. Direct substitution in (4.3) gives that the entropy derivative is exactly

$$-\frac{q 2^{1-q}}{q-1} \left(\left(\frac{1}{2} - \epsilon\right)^q - \left(\frac{1}{2} + \epsilon\right)^q + 2\epsilon \right) = -\frac{q 2^{2-2q}}{q-1} (2^q - 2q)\epsilon + O(\epsilon^3), \tag{4.6}$$

and we note that $2^q - 2q \geq 0$ for $q > 2$, so the leading coefficient is negative and so the derivative will be negative for ϵ sufficiently small.

However, we conjecture that these entropies are increasing for $0 \leq q \leq 2$, since we know that the result holds for $q = 0, 1, 2$:

Conjecture 4.4. *If all $p_i \leq 1/2$ then Tsallis entropy $H_{T,q}(\mathbf{p})$ and Rényi entropy $H_{R,q}(\mathbf{p})$ are non-decreasing functions of $\mathbf{p}$ for $0 \leq q \leq 2$.*

We use an argument similar to that which gave Proposition 2.6 to give a moment-based condition related to this conjecture.

Proposition 4.5. *Let us fix $0 < q < 2$. If, for all $r \geq 1$,*

$$\sum_{k=0}^{n} f(k)^q \beta_k^{2r+1} \leq 0, \tag{4.7}$$

then $\frac{\partial H_{T,q}}{\partial t} \geq 0$ holds.

Proof. We first add a telescoping sum in equation (4.3) to express the derivative as

$$
\begin{aligned}
&-\frac{q}{q-1} \sum_{k=0}^{n} \left(g(k-1) - g(k)\right) f(k)^{q-1} + \frac{1}{q-1} \sum_{k=0}^{n} g(k-1)^q - g(k)^q \\
&= -\frac{1}{q-1} \sum_{k=0}^{n} f(k)^q \left[\left(\frac{g(k)}{f(k)}\right)^q - \left(\frac{g(k-1)}{f(k)}\right)^q + q\left(\frac{g(k)}{f(k)} - \frac{g(k-1)}{f(k)}\right) \right] \\
&= -\frac{1}{q-1} \sum_{k=0}^{n} f(k)^q \left[(2(1-\alpha_k))^q - (2\alpha_k)^q + q\left(2(1-\alpha_k) - 2\alpha_k\right) \right] \\
&= -\frac{1}{q-1} \sum_{k=0}^{n} f(k)^q \left[(1-2\beta_k)^q - (1+2\beta_k)^q + 4\beta_k q \right] \\
&= \sum_{k=0}^{n} f(k)^q \psi_q(2\beta_k),
\end{aligned}
$$

where, using the binomial theorem, the function ψ_q can be expressed as

$$\psi_q(x) = -\frac{(1-x)^q - (1+x)^q + 2qx}{q-1} = \sum_{r=1}^{\infty} \left(q \prod_{i=2}^{2r} (q-i) \right) \frac{1}{(2r+1)!} x^{2r+1}. \tag{4.8}$$

From the assumption $0 < q < 2$, it follows that $q \prod_{i=2}^{2r}(q-i) < 0$. The proof is completed as in Proposition 2.6. □

A Proof of Proposition 2.8

Proof. Using $\alpha_{k+1} = g(k)/(g(k) + g(k+1))$, we can express the difference $\alpha_{k+1} - \alpha_k$ as

$$\frac{g(k)^2 - g(k+1)g(k-1)}{(g(k+1) + g(k))(g(k) + g(k-1))} = \frac{D_g(k)}{(g(k+1) + g(k))(g(k) + g(k-1))},$$

so the property is equivalent to

$$(g(k+2) + g(k+1))D_g(k) \geq (g(k-1) + g(k))D_g(k+1), \tag{A.1}$$

where we write $D_g(k) = g(k)^2 - g(k+1)g(k-1)$.

We write $g^{(i)}$ for the mass function of $\sum_{j \neq i} B_i$, the sum of the first $(n-1)$ Bernoulli variables with the ith one omitted, and write $D^{(i)}(k) = \left(g^{(i)}(k)\right)^2 - g^{(i)}(k+1)g^{(i)}(k-1)$ and $E^{(i)}(k) = g^{(i)}(k)g^{(i)}(k+1) - g^{(i)}(k+2)g^{(i)}(k-1)$.

The cases $q = 1$ and $q = 2$ of [7, Lemma A1] give that

$$g(k)g(k-1) = \sum_{i=1}^{n-1} p_i(1-p_i)D^{(i)}(k-1), \tag{A.2}$$

$$2g(k-1)g(k+1) = \sum_{i=1}^{n-1} p_i(1-p_i)E^{(i)}(k-1), \tag{A.3}$$

and by direct substitution we deduce that

$$g(k)^2 g(k+1) + g(k-1)g(k)g(k+2) - 2g(k-1)g(k+1)^2$$
$$= \sum_{i=1}^{n-1} p_i(1-p_i) \left(g(k)D^{(i)}(k) + g(k+2)D^{(i)}(k-1) - g(k+1)E^{(i)}(k-1) \right)$$
$$= \sum_{i=1}^{n-1} p_i(1-p_i) \left((1-p_i) \frac{\left(D^{(i)}(k) \right)^2 - D^{(i)}(k-1)D^{(i)}(k+1)}{g^{(i)}(k)} \right), \tag{A.4}$$

where the second equality follows directly by taking $g(k) = (1-p_i)g^{(i)}(k) + p_i g^{(i)}(k-1)$ for each i and simplifying inside the bracket. Similarly, we can find

$$g(k)g(k+1)^2 + g(k-1)g(k+1)g(k+2) - 2g(k)^2 g(k+2)$$
$$= \sum_{i=1}^{n-1} p_i(1-p_i) \left(g(k+1)D^{(i)}(k) + g(k-1)D^{(i)}(k+1) - g(k)E^{(i)}(k) \right)$$
$$= \sum_{i=1}^{n-1} p_i(1-p_i) \left(p_i \frac{\left(D^{(i)}(k) \right)^2 - D^{(i)}(k-1)D^{(i)}(k+1)}{g^{(i)}(k)} \right). \tag{A.5}$$

Comparing (A.4) and (A.5), using the fact that $\left(D^{(i)}(k) \right)^2 - D^{(i)}(k-1)D^{(i)}(k+1) \geq 0$ (see [3] or [8, Section 2.2]) we deduce that if all $p_i \leq 1/2$ then

$$g(k)^2 g(k+1) + g(k-1)g(k)g(k+2) - 2g(k-1)g(k+1)^2$$
$$\geq \quad g(k)g(k+1)^2 + g(k-1)g(k+1)g(k+2) - 2g(k)^2 g(k+2),$$

which by rearranging is precisely (A.1) above. $\qquad\square$

B Proof of technical lemmas

Proof of Lemma 3.2. Given $j \geq 0$, each sequence $(\beta_{k+1} - \beta_{k-j})_k$ is non-negative and non-increasing, so any product of such sequences is also non-increasing. In more detail, the

$$\nabla B_p(k) = B_p(k) - B_p(k-1) = \prod_{j=0}^{p-1}(\beta_{k+1} - \beta_{k-j}) - \prod_{j=0}^{p-1}(\beta_k - \beta_{k-j-1}),$$

and for each $j \geq 0$, the $(\beta_{k+1} - \beta_{k-j}) \leq (\beta_k - \beta_{k-j-1})$ because Proposition 2.8 gives $(\beta_{k+1} - \beta_k) \leq (\beta_{k-j} - \beta_{k-j-1})$, and each term in the product is positive by (2.3). Further, each of these terms is well-defined since $k - j - 1 \geq k - p \geq 0$. $\qquad\square$

Proof of Lemma 3.3. Using the facts that $g(n) = 0$ and that $A_{p+1}(0) = 0$, we write the RHS of (3.3) as

$$\sum_{k=p+1}^{n-1} A_{p+1}(k)g(k)v(k) - \sum_{k=p+1}^{n} A_{p+1}(k)g(k)v(k-1)$$
$$= \sum_{k=p}^{n-1} v(k) \left(g(k)A_{p+1}(k) - g(k+1)A_{p+1}(k+1) \right)$$
$$= \sum_{k=p}^{n-1} g(k)A_p(k)v(k) \left(\alpha_{k-p} - \alpha_{k+1}\frac{g(k+1)}{g(k)} \right).$$

Here, we use the facts that by definition $A_{p+1}(k) = A_p(k)\alpha_{k-p}$ and $A_{p+1}(k+1) = A_p(k)\alpha_{k+1}$. We conclude by substituting for $\alpha_{k+1}g(k+1)$ using (2.5), and noting that $\alpha_{k+1} + \alpha_{k-p} - 1 = \beta_{k+1} + \beta_{k-p}$. $\qquad\square$

To prove Lemma 3.4, we observe that equivalently, the family of polynomials $(Q_{n,p}(X_1,\ldots,X_p))_{n\geq 0}$ can be defined using the generating function:

$$\sum_{n=0}^{\infty} Q_{n,p}(X_1,\ldots,X_p)t^n = \frac{1}{(1-tX_1)(1-tX_2)\ldots(1-tX_p)} \tag{B.1}$$

Proof of Lemma 3.4. We first notice by direct calculation that

$$\frac{1}{(1-tX_2)\ldots(1-tX_{p+1})} - \frac{1}{(1-tX_1)\ldots(1-tX_p)}$$
$$= \frac{t(X_{p+1}-X_1)}{(1-tX_1)\ldots(1-tX_{p+1})}.$$

Using equation (B.1), we thus have

$$\sum_{n=0}^{\infty} Q_{n,p}(X_2,\ldots,X_{p+1})t^n - \sum_{n=0}^{\infty} Q_{n,p}(X_1,\ldots,X_p)t^n$$

$$= (X_{p+1}-X_1)\sum_{n=0}^{\infty} Q_{n,p+1}(X_1,\ldots,X_{p+1})t^{n+1}$$

$$= (X_{p+1}-X_1)\sum_{n=1}^{\infty} Q_{n-1,p+1}(X_1,\ldots,X_{p+1})t^n,$$

which yields equation (3.4) for $n \geq 1$. The case $n = 0$ is obvious. $\square$

C Heuristics in continuous case

We now explain some calculations in the continuous case that helped us to find a rigorous proof of Theorem 3.6, and that help suggest our conjecture about Renyi and Tsallis entropies. We remark that Ordentlich [13] used the original paper of Shepp and Olkin [16] to motivate conjectures concerning continuous random variables.

Let us consider a density function $f(x)$, defined for $x \in \mathbb{R}$, which is assumed to be everywhere positive, smooth and with all derivatives well-behaved at $\pm\infty$. This density will serve as a continuous analogue of both the mass functions $(f(k))_k$ and $(g(k))_k$. As a consequence, one could also see the function $\frac{1}{2} - \frac{\log(f)'(x)}{4}$ (resp. $-\frac{\log(f)'(x)}{4}$) as continuous analogues of the family $(\alpha)_k$ (resp. $(\beta)_k$). We will make the assumptions that $(\log f)'' \leq 0$ and $(\log f)''' \geq 0$, which correspond to the property $\alpha_k \leq \alpha_{k+1}$ and $\alpha_k - 2\alpha_{k+1} + \alpha_{k+2} \leq 0$. We now prove the continuous version of equation (3.5) (in the case where $q = 1$) and (4.7) (for $q \neq 1$):

Proposition C.1. *Suppose that* $(\log f)'' \leq 0$ *and* $(\log f)''' \geq 0$ *then for every real parameter* $q > 0$ *and every integer* $r \geq 1$ *we have:*

$$\int_{\mathbb{R}} f(x)^q (\log f(x)')^{2r+1} dx \geq 0.$$

Proof. For any $0 \leq p \leq r$ we set

$$I_{r,p} = (-1)^p \frac{A_p}{q^p} \int_{\mathbb{R}} f(x)^q \left(\log f(x)'\right)^{2r-2p+1} \left(\log f(x)''\right)^p dx,$$

with $A_0 = 1$ and $A_p = \prod_{k=0}^{p-1}(2r-2k)$ for $p \geq 1$. In particular $A_r = 0$ so $I_{r,r} = 0$. We now prove that the sequence $(I_{r,p})$ is non-increasing in p: for every $0 \leq p \leq r-1$, using the

fact that $f^q(\log f)' = f^q f'/f = f^{q-1}f' = (f^q)'/q$ we have:

$$
\begin{aligned}
I_{r,p} &= (-1)^p \frac{A_p}{q^{p+1}} \int_{\mathbb{R}} (f(x)^q)' (\log f(x)')^{2r-2p} (\log f(x)'')^p dx \\
&= (-1)^{p+1} \frac{A_p}{q^{p+1}} (2r - 2p) \int_{\mathbb{R}} f(x)^q (\log f(x)')^{2r-2p-1} (\log f(x)'')^{p+1} dx \\
&\quad + (-1)^{p+1} \frac{A_p}{q^{p+1}} p \int_{\mathbb{R}} f(x)^q (\log f(x)')^{2r-2p} (\log f(x)'')^{p-1} \log f(x)''' dx.
\end{aligned}
$$

Here, again we apply integration by parts, followed by the product rule. The first integral is exactly $I_{r,p+1}$, and the second one is non-negative because of the assumptions on $\log f''$ and $\log f'''$. We thus have $I_{r,p} \geq I_{r,p+1}$. We thus have

$$
\int_{\mathbb{R}} f(x)^q (\log f(x)')^{2r+1} dx = I_{r,0} \geq I_{r,r} = 0,
$$

which proves the result. □

References

[1] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der mathematischen Wissenschaften, vol. 348, Springer, 2014. MR-3155209

[2] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393. MR-1738163

[3] P. Brändén, *Iterated sequences and the geometry of zeros*, J. Reine Angew. Math. **658** (2011), 115–131. MR-2831515

[4] P. Caputo, P. Dai Pra, and G. Posta, *Convex entropy decay via the Bochner-Bakry-Emery approach*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 3, 734–753. MR-2548501

[5] P. Harremoës, *Binomial and Poisson distributions as maximum entropy distributions*, IEEE Trans. Information Theory **47** (2001), no. 5, 2039–2041. MR-1842536

[6] E. Hillion, *Concavity of entropy along binomial convolutions*, Electron. Commun. Probab. **17** (2012), no. 4, 1–9. MR-2872573

[7] E. Hillion and O. T. Johnson, *Discrete versions of the transport equation and the Shepp-Olkin conjecture*, Annals of Probability **44** (2016), no. 1, 276–306. MR-3456338

[8] E. Hillion and O. T. Johnson, *A proof of the Shepp-Olkin entropy concavity conjecture*, Bernoulli **23** (2017), no. 4B, 3638–3649. MR-3654818

[9] O. T. Johnson, *A discrete log-Sobolev inequality under a Bakry-Émery type condition*, Annales de l'Institut Henri Poincaré B (Probability and Statistics) **53** (2017), no. 4, 1952–1970. MR-3729642

[10] S. Karlin and Y. Rinott, *Entropy inequalities for classes of probability distributions. I. The univariate case*, Adv. in Appl. Probab. **13** (1981), no. 1, 93–112. MR-0595889

[11] P. Mateev, *The entropy of the multinomial distribution*, Teor. Verojatnost. i Primenen. **23** (1978), no. 1, 196–198. MR-0490451

[12] C. P. Niculescu, *A new look at Newton's inequalities*, JIPAM. J. Inequal. Pure Appl. Math. **1** (2000), Issue 2, Article 17. MR-1786404

[13] E. Ordentlich, *Maximizing the entropy of a sum of independent bounded random variables*, IEEE Transactions on Information Theory **52** (2006), no. 5, 2176–2181. MR-2234474

[14] A. Rényi, *On measures of entropy and information*, Proceedings of the 4th Berkeley Conference on Mathematical Statistics and Probability (Berkeley) (J. Neyman, ed.), University of California Press, 1961, pp. 547–561. MR-0132570

[15] L. A. Shepp and I. Olkin, *Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution*, Tech. Report 131, Stanford University, 1978.

[16] L. A. Shepp and I. Olkin, *Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution*, Contributions to probability, Academic Press, New York, 1981, pp. 201–206. MR-0618689

[17] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, Journal of Statistical Physics **52** (1988), 479–487. MR-0968597

[18] Y. Yu and O. T. Johnson, *Concavity of entropy under thinning*, Proceedings of the 2009 IEEE International Symposium on Information Theory, 28th June–3rd July 2009, Seoul, 2009, pp. 144–148.

[19] Y. Yu, *On the maximum entropy properties of the binomial distribution*, IEEE Trans. Inform. Theory **54** (2008), no. 7, 3351–3353. MR-2450793