DISCRETE VERSIONS OF THE TRANSPORT EQUATION AND THE SHEPP–OLKIN CONJECTURE[1]

By Erwan Hillion and Oliver Johnson

University of Luxembourg and University of Bristol

We introduce a framework to consider transport problems for integer-valued random variables. We introduce weighting coefficients which allow us to characterize transport problems in a gradient flow setting, and form the basis of our introduction of a discrete version of the Benamou–Brenier formula. Further, we use these coefficients to state a new form of weighted log-concavity. These results are applied to prove the monotone case of the Shepp–Olkin entropy concavity conjecture.

1. Introduction. In recent years, there has been intensive study of relationships between entropy and probabilistic inequalities. Since the work of Monge in the eighteenth century, it has been understood how one probability measure on $\mathbb{R}$ can be smoothly transformed into another along a path minimizing an appropriate cost. As described, for example, in [33, 34], use of the quadratic cost function induces the quadratic Wasserstein distance W_2 which, using the Benamou–Brenier formula [5, 6], can be understood in terms of velocity fields arising in gradient models of the kind discussed in [2, 8, 19]. In such models, concavity of entropy along the geodesic plays a central role, giving proofs of inequalities such as HWI, log-Sobolev and transport inequalities; see, for example, [9]. A key role is played by log-concavity of the underlying measures and the Ricci curvature of the underlying metric space; see, for example, [22, 31, 32].

However, this work has almost exclusively focused on continuous random variables, taking values in $\mathbb{R}^d$, or more generally on Riemannian manifolds satisfying a curvature condition. In this paper, we propose a framework for considering similar problems for integer-valued random variables. We show how many natural models of transportation of discrete random variables can be considered as gradient models and propose a discrete version of the Benamou–Brenier formula. As an example of the insights gained by this approach, we give a proof of a significant new case of the Shepp–Olkin concavity conjecture [30], which has remained unresolved for over 30 years.

Shepp and Olkin considered sums of n independent Bernoulli variables (referred to as Bernoulli sums throughout this article), with parameters $p_1, \ldots, p_n$,

respectively, where $p_i \in [0, 1]$, and n remains fixed. This sum has a probability distribution $(f_k)_{k=0,1,\ldots,n}$, and Shepp and Olkin conjectured [30] that its entropy is a concave function of its parameters.

CONJECTURE 1.1 ([30]).   *Consider the entropy of* $(f_k)_{k=0,1,\ldots,n}$, *defined by*

$$H(p_1, \ldots, p_n) := -\sum_{k=0}^{n} f_k \log(f_k),$$

where by convention $0 \log(0) = 0$. *If* $p_1, \ldots, p_n : [0, 1] \to [0, 1]$ *are affine functions, then*

$$H : [0, 1] \to \mathbb{R}, \qquad t \mapsto H\big(p_1(t), \ldots, p_n(t)\big)$$

is a concave function in t.

We emphasize that Conjecture 1.1 refers to concavity of entropy in the parameter space. This should be contrasted with concavity in the space of mass functions themselves. The result that the entropy of mixtures of mass functions $f_k(\alpha) := (1 - \alpha)f_k^{(1)} + \alpha f_k^{(2)}$ is convex in α is standard; see, for example, [10], Theorem 2.7.3. Indeed duality between this parameter representation and the distribution space is exploited in information geometry (see, e.g., [1]), where the concavity of entropy also plays a central role.

Although the Shepp–Olkin Conjecture 1.1 remains open, we briefly describe the main cases which had previously been resolved. First, Shepp and Olkin's original paper [30] showed that the entropy is a Schur-concave function, stated that Conjecture 1.1 holds for $n = 2, 3$ and proved it for interpolation between two binomials, when $p_i(t) = t$ for all i.

Second, Theorem 2 of Yu and Johnson [37] proves concavity of the entropy of $H(T_t X + T_{1-t} Y)$, for X and Y satisfying the ultra log-concavity property (see Definition 3.11 below), where T_t represents Rényi's thinning operation [28]; see equation (36) below. As remarked in [37], Corollary 1, this resolves the special case of Conjecture 1.1 where each parameter is either $p_i(t) = p_i(0)(1 - t)$ or $p_i(t) = p_i(1)t$.

Third, Theorem 1.1 of Hillion [14] resolves the case where for each i, either $p_i(t) \equiv p_i(0)$ for all t or $p_i(t) = t$ (the translation case of Example 3.10 below).

In this article, given a family of affine functions $p_1(t), \ldots, p_n(t)$, we consider the associated Bernoulli sum $(f_k(t))_{k=0,1,\ldots,n}$ as a function of the spatial variable k and the time variable t. We often omit the explicit dependence of f_k on t. Throughout this article we restrict our attention to the special case that $p_i' \geq 0$ for every $i \in \{1, \ldots, n\}$, so the random variables $f_k(t)$ satisfy a stochastic ordering property. We write the left derivative $\nabla_1 f_k = f_k - f_{k-1}$, and write $\nabla_2 = (\nabla_1)^2$ for the map taking $\nabla_2 f_k = f_k - 2f_{k-1} + f_{k-2}$.

The paper is organized as follows. In Section 2 we review properties of continuous gradient flow models and develop a framework to prove concavity of entropy. We introduce and discuss the Benamou–Brenier formula, equation (12).

In Section 3 we introduce a formalism to describe interpolation of discrete probability mass functions f_k, motivated by properties of the binomial mass functions. In Definition 3.2, we propose a discrete analogue of the Benamou–Brenier formula. A key role is played by our introduction of a family of functions $\alpha_k(t)$ which are used to generate mixtures of f_k and f_{k+1}. We write $\mathcal{A}$ for the set of measurable functions $\boldsymbol{\alpha}(t) = (\alpha_0(t), \alpha_1(t), \ldots, \alpha_n(t))$, where $\alpha_0(t) \equiv 0$ and $\alpha_n(t) \equiv 1$, and $0 \leq \alpha_k(t) \leq 1$ for all k and t.

Our formula of Benamou–Brenier type motivates the following definition:

DEFINITION 1.2.   We say that a family of probability mass functions $f_k(t)$ supported in $\{0, \ldots, n\}$ is a constant velocity path if for some v and for some family of probability mass functions $g_k^{(\boldsymbol{\alpha})}(t)$ supported in $\{0, \ldots, n-1\}$, it satisfies a modified transport equation

$$(1) \qquad \frac{\partial f_k}{\partial t}(t) = -v\nabla_1\big(g_k^{(\boldsymbol{\alpha})}(t)\big) \qquad \text{for } k = 0, 1, \ldots, n,$$

where for some $\boldsymbol{\alpha}(t) \in \mathcal{A}$,

$$(2) \quad g_k^{(\boldsymbol{\alpha})}(t) = \alpha_{k+1}(t)f_{k+1}(t) + \big(1 - \alpha_k(t)\big)f_k(t) \qquad \text{for } k = 0, 1, \ldots, n-1.$$

If $f_k(t)$ satisfies a modified transport equation, we write h for the function (not necessarily a probability mass function) satisfying a second-order modified transport equation

$$(3) \qquad \frac{\partial^2 f_k}{\partial t^2}(t) = v^2\nabla_2(h_k) \qquad \text{for } k = 0, 1, \ldots, n.$$

An explicit expression for h_k is given in equation (45).

In Section 3.3 we discuss the constant velocity path property. Lemma 3.5 shows that any such representation is unique, and we give some examples which lie within this framework. In Section 4, by examining the coefficients $\boldsymbol{\alpha} \in \mathcal{A}$ associated with a constant velocity path $f_k(t)$, called optimal coefficients for $f_k(t)$, we give sufficient conditions for the concavity of entropy along the interpolation. To be specific, we introduce the following three conditions:

CONDITION 1 (k-MON).   *Given t, we say the $\alpha_k(t)$ are k-monotone at t if*

$$(4) \qquad\qquad \alpha_k(t) \leq \alpha_{k+1}(t) \qquad \text{for all } k = 0, \ldots, n-1.$$

CONDITION 2 (t-MON).   *Given t, we say the $\alpha_k(t)$ are t-monotone at t if*

$$(5) \qquad\qquad \frac{\partial \alpha_k}{\partial t}(t) \geq 0 \qquad \text{for all } k = 0, \ldots, n.$$

CONDITION 3 (GLC). *We say probability mass function f_k supported on* $\{0, 1, \ldots, n\}$ *is $\boldsymbol{\alpha}$-generalized log-concave at t, denoted* $\mathbf{GLC}(\boldsymbol{\alpha}(t))$, *if for all* $k = 0, \ldots, n - 2$,

(6) $$\mathrm{GLC}(\boldsymbol{\alpha})_k(t) \geq 0,$$

where

(7)
$$\mathrm{GLC}(\boldsymbol{\alpha})_k(t) := \alpha_{k+1}(t)\big(1 - \alpha_{k+1}(t)\big) f_{k+1}^2(t)$$
$$- \alpha_{k+2}(t)\big(1 - \alpha_k(t)\big) f_k(t) f_{k+2}(t).$$

In Section 4 we prove the following theorem:

THEOREM 1.3. *Consider a constant velocity path $f_k(t)$ of probability mass functions and associated optimal $\boldsymbol{\alpha}(t)$. If Conditions 1, 2 and 3 hold for a given $t = t^*$, then the entropy $H(f(t))$ is concave in t at $t = t^*$.*

EXAMPLE 1.4. Theorem 1.3 gives a new perspective on Shepp and Olkin's proof [30] that the entropy of binomial $\mathrm{Bin}(n, t)$ random variables [with probability mass function $f_k(t) = \binom{n}{k} t^k (1 - t)^{n-k}$ for $k = 0, \ldots, n$] is concave in t. In this case, Example 3.1 shows that the optimal $\alpha_k(t) \equiv k/n$, so the k-monotone and t-monotone conditions, Conditions 1 and 2 are clear. Further, $\alpha_k(t) \equiv k/n$ means that GLC Condition 3 reduces to the ultra log-concavity of order n of Pemantle [27] and Liggett [21] (see Definition 3.11), which clearly holds with equality in this case. In fact, in the more general "symmetric case" where p_i' does not depend on i, Remark 5.3 shows $\alpha_k(t) = k/n$, and a similar argument applies.

Finally, we use this formalism to consider the Shepp–Olkin Conjecture 1.1, in the "monotone" setting $p_i' \geq 0$ for all i. In Section 5, we show that in the monotone case, the Shepp–Olkin interpolation is a constant velocity path in the sense of Definition 1.2. In Proposition 5.2 we show that the k-monotone Condition 1 is automatically satisfied in this case. Similarly Proposition 5.4 shows that GLC, Condition 3, also holds for all Shepp–Olkin interpolations in this context.

Unfortunately t-monotonicity, Condition 2, does not hold for all (monotone) Shepp–Olkin interpolations. However, Theorem 4.4 weakens the assumptions of Theorem 1.3, by proving that entropy remains concave if we replace Condition 2 by Condition 4, which is less restrictive, although less transparent in nature. We complete the proof of the monotone Shepp–Olkin conjecture by showing in Lemma 6.2 that Condition 4 is satisfied in this case. The proof uses properties of Bernoulli sum mass functions, including a "cubic" inequality Theorem A.2. In Section 6, we therefore prove the main result of this paper:

THEOREM 1.5 (Monotone Shepp–Olkin). *Consider the entropy of* $(f_k)_{k=0,1,...,n}$, *defined by*

$$H(p_1, \ldots, p_n) := -\sum_{k=0}^{n} f_k \log(f_k).$$

If $p_1, \ldots, p_n : [0, 1] \to [0, 1]$ *are affine functions with* $p_i' \geq 0$ *for all* i, *then the function*

$$H : [0, 1] \to \mathbb{R}, \qquad t \mapsto H\big(p_1(t), \ldots, p_n(t)\big)$$

is concave in t.

2. Geodesics for continuous random variables.

2.1. *General framework for concavity of entropy.* In Section 2 we restate results concerning entropy and geodesics for random variables on $\mathbb{R}$, using the following differential equation framework, where the form of (8) motivates equations (1) and (3):

THEOREM 2.1. *Let* $(f_t(x))_{t \in [0,1]}$ *be a smooth family of positive probability densities on* $\mathbb{R}$, *such that the entropy* $H(t) := -\int_{\mathbb{R}} f_t(x) \log(f_t(x)) \, dx$ *exists for all* t. *Consider the families of functions* $(g_t(x))_{t \in [0,1]}$ *and* $(h_t(x))_{t \in [0,1]}$ *which satisfy*

$$(8) \qquad \frac{\partial f_t(x)}{\partial t} = -\frac{\partial g_t(x)}{\partial x}, \qquad \frac{\partial^2 f_t(x)}{\partial t^2} = \frac{\partial^2 h_t(x)}{\partial x^2}.$$

Under technical conditions, such as those listed in Remark 2.2, the entropy $H(t)$ *satisfies*

$$(9) \qquad \begin{aligned} H''(t) = &-\int_{\mathbb{R}} \left(h_t(x) - \frac{g_t(x)^2}{f_t(x)} \right) \frac{\partial^2}{\partial x^2} (\log(f_t(x))) \, dx \\ &- \int_{\mathbb{R}} f_t(x) \left(\frac{\partial}{\partial x} \left(\frac{g_t(x)}{f_t(x)} \right) \right)^2 dx. \end{aligned}$$

PROOF. The two conditions listed in part (a) of Remark 2.2 are those required under Leibniz's rule for differentiation under the integral sign, yielding

$$(10) \qquad \begin{aligned} H''(t) &= -\int_{\mathbb{R}} \frac{\partial^2 f_t(x)}{\partial t^2} \log(f_t(x)) \, dx - \int_{\mathbb{R}} \frac{1}{f_t(x)} \left(\frac{\partial f_t(x)}{\partial t} \right)^2 dx \\ &= -\int_{\mathbb{R}} \frac{\partial^2 h_t(x)}{\partial x^2} \log(f_t(x)) \, dx - \int_{\mathbb{R}} \frac{1}{f_t(x)} \left(\frac{\partial g_t(x)}{\partial x} \right)^2 dx. \end{aligned}$$

Here, by part (b) of Remark 2.2 $g_t(x)$ vanishes at $x = \pm\infty$, meaning that the term $\int_{\mathbb{R}} \frac{\partial f_t(x)}{\partial t} \, dx = 0$. Using the quotient rule we can write the second term in equation (10) as

$$-\int_{\mathbb{R}} \frac{1}{f_t(x)} \left(\frac{\partial g_t(x)}{\partial x}\right)^2 dx$$

$$= -\int_{\mathbb{R}} \left(\frac{\partial f_t(x)}{\partial x}\right)^2 \frac{g_t(x)^2}{f_t(x)^3}$$

$$+ \frac{\partial f_t(x)}{\partial x}\left[2\frac{g_t(x)}{f_t(x)} \frac{\partial}{\partial x}\left(\frac{g_t(x)}{f_t(x)}\right)\right] + f_t(x)\left(\frac{\partial}{\partial x}\left(\frac{g_t(x)}{f_t(x)}\right)\right)^2 dx$$

$$= -\int_{\mathbb{R}} \left(\frac{\partial f_t(x)}{\partial x}\right)^2 \frac{g_t(x)^2}{f_t(x)^3}$$

$$- \frac{\partial^2 f_t(x)}{\partial x^2}\left(\frac{g_t(x)}{f_t(x)}\right)^2 dx + f_t(x)\left(\frac{\partial}{\partial x}\left(\frac{g_t(x)}{f_t(x)}\right)\right)^2 dx$$

since we recognize the term in square brackets as a perfect derivative, and integrate by parts. The remaining conditions listed in Remark 2.2(b) justify the necessary integrations by parts to prove the theorem. □

REMARK 2.2.   We assume, for example, that the following technical conditions hold:

(a) there exist integrable $\theta_A(x), \theta_B(x)$ such that for all t, x, $|\frac{\partial g_t(x)}{\partial x}(1 + \log(f_t(x)))| \leq \theta_A(x)$ and $|\frac{\partial^2 h_t(x)}{\partial x^2}\log(f_t(x)) + (\frac{\partial g_t(x)}{\partial x})^2\frac{1}{f_t(x)}| \leq \theta_B(x)$;

(b) for each $t \in [0, 1]$, functions $g_t(x)$, $\frac{\partial h_t(x)}{\partial x}\log(f_t(x))$, $\frac{h_t(x)}{f_t(x)}\frac{\partial f_t(x)}{\partial x}$ and $(\frac{g_t(x)}{f_t(x)})^2\frac{\partial f_t(x)}{\partial x}$ vanish at $x = \pm\infty$.

Section 2 aims to motivate results in the case where all random variables have support on a finite set, so the required differentiation formulas are automatic. For this reason, we do not discuss the question of verification of the technical conditions of Remark 2.2.

In some sense, an extreme example for which we can apply Theorem 2.1 is the following:

EXAMPLE 2.3.   Consider the translation of probability density f_0, where $f_t(x) := f_0(x - vt)$ for some constant velocity $v > 0$. It is then easy to see that $g_t(x) = vf_t(x)$ and $h_t(x) = v^2 f_t(x)$. Theorem 2.1 then confirms shift invariance and makes the entropy $H(t)$ of f_t constant.

2.2. *Benamou–Brenier formula.* The study of geodesics interpolating between continuous probability densities exploits properties of the quadratic Wasserstein distance W_2, which (see [2, 5, 6]) has a variational characterisation involving velocity fields, given by the Benamou–Brenier formula (12) below.

DEFINITION 2.4. Consider fixed smooth distribution functions F_0 and F_1. Write $\mathcal{P}_\mathbb{R}(F_0, F_1)$ for the set of probability densities $f_t(x)$, with corresponding distribution functions $F_t(x) = \int_{-\infty}^x f_t(y)\,dy$ satisfying constraints $F_t(x)|_{t=0} = F_0(x)$ and $F_t(x)|_{t=1} = F_1(x)$. Then given any sequence $f_t \in \mathcal{P}_\mathbb{R}(F_0, F_1)$, we refer to a function v_t as a velocity field if it satisfies

$$(11) \qquad \frac{\partial}{\partial t} f_t(x) + \frac{\partial}{\partial x}\big(v_t(x) f_t(x)\big) = 0.$$

Ambrosio, Gigli and Savaré [2], Section 8, give a careful analysis of conditions under which this type of continuity equation holds. They consider (see [2], Definition 1.1.1) the class of absolutely continuous curves $\mu_t \in \mathcal{P}_p(X)$, the set of probability measures with finite pth moment on separable Hilbert space X. Theorem 8.3.1 of [2] shows that for $p > 1$, a version of equation (11) holds for μ_t in this class, in fact,

$$\frac{\partial}{\partial t}\mu_t + \nabla \cdot (v_t \mu_t) = 0,$$

in the sense of distributions (using the class of smooth cylindrical test functions).

Further, Theorem 8.3.1 of [2] shows that under these conditions the resulting velocity field has $L^p(\mu_t)$ norm dominated by the metric derivative $|\mu'|(t)$ (as defined in [2], equation (1.1.3)). Using properties of so-called length spaces, this allows the following formula, first proved by Benamou and Brenier [5, 6] for probability measures on $X = \mathbb{R}^d$, to be recovered for separable Hilbert spaces X. For comparison purposes, we state this Benamou–Brenier formula for the case of $X = \mathbb{R}$:

THEOREM 2.5 ([5, 6]). *Using the notation of Definition 2.4 above, the quadratic Wasserstein distance is given by*

$$(12) \qquad W_2^2(F_0, F_1) = \inf_{f_t \in \mathcal{P}_\mathbb{R}(F_0, F_1)} \int_0^1 \left(\int_{-\infty}^\infty f_t(y) v_t(y)^2\,dy \right) dt$$

$$(13) \qquad = \inf_{f_t \in \mathcal{P}_\mathbb{R}(F_0, F_1)} \int_0^1 \left(\int_{-\infty}^\infty \left(\frac{\partial F_t}{\partial t}(y) \right)^2 \frac{1}{f_t(y)}\,dy \right) dt.$$

COROLLARY 2.6. *If $(f_t)_{t\in[0,1]}$ is a solution to the minimization problem (12), then assuming the technical conditions of Remark 2.2 hold, we can write*

$$(14) \qquad H''(f_t) = -\int_\mathbb{R} f_t(x) \left(\frac{\partial}{\partial x} v_t(x) \right)^2 dx \le 0,$$

and the inner integral of (12), $\int_{-\infty}^\infty f_t(x) v_t(x)^2\,dx$, is constant in t.

PROOF. It is shown in [5], equation (1.14), that if $(f_t)_{t\in[0,1]}$ is a solution to the minimization problem (12), then its associated velocity field $v_t(x)$ is, at least formally, a solution to the equation

$$(15) \qquad \frac{\partial v_t(x)}{\partial t} = -\frac{\partial v_t(x)}{\partial x} v_t(x).$$

Taking a further time derivative of (11) and using (15), we deduce a second-order PDE,

$$\begin{aligned}
\frac{\partial^2 f_t(x)}{\partial t^2} &= -\frac{\partial}{\partial x}\left(\frac{\partial v_t}{\partial t}(x) f_t(x) + v_t(x)\frac{\partial f_t}{\partial t}(x) \right) \\
(16) \qquad &= \frac{\partial}{\partial x}\left(\frac{\partial v_t(x)}{\partial x} v_t(x) f_t(x) + v_t(x)\frac{\partial}{\partial x}\left(v_t(x) f_t(x)\right) \right) \\
&= \frac{\partial^2}{\partial x^2}\left(v_t(x)^2 f_t(x)\right)
\end{aligned}$$

(assuming the t and x derivatives can be exchanged). In the notation of Theorem 2.1, we can rewrite equations (11) and (16) in the form $g_t(x) = v_t(x) f_t(x)$ and $h_t(x) = v_t(x)^2 f_t(x)$. This makes a clear analogy with the translation case, Example 2.3, and equation (14) follows by a straightforward application of Theorem 2.1. Similar calculations using equation (14) show that $\frac{\partial}{\partial t}(f_t(x)v_t(x)^2) = \frac{\partial}{\partial x}(f_t(x)v_t(x)^3)$, and the result follows. □

This result can be seen as a particular case of results coming from Sturm–Lott–Villani theory [22, 31, 32]. This theory establishes links between the behavior of the entropy functional along Wasserstein W_2-geodesics on a given measured length space and bounds on the Ricci curvature on this space. In particular, a Riemannian manifold (M, g) satisfies Ric ≥ 0, where Ric is the Ricci curvature tensor, if and only if for every absolutely continuous Wasserstein W_2-geodesic $(\mu_t)_{t\in[0,1]} := (f_t\, d\mathrm{vol})_{t\in[0,1]}$ the entropy function $H(t) := -\int_M f_t \log(f_t)\, d\mathrm{vol}$ is concave in t. This equivalence is used to generalize the definition of Ricci curvature bounds from the Riemannian framework to the framework of measured length spaces; that is, metric spaces (X, d) for which the distance $d(x, y)$ is the infimum of lengths of curves joining x to y.

This theory can be developed to use transportation arguments to prove probabilistic inequalities involving entropy, such as log-Sobolev, transport and HWI inequalities. For example, Cordero–Erausquin [9], Corollaries 1, 2 and 3, gives simple proofs of these inequalities, under the condition that relative density $f/\phi_{1/c}$ is log-concave (in the continuous sense), where $\phi_{1/c}$ is a normal density with variance $1/c$. This log-concavity condition is known to imply the Bakry–Émery condition [4] (see, e.g., [3, 9]), which is natural in this context. GLC, Condition 3, is introduced as a discrete version of the log-concavity condition.

2.3. *Perturbed translations.*

THEOREM 2.7. *Let $(f_t(x))_{t \in [0,1]}$ be a smooth family of positive probability densities on $\mathbb{R}$ and $(g_t(x))_{t \in [0,1]}$ be defined by equation (8). If there exists a constant v and a nondecreasing function $\alpha(t)$ such that*

$$(17) \qquad g_t(x) = v f_t(x) + \alpha(t)\frac{\partial}{\partial x} f_t(x),$$

then assuming the technical conditions of Remark 2.2 hold, the entropy $H(t)$ of f_t is a concave function of t.

PROOF. Using the facts that $\frac{\partial}{\partial t} f_t(x) = -\frac{\partial}{\partial x} g_t(x)$, and hence $\frac{\partial^2}{\partial x \, \partial t} f_t(x) = -\frac{\partial^2}{\partial x^2} g_t(x)$, we take a derivative of equation (17) to compute

$$\frac{\partial}{\partial t} g_t(x) = -v\frac{\partial}{\partial x}\left(v f_t(x) + \alpha(t)\frac{\partial}{\partial x} f_t(x)\right) + \alpha'(t)\frac{\partial}{\partial x} f_t(x)$$

$$- \alpha(t)\frac{\partial^2}{\partial x^2}\left(v f_t(x) + \alpha(t)\frac{\partial}{\partial x} f_t(x)\right),$$

so the family of functions $h_t(x)$ defined by equation (8) is equal to

$$(18) \quad h_t(x) = v^2 f_t(x) + 2v\alpha(t)\frac{\partial}{\partial x} f_t(x) + \alpha(t)^2\frac{\partial^2}{\partial x^2} f_t(x) - \alpha'(t) f_t(x).$$

It is then easy to deduce that

$$(19) \qquad h_t(x) - \frac{g_t(x)^2}{f_t(x)} = \alpha(t)^2 f_t(x)\left(\frac{\partial^2}{\partial x^2}\log(f_t(x))\right) - \alpha'(t) f_t(x).$$

Since in this case $\frac{\partial}{\partial x}(g_t(x)/f_t(x)) = \alpha(t)\frac{\partial^2}{\partial x^2}(\log f_t(x))^2$, Theorem 2.1 gives that

$$(20) \qquad \begin{aligned} H''(t) &= -2\alpha(t)^2 \int_{\mathbb{R}}\left(\frac{\partial^2}{\partial x^2}\log(f_t(x))\right)^2 f_t(x) \\ &\quad - \alpha'(t) \int_{\mathbb{R}}\left(\frac{\partial}{\partial x}\log(f_t(x))\right)^2 f_t(x), \end{aligned}$$

which shows the concavity of $H(t)$. $\quad\square$

Among the consequences of Theorem 2.7 there are the particular cases where $\alpha(t) \equiv 0$, which is the translation case of Example 2.3, and the case where $v = 0$ and $\alpha(t) = -c$ is a constant, in which case f is a solution to the heat equation $\frac{\partial f_t(x)}{\partial t} = c\frac{\partial^2 f_t(x)}{\partial x^2}$.

Theorem 2.7 can be used to study the entropy of an approximation of a Bernoulli sum by a Gaussian distribution of the same mean and variance. This motivates the Shepp–Olkin conjecture, due to the following result:

THEOREM 2.8.   *Let $p_1, \ldots, p_n : [0, 1] \to [0, 1]$ be affine functions, and let $\mu(t) := \sum_{i=1}^n p_i(t)$ and $V(t) := \sum_{i=1}^n p_i(t)(1 - p_i(t))$ be the mean and variance functions. Define*

$$(21) \qquad f_t(x) := \frac{1}{\sqrt{2\pi V(t)}} \exp\left(-\frac{(x - \mu(t))^2}{2V(t)}\right).$$

Then the entropy $H(t)$ of f_t is a concave function of t.

PROOF.   Writing $v = \mu'(t)$, since $\mu''(t) = 0$, we can use differential properties satisfied by Gaussian kernels to compute

$$g_t(x) = \mu'(t) f_t(x) - \frac{V'(t)}{2} \frac{\partial f_t(x)}{\partial x} = v f_t(x) + \alpha(t) \frac{\partial}{\partial x} f_t(x),$$

where $\alpha(t) := -\frac{1}{2} V'(t)$. Since we have $\frac{\partial}{\partial t} \alpha(t) = \sum_{i=1}^n p_i'^2 \geq 0$, we apply Theorem 2.7 to show that the entropy $H(t)$ of f_t is a concave function of t. The conditions of Remark 2.2 can all be directly verified in this case; the key is that $g_t(x)/f_t(x)$ is a linear function of x, and $h_t(x)/f_t(x)$ is quadratic in x. This argument works for any Gaussian densities of the form (21), where $V''(t) \leq 0$, and $\mu(t)$ is an affine function of t.   □

REMARK 2.9.   It is possible to use the explicit expression for the entropy of a Gaussian random variable to prove Theorem 2.8 directly. However, as there is no explicit expression for the entropy of a sum of Bernoulli variables, it is not possible to adapt such a proof in the discrete Shepp–Olkin case, and we require the assumption that all $p_i' \geq 0$ in that case.

3. Discrete gradient field models.

3.1. *Motivating example and discrete Benamou–Brenier formula.*   We now show how natural choices of paths connecting probability mass functions on the integers can be viewed in the gradient field framework of Section 2. We give a new perspective on the time derivative using a series of functions $\alpha_k(t)$, where $k = 0, \ldots, n$ and $0 \leq t \leq 1$. Recall that we use the left derivative map ∇_1 defined by $\nabla_1 f_k = f_k - f_{k-1}$ for any function f, and write $\nabla_2 = (\nabla_1)^2$ for the map taking $\nabla_2 f_k = f_k - 2 f_{k-1} + f_{k-2}$. Write ∇_1^*, defined by $\nabla_1^* f_k = f_k - f_{k+1}$, for its adjoint (with respect to counting measure). Recall $\mathcal{A}$ denotes the set of measurable functions $\boldsymbol{\alpha}(t) = (\alpha_0(t), \alpha_1(t), \ldots, \alpha_n(t))$, where $\alpha_0(t) \equiv 0$ and $\alpha_n(t) \equiv 1$, and $0 \leq \alpha_k(t) \leq 1$ for all k and t. We first give a motivating example, which is a special case of the Shepp–Olkin interpolation.

EXAMPLE 3.1.   We write $\mathrm{Bin}_k(n, p) := \binom{n}{k} p^k (1 - p)^{n-k}$ for the probability mass function of a binomial with parameters n and p. For fixed n and $0 \leq p <$

$q \leq 1$, define $p(t) = p(1-t) + qt$, and write $f_k(t) = \text{Bin}_k(n, p(t))$ for the probability mass functions which interpolate in the natural way in the parameter space. A simple calculation (see, e.g., [24] and [30]) shows that for any $k = 0, 1, \ldots, n$,

$$(22) \qquad \frac{\partial f_k}{\partial t}(t) = -\nabla_1\big(n(q-p)\,\text{Bin}_k\big(n-1, p(t)\big)\big).$$

We reformulate equation (22) using an insight of Yu [35], who defined the hypergeometric thinning operation, observing in [35], Lemma 2, that for any n, p,

$$(23) \qquad \text{Bin}_k(n-1, p) = \frac{(k+1)}{n}\,\text{Bin}_{k+1}(n, p) + \left(1 - \frac{k}{n}\right)\text{Bin}_k(n, p).$$

This suggests that we rewrite equation (22) in the form, modeled on (11),

$$(24) \qquad 0 = \frac{\partial f_k}{\partial t}(t) + \nabla_1\big((n(q-p))g_k^{(\alpha)}(t)\big) \qquad \text{for } k = 0, 1, \ldots, n,$$

for

$$(25) \quad g_k^{(\alpha)}(t) = \alpha_{k+1}(t)f_{k+1}(t) + \big(1 - \alpha_k(t)\big)f_k(t) \qquad \text{for } k = 0, 1, \ldots, n-1,$$

with $\alpha_k(t) = k/n$ for all k and t.

The form of equations (24) and (25) suggests a version of the Benamou–Brenier formula [5, 6] for discrete random variables.

DEFINITION 3.2.   We write $\mathcal{P}_{\mathbb{Z}}(f(0), f(1))$ for the set of continuous, piecewise differentiable families of probability mass functions f_k, given end constraints $f_k(t)|_{t=0} = f_k(0)$ and $f_k(t)|_{t=1} = f_k(1)$. Given $\boldsymbol{\alpha}(t) \in \mathcal{A}$, for $f_k(t) \in \mathcal{P}_{\mathbb{Z}}(f(0), f(1))$ define probability mass function $g_k^{(\alpha)}(t)$, velocity field $v_{\alpha,k}(t)$ and path length $\mathcal{I}(f)$ by

$$(26) \quad g_k^{(\alpha)}(t) = \alpha_{k+1}(t)f_{k+1}(t) + \big(1 - \alpha_k(t)\big)f_k(t) \qquad \text{for } k = 0, 1, \ldots, n-1,$$

$$0 = \frac{\partial f_k}{\partial t}(t) + \nabla_1\big(v_{\alpha,k}(t)g_k^{(\alpha)}(t)\big)$$

(27)

$$t\text{-almost everywhere for } k = 0, 1, \ldots, n,$$

$$(28) \quad \mathcal{I}(f)^2 = \int_0^1 \left(\sum_{k=0}^{n-1} g_k^{(\alpha)}(t)v_{\alpha,k}(t)^2\right)dt =: \int_0^1 \beta(t)\,dt.$$

Define V_n via

$$(29) \qquad V_n^2\big(f(0), f(1)\big) = \inf_{\substack{f_k \in \mathcal{P}_{\mathbb{Z}}(f(0), f(1)), \\ \alpha_k(t) \in \mathcal{A}}} \mathcal{I}(f)^2,$$

and refer to any path achieving the infimum in (29), if it exists, as a geodesic.

PROPOSITION 3.3. V_n *is a metric on the space of probability measures on* $\{0, \ldots, n\}$, *Moreover, for any geodesic* f *we have*

$$(30) \qquad V_n(f(s), f(t)) = |t - s| V_n(f(0), f(1)) \qquad \text{for any } 0 \le s, t \le 1.$$

PROOF. It is clear that $V_n \ge 0$ and that $V_n(f, g) = 0$ implies $f = g$. To prove V_n is symmetric, we transpose the path $f(t)$ in time $\widetilde{f}_k(t) = f_k(1 - t)$, taking $\widetilde{\alpha}_k(t) = \alpha_k(1 - t)$ gives $\widetilde{g}_k^{(\widetilde{\alpha})}(t) = g_{n-k-1}^{(\alpha)}(1 - t)$, and $v_{\widetilde{\alpha},k} = -v_{\alpha,k}$ so that $V_n^2(f(0), f(1)) = V_n^2(f(1), f(0))$.

To prove the triangle inequality, we consider three mass functions $f(0)$, f^*, $f(1)$. For any paths $(f^{(0)}(t))_{t\in[0,1]} \in \mathcal{P}_{\mathbb{Z}}(f(0), f^*)$ and $(f^{(1)}(t))_{t\in[0,1]} \in \mathcal{P}_{\mathbb{Z}}(f^*, f(1))$ we construct $(f(t))_{t\in[0,1]} \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ such that $\mathcal{I}(f) = \mathcal{I}(f^{(0)}) + \mathcal{I}(f^{(1)})$, as follows:

- If $t \le \rho$, we set $\tau_0(t) := t/\rho$ and $f_k(t) = f_k^{(0)}(\tau_0(t))$. We then have $\alpha_k(t) = \alpha_k^{(0)}(\tau_0(t))$, with $g_k^{(\alpha)}(t) = g_k^{(0),(\alpha)}(\tau_0(t))$ and $v_{\alpha,k}(t) = \frac{1}{\rho} v_{\alpha,k}^{(0)}(\tau_0(t))$.

- If $t > \rho$, we set $\tau_1(t) := (t - \rho)/(1 - \rho)$ and $f_k(t) = f_k^{(1)}(\tau_1(t))$. We have $\alpha_k(t) = \alpha_k^{(1)}(\tau_1(t))$, with $g_k^{(\alpha)}(t) = g_k^{(1),(\alpha)}(\tau_1(t))$ and $v_{\alpha,k}(t) = \frac{1}{1-\rho} v_{\alpha,k}^{(1)}(\tau_1(t))$.

The change of variables formula allows us to compute

$$\mathcal{I}(f)^2 = \int_0^\rho \sum_{k=0}^{n-1} g_k^{(0),(\alpha)}(\tau_0(t)) \frac{1}{\rho^2} v_{\alpha,k}^{(0)}(\tau_0(t))^2 \, dt$$

$$+ \int_\rho^1 \sum_{k=0}^{n-1} g_k^{(1),(\alpha)}(\tau_1(t)) \frac{1}{(1-\rho)^2} v_{\alpha,k}^{(1)}(\tau_1(t))^2 \, dt$$

$$= \frac{1}{\rho} \mathcal{I}(f_0)^2 + \frac{1}{1-\rho} \mathcal{I}(f_1)^2.$$

Choosing the optimal $\rho = \rho^* := \frac{\mathcal{I}(f^{(0)})}{\mathcal{I}(f^{(0)}) + \mathcal{I}(f^{(1)})}$ gives the result.

We can extend the same argument to prove equation (30) above. We first prove the case $t = 1$. Consider any geodesic f and $0 \le s \le 1$. We can take $f^* = f(s)$ in the argument above and decompose the geodesic into a path from $f(0)$ to $f(s)$ of length $\mathcal{I}(f^{(0)})$ and a path from $f(s)$ to $f(1)$ of length $\mathcal{I}(f^{(1)})$. We know that the optimal ρ^* is equal to s (otherwise we could reduce V_n by taking the path at a different rate, contradicting the fact that f is a geodesic). We deduce that $\mathcal{I}(f^{(0)}) = s(\mathcal{I}(f^{(0)}) + \mathcal{I}(f^{(1)})) = s V_n(f)$, or that the inner sum $\beta(t) = \sum_{k=0}^{n-1} g_k^{(\alpha)}(t) v_{\alpha,k}(t)^2$ is constant almost everywhere in t.

We can prove the more general form of equation (30) using a similar argument. We decompose the path into three parts, $f^{(0)} \in \mathcal{P}_{\mathbb{Z}}(f(0), f(s))$, $f^{(1)} \in \mathcal{P}_{\mathbb{Z}}(f(s), f(t))$ and $f^{(2)} \in \mathcal{P}_{\mathbb{Z}}(f(t), f(1))$. Let us consider some $0 < \rho_0 < \rho_1 < 1$, and take $\tau_0(t) := t/\rho_0$, $\tau_1(t) := (t - \rho_0)/(\rho_1 - \rho_0)$ and $\tau_2(t) = (t - \rho_1)/(1 - \rho_1)$.

A similar argument shows that unless $\rho_0 = s$ and $\rho_1 = t$, the length of the path can be reduced in the same way. $\quad\square$

3.2. Constant velocity paths.

LEMMA 3.4. *For any geodesic $f(t)$ between $f(0)$ and $f(1)$, the $\beta(t)$ is constant in t. Further, if there exists a geodesic between these, then writing mean $\lambda(t) = \sum_k k f_k(t)$ the*

$$V_n\big(f(0), f(1)\big) \geq \big|\lambda(0) - \lambda(1)\big|,$$

with equality if and only if $v_{\alpha,k} \equiv v$ for all k and t, for some v.

PROOF. Proposition 3.3 shows that for any geodesic, we know that $\sqrt{\beta(t)} \equiv V_n(f(0), f(1))$ for almost all t. Since $\lambda(t) = \sum_{k=0}^n k f_k(t)$, differentiating and using equation (27) gives

$$\begin{aligned}
\frac{\partial \lambda}{\partial t}(t) &= -\sum_{k=0}^n k \nabla_1\big(v_{\alpha,k}(t) g_k^{(\alpha)}(t)\big) \\
&= -\sum_{k=0}^n \nabla_1^*(k)\big(v_{\alpha,k}(t) g_k^{(\alpha)}(t)\big) = \sum_{l=0}^n g_l^{(\alpha)}(t) v_{\alpha,l}(t),
\end{aligned}$$
(31)

since $-\nabla_1^*(k) = 1$. Using equation (27) and Cauchy–Schwarz, since $g^{(\alpha)}(t)$ is a probability mass function, equation (31) gives that for any t, since

$$\left(\sum_{k=0}^n g_k^{(\alpha)}(t) v_{\alpha,k}(t)^2\right) \geq \left(\sum_{k=0}^n g_k^{(\alpha)}(t) v_{\alpha,k}(t)\right)^2 = \left(\frac{\partial \lambda(t)}{\partial t}\right)^2,$$
(32)

or that $\big|\frac{\partial \lambda(t)}{\partial t}\big| \leq \sqrt{\beta(t)} = V_n(f(0), f(1))$, and the result follows by integration.

Observe that equality holds in equation (32) if and only if $v_{\alpha,k}(t) = \frac{\partial \lambda}{\partial t}(t)$ for all k, and equality holds overall if and only if $\frac{\partial \lambda}{\partial t}(t) = V_n(f(0), f(1))$ for all t. $\quad\square$

Lemma 3.4 focuses attention on interpolations for which velocity field $v_{\alpha,k}(t) \equiv v$ for all k and t, for some $\boldsymbol{\alpha}(t) \in \mathcal{A}$. Recall that Definition 1.2 refers to such an interpolation as a "constant velocity path," and we say that $f(t)$ satisfies a modified transport equation.

LEMMA 3.5. *If $f_k(t)$ can be expressed as a constant velocity path for some choice of v and $\boldsymbol{\alpha} \in \mathcal{A}$, then this representation is unique (there is no other choice of v and $\boldsymbol{\alpha} \in \mathcal{A}$ for which it is a constant velocity path).*

PROOF. Equation (31) shows that if there exists a constant velocity path with velocity v, then $v = \lambda(1) - \lambda(0)$. Using a similar argument, we can solve for $\boldsymbol{\alpha}$.

The key is to observe that, since $\alpha_n = 1$, for any $\boldsymbol{\alpha} \in \mathcal{A}$, equation (26) means that the sum

$$(33) \qquad \sum_{l=k}^{n-1} g_l^{(\boldsymbol{\alpha})}(t) = \sum_{l=k}^{n} f_l(t) - \alpha_k(t) f_k(t).$$

Using the distribution function $F_l(t) := \sum_{k=0}^{l} f_k(t)$ and taking $v_{\alpha,l}(t) \equiv v$, we can sum equation (27) over k to obtain

$$(34) \qquad \frac{\partial F_l}{\partial t}(t) + v g_l^{(\boldsymbol{\alpha})}(t) = 0.$$

Hence, $g_l^{(\boldsymbol{\alpha})}(t)$ is also fixed by the form of the path $f_k(t)$, and on rearranging equation (33), we express

$$(35) \qquad \alpha_k(t) = \frac{\sum_{l=k}^{n} f_l(t) - \sum_{l=k}^{n-1} g_l^{(\boldsymbol{\alpha})}(t)}{f_k(t)}. \qquad \square$$

Equation (35) implies that $\alpha_k(t)$ is a smooth function of t in the case of constant velocity paths. In particular, it is legitimate to consider the derivative $\frac{\partial}{\partial t} \alpha_k(t)$, as is done, for instance, in the proof of Lemma 4.1. We now show that in certain circumstances our distance measure V_n coincides with the Wasserstein distance W_1, a metric which is known to have a natural relationship to discrete interpolations as described in Section 3.3 below.

LEMMA 3.6.   *If there exists a constant velocity path between $f(0)$ and $f(1)$ with velocity v, then the Wasserstein distance $W_1(F(0), F(1))$ and $V_n(f(0), f(1))$ coincide and are equal to $\lambda(1) - \lambda(0)$.*

PROOF.   Recall from the proof of Lemma 3.5 that if there exists a constant velocity path with velocity v, then $v = \lambda(1) - \lambda(0)$. Without loss of generality we may assume that $v \geq 0$. In this case, $V_n = v = \lambda(1) - \lambda(0)$.

Using equation (34), positivity of v means that $F_k(t)$ is decreasing in t for all t and k. This means that $F(0)$ stochastically dominates $F(1)$ in the standard sense; see, for example, [29]. Lemma 8.2 and equation (8.1) of [7] together show that for any distribution functions $F(0)$ and $F(1)$, the $W_1(F(0), F(1)) = \int |F(1)_y - F(0)_y| \, dy$, so in this stochastically ordered case we deduce that

$$W_1(F(0), F(1)) = \int (F(0)_y - F(1)_y) \, dy = \int y \, dF(1)_y - \int y \, dF(0)_y$$
$$= \lambda(1) - \lambda(0),$$

and the argument is complete. $\square$

3.3. *Binomial interpolation.*

EXAMPLE 3.7. Comparing equations (24) and (27) shows that, taking $\alpha_k(t) \equiv k/n$, the binomial interpolation (Example 3.1) has constant velocity $v_{\alpha,k}(t) \equiv n(q - p)$ and hence achieves the lower bound in Lemma 3.4, with $V_n(\text{Bin}(n, p), \text{Bin}(n, q)) = n(q - p)$.

Contrast this with the approach of Erbar and Maas [12, 23] (see also Mielke [25]), based on Markov chains with a given stationary distribution $\boldsymbol{\pi}$. In the two-point case, taking as a reference $\boldsymbol{\pi} = (q/(p + q), p/(p + q))$, Maas [23] write ρ^β for a relative density equivalent to the probability mass function $f^\beta = (f_0^\beta, f_1^\beta) = ((1 - \beta)/2, (1 + \beta)/2)$. Example 2.6 of [23] implies a distance of $\mathcal{W}(\rho^\alpha, \rho^\beta) = \frac{1}{\sqrt{2p}} \int_\alpha^\beta \sqrt{\frac{\text{arctanh} r}{r}} \, dr$, in this case, in contrast to the $|\beta - \alpha|/2$ we obtain.

Example 3.7 can be generalized considerably as follows. Given a probability mass function f, Rényi [28] introduced the thinned probability mass function $T_t f$ to be the law of the random sum

$$(36) \qquad \sum_{i=1}^{X} B_i^{(t)},$$

where $X \sim f$ and $B_i^{(t)}$ are Bernoulli(t) random variables, independent of each other and of X. Thinning interpolates between the original measure $f = T_1 f$ and a point mass at zero $T_0 f$. This operation was studied in the context of entropy of random variables in [17], and was extended by Gozlan et al. [13] and by Hillion [15] for probability measures on graphs, implying the following definition in the case of random variables supported on $\mathbb{Z}$:

DEFINITION 3.8. A coupling π of mass functions $f(0)$ and $f(1)$ supported on $\mathbb{Z}$ [i.e., a joint distribution function $\pi_{x,y}$ whose marginals satisfy $f(0)_x = \sum_y \pi_{x,y}$ and $f(1)_y = \sum_x \pi_{x,y}$], induces a path as follows. Section 2.2 of [13] defines a mass function

$$(37) \qquad f_k(t) = v_k^\pi(t) := \sum_{x,y} \pi_{x,y} \, \text{Bin}_{k-x}(|y - x|, t),$$

which we can understand as the law of the random sum $X + \sum_{i=1}^{Y-X} B_i^{(t)}$, where $(X, Y) \sim \pi$, and as before $B_i^{(t)}$ are Bernoulli(t) random variables, independent of each other and of (X, Y). Here, we use the convention that for $m \geq 0$, $\sum_{i=1}^{-m} B_i^{(t)} = -\sum_{i=1}^{m} B_i^{(t)}$.

Proposition 2.7 of [13] gives a partial differential equation showing how $f_k(t)$ evolves with t, using a mixture of left and right gradients (as in [16]). Proposition 2.5 of [13] shows that if π^* is an optimal coupling (in Wasserstein distance

W_1), then $v^{\pi^*}(t)$ defines a (constant velocity) geodesic for the W_1 distance. We relate this to the discrete Benamou–Brenier framework in the stochastically ordered context; see Lemma 3.6.

LEMMA 3.9. *If $f(0)$ is stochastically dominated by $f(1)$, then the interpolation $(f_k(t))$ defined by equation* (37) *gives a constant velocity path.*

PROOF. In this case, $x \leq y$ for all (x, y) in the support of π. Define $v = \sum_{x,y} \pi_{x,y}(y - x)$ and $\tilde{\pi}_{x,y} = \pi_{x,y}(y - x)/v$ for another "distance-biased" joint distribution function. Direct differentiation of equation (37) gives that

$$\frac{\partial f_k}{\partial t}(t) = \sum_{x,y} \pi_{x,y}(y - x)\big(\mathrm{Bin}_{k-x-1}(y - x - 1, t) - \mathrm{Bin}_{k-x}(y - x - 1, t)\big)$$

$$= -\nabla_1\big(vg_k(t)\big),$$

where $g_k(t) = \sum_{x,y} \tilde{\pi}_{x,y} \mathrm{Bin}_{k-x}(y - x - 1, t)$. Since for any x, the convolution $(1 - t)\mathrm{Bin}_x(m - 1, t) + t\,\mathrm{Bin}_{x-1}(m - 1, t) = \mathrm{Bin}_x(m, t)$, in equation (37) we can express

$$(38) \quad f_l(t) = \sum_{x,y} \pi_{x,y}\big((1 - t)\mathrm{Bin}_{l-x}(y - x - 1, t) + t\,\mathrm{Bin}_{l-x-1}(y - x - 1, t)\big),$$

and substituting in equation (35), we obtain

$$(39) \quad \alpha_k(t) = \sum_{x,y} \pi_{x,y}\left[t\,\mathrm{Bin}_{k-x-1}(y - x, t) \right.$$

$$\left. + \sum_{l=k}^{n-1} \mathrm{Bin}_{l-x}(y - x - 1, t)\big(1 - (y - x)/v\big)\right]\bigg/ f_k(t). \qquad \square$$

In future work, we hope to consider the question of which interpolations in the form of equation (37) induce coefficients satisfying $0 \leq \alpha_k(t) \leq 1$.

EXAMPLE 3.10 (Translation case). Hillion considered the translation case, where $f(1)_{k+m} = f(0)_k = f_k$ for some m. Theorem 1.1 of [14] proved that if f is log-concave (i.e., $f_k^2 \geq f_{k-1}f_{k+1}$ for all k), the entropy is concave in t. This paper generalizes Hillion's result: the conditions of Theorem 1.3 can be verified, and the concavity of entropy is reproved.

In particular we interpolate by $\pi_{x,y}$ supported only on $\{(x, y): y - x = m\}$, so that $\tilde{\pi} = \pi$, and clearly $v = m$. Then equation (38) simplifies to give $f_k(t) = (1 - t)g_k(t) + tg_{k-1}(t)$, and equation (39) becomes

$$(40) \quad \alpha_k(t) = \frac{tg_{k-1}(t)}{f_k(t)} = 1 - \frac{(1 - t)g_k(t)}{f_k(t)},$$

so that clearly $\alpha_k(t)$ lies between 0 and 1 for all k and t.

Equation (40) shows that GLC, Condition 3, holds with equality in this case. Further k-monotonicity, Condition 1, holds as a consequence of log-concavity of $g_k(t)$ (which follows from log-concavity of f). The t-monotonicity, Condition 2, is less straightforward, but can be verified using direct calculation, using the log-concavity of h.

3.4. *Generalized log-concavity.* Recall from Section 2.2 that probabilistic inequalities can be proved for densities f such that $f/\phi_{1/c}$ is log-concave. For integer-valued random variables the corresponding property of ultra log-concavity was introduced by Pemantle [27] and promoted by Liggett [21]:

DEFINITION 3.11 ([21, 27]). For any n, a probability mass function supported on $\{0, 1, \ldots, n\}$ is ultra log-concave of order n, denoted by **ULC**(n), if the ratio $f_k/\mathrm{Bin}_k(n, t)$ is a log-concave function. Equivalently we require that

$$\frac{k+1}{n}\left(1 - \frac{k+1}{n}\right) f_{k+1}^2 - \frac{k+2}{n}\left(1 - \frac{k}{n}\right) f_k f_{k+2} \geq 0 \qquad \text{for } k = 0, \ldots, n-2.$$
(41)

We include the possibility that (formally speaking) $n = \infty$, in which case we require that the ratio of f_k and a Poisson mass function be log-concave, and write **ULC**(∞).

This condition was first used to control entropy by Johnson [17], who showed that, fixing the mean, the Poisson is maximum entropy in the class **ULC**(∞). A corresponding result was proved by Yu [36], who showed that, fixing the mean, the binomial is maximum entropy in the class **ULC**(n). This generalizes the result (see [24] and [30]) that the entropy of Bernoulli sums with a given mean is maximized by the binomial, since Newton's inequalities (see, e.g., [26]) show that Bernoulli sums are **ULC**(n).

Our generalized log-concavity, Condition 3, generalizes Definition 3.11, with **ULC**(n) corresponding to **GLC**$(\boldsymbol{\alpha})$ for $\alpha_k = k/n$, as in Example 3.1. Note that GLC, Condition 3, and k-monotonicity, Condition 1, together imply that f is log-concave.

4. Framework for concavity of discrete entropy. In this section, we prove Theorem 1.3, which shows that entropy is concave if Conditions 1, 2 and 3 are satisfied. In fact, since t-monotonicity (Condition 2) is too restrictive for our purposes, we prove a more general result, Theorem 4.4, which gives concavity of entropy despite replacing Condition 2 by the weaker Condition 4. Lemma 4.2 shows that this condition is indeed weaker, and hence together with Theorem 4.4 proves Theorem 1.3.

CONDITION 4. *Consider a constant velocity path, satisfying a modified transport equation $\frac{\partial f_k}{\partial t}(t) = -v\nabla_1(g_k^{(\boldsymbol{\alpha})}(t))$ with some h satisfying $\frac{\partial^2 f_k}{\partial t^2}(t) = v^2\nabla_2(h_k)$. If we define*

$$(42) \qquad \widetilde{h}_k := \frac{2g_k^{(\boldsymbol{\alpha})} g_{k+1}^{(\boldsymbol{\alpha})} f_{k+1} - (g_k^{(\boldsymbol{\alpha})})^2 f_{k+2} - (g_{k+1}^{(\boldsymbol{\alpha})})^2 f_k}{f_{k+1}^2 - f_k f_{k+2}},$$

then we require that

$$(43) \qquad h_k \leq \widetilde{h}_k \qquad for\ k = 0, 1, \ldots, n - 2.$$

We first observe that the same coefficients $(\alpha_k)_{k=0,\ldots,n}$ introduced in equation (2) can be used to state a second-order modified transport equation:

LEMMA 4.1. *If there exist coefficients $\boldsymbol{\alpha}$ giving rise to an interpolation with constant velocity v, then*

$$(44) \qquad \frac{\partial^2 f_k}{\partial t^2} = v^2 \nabla_2(h_k),$$

where [in a result paralleling (18) in the continuous case above], for $k = 0, \ldots, n - 2$,

$$(45) \quad \begin{aligned} h_k &= (1 - \alpha_k)(1 - \alpha_{k+1}) f_k + 2\alpha_{k+1}(1 - \alpha_{k+1}) f_{k+1} + \alpha_{k+1}\alpha_{k+2} f_{k+2} \\ &\quad - f_{k+1}\frac{1}{v}\frac{\partial \alpha_{k+1}}{\partial t}. \end{aligned}$$

PROOF. Recall that we write

$$(46) \quad g_k^{(\boldsymbol{\alpha})} = \alpha_{k+1}(t) f_{k+1}(t) + \big(1 - \alpha_k(t)\big) f_k(t) \qquad for\ k = 0, 1, \ldots, n - 1.$$

Differentiating equation (46) we have

$$\begin{aligned} \frac{\partial g_k^{(\boldsymbol{\alpha})}}{\partial t} &= f_{k+1}\frac{\partial \alpha_{k+1}}{\partial t} + \alpha_{k+1}\frac{\partial f_{k+1}}{\partial t} + (1 - \alpha_k)\frac{\partial f_k}{\partial t} - f_k\frac{\partial \alpha_k}{\partial t} \\ &= \left[f_{k+1}\frac{\partial \alpha_{k+1}}{\partial t} - v(1 - \alpha_{k+1})g_k^{(\boldsymbol{\alpha})} - v\alpha_{k+1}g_{k+1}^{(\boldsymbol{\alpha})} \right] \\ &\quad - \left[f_k\frac{\partial \alpha_k}{\partial t} - v(1 - \alpha_k)g_{k-1}^{(\boldsymbol{\alpha})} - v\alpha_k g_k^{(\boldsymbol{\alpha})} \right] \\ &= \nabla_1[vh_k], \end{aligned}$$

using the expression from (45), and the proposition follows easily. □

LEMMA 4.2.   *If Conditions* 1, 2 *and* 3 *hold*, *then Condition* 4 *holds*.

PROOF.   Using $\widetilde{h}_k$ and h_k defined in equations (42) and (45), we need to prove that $h_k \le \widetilde{h}_k$ for all k. For simplicity, we write

$$(47) \qquad\qquad D_k := f_k^2 - f_{k-1}f_{k+1} \ge 0,$$

$$(48) \qquad\qquad A_k := (f_{k+1}g_k^{(\boldsymbol{\alpha})} - f_k g_{k+1}^{(\boldsymbol{\alpha})}) \ge 0,$$

$$(49) \qquad\qquad B_k := (f_{k+1}g_{k+1}^{(\boldsymbol{\alpha})} - f_{k+2}g_k^{(\boldsymbol{\alpha})}) \ge 0.$$

The positivity of A_k and B_k follows from GLC and k-monotonicity since we can write

$$(1 - \alpha_{k+1})A_k = \mathrm{GLC}(\boldsymbol{\alpha})_k + (\alpha_{k+1} - \alpha_k)f_k g_{k+1}^{(\boldsymbol{\alpha})},$$

$$\alpha_{k+1}B_k = \mathrm{GLC}(\boldsymbol{\alpha})_k + (\alpha_{k+2} - \alpha_{k+1})f_{k+2}g_k^{(\boldsymbol{\alpha})}.$$

The key is to observe that in this notation, by Lemma 4.1,

$$(50) \qquad\qquad \widetilde{h}_k = \frac{g_{k+1}^{(\boldsymbol{\alpha})}A_k + g_k^{(\boldsymbol{\alpha})}B_k}{D_{k+1}},$$

$$(51) \qquad\qquad h_k = \alpha_{k+1}g_{k+1}^{(\boldsymbol{\alpha})} + (1 - \alpha_{k+1})g_k^{(\boldsymbol{\alpha})} - f_{k+1}\frac{1}{v}\frac{\partial \alpha_{k+1}}{\partial t}.$$

Direct calculation gives

$$(52) \qquad\qquad g_k^{(\boldsymbol{\alpha})}D_{k+1} = f_{k+1}A_k + f_k B_k,$$

$$(53) \qquad\qquad g_{k+1}^{(\boldsymbol{\alpha})}D_{k+1} = f_{k+2}A_k + f_{k+1}B_k.$$

Considering the coefficients of A_k and B_k, we can substitute (52) and (53) in equation (51) to obtain

$$\widetilde{h}_k - h_k = \frac{f_{k+2}(\alpha_{k+2} - \alpha_{k+1})A_k + f_k(\alpha_{k+1} - \alpha_k)B_k}{D_{k+1}} + f_{k+1}\frac{1}{v}\frac{\partial \alpha_{k+1}}{\partial t} \ge 0,$$

(54)

where the positivity of $\frac{\partial \alpha_{k+1}}{\partial t}$ is assumed in Condition 2.   □

We need one further result, which can be directly verified by differentiation:

LEMMA 4.3.   *Writing $\theta(v) = 1/(2v) - v/2$, we have*

$$0 \le -\log v \le \theta(v) \qquad for\ v \le 1.$$

THEOREM 4.4.   *If Conditions* 1, 3 *and* 4 *hold*, *then the entropy of $H(f)$ is concave in t.*

PROOF. For simplicity, we write g_k for $g_k^{(\alpha)}$. First note that Conditions 1 and 3 together imply that f is log-concave. In fact, they imply two stronger results, that

(55)
$$\frac{f_k g_{k+1}}{f_{k+1} g_k} \leq 1 \quad \text{and} \quad \frac{f_{k+2} g_k}{f_{k+1} g_{k+1}} \leq 1.$$

This means that, using Lemma 4.3, writing $\theta(v) = 1/(2v) - v/2$, we can write

(56)
$$0 \leq -\log\left(\frac{f_k(t) f_{k+2}(t)}{f_{k+1}(t)^2}\right) = -\log\left(\frac{f_k g_{k+1}}{f_{k+1} g_k}\right) - \log\left(\frac{f_{k+2} g_k}{f_{k+1} g_{k+1}}\right)$$
$$\leq \theta\left(\frac{f_k g_{k+1}}{f_{k+1} g_k}\right) + \theta\left(\frac{f_{k+2} g_k}{f_{k+1} g_{k+1}}\right)$$
$$= \frac{D_{k+1}}{2 f_{k+1} g_k g_{k+1}}\left(\frac{g_k^2}{f_k} + \frac{g_{k+1}^2}{f_{k+2}}\right),$$

where the last identity follows by grouping together multiples of g_{k+1}/g_k and g_k/g_{k+1} and factorizing. In a standard fashion, we can write the second derivative of entropy as

$$H''(t) = -\sum_{k=0}^{n} \frac{\partial^2 f_k}{\partial t^2}(t) \log(f_k(t)) - \sum_{k=0}^{n} \frac{1}{f_k}\left(\frac{\partial f_k}{\partial t}(t)\right)^2$$

(57)
$$= -\sum_{k=0}^{n} v^2 \nabla_2(h_k) \log(f_k(t)) - \sum_{k=0}^{n} \frac{(\nabla_1(v g_k))^2}{f_k}$$

(58)
$$= v^2 \sum_{k=0}^{n} h_k\left(-\log\left(\frac{f_k(t) f_{k+2}(t)}{f_{k+1}(t)^2}\right)\right) - \sum_{k=0}^{n} \frac{(\nabla_1(v g_k))^2}{f_k}$$

(59)
$$\leq v^2 \sum_{k=0}^{n}\left[\tilde{h}_k\left(\frac{D_{k+1}}{2 f_{k+1} g_k g_{k+1}}\left(\frac{g_k^2}{f_k} + \frac{g_{k+1}^2}{f_{k+2}}\right)\right) - \left(\frac{g_k^2}{f_k} - 2\frac{g_k g_{k+1}}{f_{k+1}} + \frac{g_{k+1}^2}{f_{k+2}}\right)\right],$$

where equation (57) follows using equations (3) and (1), respectively, and equation (58) uses the adjoint of ∇_2. Finally equation (59) follows from Condition 4 and equation (56), using the fact that both terms are positive.

Using equation (50), we can write the first term in the square bracket in equation (59) as

$$\frac{(g_{k+1} A_k + g_k B_k)}{2 f_{k+1} g_k g_{k+1}}\left(\frac{g_k^2}{f_k} + \frac{g_{k+1}^2}{f_{k+2}}\right).$$

Further, since we can write

$$-\left(\frac{g_k^2}{f_k} - 2\frac{g_k g_{k+1}}{f_{k+1}} + \frac{g_{k+1}^2}{f_{k+2}}\right) = -\frac{g_k}{f_k f_{k+1}} A_k - \frac{g_{k+1}}{f_{k+1} f_{k+2}} B_k,$$

we can expand equation (59) in terms of A_k and B_k as

(60)
$$-\frac{(A_k g_{k+1} - B_k g_k)(f_{k+2} g_k^2 - f_k g_{k+1}^2)}{2 f_k f_{k+1} f_{k+2} g_k g_{k+1}}$$
$$= -\frac{f_k f_{k+1} f_{k+2}}{2 g_k g_{k+1}} \left(\frac{g_k^2}{f_k f_{k+1}} - \frac{g_{k+1}^2}{f_{k+1} f_{k+2}} \right)^2.$$

Here the final equality follows since the form of A_k and B_k means that the two bracketed terms in the first expression in equation (60) are in fact equal. $\square$

It would be of interest to understand how this remainder term (60) relates to the corresponding term found for continuous interpolations in equation (14).

5. Shepp–Olkin interpolation as a constant velocity path. Recall that in Conjecture 1.1 we are given $n \geq 1$ affine functions $p_i : [0, 1] \to [0, 1]$ where for each i, $p_i(t) = p_i(0)(1 - t) + p_i(1)t$. We denote by $(f_k(t))_{k=0,1,\ldots,n}$ the distribution of the sum of independent Bernoulli random variables of parameters $p_1(t), \ldots, p_n(t)$. Further we write $(f_k^{(i)}(t))_{k=0,\ldots,n-1}$ for the mass function of the ith "leave one out" sum—that is, the distribution of a Bernoulli sum with parameters $p_1(t), \ldots, p_{i-1}(t), p_{i+1}(t), \ldots, p_n(t)$, and $f_k^{(i,j)}(t)$ for the "leave two out" sum, involving all parameters except $p_i(t)$ and $p_j(t)$. Define

(61)
$$D_k^{(i)} = (f_k^{(i)})^2 - f_{k-1}^{(i)} f_{k+1}^{(i)},$$

(62)
$$E_k^{(i)} = f_k^{(i)} f_{k-1}^{(i)} - f_{k-2}^{(i)} f_{k+1}^{(i)},$$

with corresponding notation for D_k, $D_k^{(i,j)}$ and so on.

We now show how the Shepp–Olkin problem can be viewed in the framework we introduced in Section 3. To be specific, if each derivative p_i' is positive, the Shepp–Olkin interpolation is a constant velocity path with velocity v in the sense of Definition 1.2. That is:

PROPOSITION 5.1. *If all the $p_i' \geq 0$, then the probability mass function defined by the Shepp–Olkin interpolation satisfies a modified transport equation,*

(63)
$$\frac{\partial f_k}{\partial t}(t) + \nabla_1(v g_k) = 0.$$

Here we set $v := \sum_{i=1}^n p_i'$, and write the probability mass function

(64)
$$g_k(t) := \left(1 - \alpha_k(t)\right) f_k(t) + \alpha_{k+1}(t) f_{k+1}(t),$$

where

(65)
$$\alpha_k(t) = \frac{\sum_{i=1}^n p_i' p_i(t) f_{k-1}^{(i)}(t)}{v f_k(t)} = 1 - \frac{\sum_{i=1}^n p_i'(1 - p_i(t)) f_k^{(i)}(t)}{v f_k(t)}.$$

Observe that $\alpha_0(t) \equiv 0$ and $\alpha_n(t) \equiv 1$, and that if all $p_i' \geq 0$, then $0 \leq \alpha_k(t) \leq 1$ for all k and t. Further, this interpolation satisfies a second-order modified transport equation of the form $\frac{\partial^2 f_k}{\partial t^2}(t) = v^2 \nabla_2(h_k)$, where

$$(66) \qquad h_k = \frac{\sum_{i \neq j} p_i' p_j' f_k^{(i,j)}}{v^2}.$$

PROOF.   Observe that by definition, the f_k have probability generating function

$$\sum_{k=0}^{n} f_k(t) s^k = \prod_{i=1}^{n} (1 - p_i(t) + s p_i(t)),$$

which has derivative with respect to t given by

$$(67) \qquad \sum_{k=0}^{n} \frac{\partial f_k}{\partial t}(t) s^k = \sum_{i=1}^{n} p_i'(s-1) \prod_{j \neq i} (1 - p_j(t) + s p_j(t)).$$

Comparing coefficients of s, we see that $\frac{\partial f_k}{\partial t} + \nabla_1(v \widetilde{g}_k(t)) = 0$, where $\widetilde{g}_k(t) := \frac{1}{v} \sum_{i=1}^{n} p_i' f_k^{(i)}$. Substituting equation (65) in equation (64), we obtain that $\widetilde{g}_k(t) = g_k(t)$, and so equation (63) follows.

The values of $\alpha_0(t)$ and $\alpha_n(t)$ follow from the fact that $f_{-1}^{(i)} = f_n^{(i)} = 0$. The positivity of $\alpha_k(t)$ and $1 - \alpha_k(t)$ follow from the assumption that $p_i' \geq 0$, meaning that all the terms in both the fractions in equation (65) are positive.

The form of h_k stated in equation (66) follows by taking a further derivative of equation (67) to obtain

$$(68) \qquad \sum_{k=0}^{n} \frac{\partial^2 f_k}{\partial t^2}(t) s^k = \sum_{i=1}^{n} \sum_{j \neq i} p_i' p_j'(s-1)^2 \prod_{\ell \neq i,j} (1 - p_\ell(t) + s p_\ell(t)),$$

and again the result follows on comparing coefficients of s.   □

For clarity, we now suppress the explicit dependence of α_k on t. We first verify the k-monotonicity Condition 1 for Shepp–Olkin interpolations with $p_i' \geq 0$.

PROPOSITION 5.2.   *For the Shepp–Olkin interpolation described above, if $p_i' \geq 0$ for all i, then the coefficients $(\alpha_k)_{k=0,1,\ldots,n}$ satisfy the inequality $\alpha_k \leq \alpha_{k+1}$; that is, k-monotonicity (Condition 1) holds.*

PROOF.   If k is such that $f_k > 0$ and $f_{k+1} > 0$, then

$$\alpha_{k+1} - \alpha_k = \frac{\sum_{i=1}^{n} p_i' p_i [f_k^{(i)} f_k - f_{k-1}^{(i)} f_{k+1}]}{(\sum_{i=1}^{n} p_i') f_k f_{k+1}}.$$

Moreover, for $i \in \{1, \ldots, n\}$, we have

$$
\begin{aligned}
f_k^{(i)} f_k &- f_{k-1}^{(i)} f_{k+1} \\
&= f_k^{(i)} \big[(1 - p_i) f_k^{(i)} + p_i f_{k-1}^{(i)} \big] - f_{k-1}^{(i)} \big[(1 - p_i) f_{k+1}^{(i)} + p_i f_k^{(i)} \big] \\
&= (1 - p_i) \big[(f_k^{(i)})^2 - f_{k-1}^{(i)} f_{k+1}^{(i)} \big] \geq 0,
\end{aligned}
$$

by the log-concavity property of the Bernoulli sum $f^{(i)}$. The fact that each p_i' is nonnegative finally proves that $\alpha_k \leq \alpha_{k+1}$. $\quad\square$

REMARK 5.3. In the Shepp–Olkin case, α_k is a conditional expectation of a weighted sum, similarly to the "scaled score function" of [20]. This follows since writing B_i for a Bernoulli random variable with parameter $p_i(t)$, we obtain $\mathbb{P}(B_i = 1 | B_1 + \cdots + B_n = k) = p_i(t) f_{k-1}^{(i)}(t) / f_k(t)$ so that equation (65) can be expressed as

$$
(69) \qquad \alpha_k(t) = \mathbb{E}\left(\sum_{i=1}^{n} \lambda_i B_i \,\bigg|\, \sum_{i=1}^{n} B_i = k \right),
$$

where weights $\lambda_i = p_i' / (\sum p_i')$. Note that in particular, in the "symmetric" case where $p_i' \equiv p'$ for all i, then $\lambda_i' = 1/n$ and $\alpha_k(t) \equiv k/n$.

This conditional expectation characterization allows us to give an alternative proof of the k-monotonicity Proposition 5.2. A result of Efron [11] (see also [18], equation (3.1)) shows that if $\phi(u_1, \ldots, u_n)$ is an increasing function in each variable and $X_1, \ldots, X_n$ are independent log-concave random variables, then $\Phi(k) := \mathbb{E}[\phi(X_1, \ldots, X_n) | X_1 + \cdots + X_n = k]$ is an increasing function of k. Applying this to $\phi(B_1, \ldots, B_n) = \sum \lambda_i B_i$, the result follows.

We now prove that in the monotone Shepp–Olkin case the Bernoulli sum mass function is $\mathbf{GLC}(\boldsymbol{\alpha})$, for the natural choice of $\boldsymbol{\alpha}$, and hence Condition 3 holds.

PROPOSITION 5.4. *For the Shepp–Olkin interpolation, taking $\boldsymbol{\alpha}$ as defined in equation* (65), *if all the p_i' are positive, then the Bernoulli sum mass function $(f_k)_{k=0,1,\ldots,n}$ is $\mathbf{GLC}(\boldsymbol{\alpha})$, and Condition 3 holds.*

PROOF. Formula (65) shows that $\mathrm{GLC}(\boldsymbol{\alpha})_k$ can be written as $1/v^2$ times

$$
\begin{aligned}
(70) \qquad & \left(\sum_i p_i' p_i f_k^{(i)} \right) \cdot \left(\sum_i p_i' (1 - p_i) f_{k+1}^{(i)} \right) \\
& - \left(\sum_i p_i' (1 - p_i) f_k^{(i)} \right) \cdot \left(\sum_i p_i' p_i f_{k+1}^{(i)} \right).
\end{aligned}
$$

Expanding this expression as a quadratic form in $p'_1, \ldots, p'_n$, the coefficients of p'^2_i vanish, leaving an expression which simplifies to

$$\sum_{i<j} p'_i p'_j (p_i - p_j)^2 [(f_k^{(i,j)})^2 - f_{k-1}^{(i,j)} f_{k+1}^{(i,j)}].$$

The positivity of this expression, and hence the **GLC(α)** property, follow from the log-concavity of $(f_k^{(i,j)})_{k=0,\ldots,n-2}$ and positivity of p'_i. $\quad\square$

6. Entropy concavity for monotone Shepp–Olkin regime. We now show that entropy is concave in the monotone Shepp–Olkin regime. Having already verified Conditions 1 and 3 in Propositions 5.2 and 5.4, Theorem 4.4 shows that concavity of entropy follows if Condition 4 holds.

PROPOSITION 6.1. *For monotone Shepp–Olkin interpolations, Condition 4 holds if*

(71)
$$\sum_{i<j}(p'^2_i p_j(1 - p_j)b_{i,j} + p'^2_j p_i(1 - p_i)b_{j,i}$$
$$+ 2p'_i p'_j p_i(1 - p_i)p_j(1 - p_j)c_{i,j}) \geq 0,$$

where $b_{i,j}$ and $c_{i,j}$ are defined in equations (73) *and* (74) *below.*

PROOF. We use the fact that [in the notation of equations (48) and (49)] the numerator of $\tilde{h}_k$ can be written as $g_{k+1}A_k + g_k B_k$, where

$$A_k := (f_{k+1}g_k - f_k g_{k+1}) = \frac{1}{v}\sum_i p'_i p_i D_k^{(i)} \geq 0,$$

$$B_k := (f_{k+1}g_{k+1} - f_{k+2}g_k) = \frac{1}{v}\sum_i p'_i(1 - p_i)D_{k+1}^{(i)} \geq 0.$$

This means [using the expression for h_k from equation (66) above] that Condition 4 is equivalent to the positivity of

(72)
$$v^2(g_{k+1}A_k + g_k B_k) - D_{k+1}\sum_{i \neq j} p'_i p'_j f_k^{(i,j)}$$
$$= \sum_i (p'_i)^2 (f_{k+1}^{(i)} p_i D_k^{(i)} + f_k^{(i)}(1 - p_i)D_{k+1}^{(i)})$$
$$+ \sum_{i \neq j} p'_i p'_j (f_{k+1}^{(j)} p_i D_k^{(i)} + f_k^{(j)}(1 - p_i)D_{k+1}^{(i)} - D_{k+1}f_k^{(i,j)}).$$

We expand the two bracketed terms in equation (72) in terms of $f_k^{(i,j)}$, using Lemma A.1 below, which implies [using the notation of equations (61) and (62)]

that

$$f_k^{(i)} f_{k-1}^{(i)} = \sum_{j \neq i} p_j (1 - p_j) D_{k-1}^{(i,j)},$$

$$f_k^{(i)} f_{k-2}^{(i)} = \frac{1}{2} \sum_{j \neq i} p_j (1 - p_j) E_{k-1}^{(i,j)}.$$

First observe that

$$
\begin{aligned}
&f_{k+1}^{(i)} p_i D_k^{(i)} + f_k^{(i)} (1 - p_i) D_{k+1}^{(i)} \\
&= p_i f_k^{(i)} [f_k^{(i)} f_{k+1}^{(i)}] - p_i f_{k+1}^{(i)} [f_{k-1}^{(i)} f_{k+1}^{(i)}] + (1 - p_i) f_{k+1}^{(i)} [f_k^{(i)} f_{k+1}^{(i)}] \\
&\quad - (1 - p_i) f_k^{(i)} [f_k^{(i)} f_{k+2}^{(i)}] \\
&= \sum_{j \neq i} p_j (1 - p_j) \frac{1}{2} (2 p_i f_k^{(i)} D_k^{(i,j)} - p_i f_{k+1}^{(i)} E_k^{(i,j)} + 2(1 - p_i) f_{k+1}^{(i)} D_k^{(i,j)} \\
&\hspace{9cm} - (1 - p_i) f_k^{(i)} E_{k+1}^{(i,j)}) \\
&=: \sum_{j \neq i} p_j (1 - p_j) b_{i,j},
\end{aligned}
$$

where each term in square brackets is expanded using Lemma A.1. Further by writing $f_k^{(i)} = (1 - p_j) f_k^{(i,j)} + p_j f_{k-1}^{(i,j)}$, we can rearrange the expression for $b_{i,j}$ as

$$
\begin{aligned}
(73) \quad b_{i,j} = &\tfrac{1}{2} p_i p_j ((f_k^{(i,j)})^2 f_{k-1}^{(i,j)} - 2(f_{k-1}^{(i,j)})^2 f_{k+1}^{(i,j)} + f_{k-2}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)}) \\
&+ \tfrac{1}{2} (1 - p_i)(1 - p_j)((f_k^{(i,j)})^2 f_{k+1}^{(i,j)} - 2(f_{k+1}^{(i,j)})^2 f_{k-1}^{(i,j)} + f_{k+2}^{(i,j)} f_k^{(i,j)} f_{k-1}^{(i,j)}) \\
&+ \tfrac{1}{2} (1 - p_i) p_j (2(f_k^{(i,j)})^3 - 3 f_{k-1}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)} + (f_{k-1}^{(i,j)})^2 f_{k+2}^{(i,j)}) \\
&+ \tfrac{1}{2} p_i (1 - p_j)(2(f_k^{(i,j)})^3 - 3 f_{k-1}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)} + (f_{k+1}^{(i,j)})^2 f_{k-2}^{(i,j)}).
\end{aligned}
$$

Similarly, using simplifications such as the fact that

$$D_k = (1 - p_i)^2 D_{k-1}^{(i)} + p_i^2 D_k^{(i)} + p_i (1 - p_i) E_k^{(i)}$$

the second bracket of equation (72) can be written as $p_i (1 - p_i) p_j (1 - p_j) c_{i,j}$, where

$$
\begin{aligned}
(74) \quad c_{i,j} :=\ & (f_{k+1}^{(i,j)} E_k^{(i,j)} - f_k^{(i,j)} D_k^{(i,j)} - f_{k+2}^{(i,j)} D_{k-1}^{(i,j)}) \\
=\ & -(f_k^{(i,j)})^3 + 2 f_{k-1}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)} - (f_{k+1}^{(i,j)})^2 f_{k-2}^{(i,j)} \\
& - (f_{k-1}^{(i,j)})^2 f_{k+2}^{(i,j)} + f_{k-2}^{(i,j)} f_k^{(i,j)} f_{k+2}^{(i,j)}. \qquad \Box
\end{aligned}
$$

LEMMA 6.2. *For the monotone Shepp–Olkin interpolation, for each $i \neq j$, the $b_{i,j} \geq 0$ and*

$$(75) \qquad b_{i,j} \geq -\frac{p_i(1-p_j) + p_j(1-p_i)}{2} c_{i,j},$$

and hence Condition 4 is satisfied, and so the entropy is concave.

PROOF. To verify the positivity of $b_{i,j}$, we simply observe that each of the brackets in equation (73) is positive, by an application of equations (78), (79), (82) and (83) proved in the Appendix below.

To verify that $b_{i,j} + \frac{1}{2}(p_i(1-p_j) + p_j(1-p_i))c_{i,j}$ is positive, we consider adding the final two terms of equation (73) to the expression given in (74), to obtain

$$(1-p_i)p_j\big((f_k^{(i,j)})^3 - f_{k-1}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)} - (f_{k+1}^{(i,j)})^2 f_{k-2}^{(i,j)} + f_{k-2}^{(i,j)} f_k^{(i,j)} f_{k+2}^{(i,j)}\big)$$
$$+ p_i(1-p_j)$$
$$\times \big((f_k^{(i,j)})^3 - f_{k-1}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)} - (f_{k-1}^{(i,j)})^2 f_{k+2}^{(i,j)} + f_{k-2}^{(i,j)} f_k^{(i,j)} f_{k+2}^{(i,j)}\big),$$

where the positivity of the two final terms is guaranteed by equations (80) and (81) below.

Condition 4 is verified by considering two separate cases. If $c_{i,j} \geq 0$, then all the terms in equation (71) are positive. Otherwise, if $c_{i,j} \leq 0$, then the bracketed term in equation (71) has negative discriminant as a function of p_i' and p_j',

$$4p_i(1-p_i)p_j(1-p_j)c_{i,j}^2 - 4b_{i,j}b_{j,i}$$
$$\leq c_{i,j}^2\big(4p_i(1-p_i)p_j(1-p_j) - (p_i(1-p_j) + p_j(1-p_i))^2\big)$$
$$= -c_{i,j}^2(p_i - p_j)^2 \leq 0,$$

since under this assumption both sides of equation (75) are positive, so it can be squared. In either case, we conclude that equation (71) is positive, and Condition 4 is satisfied. □

Since Condition 4 has been verified, the proof of Theorem 1.5 is complete.

APPENDIX: TECHNICAL RESULTS REGARDING BERNOULLI SUMS

In this section, we prove some technical results regarding the mass functions of Bernoulli sum random variables, required to prove the monotone Shepp–Olkin Theorem 1.5.

LEMMA A.1. *Let $(f_k)_{k \in \mathbb{Z}}$ be the Bernoulli sum of parameters $p_1, \ldots, p_m$. Then for every $k \in \mathbb{Z}$ and $q \geq 1$, we have the identity*

$$(76) \qquad q f_k f_{k-q} = \sum_{j=1}^{m} p_j(1-p_j)[f_{k-1}^{(j)} f_{k-q}^{(j)} - f_k^{(j)} f_{k-q-1}^{(j)}].$$

PROOF.   We use induction on the number m of parameters, the case where $m = 1$ being obvious. If $(f_k)_{k \in \mathbb{Z}}$ is the Bernoulli sum of parameters $p_1, \ldots, p_m$, we set for $p \in [0, 1]$

$$\tilde{f}_k := (1 - p)f_k + pf_{k-1},$$

and we want to prove, given $k \in \mathbb{Z}$ and $q \geq 1$, that

$$q\tilde{f}_k\tilde{f}_{k-q} = \sum_{j=1}^{m} p_j(1 - p_j)[\tilde{f}_{k-1}^{(j)}\tilde{f}_{k-q}^{(j)} - \tilde{f}_k^{(j)}\tilde{f}_{k-q-1}^{(j)}]$$

(77)

$$+ p(1 - p)[f_{k-1}f_{k-q} - f_kf_{k-q-1}].$$

Expanding each side with the respect to the basis $(1 - p)^2, 2p(1 - p), p^2$, using the fact that $\tilde{f}_k^{(j)} := (1 - p)f_k^{(j)} + pf_{k-1}^{(j)}$, it is easy to show that equation (77) is satisfied for some $k \in \mathbb{Z}$ and $q \geq 1$ if (76) holds true for the pairs $(k, q),(k, q + 1)$ and $(k - 1, q)$. $\square$

Next we prove the following technical inequality, which may be of independent interest:

THEOREM A.2.   Property(m) *holds*: *that is, for every* $(g_k^{[m]})_{k \in \mathbb{Z}}$ *which is the probability mass function of a sum of m independent Bernoulli variables*

$$C_1(k) := g_{k-1}^{[m]}(g_k^{[m]})^2 - 2(g_{k-1}^{[m]})^2 g_{k+1}^{[m]} + g_k^{[m]} g_{k+1}^{[m]} g_{k-2}^{[m]} \geq 0 \qquad \text{for all } k \in \mathbb{Z}.$$

(78)

We first show that Property(m) implies a number of related inequalities, which are of use elsewhere in the paper:

COROLLARY A.3.   *If* Property(m) *holds, then for any* $g^{[m]}$, *the probability mass function for the sum of m independent Bernoulli random variables, for all* $k \in \mathbb{Z}$,

$$(79) \quad \overline{C}_1(k) := (g_k^{[m]})^2 g_{k+1}^{[m]} - 2(g_{k+1}^{[m]})^2 g_{k-1}^{[m]} + g_{k+2}^{[m]} g_k^{[m]} g_{k-1}^{[m]} \geq 0,$$

$$(80) \quad C_2(k) := (g_k^{[m]})^3 - g_{k-1}^{[m]} g_k^{[m]} g_{k+1}^{[m]} - (g_{k-1}^{[m]})^2 g_{k+2}^{[m]} + g_{k-2}^{[m]} g_k^{[m]} g_{k+2}^{[m]} \geq 0,$$

$$(81) \quad \overline{C}_2(k) := (g_k^{[m]})^3 - g_{k-1}^{[m]} g_k^{[m]} g_{k+1}^{[m]} - (g_{k+1}^{[m]})^2 g_{k-2}^{[m]} + g_{k-2}^{[m]} g_k^{[m]} g_{k+2}^{[m]} \geq 0,$$

$$(82) \quad C_3(k) := 2(g_k^{[m]})^3 - 3g_{k-1}^{[m]} g_k^{[m]} g_{k+1}^{[m]} + (g_{k-1}^{[m]})^2 g_{k+2}^{[m]} \geq 0,$$

$$(83) \quad \overline{C}_3(k) := 2(g_k^{[m]})^3 - 3g_{k-1}^{[m]} g_k^{[m]} g_{k+1}^{[m]} + (g_{k+1}^{[m]})^2 g_{k-2}^{[m]} \geq 0,$$

$$(84) \quad D_1(k) := 2(g_k^{[m]})^2 g_{k-2}^{[m]} - 3g_{k-2}^{[m]} g_{k-1}^{[m]} g_{k+1}^{[m]} + g_{k+1}^{[m]} g_k^{[m]} g_{k-3}^{[m]} \geq 0.$$

PROOF. First note that these inequalities can be paired up by a duality argument. That is, if Property(m) holds for every Bernoulli sum $g^{[m]}$, it is true for $\overline{g}_k := g^{[m]}_{m-k}$ with parameters $1 - p_1, 1 - p_2, \ldots, 1 - p_m$, which implies equation (79). Similarly (80) implies (81) and (82) implies (83). We write

$$D^{[m]}_k = (g^{[m]}_k)^2 - g^{[m]}_{k-1} g^{[m]}_{k+1},$$

which is positive because $g^{[m]}$ is a sum of independent Bernoulli random variables, and therefore log-concave. In this notation, equations (80), (82) and (84) are a consequence of (78), since simple calculations show that

$$g^{[m]}_{k+1} C_2(k) = 2(g^{[m]}_k g^{[m]}_{k+1} - g^{[m]}_{k-1} g^{[m]}_{k+2}) D^{[m]}_k + g^{[m]}_{k+2} C_1(k) \geq 0,$$

$$g^{[m]}_k C_3(k) = 2(D^{[m]}_k)^2 + g^{[m]}_{k-1} \overline{C}_1(k) \geq 0,$$

$$g^{[m]}_{k-1} D_1(k) = 2g^{[m]}_{k-2} C_1(k) + g^{[m]}_{k+1} C_1(k-1) \geq 0.$$

Here, positivity of $(g^{[m]}_k g^{[m]}_{k+1} - g^{[m]}_{k-1} g^{[m]}_{k+2})$ is again a consequence of log-concavity of $g^{[m]}$. $\square$

In a similar way, we can argue that:

PROPOSITION A.4. *If* Property(m) *is satisfied, then for every sum of m independent Bernoulli variables we have*

(85) $$g^{[m]}_k D^{[m]}_k + g^{[m]}_{k-2} D^{[m]}_{k+1} \geq 2g^{[m]}_{k+2} D^{[m]}_{k-1} \qquad \text{for every } k \in \mathbb{Z}.$$

PROOF. We can restate Property(m) as being equivalent to the inequality

$$D^{[m]}_k \geq \frac{g^{[m]}_{k+1}}{g^{[m]}_{k-1}} D^{[m]}_{k-1},$$

the iteration of which gives

$$D^{[m]}_{k+1} \geq \frac{g^{[m]}_{k+2}}{g^{[m]}_k} \frac{g^{[m]}_{k+1}}{g^{[m]}_{k-1}} D^{[m]}_{k-1},$$

so equation (85) holds if we have

$$\frac{g^{[m]}_k g^{[m]}_{k+1}}{g^{[m]}_{k-1} g^{[m]}_{k+2}} + \frac{g^{[m]}_{k-2} g^{[m]}_{k+1}}{g^{[m]}_k g^{[m]}_{k-1}} - 2 \geq 0,$$

which can be rewritten as

$$\frac{C_1(k+1)}{g^{[m]}_{k-1} g^{[m]}_{k+1} g^{[m]}_{k+2}} + 2\frac{(D^{[m]}_k)^2}{(g^{[m]}_k)^2 g^{[m]}_{k-1} g^{[m]}_{k+1}} + \frac{C_1(k)}{(g^{[m]}_k)^2 g^{[m]}_{k-1}} \geq 0,$$

which is positive by assumption, which proves the proposition.   □

PROOF OF THEOREM A.2.   We prove Property(m) by induction on the number of parameters m. It is clear that Property(1) is true. Let us suppose Property(m) holds for some $m \geq 1$. In order to prove Property($m + 1$), it suffices to show that, for every $k \in \mathbb{Z}$,

$$(86) \qquad g_{k-1}^{[m+1]}(g_k^{[m+1]})^2 - 2(g_{k-1}^{[m+1]})^2 g_{k+1}^{[m+1]} + g_{k-2}^{[m+1]} g_k^{[m+1]} g_{k+1}^{[m+1]} \geq 0,$$

where $g^{[m+1]}$ is the distribution of a sum of $m + 1$ Bernoulli variables. For $p = p_{m+1} \in [0, 1]$, we can write $g_k^{[m+1]} := (1 - p)g_k^{[m]} + pg_{k-1}^{[m]}$. To prove that (86) is positive, we expand it as a polynomial in p, of order 3, in the basis $(1 - p)^3$, $p(1 - p)^2$, $p^2(1 - p)$, p^3, and show that the coefficients of each of these terms are positive:

(1) The coefficient of $(1 - p)^3$ is $C_1(k)$, which is clearly positive, by Property(m).
(2) The coefficient of p^3 is $C_1(k - 1)$, which is also positive, by Property(m).
(3) The coefficient of $p(1 - p)^2$ is $D_1(k)$, which is positive, since Property(m) implies equation (84).
(4) The coefficient of $p^2(1 - p)$ can be written $g_{k-1}^{[m]} D_{k-1}^{[m]} + g_{k-3}^{[m]} D_k^{[m]} - 2g_{k+1}^{[m]} D_{k-2}^{[m]}$, which is positive by Proposition A.4.

Since each coefficient is positive, we deduce that equation (86) is satisfied, which shows that Property($m + 1$) holds.   □

REFERENCES

[1] AMARI, S.-I. and NAGAOKA, H. (2000). *Methods of Information Geometry. Translations of Mathematical Monographs* **191**. Amer. Math. Soc., Providence, RI. MR1800071

[2] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. Birkhäuser, Basel. MR2401600

[3] ANÉ, C., BLACHÈRE, S., CHAFAÏ, D., FOUGÈRES, P., GENTIL, I., MALRIEU, F., ROBERTO, C. and SCHEFFER, G. (2000). *Sur les Inégalités de Sobolev Logarithmiques. Panoramas et Synthèses* **10**. Société Mathématique de France, Paris. MR1845806

[4] BAKRY, D. and ÉMERY, M. (1985). Diffusions hypercontractives. In *Séminaire de Probabilités, XIX*, 1983/84. *Lecture Notes in Math.* **1123** 177–206. Springer, Berlin. MR0889476

[5] BENAMOU, J.-D. and BRENIER, Y. (1999). A numerical method for the optimal time-continuous mass transport problem and related problems. In *Monge Ampère Equation*: *Applications to Geometry and Optimization* (*Deerfield Beach, FL*, 1997). *Contemp. Math.* **226** 1–11. Amer. Math. Soc., Providence, RI. MR1660739

[6] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163

[7] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[8] CARLEN, E. A. and GANGBO, W. (2003). Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math.* (2) **157** 807–846. MR1983782

[9] CORDERO-ERAUSQUIN, D. (2002). Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* **161** 257–269. MR1894593

[10] COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory*. Wiley, New York. MR1122806

[11] EFRON, B. (1965). Increasing properties of Pólya frequency functions. *Ann. Math. Statist.* **36** 272–279. MR0171335

[12] ERBAR, M. and MAAS, J. (2012). Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.* **206** 997–1038. MR2989449

[13] GOZLAN, N., ROBERTO, C., SAMSON, P.-M. and TETALI, P. (2014). Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields* **160** 47–94. MR3256809

[14] HILLION, E. (2012). Concavity of entropy along binomial convolutions. *Electron. Commun. Probab.* **17** 9. MR2872573

[15] HILLION, E. (2014). Contraction of Measures on Graphs. *Potential Anal.* **41** 679–698. MR3264815

[16] HILLION, E., JOHNSON, O. T. and YU, Y. (2014). A natural derivative on $[0, n]$ and a binomial Poincaré inequality. *ESAIM Probab. Statist.* **18** 703–712. MR3334010

[17] JOHNSON, O. (2007). Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.* **117** 791–802. MR2327839

[18] JOHNSON, O. T. and SUHOV, Y. M. (2003). The von Neumann entropy and information rate for integrable quantum Gibbs ensembles 2. *Quantum Computers and Computing* **4** 128–143.

[19] JORDAN, R., KINDERLEHRER, D. and OTTO, F. (1998). The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** 1–17. MR1617171

[20] KONTOYIANNIS, I., HARREMOËS, P. and JOHNSON, O. (2005). Entropy and the law of small numbers. *IEEE Trans. Inform. Theory* **51** 466–472. MR2236061

[21] LIGGETT, T. M. (1997). Ultra logconcave sequences and negative dependence. *J. Combin. Theory Ser. A* **79** 315–325. MR1462561

[22] LOTT, J. and VILLANI, C. (2009). Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* (2) **169** 903–991. MR2480619

[23] MAAS, J. (2011). Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.* **261** 2250–2292. MR2824578

[24] MATEEV, P. (1978). The entropy of the multinomial distribution. *Teor. Verojatnost. i Primenen.* **23** 196–198. MR0490451

[25] MIELKE, A. (2013). Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations* **48** 1–31. MR3090532

[26] NICULESCU, C. P. (2000). A new look at Newton's inequalities. *JIPAM. J. Inequal. Pure Appl. Math.* **1** Article 170. MR1786404

[27] PEMANTLE, R. (2000). Towards a theory of negative dependence. *J. Math. Phys.* **41** 1371–1390. MR1757964

[28] RÉNYI, A. (1957). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1** 519–527. MR0094861

[29] SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. Springer, New York. MR2265633

[30] SHEPP, L. A. and OLKIN, I. (1981). Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution. In *Contributions to Probability* 201–206. Academic Press, New York. MR0618689

[31] STURM, K.-T. (2006). On the geometry of metric measure spaces. I. *Acta Math.* **196** 65–131. MR2237206

[32] STURM, K.-T. (2006). On the geometry of metric measure spaces. II. *Acta Math.* **196** 133–177. MR2237207

[33] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[34] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin. MR2459454

[35] YU, Y. (2008). On the maximum entropy properties of the binomial distribution. *IEEE Trans. Inform. Theory* **54** 3351–3353. MR2450793

[36] YU, Y. (2009). Monotonic convergence in an information-theoretic law of small numbers. *IEEE Trans. Inform. Theory* **55** 5412–5422. MR2597172

[37] YU, Y. and JOHNSON, O. T. (2009). Concavity of entropy under thinning. In *Proceedings of ISIT* 2009, 28*th June*–3*rd July* 2009, *Seoul* 144–148. IEEE, New York.

INSTITUT MATÉMATIQUES DE MARSEILLE—UMR 7373
TECHNOPÔLE CHÂTEAU-GOMBERT
39, RUE FRÉDÉRIC JOLIOT-CURIE
13453 MARSEILLE CEDEX 13
FRANCE
E-MAIL: erwan.hillion@univ-amu.fr

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
UNIVERSITY WALK
BRISTOL BS8 1TW
UNITED KINGDOM
E-MAIL: o.johnson@bristol.ac.uk