

Concavity of entropy along binomial convolutions

Erwan Hillion*

Abstract

Motivated by a generalization of Sturm-Lott-Villani theory to discrete spaces and by a conjecture stated by Shepp and Olkin about the entropy of sums of Bernoulli random variables, we prove the concavity in t of the entropy of the convolution of a probability measure a , which has the law of a sum of independent Bernoulli variables, by the binomial measure of parameters $n \geq 1$ and t .

Keywords: Olkin-Shepp conjecture ; concavity of entropy ; binomial distribution.

AMS MSC 2010: 60E15 ; 94A17.

Submitted to ECP on May 15, 2011, final version accepted on January 4, 2012.

1 Introduction

Throughout this article, a probability measure a on \mathbb{Z} will be considered as a family of non-negative real numbers $(a(k))_{k \in \mathbb{Z}}$ satisfying $\sum_{k \in \mathbb{Z}} a(k) = 1$. The entropy of a probability measure a is defined by the formula

$$H(a) := \sum_{k \in \mathbb{Z}} U(a(k)),$$

where the function $U : \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined by $U(0) := 0$ and $U(x) := -x \log(x)$ for $x > 0$.

Given an integer $n \geq 1$ and a real parameter $t \in [0, 1]$, the binomial measure $(b_{n,t}(k))_{k \in \mathbb{Z}}$ is defined by

$$\forall k \in \{0, \dots, n\}, b_{n,t}(k) := \binom{n}{k} t^k (1-t)^{n-k}$$

and by $b_{n,t}(k) := 0$ elsewhere.

Given a probability measure $a = (a(k))_{k \in \mathbb{Z}}$ and an integer $n \geq 1$, the binomial convolution of a (of order n) is the family of measures $(a_{n,t}(k))_{k \in \mathbb{Z}}$ defined for $t \in [0, 1]$ by

$$a_{n,t}(k) := \sum_{l \in \mathbb{Z}} a(k-l) b_{n,t}(l) = \sum_{l \in \mathbb{Z}} a(l) b_{n,t}(k-l).$$

It is easy to see that $a_{n,0}$ is exactly the measure a and $a_{n,1}$ is the measure a shifted n times on its right.

The goal of this paper is to prove the following:

*University of Bristol, United Kingdom. E-mail: Erwan.Hillion@bris.ac.uk.

Theorem 1.1. *Let a be a probability measure on \mathbb{Z} which has the law of a sum of independent Bernoulli variables, and $a_{n,t}$ its binomial convolution for some $n \geq 1$. Then the function $H(t) := H(a_{n,t})$ is concave in t .*

Let us recall what a sum of independent Bernoulli variables (which will often be abbreviated in s.o.i.B.) is:

Definition 1.2. *Given a family of numbers $(p_1, \dots, p_r) \in [0, 1]^r$, the probability measure $\mathcal{B}_{p_1, \dots, p_r}$ defined recursively by $\mathcal{B}_\emptyset := \delta_0$ and*

$$\mathcal{B}_{p_1, \dots, p_{l+1}}(k) := (1 - p_{l+1})\mathcal{B}_{p_1, \dots, p_l}(k) + p_{l+1}\mathcal{B}_{p_1, \dots, p_l}(k - 1)$$

is called the s.o.i.B. of parameters p_1, \dots, p_r .

It follows from the definition that if a is the s.o.i.B. of parameters (p_1, \dots, p_r) , then its binomial convolution $a_{n,t}$ is the s.o.i.B. of parameters $(p_1, \dots, p_r, t, \dots, t)$, where the variable t is repeated n times. In this special case, Theorem 1.1 can be seen as a particular case of the Shepp-Olkin conjecture:

Conjecture 1.3. *Given some $r \geq 1$, the function $(p_1, \dots, p_r) \mapsto H(\mathcal{B}_{p_1, \dots, p_r})$ is concave in the r variables (p_1, \dots, p_r) .*

This conjecture was stated in [6], and the same article gives a proof of the concavity of $t \mapsto H(\mathcal{B}_{p, \dots, p})$, which can also be seen as a consequence of Theorem 1.1 when the measure a is the Dirac measure δ_0 .

Shepp-Olkin conjecture has later been proven true in other cases; for instance Johnson and Yu (see [9]) proved the concavity of $t \mapsto H(\mathcal{B}_{t, \dots, t, 1-t, \dots, 1-t})$, where the variables t and $1 - t$ are not necessarily repeated the same number of times. It seems that Theorem 1.1 gives a proof of new special cases of Conjecture 1.3.

Another motivation for Theorem 1.1 comes from the fact that binomial measures satisfy a discrete version of the transport equation:

$$\forall f : \{0, \dots, n\} \rightarrow \mathbb{R}, \quad \frac{\partial}{\partial t} \sum_{k=0}^n f(k)b_{n,t}(k) = n \sum_{k=0}^n \nabla_{[0,n]} f(k)b_{n,t}(k), \quad (1.1)$$

where the "natural derivation" operator $\nabla_{[0,n]}$ is defined on the set of functions $f : \{0, \dots, n\} \rightarrow \mathbb{R}$ by

$$\nabla_{[0,n]} f(k) := \frac{k}{n}(f(k) - f(k - 1)) + \frac{n - k}{n}(f(k + 1) - f(k)).$$

Some properties of this operator can be found in the (forthcoming) article [2].

Equation (1.1) can be seen as a discrete version of the continuous transport equation

$$\frac{\partial}{\partial t} a_t(x) = -n \frac{\partial}{\partial x} a_t(x) \quad (1.2)$$

(where n can be any real number) whose solutions are defined by an initial measure a_0 and by the formula $a_t := a_0 * \delta_{nt}$. In other words, the family of measures $(a_t)_{t \in [0,1]}$ describes the translation of the measure a_0 on a distance n on its right. Consequently, binomial convolutions can be seen as a canonical way to translate smoothly a probability measure on the discrete line on a distance n on its right.

A curve $(a_t)_{t \in [0,1]}$ in the space of probability measures on \mathbb{R} is a translation if and only if it is a Wasserstein geodesic between two measures a_0 and a_1 such that $a_1 = a_0 * \delta_n$ for some $n \in \mathbb{R}$.

Since the works of Sturm ([7], [8]) and Lott-Villani ([4]), geometric properties of a measured length space (X, d, ν) are linked to concavity properties of the entropy (with respect to ν) functional along the geodesics of the Wasserstein space associated to (X, d) . In particular, these authors have proven the concavity of the entropy functional along Wasserstein geodesics on the real line. Moreover, the worst cases in this concavity theorem correspond to translations of measures, along which the entropy is constant.

This shows that a proof of the concavity of entropy along binomial convolutions could be an important first step in the description of the "non-negatively curved" behavior of the discrete line.

The paper is organized as follows. We begin by stating some non-rigorous arguments, based on a classical Gaussian approximation of binomial laws, that let us hope that Theorem 1.1 should be true. The next section is technical, and presents the two main tools used in the proof of Theorem 1.1, which are a re-statement of the transport equation and the "modulus of log-concavity" whose non-negativity will be used in controlling some Taylor expansions. The last section is devoted to the proof itself of Theorem 1.1.

2 Heuristics

In this section, we provide some non-rigorous arguments that explain why we should expect Theorem 1.1 to be true. More precisely, we prove two results about concavity of entropy: the first for solutions of the transport equation and the second for convolutions of functions by a well-chosen Gaussian kernel.

Let $f(x, t)$ be a positive function satisfying the transport equation

$$\frac{\partial}{\partial t} f(x, t) = -n \frac{\partial}{\partial x} f(x, t). \tag{2.1}$$

We want to prove that its entropy function $H(t) := \int_{\mathbb{R}} U(\rho(x, t)) dx$ is constant (and thus concave). Using the change of variable $y = x + nt$ gives that $H(t) = H(0)$, so H is constant.

The major drawback of this argument is that it does not seem possible to adapt it directly in the discrete case. But it is possible to give another proof of this fact avoiding the change of variable, by writing

$$\begin{aligned} \frac{\partial^2}{\partial t^2} H(t) &= \int_{\mathbb{R}} \left(\frac{\partial^2}{\partial t^2} f(x, t) \right) U'(f(x, t)) + \left(\frac{\partial}{\partial t} f(x, t) \right)^2 U''(x, t) dx \\ &= n^2 \int_{\mathbb{R}} \left(\frac{\partial^2}{\partial x^2} f(x, t) \right) U'(f(x, t)) + \left(\frac{\partial}{\partial x} f(x, t) \right)^2 U''(x, t) dx \\ &= n^2 \int_{\mathbb{R}} \frac{\partial^2}{\partial x^2} U(f(x, t)) dx \\ &= 0. \end{aligned}$$

Remark: Here we used an integration by parts, and this method has a discrete counterpart which is the use of telescopic series. We will use this fact in the proof of Theorem 1.1.

So far, the heuristic discussion suggests us that the entropy of a binomial convolution is almost constant. In the rest of this section, we will show that a better continuous approximation let us hope that this entropy function is actually concave.

In order to construct this “better approximation”, we notice that, although constructed to be a discrete version of δ_{nt} , the binomial measure $\text{bin}(n, t)$ is known to be better approximated by the continuous Gaussian measure $\gamma(nt, nt(1-t))$ whose density $\rho(x, t)$ is the fundamental solution of a “modified transport equation”

$$\frac{\partial}{\partial t} \rho(x, t) = -n \frac{\partial}{\partial x} \rho(x, t) + n \frac{1-2t}{2} \frac{\partial^2}{\partial x^2} \rho(x, t). \tag{2.2}$$

Now, let $f(x, t)$ denote the convolution of a smooth probability density $f(0, x)$ by $\rho(x, t)$. Then $f(x, t)$ is a family of smooth probability densities satisfying the equation (2.2). The next proposition establishes the concavity of the entropy $H(t) := H(f(\cdot, t)) := \int_{\mathbb{R}} U(f(x, t)) dx$.

Proposition 2.1. *The function $t \mapsto H(t)$ satisfies*

$$\frac{\partial^2}{\partial t^2} H(t) = -n^2 \left(\frac{1-2t}{2} \right)^2 \int_{\mathbb{R}} f(x, t) \left(\frac{\partial}{\partial x^2} \log(f(x, t)) \right)^2 dx \leq 0.$$

Remark: The statement of Proposition 2.1 is similar to de Bruijn’s identity (see for example [1, Lem.1]), but studies the second derivative of $H(t)$ instead of the first derivative.

Proof of Proposition 2.1: Let us first consider the auxilliary family of densities $g(x, t) := f(x + nt, t)$, which satisfies

$$H(g(\cdot, t)) = H(f(\cdot, t)) = H(t)$$

and is solution to the PDE

$$\frac{\partial}{\partial t} g(x, t) = n \frac{1-2t}{2} \frac{\partial^2}{\partial x^2} g(x, t). \tag{2.3}$$

Until the end of the proof, we simplify the notations, denoting by g', g'', \dots the successive partial differentials of g with respect to x .

$$\begin{aligned} \frac{\partial^2}{\partial t^2} H(t) &= \int_{\mathbb{R}} -g'' \log(g) - \frac{(g')^2}{g} dx \\ &= \int_{\mathbb{R}} \left(-n^2 \left(\frac{1-2t}{2} \right)^2 g^{(4)} \right) \log(g) - \left(n \frac{1-2t}{2} g'' \right)^2 \frac{1}{g} dx \\ &= -n^2 \left(\frac{1-2t}{2} \right)^2 \int_{\mathbb{R}} g^{(4)} \log(g) + \frac{(g'')^2}{g} dx. \end{aligned}$$

To prove the concavity of H , it thus suffices to show that the last integral is non-negative, and this can be done by tricky integrations by parts:

$$\begin{aligned} \int_{\mathbb{R}} g^{(4)} \log(g) + \frac{(g'')^2}{g} dx &= \int_{\mathbb{R}} g'' (\log(g))'' dx + \int_{\mathbb{R}} \frac{(g'')^2}{g} dx \\ &= \int_{\mathbb{R}} 2 \frac{(g'')^2}{g} - \frac{(g')^2 g''}{g^2} dx \\ &= \int_{\mathbb{R}} 2 \frac{(g'')^2}{g} - \frac{(g')^2 g''}{g^2} dx - \int_{\mathbb{R}} \left(\frac{(g')^3}{g^2} \right)' dx \\ &= \int_{\mathbb{R}} 2 \frac{(g'')^2}{g} - \frac{g'^2 g''}{g^2} dx - \int_{\mathbb{R}} \left(3 \frac{(g')^2 g''}{g^2} - 2 \frac{(g')^4}{g^3} \right) dx \\ &= 2 \int_{\mathbb{R}} g \left(\frac{g''}{g} - \frac{(g')^2}{g^2} \right)^2 dx \\ &= 2 \int_{\mathbb{R}} g (\log(g)'')^2 dx. \end{aligned}$$

This proves the concavity of $H(t)$ and gives an explicit value for its second derivative:

$$\frac{\partial^2}{\partial t^2} H(t) = -n^2 \frac{(1-2t)^2}{2} \int_{\mathbb{R}} f(x,t) \left(\frac{\partial}{\partial x^2} \log(f(x,t)) \right)^2 dx. \square$$

Remark: Actually, the proof of Theorem 1.1 will provide an upper bound on $\frac{\partial^2}{\partial t^2} H(t)$ which will be very similar to the exact value of $\frac{\partial^2}{\partial t^2} H(t)$ found in this continuous approximation (equation (4.7)).

3 Technical tools

In this chapter, we isolate some technical lemmas in order to make the proof of Theorem 1.1 more readable. We first study some "differential" properties of binomial convolutions. Then we introduce the modulus of log-concavity and see how it can be recovered in various formulas involving binomial convolutions.

From now, we will often omit the dependence in t of the measures $(a_{n,t}(k))_{k \in \mathbb{Z}}$.

3.1 The transport equation revisited

The discrete transport equation (1.1) satisfied by binomial measures is in fact not expressed in a way that can be easily used for our purposes. Moreover, there is no nice way to generalize this equation to binomial convolutions. It is more convenient to use instead the following well-known formula.

Definition 3.1. We denote by ∇_1 (resp. ∇_2) the left derivation (resp. the "twice left" derivation) operator, which satisfies, for every function $f : \mathbb{Z} \rightarrow \mathbb{R}$,

$$\nabla_1 f(k) := f(k) - f(k-1), \text{ resp. } \nabla_2 f(k) := f(k) - 2f(k-1) + f(k-2).$$

Let us fix an integer $n \geq 2$ and a probability measure $a = (a(k))_{k \in \mathbb{Z}}$. The next proposition can be seen as another way to express the transport equation, which will be easier to use in the proof of Theorem 1.1:

Proposition 3.2. For any binomial convolution a_n ,

- i : $\frac{\partial}{\partial t} a_n(k) = -n \nabla_1 a_{n-1}(k)$,
- ii : $\frac{\partial^2}{\partial t^2} a_n(k) = n(n-1) \nabla_2 a_{n-2}(k)$.

Proof: By elementary properties of convolution products, it suffices to prove the first point of Proposition 3.2 in the special case where $a_{n,t}(k) = b_{n,t}(k)$, and this point follows from a direct calculation and the elementary formulas $k \binom{n}{k} = n \binom{n-1}{k-1}$ and $(n-k) \binom{n}{k} = n \binom{n-1}{k}$.

The second point of the proposition follows from a double application of the first point. \square

Remark: It is easy to link the measures a_n , a_{n-1} and a_{n-2} by the formulas

$$a_n(k) = (1-t)a_{n-1}(k) + ta_{n-1}(k-1), \tag{3.1}$$

$$a_n(k) = (1-t)^2 a_{n-2}(k) + 2t(1-t)a_{n-2}(k-1) + t^2 a_{n-2}(k-2). \tag{3.2}$$

Remark: We recover in some twisted way the modified transport equation (see Equation (2.2)) in the first point of Proposition 3.2 by noticing that

$$\nabla_1 a_{n-1}(k) = \frac{a_n(k+1) - a_n(k-1)}{2} - \frac{1}{2} ((1-t)\nabla_2 a_{n-1}(k+1) - t\nabla_2 a_{n-1}(k)).$$

3.2 The modulus of log-concavity

A measure $(a(k))_{k \in \mathbb{Z}}$ is said to be log-concave if it satisfies, for every $k \in \mathbb{Z}$, $a(k)^2 \geq a(k-1)a(k+1)$. A well-known property states that any s.o.i.B. (and thus any binomial convolution of a s.o.i.B.) is a log-concave measure (see for instance the introduction of [3]). This suggests us to introduce the modulus of log-concavity:

Definition 3.3. For any $n \geq 2$ and any binomial convolution $(a_n(k))_{k \in \mathbb{Z}}$, the modulus of log-concavity is the function $(v_n(k))_{k \in \mathbb{Z}}$ defined by

$$v_n(k) := a_{n-2}(k-1)^2 - a_{n-2}(k)a_{n-2}(k-2).$$

Remark: The choice of defining $v_n(k)$ in terms of the $(a_{n-2}(k))$ instead of the $(a_n(k))$ will be justified by the simpler forms the next equations will take.

The main interest in introducing the modulus of log-concavity is the fact that, when $(a(k))_{k \in \mathbb{Z}}$ is a s.o.i.B., the log-concavity of the binomial convolution a_{n-2} implies that $v_n(k)$ is always non-negative. A closer look at this fact allows us to be more precise. The formula

$$\nabla_2 \log(a_{n-2}(k)) = \log \left(1 - \frac{v_n(k)}{(a_{n-2}(k-1))^2} \right)$$

can be compared to its continuous analogue $\log(f)'' = \frac{ff'' - f'^2}{f^2}$, which let us hope that, at least when $v_n(k) \ll (a_{n-2}(k-1))^2$, the modulus $v_n(k)$ is close to some expression of the form $(\nabla_1 a_n(k))^2 - a_n(k)\nabla_2 a_n(k)$. This intuition is made rigorous by the following:

Proposition 3.4. For any binomial convolution $(a_n(k))_{k \in \mathbb{Z}}$ with $n \geq 2$, we have:

$$v_n(k) := (\nabla_1 a_{n-1}(k))^2 - a_n(k)\nabla_2 a_{n-2}(k).$$

Proof: The idea of the proof consists in writing $(\nabla_1 a_{n-1}(k))^2 - a_n(k)\nabla_2 a_{n-2}(k)$ only in terms of the $(a_{n-2}(k))_{k \in \mathbb{Z}}$, using formulas (3.1) and (3.2). The equality is reached after a few lines of calculations. \square

The same method can be used to find another formulas which are more difficult to be interpreted, but will be useful in the proof of Theorem 1.1:

Proposition 3.5. For any binomial convolution $a_n(t)$,

- i: $a_{n-1}(k)^2 - a_n(k)a_{n-2}(k) = t^2 v_n(k)$.
- ii: $a_{n-1}(k)a_{n-1}(k-1) - a_n(k)a_{n-2}(k-1) = -t(1-t)v_n(k)$.
- iii: $a_{n-1}(k-1)^2 - a_n(k)a_{n-2}(k-2) = (1-t)^2 v_n(k)$.

4 Proof of Theorem 1.1

The proof of Theorem 1.1 is inspired by the heuristic proof given at the beginning of the second section. This proof was based on two ingredients: the transport equation (2.1) and the use of integrations by parts which lead to the identity

$$\int_{\mathbb{R}} \left(\frac{\partial^2}{\partial x^2} f(x, t) \right) U'(f(x, t)) dx = - \int_{\mathbb{R}} \left(\frac{\partial}{\partial x} f(x, t) \right)^2 U''(f(x, t)) dx. \quad (4.1)$$

As in the heuristic proof, Theorem 1.1 will be proven by a tricky use of integrations by parts, that in the discrete case will take the form of sums of telescopic series. Equation (4.1) suggests the use of the two following telescopic sums (where by convention $0 \cdot \log(0) = 0$):

$$\sum_{k \in \mathbb{Z}} \nabla_2(a_{n-2}(k) \log(a_{n-2}(k))) = 0 \tag{4.2}$$

and

$$\sum_{k \in \mathbb{Z}} \nabla_1(\nabla_1(a_{n-2}(k) \log(a_{n-1}(k)))) = 0. \tag{4.3}$$

The next proposition shows that a tricky use of these two sums gives a very useful inequality:

Proposition 4.1. *For any $k \in \mathbb{Z}$ such that $a_n(k) \neq 0$ we have the inequality:*

$$\begin{aligned} & \nabla_2(a_{n-2}(k) \log(a_n(k))) \\ & -2\nabla_1(\nabla_1(a_{n-2}(k) \log(a_{n-1}(k)))) + \nabla_2(a_{n-2}(k) \log(a_{n-2}(k))) \\ & \geq -\frac{v_n(k)}{a_n(k)}. \end{aligned}$$

Proof: We begin the proof by writing:

$$\begin{aligned} & \nabla_2(a_{n-2}(k) \log(a_n(k))) \\ & -2\nabla_1(\nabla_1(a_{n-2}(k) \log(a_{n-1}(k)))) + \nabla_2(a_{n-2}(k) \log(a_{n-2}(k))) \\ & = a_{n-2}(k) \log\left(\frac{a_{n-2}(k)a_n(k)}{a_{n-1}(k)^2}\right) \tag{4.4} \end{aligned}$$

$$-2a_{n-2}(k-1) \log\left(\frac{a_{n-2}(k-1)a_n(k)}{a_{n-1}(k)a_{n-1}(k-1)}\right) \tag{4.5}$$

$$+a_{n-2}(k-2) \log\left(\frac{a_{n-2}(k-2)a_n(k)}{a_{n-1}(k-1)^2}\right). \tag{4.6}$$

The next step of the proof consists in studying (and bounding by below) each term (4.4), (4.5) and (4.6) of the right-hand side of the above equation separately. In each case, the method consists in using Proposition 3.5 to make appear the modulus of log-concavity and then use some elementary inequalities involving the function U .

For the first term, we write:

$$\begin{aligned} a_{n-2}(k) \log\left(\frac{a_{n-2}(k)a_n(k)}{a_{n-1}(k)^2}\right) &= \frac{a_{n-1}(k)^2}{a_n(k)} U\left(\frac{a_{n-2}(k)a_n(k)}{a_{n-1}(k)^2}\right) \\ &= \frac{a_{n-1}(k)^2}{a_n(k)} U\left(1 - t^2 \frac{v_n(k)}{a_{n-1}(k)^2}\right) \\ &\geq \frac{a_{n-1}(k)^2}{a_n(k)} \left(-t^2 \frac{v_n(k)}{a_{n-1}(k)^2} + \frac{t^4}{2} \frac{v_n(k)^2}{a_{n-1}(k)^4}\right) \\ &= -t^2 \frac{v_n(k)}{a_n(k)} + \frac{t^4}{2} \frac{v_n(k)^2}{a_{n-1}(k)^2}. \end{aligned}$$

We used here the non-negativity of $v_n(k)$ and the elementary inequality

$$\forall x \geq 0, U(1-x) \geq -x + \frac{x^2}{2}.$$

The same method leads to similar bounds for the term (4.6):

$$a_{n-2}(k-2) \log \left(\frac{a_{n-2}(k-2)a_n(k)}{a_{n-1}(k-1)^2} \right) \geq -(1-t)^2 \frac{v_n(k)}{a_n(k)} + \frac{(1-t)^4}{2} \frac{v_n(k)^2}{a_{n-1}(k-1)^2}.$$

The term (4.5) is bounded by using the inequality

$$\forall x \geq 0, -U(1+x) \geq -x - \frac{x^2}{2},$$

which gives

$$a_{n-2}(k-1) \log \left(\frac{a_{n-2}(k-1)a_n(k)}{a_{n-1}(k)a_{n-1}(k-1)} \right) \leq t(1-t) \frac{v_n(k)}{a_n(k)} + \frac{t^2(1-t)^2}{2} \frac{v_n(k)^2}{a_{n-1}(k)a_{n-1}(k-1)}.$$

Combining these three inequalities gives

$$\begin{aligned} & \nabla_2 a_{n-2}(k) \log(a_n(k)) \\ & -2\nabla_1 (\nabla_1(a_{n-2}(k)) \log(a_{n-1}(k))) + \nabla_2(a_{n-2}(k) \log(a_{n-2}(k))) \\ & \geq -\frac{v_n(k)}{a_n(k)} + \frac{1}{2}(v_n(k))^2 \left(\frac{t^2}{a_{n-1}(k)} - \frac{(1-t)^2}{a_{n-1}} \right)^2 \geq -\frac{v_n(k)}{a_n(k)}. \square \end{aligned}$$

The rest of the proof of Theorem 1.1 is now straightforward.

Proof of Theorem 1.1:

Let $(p_1, \dots, p_r) \in [0, 1]^r$ be such that a is the s.o.i.B. of parameters p_1, \dots, p_r . We can suppose that each p_i is different from 0 and 1. In this case, for every $t \in]0, 1[$, the support of $a_{n,t}$ is exactly $\{0, \dots, N\}$, where $N := n + r$.

Moreover, the telescopic series (4.2) and (4.3) can be rewritten (still with the convention that $0 \cdot \log(0) = 0$)

$$\sum_{k=0}^N \nabla_2(a_{n-2}(k) \log(a_{n-2}(k))) = 0$$

and

$$\sum_{k=0}^N \nabla_1 (\nabla_1(a_{n-2}(k)) \log(a_{n-1}(k))) = 0.$$

As each function $t \mapsto a_{n,t}(k)$ is smooth, we have

$$\begin{aligned} \frac{\partial^2}{\partial t^2} H(t) &= -\sum_{k=0}^N \frac{\partial^2}{\partial t^2} a_n(k) U'(a_n(k)) + \left(\frac{\partial}{\partial t} a_n(k) \right)^2 U''(a_n(k)) \\ &= -\sum_{k=0}^N n(n-1) \nabla_2 a_{n-2}(k) U'(a_n(k)) + n^2 (\nabla_1 a_{n-1}(k))^2 U''(a_n(k)) \\ &\leq -n(n-1) \sum_{k=0}^N \nabla_2 a_{n-2}(k) U'(a_n(k)) + (\nabla_1 a_{n-1}(k))^2 U''(a_n(k)), \end{aligned}$$

The inequality comes from the fact that $U''(a_n(k)) = \frac{1}{a_n(k)} \geq 0$. Theorem 1.1 will thus be proven if the last sum is non-negative.

Using Proposition 4.1 gives

$$\begin{aligned} & \sum_{k=0}^N \nabla_2 a_{n-2}(k) U'(a_n(k)) + (\nabla_1 a_{n-1}(k))^2 U''(a_n^k) = \\ & \sum_{k=0}^N \nabla_2 a_{n-2}(k) \log(a_n(k)) - 2 \nabla_1 (\nabla_1(a_{n-2}(k)) \log(a_{n-1}(k))) \\ & \quad + \nabla_2(a_{n-2}(k) \log(a_{n-2}(k))) + \sum_{k=0}^N \frac{(\nabla_1 a_{n-1}(k))^2}{a_n^k} \\ & \geq \sum_{k=0}^N -\frac{v_n(k)}{a_n(k)} + \frac{(\nabla_1 a_{n-1}(k))^2}{a_n(k)}. \end{aligned}$$

By Lemma 3.4,

$$\sum_{k=0}^N -\frac{v_n(k)}{a_n(k)} + \frac{(\nabla_1 a_{n-1}(k))^2}{a_n(k)} = \sum_{k=0}^N \nabla_2 a_{n-2}(k) = 0.$$

Then

$$\sum_{k=0}^N \nabla_2 a_{n-2}(k) U'(a_n(k)) + (\nabla_1 a_{n-1}(k))^2 U''(a_n^k) \geq 0,$$

which proves Theorem 1.1. \square

Remark: A careful look at the proof of Proposition 4.1 gives the bound

$$\frac{\partial^2}{\partial t^2} H(t) \leq -\frac{n(n-1)}{2} \sum_{k=0}^N (v_n(k))^2 \left(\frac{t^2}{a_{n-1}(k)} - \frac{(1-t)^2}{a_{n-1}(k-1)} \right)^2, \quad (4.7)$$

which is very similar to the exact value of $\frac{\partial^2}{\partial t^2} H(t)$ found in Proposition 2.1.

References

- [1] A.R. Barron. Entropy and the central limit theorem *The Annals of probability*, 14(1):336–342, 1986. MR-0815975
- [2] E. Hillion, O. Johnson, and Y. Yu. Translations of probability measures on \mathbb{Z} .
- [3] O. Johnson and C. Goldschmidt. Preservation of log-concavity on summation. *Arxiv preprint math/0502548*, 2005. MR-2219340
- [4] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, 169:903–991. MR-2480619
- [5] P.S. Mateev. On the entropy of a multinomial distribution. *Teoriya Veroyatnostei i ee Primeneniya*, 23(1):196–198, 1978. MR-0490451
- [6] LA Shepp and J. Olkin. Entropy of the sum of independent bernoulli random variables and of the multidimensional distribution. 1981.
- [7] K.T. Sturm. On the geometry of metric measure spaces I. *Acta mathematica*, 196(1):65–131, 2006. MR-2237206
- [8] K.T. Sturm. On the geometry of metric measure spaces II. *Acta mathematica*, 196(1):133–177, 2006. MR-2237207
- [9] Y. Yu and O. Johnson. Concavity of entropy under thinning. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 144–148. IEEE, 2009.

Acknowledgments. We wish to thank the referees for useful comments and remarks, especially about the de Bruijn’s identity.