Stam inequality on $\mathbb{Z}_n$

Paolo Gibilisco[*], Daniele Imparato[†] and Tommaso Isola[‡]

January 18, 2008

Abstract

We prove a discrete version of Stam inequality for random variables taking values on $\mathbb{Z}_n$.

2000 *Mathematics Subject Classification:* 94A17, 62B10
Key words and phrases. Fisher score, Fisher information, Stam inequality.

1 Introduction

If X is a real random variable with differentiable strictly positive density f, then the Fisher information I_X is defined as

$$I_X := \int (f'(x)/f(x))^2 f(x)dx. \tag{1.1}$$

Stam (1959) proved a convolution inequality for I_X. More precisely, if X, Y are independent random variables such that I_X, $I_Y < \infty$, then

$$\frac{1}{I_{X+Y}} \geq \frac{1}{I_X} + \frac{1}{I_Y}, \tag{1.2}$$

with equality if and only if X, Y are Gaussian.

From this inequality important results like the entropy power inequality and the log-Sobolev inequality follow, see Blachman (1965), Carlen (1991), Kagan and Landsman (1997), Zamir (1998). For a recent application in statistical mechanics see Villani (2003). Recently, the inequality has been greatly generalized in Madiman and Barron (2007).

A free analogue of Stam inequality has been proved in free probability by the introduction of the free Fisher information. In this case the equality in (1.2) characterizes the free analogue of the Gaussian distribution, namely the semicircular Wigner distribution, see Voiculescu (1998).

Discrete analogues of the Fisher score and Fisher information have been already discussed in the literature, see Johnstone and MacGibbon (1987), Kontoyiannis, Harremoes and Johnson (2005), Madiman, Johnson and Kontoyiannis (2007). A version of Stam inequality on the set of integers $\mathbb{Z}$ has been proved by Papathanasiou (1993) and rediscovered by Kagan (2001). In this case the equality characterizes the Poisson distribution. As noted in the final comments in Kagan (2001a), some features of Stam inequality appear group theoretical in character.

In this paper we show that, after suitable modifications, the proof for $\mathbb{Z}$ applies to the cyclic group $\mathbb{Z}_n$ giving a further version of Stam inequality. In the present case we show that the equality characterizes the uniform distribution. This is in some sense natural because the uniform distribution maximizes entropy on $\mathbb{Z}_n$; in the appropriate contexts this is true also for the Gaussian, Poisson and Wigner distributions, see Voiculescu (1998), Harremoes (2001), Johnson(2007).

[*]Dipartimento SEFEMEQ, Facoltà di Economia, Università di Roma "Tor Vergata", Via Columbia 2, 00133 Rome, Italy. Email: gibilisco@volterra.uniroma2.it – URL: http://www.economia.uniroma2.it/sefemeq/professori/gibilisco
[†]Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy. Email: daniele.imparato@polito.it
[‡]Dipartimento di Matematica, Università di Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy. Email: isola@mat.uniroma2.it – URL: http://www.mat.uniroma2.it/∼isola

2 Preliminaries

We recall the formulation of Stam inequality in two cases, where it has already been proved.

2.1 Stam inequality on $\mathbb{R}$

Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable, strictly positive density. One may define the Fisher f-score $J_f : \mathbb{R} \to \mathbb{R}$ by

$$J_f := \frac{f'}{f}.$$

Let $(\Omega, \mathcal{F}, p)$ be a probability space. In general, if $X : (\Omega, \mathcal{F}, p) \to \mathbb{R}$ is a random variable with density f we write $J_X = J_f$ and define the Fisher information (about a shift parameter) as

$$I_X := \mathrm{Var}_f(J_f) = E_f(J_f^2);$$

namely,

$$I_X = \int_{\mathbb{R}} (f'(x)/f(x))^2 f(x) dx. \tag{2.1}$$

Theorem 2.1. *(see Stam (1959), Blachman (1965))*
If $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{R}$ are independent random variables such that I_X, $I_Y < \infty$, then

$$\frac{1}{I_{X+Y}} \geq \frac{1}{I_X} + \frac{1}{I_Y}, \tag{2.2}$$

with strict equality if and only if X, Y are Gaussian.

2.2 Stam inequality on $\mathbb{Z}$

Let $f : \mathbb{Z} \to \mathbb{R}$ be a (discrete) density. We say that f belongs to the class RSP (right side positivity) if

$$f(k) > 0 \Longrightarrow f(k+1) > 0.$$

If $f \in RSP$, then we may define the Fisher f-score by

$$J_f(k) = \begin{cases} \frac{f(k)-f(k-1)}{f(k)} & f(k) > 0, \\ 0 & f(k) = 0. \end{cases}$$

If $X : (\Omega, \mathcal{F}, p) \to \mathbb{Z}$ is a random variable with (discrete) density $f \in \mathrm{RSP}$ we write $J_X = J_f$ and define the Fisher information (about a shift parameter) as

$$I_X := \mathrm{Var}_f(J_f) = E_f(J_f^2).$$

Theorem 2.2. *(see Papathanasiou (1993), Kagan (2001a))*
If $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{Z}$ are independent random variables with densities in RSP and such that I_X, $I_Y < \infty$, then

$$\frac{1}{I_{X+Y}} \geq \frac{1}{I_X} + \frac{1}{I_Y}, \tag{2.3}$$

with strict equality if and only if X, Y have (possibly shifted) Poisson distribution.

3 Stam inequality on $\mathbb{Z}_n$

We denote by $\mathbb{Z}_n$ the cyclic group $\{0, 1, ..., n-1\}$. Introduce the class $\mathcal{P}$ of strictly positive densities, that is

$$\mathcal{P} := \left\{ f : \mathbb{Z}_n \to \mathbb{R} \Big| \sum_{j \in \mathbb{Z}_n} f(j) = 1, \quad f(k) > 0 \ \ \forall k \in \mathbb{Z}_n \right\}.$$

We assume, from now on, that all densities are strictly positive.

Let $f \in \mathcal{P}$. In analogy with the previous definitions, we may introduce the Fisher f-score $J_f : \mathbb{Z}_n \to \mathbb{R}$ by

$$J_f(k) := \frac{f(k) - f(k-1)}{f(k)}.$$

By definition of expectation and since $k \in \mathbb{Z}_n$, it is straightforward to see that J_f is an f-centered random variable; namely,

$$\mathbb{E}_f[J_f] = 0. \tag{3.1}$$

Next, if $X : (\Omega, \mathcal{F}, p) \to \mathbb{Z}_n$ is a random variable with density $f(k) = f_X(k) := p(X = k)$, where $f_X \in \mathcal{P}$, from the score $J_X := J_f$ it is possible to define the Fisher information (about a shift parameter) as

$$I_X := \mathrm{Var}_f(J_f) = E_f(J_f^2).$$

Note that, due to the finiteness of $\mathbb{Z}_n$, in this case the condition $I_X < \infty$ always holds. However, we cannot ensure in general that $I_X \neq 0$. In fact, it is easy to characterize this degenerate case.

Lemma 3.1. *The following conditions are equivalent*

(1) X *has uniform distribution;*

(2) $J_X = 0$*;*

(3) $I_X = 0$*;*

(4) J_X *has constant increments, namely* $J_X(x+1) - J_X(x) = \alpha = constant$.

Proof. The equivalence of $(1), (2), (3)$ is immediate by definitions. It is also obvious that (2) implies (4). Therefore, it is enough to prove that (4) implies (1).

Let

$$J_X(x+1) - J_X(x) = \alpha = constant.$$

We have

$$J_X(n) = J_X(0) + n\alpha = J_X(0).$$

This implies $n\alpha = 0$, namely $\alpha = 0$, that is, $J_X = constant$. Therefore, there exists a constant $\beta > 0$ such that

$$\frac{f_X(x-1)}{f_X(x)} = \frac{1}{\beta}.$$

We have

$$f_X(x) = \beta f_X(x-1),$$

namely

$$f_X(n) = \beta^n f_X(0) = f_X(0),$$

so that $\beta^n = 1$. This implies $\beta = 1$, that is, X is uniform. This concludes the proof. $\square$

Let us recall also the following result that is immediate by using the convolution formula.

Proposition 3.2. *If* $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{Z}_n$ *are independent random variables and* X *is uniform then also* $Z = X + Y$ *is uniform.*

Proposition 3.3. *Let $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{Z}_n$ be independent random variables such that their densities belong to $\mathcal{P}$. If X or Y has uniform distribution, then*

$$\frac{1}{I_{X+Y}} = \frac{1}{I_X} + \frac{1}{I_Y},$$

in the sense that both sides of equality are equal to infinity.

Proof. Let $Z = X + Y$. If X is uniform, then Z is uniform by Proposition 3.2 and we are done by Lemma 3.1. □

Because of the above proposition, in what follows we consider random variables with strictly positive Fisher information.

Before the proof of the main result, we need the following lemma.

Lemma 3.4. *Let $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{Z}_n$ be two independent random variables with densities $f_X, f_Y \in \mathcal{P}$ and let $Z := X + Y$. Then*

$$J_Z(Z) = \mathbb{E}_p[J_X(X)|Z] = \mathbb{E}_p[J_Y(Y)|Z]. \tag{3.2}$$

Proof. Let f_Z be the density of Z; namely,

$$f_Z(k) = \sum_{j=0}^{n-1} f_X(k-j) f_Y(j), \quad k \in \mathbb{Z}_n,$$

with $f_Z \in \mathcal{P}$. Then,

$$\begin{array}{rcl} f_Z(k) - f_Z(k-1) & = & \displaystyle\sum_{j=0}^{n-1} [f_X(k-j) - f_X(k-j-1)] f_Y(j) \\ & = & \displaystyle\sum_{j=0}^{n-1} f_Y(k-j)[f_X(j) - f_X(j-1)] \end{array}.$$

Therefore, given $k \in \mathbb{Z}_n$,

$$\begin{array}{rcl} J_Z(k) & = & \dfrac{f_Z(k) - f_Z(k-1)}{f_Z(k)} \\ & = & \displaystyle\sum_{j=0}^{n-1} \frac{f_X(j) f_Y(k-j)}{f_Z(k)} \frac{[f_X(j) - f_X(j-1)]}{f_X(j)} \\ & = & \displaystyle\sum_{j=0}^{n-1} J_X(j) p(X = j | Z = k) \\ & = & \mathbb{E}_{f_X}[J_X | Z = k] \\ & = & \mathbb{E}_p[J_X(X) | Z = k]. \end{array}$$

Similarly, by symmetry of the convolution formula one can obtain

$$J_Z(k) = \mathbb{E}_p[J_Y(Y)|Z = k], \quad k \in \mathbb{Z}_n,$$

proving Lemma 3.4.

□

We are ready to prove the main result. As already said in the Introduction, note that not only the statement but also the proof of the following theorem exactly mimics the cases $\mathbb{R}$ and $\mathbb{Z}$.

Theorem 3.5. *Let $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{Z}_n$ be two independent random variables such that $I_X, I_Y > 0$. Then*

$$\frac{1}{I_{X+Y}} > \frac{1}{I_X} + \frac{1}{I_Y}. \tag{3.3}$$

Proof. Define $Z := X + Y$. Let $a, b \in \mathbb{R}$; then, by Lemma 3.4

$$\begin{aligned} \mathbb{E}_p[aJ_X(X) + bJ_Y(Y)|Z] &= a\mathbb{E}_p[J_X(X)|Z] + b\mathbb{E}_p[J_Y(Y)|Z] \\ &= (a+b)J_Z(Z). \end{aligned} \tag{3.4}$$

Hence, by applying Jensen's inequality it holds

$$\begin{aligned} \mathbb{E}_p[(aJ_X(X) + bJ_Y(Y))^2] &= \mathbb{E}_p[\mathbb{E}_p[(aJ_X(X) + bJ_Y(Y))^2|Z]] \\ &\geq \mathbb{E}_p[\mathbb{E}_p[aJ_X(X) + bJ_Y(Y)|Z]^2] \\ &= \mathbb{E}_p[(a+b)^2 J_Z(Z)^2] \\ &= (a+b)^2 I_Z, \end{aligned} \tag{3.5}$$

and thus

$$\begin{aligned} (a+b)^2 I_Z &\leq \mathbb{E}_p[(aJ_X(X) + bJ_Y(Y))^2] \\ &= a^2\mathbb{E}_p[J_X(X)^2] + 2ab\mathbb{E}_p[J_X(X)J_Y(Y)] + b^2\mathbb{E}_p[J_X(X)^2] \\ &= a^2 I_X + b^2 I_Y + 2ab\mathbb{E}_p[J_X(X)J_Y(Y)] \\ &= a^2 I_X + b^2 I_Y, \end{aligned}$$

where the last equality follows from independence and (3.1).

Now, take $a := 1/I_X$ and $b := 1/I_Y$; then we obtain

$$\left(\frac{1}{I_X} + \frac{1}{I_Y}\right)^2 I_Z \leq \frac{1}{I_X} + \frac{1}{I_Y}.$$

It remains to be proved that the equality sign cannot hold in (3.3). To this purpose, define $c := a + b$, where, again, $a = 1/I_X$ and $b = 1/I_Y$; then equality holds if and only if

$$c^2 I_Z = a^2 I_X + b^2 I_Y. \tag{3.6}$$

Let us first prove that (3.6) is equivalent to

$$aJ_X(X) + bJ_Y(Y) = cJ_Z(X+Y). \tag{3.7}$$

Indeed, let $H := aJ_X(X) + bJ_Y(Y)$; then equality in (3.5) occurs if and only if

$$\mathbb{E}_p[H^2|Z] = (\mathbb{E}_p[H|Z])^2,$$

i.e.

$$\mathbb{E}_p[(H - \mathbb{E}_p[H|Z])^2|Z] = 0.$$

Therefore, $H = \mathbb{E}_p[H|Z]$, so that

$$\begin{aligned} \mathbb{E}_p[aJ_X(X) + bJ_Y(Y)|Z] &= aJ_X(X) + bJ_Y(Y) \\ &= cJ_Z(Z), \end{aligned}$$

due to (3.4). Conversely, if (3.7) holds, then by applying the squared power and the expectation operator we obtain (3.6).

Let $x, y \in \mathbb{Z}_n$; because of independence

$$p(X = x+1 \cap Y = y) = p(X = x+1) \cdot p(Y = y) \neq 0.$$

Thus, it makes sense to write equality (3.7) on $A := (X = x+1) \cap (Y = y)$, so that

$$aJ_X(x+1) + bJ_Y(y) = cJ_Z(x+y+1)$$

and analogously

$$aJ_X(x) + bJ_Y(y+1) = cJ_Z(x+y+1).$$

Subtracting these relations one has

$$a[J_X(x+1) - J_X(x)] = b[J_Y(y+1) - J_Y(y)], \quad \forall x, y \in \mathbb{Z}_n.$$

Therefore, the increments of each score are constant, so that $I_X = 0$ from Lemma 3.1. In the hypotheses we assumed $I_X > 0$ and this contradiction ends the proof. □

References

Blachman, N. M., 1965. The convolution inequality for entropy powers. IEEE Trans. Inform. theory IT-11, 267–271.

Carlen, E., 1991. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. 101, 194-211.

Harremoes, P., 2001. Binomial and Poisson distribution as maximum entropy distributions. IEEE Trans. Inform. Theory 47, 2039-2041.

Johnson, O. T., 2007. Log-concavity and the maximum entropy property of the Poisson distribution. Stoch Proc. Appl. 117, 791-802.

Johnstone, I.M., MacGibbon, B., 1987. Une mesure d'information caractérisant la loi de Poisson, in Séminaire de Probabilités, XXI, vol. 1247 of Lecture Notes in Math., Springer, Berlin, pp. 563–573.

Kagan, A., Landsman Z., 1997. Statistical meaning of Carlen's superadditivity of the Fisher information. Statist. Prob. Letters 32, 175–179.

Kagan, A., 2001. A discrete version of Stam inequality and a characterization of the Poisson distribution. J. Statist. Plann. Inference 92, 7–12.

Kagan, A., 2001. Letter to the editor: "A discrete version of Stam inequality and a characterization of the Poisson distribution". [J. Statist. Plann. Inference 92, 7–12, 2001]. J. Statist. Plann. Inference 99, 1.

Kontoyiannis, I., Harremoës, P., Johnson, O., 2005. Entropy and the law of small numbers. IEEE Trans. Inform. Theory 51, 466-472.

Madiman, M., Barron, R. A., 2007. Generalized entropy power inequalities and monotonicity properties of information. IEEE Trans. Inform. Theory 53, 2317-2329.

Madiman, M., Johnson, O., Kontoyiannis, I., 2007. Fisher information, compound Poisson approximation and the Poisson channel, in Proc. IEEE Intl. Symp. Inform. Theory, Nice, France.

Papathanasiou, V., 1993. Some characteristic properties of the Fisher information matrix via Cacoullo-type inequalities. J. Multivariate Anal. 44, 256-265.

Stam, A.J., 1959. Some inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control 2, 101-112.

Villani, C., 2003. Cercignani's conjecture is sometimes true and always almost true. Comm. Math. Phys. 234, 455-490.

Voiculescu, D., 1998. The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms. Invent. Math. 132, 189–227.

Zamir, R., 1998. A proof of the Fisher information inequality via a data processing argument. IEEE Trans. Inform. Theory 44, 1246–1250.