Fisher information and Stam inequality on a finite group

Paolo Gibilisco[*] and Tommaso Isola[†]

February 11, 2008

Abstract

We prove a discrete version of Stam inequality for random variables taking values on a finite group.

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

1 Introduction

The Fisher information I_X of a real random variable (with strictly positive differentiable density function f) is defined as

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

If X, Y are independent random variables such that I_X, $I_Y < \infty$, Stam was able to prove the inequality

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

where equality holds iff X, Y are Gaussian (see [17, 1]).

This result has been very useful in a manifold of different areas: analysis, probability, statistics, information theory, statistical mechanics and so on (see [2, 4, 10, 5, 18]). Therefore it is not surprising that different proofs and generalizations appear in the recent literature on the subject (see for example [20, 14]).

A free analogue of Fisher information has been introduced in free probability. Also in this case one can prove a Stam-like inequality. The equality case characterizes the Wigner distribution that, in many respects, is the free analogue of the Gaussian distribution (see [19]).

Discrete versions of Fisher information and the Stam inequality are well-known. On the integers $\mathbb{Z}$, equality characterizes the Poisson distribution, while on the cyclic group $\mathbb{Z}_n$, equality occurs for the uniform distribution (see [9, 16, 11, 12, 13, 15, 6]).

It has been observed that there are group-theoretical features in the proof of Stam inequality (see [11]). Nevertheless, up to now, inequality (1.2) has been proved only on specific groups like $\mathbb{R}, \mathbb{Z}$ and $\mathbb{Z}_n$. In this paper we consider the family of all finite groups and we show that, *mutatis mutandis* one can introduce Fisher information and prove Stam inequality on an arbitrary finite group.

2 Preliminaries

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

[*]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, Università di Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy. Email: isola@mat.uniroma2.it – URL: http://www.mat.uniroma2.it/∼isola

2.1 Stam inequality on $\mathbb{R}$ and $\mathbb{R}^n$

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 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. [17, 1] *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 equality if and only if X, Y are Gaussian (with the same covariance).

The same result holds for random vectors. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable, strictly positive density. We use the notation $f_{x_i} = \frac{\partial f}{\partial x_i}$. One may define the Fisher f-score (in the direction x_i) $J_f^{x_i} : \mathbb{R}^n \to \mathbb{R}$ by

$$J_f^{x_i} := \frac{f_{x_i}}{f}.$$

Let $(\Omega, \mathcal{F}, p)$ be a probability space. In general, if $X : (\Omega, \mathcal{F}, p) \to \mathbb{R}^n$ is a random vector with density f we write $J_X^{x_i} = J_f^{x_i}$ and define the Fisher information as (see p.201 in [2] and p. 838 in [3])

$$I_X := \sum_{i=1}^{n} E_f[(J_f^{x_i})^2].$$

Note that in this case I_X is the trace of the Fisher information matrix

$$E_f\left[\frac{\partial \log(f)}{\partial x_i}\frac{\partial \log(f)}{\partial x_j}\right] \qquad i, j = 1, ..., n$$

Theorem 2.2. *If $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{R}^n$ are independent random vectors 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 equality if and only if X, Y are Gaussian (with the same covariance matrix).

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 $f \in$ RSP we write $J_X = J_f$ and define the Fisher information as

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

Theorem 2.3. [16, 11] *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*

$$\frac{1}{I_{X+Y}} \geq \frac{1}{I_X} + \frac{1}{I_Y}, \tag{2.4}$$

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

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

Let $f : \mathbb{Z}_n \to \mathbb{R}$ be a (discrete) density that is strictly positive. We define the Fisher f-score by

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

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

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

Theorem 2.4. [6] *If $X, Y : (\Omega, \mathcal{F}, p) \to \mathbb{Z}$ are independent random variables with positive densities then*

$$\frac{1}{I_{X+Y}} \geq \frac{1}{I_X} + \frac{1}{I_Y}, \tag{2.5}$$

with equality if and only if X and Y have uniform distribution.

In the following section we generalize this result to an arbitrary finite group.

3 Stam inequality on a finite group

Let G be a finite group. Introduce the class $\mathcal{P}$ of strictly positive densities, that is

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

We assume, from now on, that all densities belong to $\mathcal{P}$.

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

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

Then, J_f^g is an f-centered random variable

$$\mathbb{E}_f[J_f^g] := \sum_{k \in G} J_f^g(k) f(k) = 0. \tag{3.1}$$

If $X : (\Omega, \mathcal{F}, p) \to G$ is a random variable with density $f_X(k) := p(X = k)$, and if $f_X \in \mathcal{P}$, define the score $J_X^g := J_f^g$.

Lemma 3.1. *Let $g \in G$, $X, Y : (\Omega, \mathcal{F}, p) \to G$ be two independent random variables with densities $f_X, f_Y \in \mathcal{P}$ and let $Z := XY$. Then,*

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

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

$$f_Z(k) = \sum_{j \in G} f_X(j) f_Y(j^{-1}k), \quad k \in G,$$

so that $f_Z \in \mathcal{P}$. Then,

$$\begin{aligned} f_Z(k) - f_Z(g^{-1}k) &= \sum_{j\in G} f_X(j) f_Y(j^{-1}k) - \sum_{j\in G} f_X(j) f_Y(j^{-1}g^{-1}k) \\ &= \sum_{j\in G} \big(f_X(j) - f_X(g^{-1}j)\big) f_Y(j^{-1}k). \end{aligned}$$

Therefore, for $k \in G$,

$$\begin{aligned} J_Z^g(k) &= \frac{f_Z(k) - f_Z(g^{-1}k)}{f_Z(k)} \\ &= \sum_{j\in G} \frac{f_X(j) f_Y(j^{-1}k)}{f_Z(k)} \frac{f_X(j) - f_X(g^{-1}j)}{f_X(j)} \\ &= \sum_{j\in G} J_X^g(j) p(X=j|Z=k) \\ &= \mathbb{E}_{f_X}[J_X^g | Z=k] \\ &= \mathbb{E}_p[J_X^g(X)|Z=k]. \end{aligned}$$

Similarly, by symmetry of the convolution formula one obtains

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

proving the claim. □

Lemma 3.2. *Let $g \in G$, $X, Y : (\Omega, \mathcal{F}, p) \to G$ be two independent random variables with densities $f_X, f_Y \in \mathcal{P}$ and let $Z := XY$. Then, for any $a, b \in \mathbb{R}$, we have*

$$(a+b)^2 \mathbb{E}_p[J_Z^g(Z)^2] \le a^2 \mathbb{E}_p[J_X^g(X)^2] + b^2 \mathbb{E}_p[J_Y^g(Y)^2]. \tag{3.2}$$

Moreover, if equality holds in (3.2) *then J_X^g, J_Y^g are constant on G.*

Proof. By Lemma 3.1

$$\mathbb{E}_p[aJ_X^g(X) + bJ_Y^g(Y)|Z] = a\mathbb{E}_p[J_X^g(X)|Z] + b\mathbb{E}_p[J_Y^g(Y)|Z] = (a+b)J_Z^g(Z). \tag{3.3}$$

Hence, by applying Jensen's inequality it holds

$$\begin{aligned} \mathbb{E}_p[(aJ_X^g(X) + bJ_Y^g(Y))^2] &= \mathbb{E}_p[\mathbb{E}_p[(aJ_X^g(X) + bJ_Y^g(Y))^2|Z]] \\ &\ge \mathbb{E}_p[\mathbb{E}_p[aJ_X^g(X) + bJ_Y^g(Y)|Z]^2] \\ &= \mathbb{E}_p[(a+b)^2 J_Z^g(Z)^2] \\ &= (a+b)^2 \mathbb{E}_p[J_Z^g(Z)^2], \end{aligned} \tag{3.4}$$

and thus

$$\begin{aligned} (a+b)^2 \mathbb{E}_p[J_Z^g(Z)^2] &\le \mathbb{E}_p[(aJ_X^g(X) + bJ_Y^g(Y))^2] \\ &= a^2 \mathbb{E}_p[J_X^g(X)^2] + 2ab\mathbb{E}_p[J_X^g(X)J_Y^g(Y)] + b^2\mathbb{E}_p[J_Y^g(Y)^2] \\ &= a^2 \mathbb{E}_p[J_X^g(X)^2] + b^2\mathbb{E}_p[J_Y^g(Y)^2], \end{aligned}$$

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

We now consider the case of equality in (3.2). Set $c = a+b$, and let us prove that equality in (3.2) holds iff

$$aJ_X^g(X) + bJ_Y^g(Y) = cJ_Z^g(XY). \tag{3.5}$$

Indeed, let $H := aJ_X^g(X) + bJ_Y^g(Y)$; then equality in (3.4) 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

$$\mathbb{E}_p[aJ_X^g(X) + bJ_Y^g(Y)|Z] = aJ_X^g(X) + bJ_Y^g(Y) = cJ_Z^g(Z),$$

due to (3.3). Conversely, if (3.5) holds, then by applying the squared power and the expectation operator we obtain equality in (3.2).

Using (3.5), we now prove the last statement of the Lemma. Let us choose a set of generators of the group G, *i.e.* $\Gamma := \{\gamma_1, \dots, \gamma_n\} \subset G$ such that the subgroup generated by them is the whole G. Let us also denote by $\Gamma^{-1} := \{\gamma_1^{-1}, \dots, \gamma_n^{-1}\}$. Let now $x, y \in G$; because of independence, for any $\gamma \in \Gamma \cup \Gamma^{-1}$,

$$\begin{cases} p(X = x\gamma, Y = y) = p(X = x\gamma) \cdot p(Y = y) > 0 \\ p(X = x, Y = \gamma y) = p(X = x) \cdot p(Y = \gamma y) > 0. \end{cases}$$

Thus, it makes sense to write equality (3.5) on $A := \{X = x\gamma\} \cap \{Y = y\}$ and on $B := \{X = x\} \cap \{Y = \gamma y\}$, so that

$$\begin{cases} aJ_X^g(x\gamma) + bJ_Y^g(y) = cJ_Z^g(x\gamma y) \\ aJ_X^g(x) + bJ_Y^g(\gamma y) = cJ_Z^g(x\gamma y). \end{cases}$$

Subtracting these relations one has

$$a[J_X^g(x\gamma) - J_X^g(x)] = b[J_Y^g(\gamma y) - J_Y^g(y)], \quad \forall x, y \in G.$$

Therefore, for any $\gamma \in \Gamma \cup \Gamma^{-1}$, there is $a(\gamma) \in \mathbb{R}$ such that $J_X^g(x\gamma) - J_X^g(x) = a(\gamma)$, for any $x \in G$. Thus, if n is the order of γ in G, *i.e.* $\gamma^n = e$, the identity of G, then $J_X^g(x) = J_X^g(x\gamma^n) = J_X^g(x) + na(\gamma)$, for any $x \in G$, which implies $a(\gamma) = 0$. Therefore

$$J_X^g(x\gamma) = J_X^g(x), \quad x \in G, \gamma \in \Gamma \cup \Gamma^{-1}. \tag{3.6}$$

Since any $k \in G$, $k \neq e$, can be written as a product of elements in $\Gamma \cup \Gamma^{-1}$, *i.e.* $k = \gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_\ell}$, for $\gamma_{i_j} \in \Gamma \cup \Gamma^{-1}$, we can use (3.6) iteratively, and obtain

$$J_X^g(xk) = J_X^g(x\gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_\ell}) = J_X^g(x\gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_{\ell-1}}) = \ldots = J_X^g(x), \quad x \in G.$$

In particular, for $x = e$ we obtain $J_X^g(k) = J_X^g(e)$, for any $k \in G$, *i.e.* J_X^g is constant on G. The proof for J_Y^g is analogous. $\square$

Let us now fix a set of generators of the group G, *i.e.* $\Gamma := \{\gamma_1, \dots, \gamma_n\} \subset G$ such that the subgroup generated by them is the whole G. If $X : (\Omega, \mathcal{F}, p) \to G$ is a random variable with density $f_X \in \mathcal{P}$, define the Fisher information

$$I_X := \sum_{\gamma \in \Gamma} \mathbb{E}_f[(J_f^\gamma)^2] = \sum_{\gamma \in \Gamma} \sum_{k \in G} \Big(\frac{f(k) - f(\gamma^{-1}k)}{f(k)}\Big)^2 f(k).$$

Note that, due to the finiteness of G, 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 3.3. *The following conditions are equivalent*

(1) *X has uniform distribution;*

(2) *$J_X^\gamma = 0$, for any $\gamma \in \Gamma$;*

(3) *$I_X = 0$;*

(4) *J_X^γ is constant, for any $\gamma \in \Gamma$.*

Proof. The implications (1) $\implies$ (2) $\implies$ (4) are immediately proved. The equivalence of (2) and (3) is also easy to show.
Therefore, it is enough to prove that (4) implies (1). So, for any $\gamma \in \Gamma$ there is $a(\gamma) \in \mathbb{R}$ such that $J_X^\gamma(x) = a(\gamma)$, for any $x \in G$, *i.e.*, with $f \equiv f_X$, $\frac{f(x)-f(\gamma^{-1}x)}{f(x)} = a(\gamma)$, for any $x \in G$, which is equivalent to

$$(1 - a(\gamma))f(x) = f(\gamma^{-1}x), \quad x \in G. \tag{3.7}$$

Thus, if n is the order of γ^{-1} in G, *i.e.*$\gamma^{-n} = e$, the identity of G, then

$$(1 - a(\gamma))^n f(x) = (1 - a(\gamma))^{n-1} f(\gamma^{-1}x) = \ldots = f(\gamma^{-n}x) = f(x),$$

for any $x \in G$, which implies $a(\gamma) = 0$. Therefore,

$$f(\gamma^{-1}x) = f(x), \quad x \in G, \gamma \in \Gamma. \tag{3.8}$$

From this it also follows

$$f(\gamma x) = f(x), \quad x \in G, \gamma \in \Gamma. \tag{3.9}$$

Since any $k \in G$, $k \neq e$, can be written as a product of elements in $\Gamma \cup \Gamma^{-1}$, *i.e.* $k = \gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_\ell}$, for $\gamma_{i_j} \in \Gamma \cup \Gamma^{-1}$, we can use (3.8) and (3.9) iteratively, and obtain

$$f(kx) = f(\gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_\ell}x) = f(\gamma_{i_2}\cdots\gamma_{i_\ell}x) = \ldots = f(x), \quad x \in G.$$

In particular, for $x = e$ we obtain $f(k) = f(e)$, for any $k \in G$, *i.e.* f is constant on G, that is, X is uniform. This concludes the proof. □

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

Proposition 3.4. *If $X, Y : (\Omega, \mathcal{F}, p) \to G$ are independent random variables and X is uniform then also $Z = XY$ is uniform.*

Proposition 3.5. *Let $X, Y : (\Omega, \mathcal{F}, p) \to$ be independent random variables such that their densities belong to $\mathcal{P}$. If X or Y has uniform distribution, then*

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

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

Proof. Let $Z = XY$. If X is uniform, then Z is uniform by Proposition 3.4 and we are done by Lemma 3.3. □

Because of the above proposition, it remains to consider random variables with strictly positive Fisher information.

We are ready to prove the main result.

Theorem 3.6. *Let $X, Y : (\Omega, \mathcal{F}, p) \to G$ be two independent random variables such that $I_X, I_Y > 0$. Then*

$$\frac{1}{I_{XY}} > \frac{1}{I_X} + \frac{1}{I_Y}. \tag{3.10}$$

Proof. Define $Z := XY$, and let $a, b \in \mathbb{R}$. Then, for any $\gamma \in \Gamma$ we have from (3.2)

$$(a+b)^2 \mathbb{E}_p[J_Z^\gamma(Z)^2] \leq a^2 \mathbb{E}_p[J_X^\gamma(X)^2] + b^2 \mathbb{E}_p[J_Y^\gamma(Y)^2]. \tag{3.11}$$

Summing up over $\gamma \in \Gamma$, we obtain

$$(a+b)^2 I_Z = (a+b)^2 \sum_{\gamma\in\Gamma} \mathbb{E}_p[J_Z^\gamma(Z)^2] \leq a^2 \sum_{\gamma\in\Gamma} \mathbb{E}_p[J_X^\gamma(X)^2] + b^2 \sum_{\gamma\in\Gamma} \mathbb{E}_p[J_Y^\gamma(Y)^2] = a^2 I_X + b^2 I_Y.$$

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 (3.10). 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

$$c^2 I_Z = a^2 I_X + b^2 I_Y.$$

This implies equality in (3.11), for any $\gamma \in \Gamma$. From Lemma 3.2 it follows that J_X^γ, J_Y^γ are constant on G, for any $\gamma \in \Gamma$. But then, from Lemma 3.3, $I_X = I_Y = 0$, which is absurd. So equality cannot hold in (3.10). $\square$

References

[1] Blachman N. M., The convolution inequality for entropy powers. *IEEE Trans. Inform. theory*, 11:267–271, 1965.

[2] Carlen E., Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 101(1): 194-211, 1991.

[3] Costa M.H.M. and Cover T., On the similarity of the entropy power inequality and the Brunn-Minkowski inequality *IEEE Trans. Inform. Theory*, vol. IT-30: 837-839, 1984.

[4] Dembo A., Cover T. and Thomas J., Information theoretic inequalities, *IEEE Trans. Inform. Theory*, 37(6): 1501–1518, 1991.

[5] Gardner, R.J., The Brunn-Minkowski inequality, *Bull. Amer. Math. Soc.*, 39(3): 355-405, 2002.

[6] Gibilisco P., Imparato D., Isola T., Stam inequality on $\mathbb{Z}_n$, *Statis. Probab. Lett.* (to appear), 2008.

[7] Harremoes P., Binomial and Poisson distribution as maximum entropy distributions. *IEEE Trans. Inform. Theory*, 47(5): 2039-2041, 2001.

[8] Johnson O. T., Log-concavity and the maximum entropy property of the Poisson distribution. *Stoch. Proc. Appl.*, 117(6): 791-802, 2007.

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

[10] Kagan A., Landsman Z., Statistical meaning of Carlen's superadditivity of the Fisher information. *Statis. Probab. Lett.*, 32:175–179, 1997.

[11] Kagan A., A discrete version of Stam inequality and a characterization of the Poisson distribution. *J. Statist. Plann. Inference*, 92(1-2):7–12, 2001.

[12] Kagan A., Letter to the editor: "A discrete version of Stam inequality and a characterization of the Poisson distribution". [*J. Statist. Plann. Inference*, 92(1-2):7–12, 2001]. *J. Statist. Plann. Inference*, 99(1):1, 2001.

[13] Kontoyiannis I., Harremoës P. and Johnson O., Entropy and the law of small numbers. *IEEE Trans. Inform. Theory*, 51(2):466-472, 2005.

[14] Madiman M., and Barron R. A., Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory*, 53(7): 2317-2329, 2007.

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

[16] Papathanasiou V., Some characteristic properties of the Fisher information matrix via Cacoullo-type inequalities. *J. Multivariate Anal.*, 44(2): 256-265, 1993.

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

[18] Villani C., Cercignani's conjecture is sometimes true and always almost true. *Comm. Math. Phys.*, 234(3):455-490, 2003.

[19] Voiculescu D., The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms. *Invent. Math.*, 132(1): 189–227, 1998.

[20] Zamir R., A proof of the Fisher information inequality via a data processing argument. *IEEE Trans. Inform. Theory*, 44(3): 1246–1250, 1998.