Abram Kagan; Tinghui Yu

Some inequalities related to the Stam inequality

Applications of Mathematics, Vol. 53 (2008), No. 3, 195–205

Persistent URL: http://dml.cz/dmlcz/140315

SOME INEQUALITIES RELATED TO THE STAM INEQUALITY

Abram Kagan, Tinghui Yu, College Park

(Invited)

Abstract. Zamir showed in 1998 that the Stam classical inequality for the Fisher information (about a location parameter)

$$1/I(X+Y) \geqslant 1/I(X) + 1/I(Y)$$

for independent random variables X, Y is a simple corollary of basic properties of the Fisher information (monotonicity, additivity and a reparametrization formula). The idea of his proof works for a special case of a general (not necessarily location) parameter. Stam type inequalities are obtained for the Fisher information in a multivariate observation depending on a univariate location parameter and for the variance of the Pitman estimator of the latter.

Keywords: Fisher information, location parameter, Pitman estimators

MSC 2010: 62F11, 62B10

1. Introduction

Here basic properties (monotonicity, additivity and a reparametrization formula) of the Fisher information are presented and, following Zamir [10], the Stam inequality is obtained as a direct corollary of these properties.

Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a parametric family of probability distributions of a random element X taking values in a measurable space $(\mathcal{X}, \mathcal{A})$, the parameter space Θ being an open set of $\mathbb{R}$. For the purpose of this paper, the following simplified version of the concept of a regular statistical experiment suffices. A triple $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mathcal{P})$ is called a *regular statistical experiment* (consisting in an observation of X) if

(a) all P_θ are given by densities $p(x;\theta) = \mathrm{d}P_\theta/\mathrm{d}\mu$ with respect to a measure μ,

(b) $p(x;\theta)$ is continuously differentiable in $\theta \in \Theta$ for μ-almost all $x \in \mathcal{X}$ and

(c) the Fisher information on θ in X (or in $\mathcal{E}$),

$$I(X;\theta) = I_X(\theta) = \int \Big(\frac{\partial p(x;\theta)}{\partial\theta}\Big)^2 / p(x;\theta)\,\mathrm{d}\mu(x),$$

is finite (the integration is over the set $\{x\colon\ p(x;\theta) > 0\}$). In Ibragimov and Khas'minskij [2], the class of regular statistical experiments is larger than the one we have defined. In particular, they need only mean square differentiability in θ of the density $p(x;\theta)$.

The following well-known properties of the Fisher information hold for regular experiments.

1) *Monotonicity.* If $S\colon (\mathcal{X},\mathcal{A}) \to (\mathcal{S},\mathcal{B})$ is a statistic, $Q_\theta(B) = P_\theta(S \in B) = P_\theta(S^{-1}B)$, $B \in \mathcal{B}$ (or, in other terms, $\mathcal{E}_S = (\mathcal{S},\mathcal{B},\ \mathcal{Q} = \{Q_\theta,\ \theta \in \Theta\})$ is a subexperiment of $\mathcal{E}$), then

$$I(S;\theta) \leqslant I(X;\theta), \quad \theta \in \Theta.$$

2) *Additivity.* If X_i, $i = 1,2$, are random elements taking values in $(\mathcal{X}_i,\mathcal{A}_i)$ which are independent for each θ, i.e., for all $A_i \in \mathcal{A}_i$, $i = 1,2$,

$$P_\theta(X_1 \in A_1,\ X_2 \in A_2) = P_\theta(X_1 \in A_1)P_\theta(X_2 \in A_2), \quad \theta \in \Theta,$$

and $X = (X_1, X_2)$, then

$$I(X;\theta) = I(X_1;\theta) + I(X_2;\theta).$$

3) *Reparametrization formula.* If g is a differentiable function, then for $\xi = g(\theta)$

$$I(X;\theta) = |g'(\theta)|^2 I(X;\xi)\big|_{\xi=g(\theta)}.$$

Note in passing that if $p(x;\theta) > 0$, then $I(T;\theta) = I(X;\theta)$ implies sufficiency of a statistic T; without positivity of $p(x;\theta)$ this does not hold in general, as shown in Kagan and Shepp [5].

Multivariate versions of 1)–3) are also well known.

1′) If $\boldsymbol{\theta}$ is an m-variate parameter, $\boldsymbol{\theta} \in \Theta$, an open set in $\mathbb{R}^m$, and $I(X;\boldsymbol{\theta})$ is the $m \times m$ matrix of Fisher information on $\boldsymbol{\theta}$ in X, then for any statistic S,

$$I(S;\boldsymbol{\theta}) \leqslant I(X;\boldsymbol{\theta}),$$

i.e., $I(X;\boldsymbol{\theta}) - I(S;\boldsymbol{\theta})$ is a positive semi-definite matrix.

2′) The additivity property has the same form as in the case of a univariate parameter.

3′) If $\boldsymbol{\xi} = g(\boldsymbol{\theta})$ where g is a differentiable mapping of an open set $\Theta \subset \mathbb{R}^m$ into an open set $\boldsymbol{\xi} \subset \mathbb{R}^k$ with Jacobian

$$H = \left(\frac{\partial g_i}{\partial \theta_j}\right), \quad i = 1, \dots, k; \;\; j = 1, \dots, m,$$

then

$$I(X; \boldsymbol{\theta}) = H^{\mathrm{T}} I(X; \boldsymbol{\xi})\big|_{\boldsymbol{\xi} = g(\boldsymbol{\theta})} H$$

where T stands for transposition.

Let us turn to the case when the distribution of X is absolutely continuous and θ a location parameter so that the density $p(x; \theta) = p(x - \theta)$. Now the Fisher information does not depend on θ,

$$\begin{aligned} I(X; \theta) &= \int_{x:\, p(x-\theta)>0} \{p'(x-\theta)/p(x-\theta)\}^2 p(x-\theta)\,\mathrm{d}x \\ &= \int_{x:\, p(x)>0} \{p'(x)/p(x)\}^2 p(x)\,\mathrm{d}x. \end{aligned} \tag{1}$$

In what follows, $I(X)$ will denote the Fisher information on θ in an observation with density $p(x - \theta)$. For independent X_1, X_2 with densities $p_1(x)$, $p_2(x)$, respectively, $I(X_1 + X_2)$ denotes the Fisher information on θ in an observation with density $p(x - \theta)$ where $p(x) = (p_1 * p_2)(x)$.

As a direct corollary of 1), for independent X_1, X_2,

$$I(X_1 + X_2) \leqslant \min\{I(X_1), I(X_2)\}.$$

In Stam [9] a much stronger inequality was proved,

$$\frac{1}{I(X_1 + X_2)} \geqslant \frac{1}{I(X_1)} + \frac{1}{I(X_2)} \tag{2}$$

that is closely linked to the Shannon classical inequality for the differential entropy $H(X)$: for independent X_1, X_2,

$$\mathrm{e}^{2H(X_1+X_2)} \geqslant \mathrm{e}^{2H(X_1)} + \mathrm{e}^{2H(X_2)}.$$

Recently, Madiman and Barron [7] proved a much stronger version of (2): for independent $X_1, \dots, X_n$,

$$\frac{1}{I(X_1 + \ldots + X_n)} \geqslant \frac{1}{\binom{n-1}{m-1}} \sum_{\mathbf{s}} \frac{1}{I(\sum_{i \in \mathbf{s}} X_i)}, \tag{3}$$

where the summation is over all combinations $\mathbf{s}$ of m elements chosen from $\{1,\dots,n\}$.

One of the corollaries of (3) is the monotone decreasing in n of the information $I((X_1+\ldots+X_n)/\sqrt{n}) = nI(X_1+\ldots+X_n)$ contained in the normalized sum of independent identically distributed $X_1, X_2, \ldots$.

Let us turn now to Zamir's proof of (2) based on properties 1)–3) of the Fisher information.

Let w_1, w_2 be positive numbers with $w_1+w_2=1$ and let observations X_i' be of the form

$$X_i' = w_i\theta + X_i, \quad i=1,2$$

with $\theta\in\mathbb{R}$ as a parameter and X_1, X_2 independent with $X_i\sim p_i(x)$, $i=1,2$. By virtue of 3),

$$I(X_i';\theta) = w_i^2 I(X_i), \quad i=1,2.$$

Consider now a statistic

$$S(X_1',X_2') = X_1'+X_2' = \theta + X_1 + X_2.$$

Due to 1) and 2),

$$(4) \qquad I(X_1+X_2) = I(X_1'+X_2') \leqslant I(X_1') + I(X_2') = w_1^2 I(X_1) + w_2^2 I(X_2).$$

Choosing

$$w_i = \frac{1/I(X_i)}{1/I(X_1)+1/I(X_2)}, \quad i=1,2,$$

one immediately gets from (4) the Stam inequality

$$\frac{1}{I(X_1+X_2)} \geqslant \frac{1}{I(X_1)} + \frac{1}{I(X_2)}.$$

If $\mathbf{X}$, $\mathbf{X}^{\mathrm{T}} = (X_1,\dots,X_s)$ is an m-variate random vector with density $p(\mathbf{x}-\boldsymbol{\theta}) = p(x_1-\theta_1,\dots,x_m-\theta_m)$ depending on an m-variate location parameter $\boldsymbol{\theta}\in\mathbb{R}^m$, the matrix $I(\mathbf{X})$ of the Fisher information on $\boldsymbol{\theta}$ in $\mathbf{X}$ does not depend on $\boldsymbol{\theta}$,

$$I(\mathbf{X}) = (I_{ij})_{i,j=1,\dots,m}, \quad I_{ij} = \int_{\mathbf{x}:\, p(\mathbf{x})>0} \frac{1}{p}\Big(\frac{\partial p}{\partial x_i}\Big)\Big(\frac{\partial p}{\partial x_j}\Big)\,\mathrm{d}\mathbf{x},$$

and is positive definite (the matrix $I(X;\theta)$ of the Fisher information on a general m-variate parameter, not necessarily location, is only non-negative definite). Indeed, take a nonzero $\mathbf{c}\in\mathbb{R}^m$ and consider a random vector $\tilde{\mathbf{X}}$ with density $p(x_1-c_1\theta,\dots,x_m-c_m\theta)$. Plainly, $I(\tilde{\mathbf{X}};\theta) = \mathbf{c}^{\mathrm{T}} I(\mathbf{X})\mathbf{c}$ and due to 1), $I(\tilde{\mathbf{X}};\theta) \geqslant$

$I(\tilde{X}_j;\theta)$. The density of the jth component $\tilde{X}_j$ of $\mathbf{X}$ is $p_j(x_j - c_j\theta)$ so that $I(\tilde{X}_j;\theta) > 0$ if $c_j \neq 0$. Hence $I(\mathbf{X})$ is positive definite.

Now let W_1, W_2 be $(m \times m)$ matrices with $W_1 + W_2 = I_m$, the $(m \times m)$ identity matrix. Set

$$\mathbf{X}_i' = W_i\theta + \mathbf{X}_i, \quad i = 1, 2,$$

where $\mathbf{X}_1$, $\mathbf{X}_2$ are independent m-variate random vectors, $\mathbf{X}_i \sim p_i(\mathbf{x})$, $i = 1, 2$ and $\boldsymbol{\theta} \in \mathbb{R}^m$. By virtue of $1')$–$3')$,

$$\begin{aligned} I(\mathbf{X}_1 + \mathbf{X}_2) = I(\mathbf{X}_1' + \mathbf{X}_2') &\leqslant I(\mathbf{X}_1'; \boldsymbol{\theta}) + I(\mathbf{X}_2'; \boldsymbol{\theta}) \\ &= W_1^{\mathrm{T}} I(\mathbf{X}_1) W_1 + W_2^{\mathrm{T}} I(\mathbf{X}_2) W_2. \end{aligned} \tag{5}$$

Choosing in (5)

$$W_i = (I(\mathbf{X}_i))^{-1}\{(I(\mathbf{X}_1))^{-1} + (I(\mathbf{X}_2))^{-1}\}^{-1}, \quad i = 1, 2$$

one gets

$$I(\mathbf{X}_1 + \mathbf{X}_2) \leqslant \{(I(\mathbf{X}_1))^{-1} + (I(\mathbf{X}_2))^{-1}\}^{-1}$$

whence, by taking the inverse of both sides, the multivariate Stam inequality follows:

$$(I(\mathbf{X}_1 + \mathbf{X}_2))^{-1} \geqslant (I(\mathbf{X}_1))^{-1} + (I(\mathbf{X}_2))^{-1}. \tag{6}$$

The matrices $I(\mathbf{X}_1)$ and $I(\mathbf{X}_2)$ are not assumed commutative. This proof of (6) is due to Zamir [10]. The authors' contribution is an observation that the matrix of the Fisher information on a multivariate location parameter is positive definite so that there is no need in assuming the information matrices nonsingular.

Note that in Kagan and Landsman [3] another inequality for the matrices of the Fisher information first proved analytically in Carlen [1], was shown to be a direct corollary of 1) and 2).

2. The case of a general parameter

Let X_1, X_2 be independent random variables with densities $p_1(x;\theta_1)$, $p_2(x;\theta_2)$ depending on general (not necessarily location) parameters θ_1, θ_2 belonging to the same parameter set $\Theta = (a, b)$, $a \leqslant 0$, $b > 0$ such that $\alpha\Theta \subset \Theta$ for any α, $0 < \alpha < 1$.

To get a version of the Stam inequality for $X_1 \sim p_1(x;\theta_1)$, $X_2 \sim p_2(x;\theta_2)$, we need a number of assumptions.

First, the Fisher information $I(X_1;\theta_1)$ on θ_1 in X_1 and $I(X_2;\theta_2)$ on θ_2 in X_2 is assumed finite, positive and constant in the parameters,

$$0 < I(X_i;\theta_i) = I_i < \infty, \quad i = 1, 2. \tag{7}$$

The condition (7) plainly holds in the case of location parameters θ_1, θ_2 but it is much more general. If X has a density $p(x;\eta)$ and a new parameter η is introduced by $\eta = g(\theta)$ so that $\tilde{p}(x;\theta) = p(x;g(\theta))$, then $I(X;\theta) = |g'(\theta)|^2 I(X;\eta)\big|_{\eta=g(\theta)}$ whence one can construct many families with a constant Fisher information. For example, if X has a Pareto density

$$p(x;\eta) = (\eta - 1)/x^\eta, \quad x \geqslant 1$$

with $\eta > 1$ as a parameter, the reparametrization $\eta = \mathrm{e}^\theta + 1$ stabilizes the information on θ.

Second, let $S = S(X_1, X_2)$ be a statistic taking values in a measurable space $(\mathcal{S}, \mathcal{B})$. It is assumed that the density $p(s;\theta_1,\theta_2)$ of S depends on the parameters only through $\theta_1 + \theta_2$,

$$p(s;\theta_1,\theta_2) = p(s;\theta_1 + \theta_2), \quad s \in \mathcal{S}, \tag{8}$$

so that the distribution of S depends on a univariate parameter $\theta = \theta_1 + \theta_2$. If $p_i(x;\theta_i) = p_i(x - \theta_i)$, $i = 1, 2$ and $S(X_1 + X_2) = X_1 + X_2$, (8) is plainly satisfied.

Theorem 1. *Under the conditions* (7), (8), *the Fisher information* $I(S;\theta)$ *on* θ *in* S *satisfies the inequality*

$$\frac{1}{I(S;\theta)} \geqslant \frac{1}{I_1} + \frac{1}{I_2}. \tag{9}$$

P r o o f. Take positive w_1, w_2 with $w_1 + w_2 = 1$ and set $\theta_1 = w_1\theta$, $\theta_2 = w_2\theta$. Then $\theta_1 + \theta_2 = \theta$. By 3), $I(X_i;\theta) = w_i^2 I_i$, $i = 1, 2$ and by 1) and 2),

$$I(S;\theta) \leqslant I(X_1;\theta) + I(X_2;\theta) = w_1^2 I_1 + w_2^2 I_2.$$

Choosing

$$w_i = \frac{1/I_i}{1/I_1 + 1/I_2}, \quad i = 1, 2$$

leads to (9). □

R e m a r k. Zamir's idea works in some cases when versions of (7), (8) hold. Here is an example in which the dependence of the distribution of S on $\theta_1 + \theta_2$ is replaced with the dependence of its distribution on $\theta_1\theta_2$ where both θ_1 and θ_2 are positive.

Let independent random variables X_1, X_2 have densities $\theta_1 p_1(\theta_1 x)$, $\theta_2 p_2(\theta_2 x)$ depending on scale parameters $\theta_1, \theta_2 \in \mathbb{R}_+$. If the distributions of X_1 and X_2 are concentrated on $\mathbb{R}_+$ or $\mathbb{R}_-$, the setup is reduced to that of location parameters. This assumption is not made here.

Let $T(X_1, X_2) = X_1 X_2$. It is easily seen that the distribution of T depends on θ_1, θ_2 only through the scale parameter $\theta = \theta_1\theta_2$,

$$p(t;\theta) = \theta p(\theta x).$$

Simple calculations show that

$$I(X_i;\theta_i) = \theta_i^{-2} I(X_i;1), \;\; i = 1,2; \quad I(T;\theta) = \theta^{-2} I(T;1).$$

Now set $\theta_1 = \theta^{\gamma_1}$, $\theta_2 = \theta^{\gamma_2}$ with $\gamma_i > 0$, $\gamma_1 + \gamma_2 = 1$. Then $\theta_1\theta_2 = \theta$ and

$$I(X_i;\theta) = (\gamma_i \theta^{\gamma_i - 1})^2 I(X_i;\theta_i) = \gamma_i^2 \theta^{-2} I(X_i;1), \quad i = 1,2.$$

One has

$$I(T;\theta) \leqslant I(X_1;\theta) + I(X_2;\theta)$$

whence

$$I(T;1) \leqslant \gamma_1^2 I(X_1;1) + \gamma_2^2 I(X_2;1).$$

Choosing

$$\gamma_i = \frac{(I(X_i;1))^{-1}}{(I(X_1;1))^{-1} + (I(X_2;1))^{-1}}$$

one gets a Stam type inequality for the Fisher information on a scale parameter: for independent X_1, X_2 one has

$$\frac{1}{I(X_1X_2;\theta)} \geqslant \frac{1}{I(X_1;\theta)} + \frac{1}{I(X_2;\theta).}$$

Unfortunately, the proof does not work when the distribution of S depends on an arbitrary (univariate) function $h(\theta_1, \theta_2)$.

3. Relation to the Pitman estimators

Let $\mathbf{X}^{\mathrm{T}} = (X_1, \ldots, X_m) \sim p(x_1 - \theta, \ldots, x_m - \theta) = p(\mathbf{x} - \theta \cdot \mathbf{1})$ where $\mathbf{1}^{\mathrm{T}} = (1, \ldots, 1)$ is an m-variate vector with all the components 1, be an m-variate random vector whose distribution depends on a univariate location parameter θ. If $\mathbf{I}$ is the matrix of the Fisher information on $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \ldots, \theta_m)$ in an observation with density $p(\mathbf{x} - \boldsymbol{\theta})$, then the Fisher information I on θ in $\mathbf{X}$ is

$$I = \mathbf{1}^{\mathrm{T}}\mathbf{I}\mathbf{1}.$$

Let now $\mathbf{X}_1$, $\mathbf{X}_2$ be independent random vectors, $\mathbf{X}_1 \sim p_1(\mathbf{x} - \theta \cdot \mathbf{1})$, $\mathbf{X}_2 \sim p_2(\mathbf{x} - \theta \cdot \mathbf{1})$, $p(\mathbf{x}) = (p_1 * p_2)(\mathbf{x})$ and let I_1, I_2, I denote the Fisher observation on the univariate parameter θ contained in $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}$, respectively. As an immediate corollary of Theorem 1, one gets

$$\frac{1}{I} \geqslant \frac{1}{I_1} + \frac{1}{I_2}. \tag{10}$$

This inequality is independent of the multivariate Stam inequality

$$\mathbf{I}^{-1} \geqslant \mathbf{I}_1^{-1} + \mathbf{I}_2^{-1} \tag{11}$$

where $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{I}$ are the matrices of the Fisher information on the m-variate parameter $\boldsymbol{\theta}$ contained in $\mathbf{X}_1 \sim p_1(\mathbf{x} - \boldsymbol{\theta})$, $\mathbf{X}_2 \sim p_2(\mathbf{x} - \boldsymbol{\theta})$, $\mathbf{X} \sim p(\mathbf{x} - \boldsymbol{\theta})$.

Inequality (10) has its analog in terms of Pitman estimators; no regularity type conditions, even absolute continuity, are required from the distributions.

Let

$$\begin{aligned} \mathbf{x'}_1^{\mathrm{T}} &= (x'_{11}, \ldots, x'_{1m}), \ \ldots, \ \mathbf{x'}_n^{\mathrm{T}} = (x'_{n1}, \ldots, x'_{nm}), \\ \mathbf{x''}_1^{\mathrm{T}} &= (x''_{11}, \ldots, x''_{1m}), \ \ldots, \ \mathbf{x''}_n^{\mathrm{T}} = (x''_{n1}, \ldots, x''_{nm}) \end{aligned}$$

be independent samples from distributions $F_1(\mathbf{x} - \theta \cdot \mathbf{1})$ and $F_2(\mathbf{x} - \theta \cdot \mathbf{1})$ and let

$$\mathbf{x}_1^{\mathrm{T}} = (x_{11}, \ldots, x_{1m}), \ \ldots, \ \mathbf{x}_n^{\mathrm{T}} = (x_{n1}, \ldots, x_{nm})$$

be a sample from $F(\mathbf{x} - \theta \cdot \mathbf{1})$ where $F = F_1 * F_2$.

Set

$$\bar{x}'_1 = (x'_{11} + \ldots + x'_{n1})/n, \ \ldots, \ \bar{x}'_m = (x'_{1m} + \ldots + x'_{nm})/n$$

and

$$\bar{x}' = (\bar{x}'_1 + \ldots + \bar{x}'_m)/m$$

with

$$\bar{x}''_1, \ldots, \bar{x}''_m,\ \bar{x}'',\ \bar{x}_1, \ldots, \bar{x}_m,\ \bar{x}$$

defined similarly for the other two samples.

Let σ', σ'', σ be the σ-algebras generated by $x'_{11} - \bar{x}'$, ..., $x'_{nm} - \bar{x}'$; $x''_{11} - \bar{x}''$, ..., $x''_{nm} - \bar{x}''$; $x'_{11} - \bar{x}' + x''_{11} - \bar{x}''$, ..., $x'_{nm} - \bar{x}' + x''_{nm} - \bar{x}''$, respectively. Plainly, σ is a subalgebra of the σ-algebra generated by

$$x'_{11} - \bar{x}',\ \ldots,\ x'_{nm} - \bar{x}',\ x''_{11} - \bar{x}'',\ \ldots,\ x''_{nm} - \bar{x}''.$$

The latter is usually denoted $\sigma' \vee \sigma''$ so that $\sigma \subset \sigma' \vee \sigma''$.

An estimator $\tilde{\theta}(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ of θ from a sample from $G(\mathbf{y} - \theta \cdot \mathbf{1})$ is called equivariant if for any $c \in \mathbb{R}$

$$\tilde{\theta}(\mathbf{y}_1 + c \cdot \mathbf{1}, \ldots, \mathbf{y}_n + c \cdot \mathbf{1}) = \tilde{\theta}(\mathbf{y}_1, \ldots, \mathbf{y}_n) + c. \tag{12}$$

Assuming $\int |\mathbf{x}|^2 \, \mathrm{d}F_i(\mathbf{x}) < \infty$, $i = 1, 2$, the Pitman estimators t'_n, t''_n of θ (with respect to the quadratic loss function) from samples of size n from $p_1(\mathbf{x} - \theta \cdot \mathbf{1})$ and $p_2(\mathbf{x} - \theta \cdot \mathbf{1})$, i.e., the minimum variance equivariant estimators, can be written as

$$t'_n = \bar{x}' - E(\bar{x}' \mid \sigma'), \quad t''_n = \bar{x}'' - E(\bar{x}'' \mid \sigma'')$$

and their variances as

$$\operatorname{var}(t'_n) = \operatorname{var}(\bar{x}') - \operatorname{var}\{E(\bar{x}' \mid \sigma')\}, \quad \operatorname{var}(t''_n) = \operatorname{var}(\bar{x}'') - \operatorname{var}\{E(\bar{x}'' \mid \sigma'')\}.$$

(All the expectations are taken at $\theta = 0$, though the variances do not depend on θ.) Now

$$\operatorname{var}(t_n) = \operatorname{var}(\bar{x}) - \operatorname{var}\{E(\bar{x} \mid \sigma)\}$$

and using the fact that $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is equidistributed with $(\mathbf{x}'_1 + \mathbf{x}''_1, \ldots, \mathbf{x}'_n + \mathbf{x}''_n)$, one gets

$$\operatorname{var}(t_n) = \operatorname{var}(\bar{x}' + \bar{x}'') - \operatorname{var}\{E(\bar{x}' + \bar{x}'' \mid \sigma)\}.$$

Since $\sigma \subset \sigma' \vee \sigma''$, one has

$$\operatorname{var}\{E(\bar{x}' + \bar{x}'' \mid \sigma)\} \leqslant \operatorname{var}\{E(\bar{x}' + \bar{x}'' \mid \sigma' \vee \sigma'')\}.$$

Furthermore, $\bar{x}', x'_{11} - \bar{x}', \ldots, x'_{nm} - \bar{x}'$ is independent of $x''_{11} - \bar{x}'', \ldots, x''_{nm} - \bar{x}''$ implying

$$E(\bar{x}' \mid \sigma' \vee \sigma'') = E(\bar{x}' \mid \sigma'), \quad E(\bar{x}'' \mid \sigma' \vee \sigma'') = E(\bar{x}'' \mid \sigma'')$$

(see, e.g., Shao [8]). Thus,

$$\operatorname{var}(t_n) \geqslant \operatorname{var}(t_n') + \operatorname{var}(t_n'').$$

This inequality, holding for any n, may be considered a small sample version of (10). In the case of $m = 1$ it was proved in Kagan [4]. For other connections between the variance of Pitman estimators and the Fisher information see Kagan et al. [6].

As said above, (10) and (11) are independent in the sense that neither is a corollary of the other. However, an inequality connecting I and $\mathbf{I}$ has a simple statistical interpretation.

Let $\mathbf{w}^{\mathrm{T}} = (w_1, \ldots, w_m)$ be a vector with $\mathbf{w}^{\mathrm{T}}\mathbf{1} = 1$. Then, by virtue of the Cauchy inequality,

$$1 = (\mathbf{w}^{\mathrm{T}}\mathbf{1})^2 = (\mathbf{w}^{\mathrm{T}}\mathbf{I}^{-1/2}\mathbf{I}^{1/2}\mathbf{1})^2 \leqslant (\mathbf{w}^{\mathrm{T}}\mathbf{I}^{-1}\mathbf{w})(\mathbf{1}^{\mathrm{T}}\mathbf{I}\mathbf{1})$$

so that

$$I \geqslant \frac{1}{\mathbf{w}^{\mathrm{T}}\mathbf{I}^{-1}\mathbf{w}}. \tag{13}$$

Let now $\mathbf{t}_n$ be the Pitman estimator of an m-variate $\boldsymbol{\theta}$ from a sample $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ from $F(\mathbf{x} - \boldsymbol{\theta})$. If $\int \mathbf{x}^2 \, \mathrm{d}F(\mathbf{x}) < \infty$, $\mathbf{t}_n$ can be written (componentwise) as

$$\mathbf{t}_n = \bar{\mathbf{x}} - E(\bar{\mathbf{x}} \mid x_{11} - \bar{x}_1, \ldots, x_{n1} - \bar{x}_1, \ldots, x_{1m} - \bar{x}_m, \ldots, x_{nm} - \bar{x}_m). \tag{14}$$

Note that the σ-algebra generated by the residuals in (14) is smaller than the σ-algebra generated by $x_{11} - \bar{x}, \ldots, x_{nm} - \bar{x}$ where $\bar{x} = (\bar{x}_1 + \ldots + \bar{x}_m)/m$ (mind the difference between $\bar{\mathbf{x}}$ and $\bar{x}$). The latter occurs in the representation

$$t_n = \bar{x} - E(\bar{x} \mid x_{11} - \bar{x}, \ldots, x_{nm} - \bar{x}) \tag{15}$$

of the Pitman estimator of a univariate θ from a sample $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ from $F(\mathbf{x} - \theta \cdot \mathbf{1})$ when $\mathbf{w}^{\mathrm{T}}\mathbf{t}_n$ is an equivariant estimator of θ and, thus,

$$\mathbf{w}^{\mathrm{T}} \operatorname{var}(\mathbf{t}_n)\mathbf{w} = \operatorname{var}(\mathbf{w}^{\mathrm{T}}\mathbf{t}_n) \geqslant \operatorname{var}(t_n). \tag{16}$$

This inequality is, in a sense, a small sample version of (13). Indeed, as $n \to \infty$,

$$n \operatorname{var}(\mathbf{t}_n) = \mathbf{I}^{-1}(1 + o(1)), \quad n \operatorname{var}(t_n) = \frac{1}{\mathbf{1}^{\mathrm{T}}\mathbf{I}\mathbf{1}}(1 + o(1)),$$

so that (16) becomes (13). The relation between these two equations is one more illustration of that many results for the Fisher information/information matrix have direct analogs in terms of the variances of the Pitman estimators in small samples.

Acknowledgement. The authors are grateful to Andrew Barron and Mokshay Madiman for references and useful discussions. Very detailed comments of an anonymous referee helped to improve the presentation.

References

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

[2] *I. A. Ibragimov, R. Z. Khas'minskij*: Statistical Estimation. Asymptotic Theory. Springer, New York, 1981.

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

[4] *A. Kagan*: An inequality for the Pitman estimators related to the Stam inequality. Sankhya *A64* (2002), 282–292.

[5] *A. Kagan, L. A. Shepp*: A sufficiency paradox: an insufficient statistic preserving the Fisher information. Amer. Statist. *59* (2005), 54–56.

[6] *A. Kagan, T. Yu, A. Barron, M. Madiman*: Contribution to the theory of Pitman estimators. Submitted.

[7] *M. Madiman, A. Barron*: The monotonicity of information in the central limit theorem and entropy power inequalities. Preprint. Dept. of Statistics, Yale University, 2006.

[8] *J. Shao*: Mathematical Statistics, 2nd ed. Springer, New York, 2003.

[9] *A. J. Stam*: Some inequalities satisfied by the quantities of information of Fisher and Shannon. Inform. and Control *2* (1959), 101–112.

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

Author's address: *A. Kagan*, *T. Yu*, Department of Mathematics, University of Maryland, College Park, MD 20742, U.S.A., e-mail: amk@math.umd.edu, yuth@math.umd.edu.