Записки научных
семинаров ПОМИ
Том 408, 2012 г.

A. M. Kagan, Tinghui Yu, A. Barron, M. Madiman

CONTRIBUTION TO THE THEORY OF PITMAN ESTIMATORS

Abstract. New inequalities are proved for the variance of the Pitman estimators (minimum variance equivariant estimators) of θ constructed from samples of fixed size from populations $F(x-\theta)$. The inequalities are closely related to the classical Stam inequality for the Fisher information, its analog in small samples, and a powerful variance drop inequality. The only condition required is finite variance of F; even the absolute continuity of F is not assumed. As corollaries of the main inequalities for small samples, one obtains alternate proofs of known properties of the Fisher information, as well as interesting new observations like the fact that the variance of the Pitman estimator based on a sample of size n scaled by n monotonically decreases in n. Extensions of the results to the polynomial versions of the Pitman estimators and a multivariate location parameter are given. Also, the search for characterization of equality conditions for one of the inequalities leads to a Cauchy-type functional equation for independent random variables, and an interesting new behavior of its solutions is described.

§1. Introduction

Our goal is to present some new inequalities for the variance of the Pitman estimators of a location parameter from different related samples.

Denote by t_n the Pitman estimator (i.e., the minimum variance equivariant estimator) of θ from a sample $(x_1,\dots,x_n)$ of size n from population $F(x-\theta)$. For simplicity, we first focus on the univariate case, i.e., $x_i\in\mathbb{R}$. If $\int x^2dF(x)<\infty$, the Pitman estimator can be written as

$$t_n=\overline{x}-\mathbf{E}(\overline{x}|x_1-\overline{x},\dots,x_n-\overline{x}) \qquad (1)$$

where $\overline{x}$ is the sample mean and E denotes the expectation with respect to $F(x)$ (i.e., when $\theta=0$).

Key words and phrases: Fisher information, location parameter, monotonicity of the variance, Stam inequality.

For the univariate case, if $F' = f$ exists, t_n can be also written as

$$t_n = \frac{\int u \prod_1^n f(x_i - u)\,du}{\int \prod_1^n f(x_i - u)\,du}$$

showing that t_n is a generalized Bayes estimator corresponding to an improper prior (uniform on the whole $\mathbb{R}$). In this paper the representation (1) crucial in studying the behavior of t_n in large samples will not be used.

In Sec. 2, we obtain a relationship between the variances of the Pitman estimators based on data obtained by adding (convolving) the initial samples. As an application of this inequality, one obtains a new proof of a Fisher information inequality related to the central limit theorem. Another application, to distributed estimation using sensor networks, is described elsewhere [15].

If $t_n^{(1)}, \dots, t_n^{(N)}$ denote the Pitman estimators from samples of size n from $F_1(x-\theta), \dots, F_N(x-\theta)$, and t_n is the Pitman estimator from a sample of size n from $F(x-\theta)$ where $F = F_1 * \ldots * F_N$, Kagan [10] showed the superadditivity property

$$\operatorname{var}(t_n) \geqslant \operatorname{var}(t_n^{(1)}) + \cdots + \operatorname{var}(t_n^{(N)}). \tag{2}$$

In Sec., 3, we obtain this as a corollary of the main inequality in Sec. 2, and study an analytic problem arising in connection with identifying its equality conditions. In particular, a version of the classical Cauchy functional equation for independent random variables is studied; the behavior of this equation turns out to be more subtle than in the usual settings.

In Sec. 4, various inequalities relevant to estimation from a combination of samples are given. For instance, for the Pitman estimator t_{m+n} constructed from observations $x_1, \dots, x_m, y_1, \dots, y_n$ where the first m observations come from $F(x-\theta)$ and the last n from $G(x-\theta)$,

$$\frac{1}{\operatorname{var}(t_{m+n})} \geqslant \frac{1}{\operatorname{var}(t_m)} + \frac{1}{\operatorname{var}(t_n)}, \tag{3}$$

where t_m and t_n denote the Pitman estimators constructed from $x_1, \dots, x_m$ and $y_1, \dots, y_n$, respectively. A generalization of this inequality has an interesting application to a data pricing problem (where datasets are to be sold, and the value of a dataset comes from the information it yields about an unknown location parameter); this application is described by the authors elsewhere [16].

As an application of the inequalities proved in Sec. 4, we prove in Sec. 5 that for any $n \geqslant 1$, with t_n now denoting the Pitman estimator constructed from $x_1, \dots, x_n$ for any n,

$$n \operatorname{var}(t_n) \geqslant (n+1) \operatorname{var}(t_{n+1}) \tag{4}$$

with the equality sign holding for an $n \geqslant 2$ only for a sample from Gaussian population (in which case $n \operatorname{var}(t_n)$ is constant in n).

If $(x_1, \dots, x_n)$ is a sample from s-variate population $F(x-\theta)$, $x, \theta \in \mathbb{R}^s$ with $\int_{\mathbb{R}^s} |x|^2 dF(x) < \infty$, the Pitman estimator is defined as the minimum covariance matrix equivariant estimator. Though there is only partial ordering in the set of covariance matrices, the set of covariance matrices of equivariant estimators has a minimal element which is the covariance matrix of the Pitman estimator (1) of the s-variate location parameter. Multivariate extensions of most of the inequalities mentioned above are given in Sec. 6.

Assuming $\int x^{2k} dF(x) < \infty$ for some integer $k \geqslant 1$, the polynomial Pitman estimator $\widehat{t}_n^{(k)}$ of degree k is, by definition, the minimum variance equivariant polynomial estimator (see Kagan [11]). An advantage of the polynomial Pitman estimator is that it depends only on the first $2k$ moments of F. In Sec. 7, it is shown that the polynomial Pitman estimator preserves almost all the properties of t_n that are studied here.

In Sec. 8 the setup of observations $x_1, \dots, x_n$ additively perturbed by independent $y_1, \dots, y_n$ with self-decomposable distribution function $G(y/\lambda)$ is considered. For the Pitman estimator $t_{n,\lambda}$ from a sample of size n from $F_\lambda(x-\theta)$ where $F_\lambda(x) = \int F(x-u) dG(u/\lambda)$ we prove that $\operatorname{var}(t_{n,\lambda})$ as a function of λ, monotonically decreases on $(-\infty, 0)$ and increases on $(0, +\infty)$. This makes rigorous the intuition that adding "noise" makes estimation harder.

Section 9 concludes with some discussion of the issues that arise in considering various possible generalizations of the results presented in this paper.

1.1. Related literature. All our results have direct counterparts in terms of the Fisher information, and demonstrate very close similarities between properties of the inverse Fisher information and the variance of Pitman estimators.

Denote by $I(X)$ the Fisher information on a parameter $\theta \in \mathbb{R}$ contained in an observation $X + \theta$. Plainly, the information depends only on the distribution F of the noise X but not on θ.

For independent X, Y the inequality $I(X+Y) \leqslant I(X)$ is almost trivial (an observation $X+Y+\theta$ is "more noisy" than $X+\theta$). A much less trivial inequality was proved in Stam [20]:

$$\frac{1}{I(X+Y)} \geqslant \frac{1}{I(X)} + \frac{1}{I(Y)}. \tag{5}$$

In Zamir [21], the Stam inequality is obtained as a direct corollary of the basic properties of the Fisher information: additivity, monotonicity and reparameterization formula.

The main inequality in Sec. 2 is closely related to the classical Stam inequality for the Fisher information, its version in estimation and a powerful variance drop inequality proved in a general form in Madiman and Barron [17] (described below). In Artstein *et. al.* [1] and Madiman and Barron [17] the variance drop inequality led to improvements of the Stam inequality.

Let now $F(x) = (F_1 * F_2)(x) = \int F_1(y)dF_2(x-y)$ and t_n', t_n'', t_n be the Pitman estimators from samples of size n from $F_1(x-\theta)$, $F_2(x-\theta)$ and $F(x-\theta)$, respectively. If $\int x^2 dF(x) < \infty$, the following inequality holds for the variances (Kagan [10]):

$$\operatorname{var}(t_n) \geqslant \operatorname{var}(t_n') + \operatorname{var}(t_n''). \tag{6}$$

This inequality is, in a sense, a finite sample version of (5), as discussed in Kagan [10]. It is generalized in Sec. 2, and its equality conditions are obtained in Sec. 3.

Several of the results in this paper rely on the following *variance drop lemma.*

Lemma 1. *Let $X_1, \ldots, X_N$ be independent (not necessarily identically distributed) random vectors. For $\mathbf{s} = \{i_1, \ldots, i_m\} \subset \{1, \ldots, N\}$ set $\mathbf{X_s} = (X_{i_1}, \ldots, X_{i_m})$, with $i_1 < i_2 < \ldots < i_m$ without loss of generality. For arbitrary functions $\phi_\mathbf{s}(\mathbf{X_s})$ with $\operatorname{var}\{\phi_\mathbf{s}(\mathbf{X_s})\} < \infty$ and any weights $w_\mathbf{s} > 0, \sum_\mathbf{s} w_\mathbf{s} = 1$,*

$$\operatorname{var}\left\{\sum_\mathbf{s} w_\mathbf{s}\phi_\mathbf{s}(\mathbf{X_s})\right\} \leqslant \binom{N-1}{m-1} \sum_\mathbf{s} w_\mathbf{s}^2 \operatorname{var}\{\phi_\mathbf{s}(\mathbf{X_s})\}, \tag{7}$$

where the summation in both sides is extended over all unordered sets (combinations) $\mathbf{s}$ of m elements from $\{1, \ldots, N\}$. The equality sign in (7) holds *if and only if all* $\phi_{\mathbf{s}}(\mathbf{X}_{\mathbf{s}})$ *are additively decomposable, i.e.,*

$$\phi_{\mathbf{s}}(\mathbf{X}_{\mathbf{s}}) = \sum_{i \in \mathbf{s}} \phi_{\mathbf{s}i}(X_i). \tag{8}$$

The main idea of the proof goes back to Hoeffding [5] and is based on an ANOVA type decomposition, see also Efron and Stein [4]. See Artstein *et. al.* [1] for the proof of Lemma 1 in case of $m = N - 1$, and Madiman and Barron [17] for the general case. In Sec. 6, we observe that this lemma has a multivariate extension, and use it to prove various inequalities for Pitman estimation of a multivariate location parameter.

The main inequality of Sec. 4 is also related to Carlen's [3] superadditivity of Fisher information, as touched upon there. See [8] for the statistical meaning and proof of Carlen's superadditivity.

§2. Convolving independent samples from different populations

Here we first prove a stronger version of superadditivity (6).

Let $\mathbf{x_k} = (x_{k1}, \dots, x_{kn})$, $k = 1, \dots, N$ be a sample of size n from population $F_k(x - \theta)$. Set

$$\overline{x}_k = \frac{x_{k1} + \dots + x_{kn}}{n}, \quad R_k = (x_{k1} - \overline{x}_k, \dots, x_{kn} - \overline{x}_k), \quad \sigma_k^2 = \operatorname{var}(x_{ki}),$$

and for $\mathbf{s} = \{i_1, \dots, i_m\} \subset \{1, \dots, N\}$,

$$F_{\mathbf{s}}(x) = (F_{i_1} * \dots * F_{i_m})(x), \quad \overline{x}_{\mathbf{s}} = \sum_{k \in \mathbf{s}} \overline{x}_k, \quad R_{\mathbf{s}} = \sum_{k \in \mathbf{s}} R_k \text{(componentwise)}.$$

Also set

$$\begin{aligned} F(x) &= (F_1 * \dots * F_N)(x), \\ \overline{x} &= \overline{x}_1 + \dots + \overline{x}_N, \\ R &= R_1 + \dots + R_N, \\ \sigma^2 &= \sigma_1^2 + \dots + \sigma_N^2. \end{aligned}$$

We will need the following well known lemma (see, e.g., [18, **?**].

Lemma 2. *Let* ξ *be a random variable with* $\mathbf{E}|\xi| < \infty$ *and* η_1, η_2 *arbitrary random elements. If* (ξ, η_1) *and* η_2 *are independent then*

$$\mathbf{E}(\xi|\eta_1, \eta_2) = \mathbf{E}(\xi|\eta_1) \quad \text{a.s.} \tag{9}$$

Theorem 1. *Let $t_{\mathbf{s},n}$ denote the Pitman estimator of θ from a sample of size n from $F_{\mathbf{s}}(x-\theta)$, and t_n denote the Pitman estimator from a sample of size n from $F(x-\theta)$. Under the only condition $\sigma^2<\infty$, for any $n\geqslant 1$ and any m with $1\leqslant m\leqslant N$,*

$$\operatorname{var}(t_n)\geqslant\frac{1}{\binom{N-1}{m-1}}\sum_{\mathbf{s}}\operatorname{var}(t_{\mathbf{s},n})\tag{10}$$

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

Proof. Set $r=\binom{N-1}{m-1}$. From the definition (1) one has

$$\operatorname{var}(t_n)=\sigma^2/n-\operatorname{var}\{\mathbf{E}(\overline{x}|R)\}=\sum_1^N(\sigma_k^2/n)-\operatorname{var}\left\{\mathbf{E}\left(\sum_1^N\overline{x}_k|R\right)\right\}.$$

Similarly,

$$\begin{aligned}(1/r)\sum_{\mathbf{s}}\operatorname{var}(t_{\mathbf{s},n}) &= (1/r)\sum_{\mathbf{s}}\sum_{k\in\mathbf{s}}\sigma_k^2/n-(1/r)\sum_{\mathbf{s}}\operatorname{var}\{\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})\}\\ &= \sum_1^N(\sigma_k^2/n)-(1/r)\sum_{\mathbf{s}}\operatorname{var}\{\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})\},\end{aligned}$$

where the last equality is due to the fact that each $k\in\{1,\dots,N\}$ appears exactly r times in $\mathbf{s}$. On setting $\phi_{\mathbf{s}}=\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})$ and $w_{\mathbf{s}}=\binom{N}{m}^{-1}$ for all $\mathbf{s}$ and noticing that so defined $\phi_{\mathbf{s}}$ depends only on $\mathbf{x}_k$, $k\in\mathbf{s}$, one has by virtue of Lemma 1

$$r\sum_s\operatorname{var}\{\mathbf{E}(\overline{x}_s|R_s)\}\geqslant\operatorname{var}\left\{\sum_s\mathbf{E}(\overline{x}_s|R_s)\right\}.\tag{11}$$

Denote by $\overline{\mathbf{s}}$ the complement of $\mathbf{s}$ in $\{1,\dots,N\}$. Then $R_{\mathbf{s}}$ and $R_{\overline{\mathbf{s}}}$ depend on disjoint sets of independent random vectors $\mathbf{x}_1,\dots,\mathbf{x}_N$ and thus are independent.

By virtue of Lemma 2,

$$\phi_{\mathbf{s}}=\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}},R_{\overline{\mathbf{s}}}).$$

From the definition of the n-variate vectors $R_{\mathbf{s}}$ and $R_{\overline{\mathbf{s}}}$ one has $R=R_{\mathbf{s}}+R_{\overline{\mathbf{s}}}$. Now due to a well known property of the conditional expectation,

$$\mathbf{E}(\overline{x}_{\mathbf{s}}|R)=\mathbf{E}[\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}},R_{\overline{\mathbf{s}}})|R]=\mathbf{E}[\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})|R].$$

Since for any random variable ξ and random element η

$$\operatorname{var}(\xi) \geqslant \operatorname{var}\{\mathbf{E}(\xi|\eta)\},$$

the previous relation results in

$$\begin{aligned}
\operatorname{var}\Big\{\sum_{\mathbf{s}} \mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})\Big\} &\geqslant \operatorname{var}\Big\{\mathbf{E}\Big(\sum_{\mathbf{s}} \mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})|R\Big)\Big\} \\
&= \operatorname{var}\Big\{\mathbf{E}\Big(\sum_{\mathbf{s}} \mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}}, R_{\overline{\mathbf{s}}})|R\Big)\Big\} \\
&= \operatorname{var}\Big\{\sum_{\mathbf{s}} \mathbf{E}(\overline{x}_{\mathbf{s}}|R)\Big\} = \operatorname{var}\Big\{\mathbf{E}\Big(\sum_{\mathbf{s}} \overline{x}_{\mathbf{s}}|R\Big)\Big\} \\
&= \operatorname{var}\Big\{\mathbf{E}\Big(r\sum_{k=1}^{N} \overline{x}_k|R\Big)\Big\} = r^2 \operatorname{var}\{\mathbf{E}(\overline{x}|R)\}. \quad (12)
\end{aligned}$$

Combining (11) with (12) leads to

$$\operatorname{var}\Big\{\mathbf{E}\Big(\sum_{k=1}^{N} \overline{x}_k|R\Big)\Big\} \leqslant \frac{1}{r}\sum_{\mathbf{s}} \operatorname{var}\{\mathbf{E}(\overline{x}_{\mathbf{s}}|R_{\mathbf{s}})\}, \tag{13}$$

which is equivalent to the claimed result (10). □

It is of special interest to study the simple case where $F_1 = \cdots = F_N = H$. This gives the monotonicity of $\operatorname{var}(t_n^{*N})$ with respect to the group number N, in contrast to (27) in Sec. 5, whose monotonicity is with respect to the sample size n.

Corollary 1. *For any $N > 1$, if t_n^{*N} is the Pitman estimator of θ from a sample of size n from $H^{*N}(x-\theta)$ where $H^{*N} = H * \cdots * H$, then*

$$\frac{\operatorname{var}(t_n^{*N})}{N} \geqslant \frac{\operatorname{var}(t_n^{*(N-1)})}{N-1}. \tag{14}$$

Here n and N are independent parameters.

Proof. Choose $m = N-1$ in Theorem 1. Under the conditions of Corollary, $(t_{n,\mathbf{s}})$ are equidistributed for all N combinations $\mathbf{s}$ of $N-1$ elements so that (10) becomes

$$\operatorname{var}(t_n^{*N}) \geqslant \frac{N}{N-1}\operatorname{var}(t_n^{*(N-1)}). \quad \square$$

Recall that for independent identically distributed $X_1, \dots, X_N$, Artstein *et. al.* [1] showed that

$$NI(X_1 + \cdots + X_N) \leqslant (N-1)I(X_1 + \cdots + X_{N-1}) \tag{15}$$

for any $N \geqslant 1$. As shown in Ibragimov and Has'minskii [7], if $I(X) < \infty$ and $\int |x|^\delta dF(x) < \infty$ for some $\delta > 0$,

$$\mathrm{var}(t_n) = \frac{1}{nI(X)}(1 + o(1)), \quad n \to \infty. \tag{16}$$

Thus the inequality (14) may be considered a small sample version of inequality (15) for the Fisher information. Furthermore, note that the monotonicity (15) of Fisher information follows from (14) and (16).

Another corollary of Theorem 1 is a dissipative property of the conditional expectation of the sample mean.

Corollary 2. *If $F_1 = \cdots = F_N = H$, then for any $N > 1$*

$$(N-1)\mathrm{var}\{\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_{N-1})\} \geqslant N \mathrm{var}\{\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_N)\}. \tag{17}$$

Proof. Since $x_{11}, \dots, x_{Nn}$ are independent identically distributed random variables, one has for any n and N

$$\begin{aligned} \mathrm{var}(t_n^{(N)}) &= \mathrm{var}\left\{ \sum_{k=1}^N \overline{x}_k - \mathbf{E}\left(\sum_{k=1}^N \overline{x}_k | R_1 + \cdots + R_N \right) \right\} \\ &= \mathrm{var}\left\{ \sum_{k=1}^N \overline{x}_k \right\} - \mathrm{var}\left\{ \mathbf{E}\left(\sum_{k=1}^N \overline{x}_k | R_1 + \cdots + R_N \right) \right\} \\ &= \sigma^2/n - \mathrm{var}\left\{ N\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_N) \right\} \end{aligned}$$

that combined with (14) immediately leads to (17). □

Notice that (17) is much stronger than monotonicity of $\mathrm{var}\{\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_N)\}$ that follows directly from

$$\begin{aligned} \mathrm{var}\{\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_{N-1})\} &= \mathrm{var}\{\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_{N-1}, R_N)\} \\ &\geqslant \mathrm{var}\{\mathbf{E}(\overline{x}_1 | R_1 + \cdots + R_N)\}, \end{aligned}$$

due to independence of $(\overline{x}_1, R_1, \dots, R_{N-1})$ and $\mathbf{x}_N$.

§3. A corollary and an analytical characterization problem related to the Pitman estimators

Turn now to an elegant corollary of Theorem 1. On setting $m = 1$ in Theorem 1, the subsets $\mathbf{s}$ are reduced to one element each, $\mathbf{s} = \{k\}$, $k = 1, \ldots, N$ and one gets the superadditivity inequality from Kagan [10]:

Corollary 3. *If $t_n^{(1)}, \ldots, t_n^{(N)}$ are the Pitman estimators from samples of size n from $F_1(x-\theta), \ldots, F_N(x-\theta)$, and t_n is the Pitman estimator from a sample of size n from $F(x-\theta)$ where $F = F_1 * \ldots * F_N$, then*

$$\operatorname{var}(t_n) \geqslant \sum_{k=1}^{N} \operatorname{var}(t_n^{(k)}). \tag{18}$$

An interesting analytic problem, a Cauchy type functional equation for independent random variables, arises in connection to the relation

$$\operatorname{var}(t_n) = \sum_{k=1}^{N} \operatorname{var}(t_n^{(k)}). \tag{19}$$

We will show below that with some conditions on $F_1, \ldots, F_N$, (19) is a characteristic property of Gaussian distributions. Note that to study the relation (19), it suffices to consider the case of $N = 2$.

Let $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$ be independent samples from populations $F_1(x-\theta_1)$, $F_2(y-\theta_2)$, respectively, and let t_n' and t_n'' be the Pitman estimators of θ_1 and θ_2. The Pitman estimator of $\theta_1+\theta_2$ from the combined sample $(x_1, \ldots, y_n)$ is $t_n' + t_n''$.

For the Pitman estimator t_n of θ from a sample of size n from population $(F_1 * F_2)(x-\theta)$, consider $t_n(x_1+y_1, \ldots, x_n+y_n)$. This is an equivariant estimator of $\theta_1+\theta_2$ from the above combined sample, so that

$$\operatorname{var}(t_n) \geqslant \operatorname{var}(t_n') + \operatorname{var}(t_n''). \tag{20}$$

Due to the uniqueness of the Pitman estimator, the equality sign in (20) holds if and only if

$$t_n(x_1+y_1, \ldots, x_n+y_n) = t_n'(x_1, \ldots, x_n) + t_n''(y_1, \ldots, y_n) \tag{21}$$

with probability one. This is a Cauchy type functional equation holding for random variables and is different from the classical Cauchy equation.

It turns out that even in the simplest case of $n = 1$ when the equation is of the form

$$f(X) + g(Y) = h(X+Y) \tag{22}$$

and X, Y are independent continuous random variables, solutions f, g of (22) may be nonlinear.

Indeed, let ξ be a uniform random variable on $(0,1)$. Consider its diadic representation

$$\xi = \sum_{k=1}^{\infty} \frac{\xi_k}{2^k},$$

where ξ_1, ξ_2, $\ldots$ are independent binary random variables with $\mathbf{P}(\xi_k = 0) = \mathbf{P}(\xi_k = 1) = .5$. Now set

$$X = \sum_{k \text{ even}} \frac{\xi_k}{2^k}, \quad Y = \sum_{k \text{ odd}} \frac{\xi_k}{2^k}.$$

Then X and Y are independent random variables with continuous (though singular) distributions and they both are functions of $X+Y=\xi$ (X and Y are strong components of ξ, in terminology of Hoffmann-Jorgensen *et. al.* [6]). Thus, for any measurable functions f and g, the relation (22) holds.

On the other hand, if both X and Y have positive almost everywhere (with respect to the Lebesgue measure) densities and f, g are locally integrable functions, then the equation (22) has only linear solutions f, g (and certainly h).

From positivity of the densities, one has

$$f(x) + g(y) = h(x+y) \tag{23}$$

almost everywhere (with respect to the plane Lebesgue measure). On taking a smooth function $k(x)$ with compact support, multiplying both sides of (23) by $k(x)$ and integrating over x, results in

$$\int_{-\infty}^{+\infty} f(x)k(x)\,dx + g(y)\int_{-\infty}^{+\infty} k(x)\,dx = \int_{-\infty}^{+\infty} h(x+y)k(x)\,dx = \int_{-\infty}^{+\infty} h(u)k(u-y)\,du,$$

where the right hand side is continuous in y. Thus, $g(y)$ is continuous and so is $f(x)$ implying that (23) holds for all (and not almost all) x, y (the idea is due to Hillel Furstenberg).

Now (23) becomes the Cauchy classical equation that has only linear solutions.

Returning to (21) and noticing that $\mathbf{E}|t_n'| < \infty$, $\mathbf{E}|t_n''| < \infty$, one concludes that if F_1 and F_2 are given by almost everywhere positive densities, then for almost all (with respect to the Lebesgue measure in $\mathbb{R}^{2n}$)

$$t_n'(x_1 \ldots, x_n) + t_n''(y_1, \ldots, y_n) = t_n(x_1+y_1, \ldots, x_n+y_n). \tag{24}$$

Treating (24) as a Cauchy type equation in x_i, y_i with the remaining $n-1$ pairs of the arguments fixed, one gets the linearity of t'_n, t''_n in each of their arguments whence due to the symmetry $t'_n = \overline{x}$, $t''_n = \overline{y}$ implying for $n \geqslant 3$ that F_1 and F_2 are Gaussian. Thus, the following result is proved.

Theorem 2. *Let $t_n^{(1)}, \dots, t_n^{(N)}$, $N > 1$ are the Pitman estimators of θ from samples of size $n \geqslant 3$ from populations $F_1(x-\theta), \dots, F_N(x-\theta)$ with finite second moments and almost everywhere positive densities, and t_n the Pitman estimator form a sample of size n from $(F_1 * \dots * F_N)(x-\theta)$. Then*

$$\operatorname{var}(t_n) = \sum_1^N \operatorname{var}(t_n^{(k)})$$

if and only if all the populations are Gaussian.

§4. Combining independent samples from different populations

Let $(x_1^{(k)}, \dots, x_{n_k}^{(k)})$, $k = 1, \dots, N$ be independent samples of size $n_1, \dots, n_N$ from populations $F_1(x-\theta), \dots, F_N(x-\theta)$ with finite variances and $t_{n_k}^{(k)}$ be the Pitman estimator of θ from the sample $(x_1^{(k)}, \dots, x_{n_k}^{(k)})$ of size n_k. For $\mathbf{s} = \{i_1, \dots, i_m\}$, we denote by $t_{n(\mathbf{s})}^{(\mathbf{s})}$ the Pitman estimator of θ from the sample of size $n(\mathbf{s}) = n_{i_1} + \dots + n_{i_m}$ that is obtained from pooling samples with superindices from $\mathbf{s}$. By $t_n^{(1,\dots,N)}$ we denote the Pitman estimator of θ from the sample $(x_1^{(1)}, \dots, x_{n_N}^{(N)})$ of size $n = n_1 + \dots + n_N$. Trivially, $\operatorname{var}(t_n^{(1,\dots,N)})$ is the smallest among $\operatorname{var}(t_{n(\mathbf{s})}^{(\mathbf{s})})$. Here a stronger result is proved.

Theorem 3. *The following inequality holds:*

$$\frac{1}{\operatorname{var}(t_n^{(1,\dots,N)})} \geqslant \frac{1}{\binom{N-1}{m-1}} \sum_{\mathbf{s}} \frac{1}{\operatorname{var}(t_{n(\mathbf{s})}^{(\mathbf{s})})} \tag{25}$$

where the summation in (25) *is over all combinations* $\mathbf{s}$ *of* m *elements from* $\{1, \dots, N\}$.

Proof. On setting in Lemma 1 $\psi_{\mathbf{s}} = t_{n(\mathbf{s})}^{(\mathbf{s})}$ and choosing the weights $w_{\mathbf{s}}$ minimizing the right hand side of (7),

$$w_{\mathbf{s}} = \pi_{\mathbf{s}} / \sum_{\mathbf{s}} \pi_{\mathbf{s}},$$

where $\pi_{\mathbf{s}} = 1/\operatorname{var}(t_{n(\mathbf{s})}^{(\mathbf{s})})$, one gets

$$\binom{N-1}{m-1} \frac{1}{\sum_{\mathbf{s}} \frac{1}{\operatorname{var}(t_{n(\mathbf{s})}^{(\mathbf{s})})}} \geqslant \operatorname{var}\left(\sum_{\mathbf{s}} w_{\mathbf{s}} t_{n(\mathbf{s})}^{(\mathbf{s})}\right).$$

For sample $(x_1^{(1)}, \ldots, x_{n_N}^{(N)})$, $\sum_{\mathbf{s}} w_{\mathbf{s}} t_{n(\mathbf{s})}^{(\mathbf{s})}$ is an equivariant estimator while $t_n^{(1,\ldots,N)}$ is the Pitman estimator. Thus

$$\operatorname{var}\left(\sum_{\mathbf{s}} w_{\mathbf{s}} t_{n(\mathbf{s})}^{(\mathbf{s})}\right) \geqslant \operatorname{var}(t_n^{(1,\ldots,N)})$$

which, combined with the previous inequality, is exactly (25). □

In a special case, when the subsets $\mathbf{s}$ consist of one element and $t_{n_k}^{(k)}$ is the Pitman estimator from $(x_1^{(k)}, \ldots, x_{n_k}^{(k)})$, Theorem 3 becomes

$$\frac{1}{\operatorname{var}(t_n^{(1,\ldots,N)})} \geqslant \frac{1}{\operatorname{var}(t_{n_1}^{(1)})} + \cdots + \frac{1}{\operatorname{var}(t_{n_N}^{(N)})}. \tag{26}$$

This inequality is reminiscent of Carlen's superadditivity for the trace of the Fisher information matrix, which involves the Fisher informations obtained by taking the limit as sample sizes go to infinity. However, Carlen's superadditivity is true for random variables with arbitrary dependence, whereas (26) has only been proved under assumption of independence of samples.

§5. Some corollaries, including the monotonicity of $n \operatorname{var}(t_n)$

Notice that if for a sample of size m from $F(x-\theta)$, $\operatorname{var}(t_m) < \infty$, then $\operatorname{var}(t_n) < \infty$ for samples $(x_1, \ldots, x_n)$ of any size $n > m$.

Set $F_1 = \ldots = F_N = F$, $n_1 = \ldots = n_N = 1$, and $m = N-1$ in Theorem 3. Then $n(\mathbf{s}) = N-1$ for each $\mathbf{s}$ with m elements, and $n = N$, and Theorem 3 reads

$$\frac{1}{\operatorname{var}(t_N^{(1,\ldots,N)})} \geqslant \frac{1}{N-1} \sum_j \frac{1}{\operatorname{var}(t_{N-1}^{(1,\ldots,j-1,j+1,\ldots,N)})} = \frac{N}{N-1} \frac{1}{\operatorname{var}(t_{N-1}^{(1,\ldots,N-1)})},$$

where the last equality is due to symmetry. Now $t_N^{(1,\dots,N)}$ is just the Pitman estimator of θ from a sample of size N from $F(x-\theta)$. Thus, interpreting N as sample size instead of group size, we have the following result.

Theorem 4. *Let t_n be the Pitman estimator of θ from a sample of size n from a population $F(x-\theta)$. If for some m, $\operatorname{var}(t_m) < \infty$, then for all $n \geqslant m+1$*

$$(n+1)\operatorname{var}(t_{n+1}) \leqslant n \operatorname{var}(t_n). \tag{27}$$

For $n \geqslant 2$, the equality sign holds if and only if F is Gaussian.

Remarks.

(1) If F is Gaussian $N(0,\sigma^2)$, then clearly $n\operatorname{var}(t_n) = \sigma^2$ for all n. In fact, the equality

$$n\operatorname{var}(t_n) = (n+1)\operatorname{var}(t_{n+1})$$

holding for any $n \geqslant 2$ characterizes the Gaussian distribution since it implies the additive decomposability of t_n. If an equivariant estimator is additively decomposable, it is linear and due to the symmetry of t_n one has $t_n = \overline{x}$.

(2) The condition of Theorem 4 is fulfilled for $m = 1$ (and thus for any m) if $\int x^2 dF(x) < \infty$. However, for many F with infinite second moment (e.g., Cauchy), $\operatorname{var}(t_m) < \infty$ for some m and Theorem 4 holds.

(3) Note that even absolute continuity of F is not required, not to mention the finiteness of the Fisher information.

(4) If F is the distribution function of an exponential distribution with parameter $1/\lambda$,

$$n\operatorname{var}(t_n) = \frac{2\lambda n}{(n+1)(n+2)}.$$

If F is the distribution function of a uniform distribution on $(-1,1)$,

$$n\operatorname{var}(t_n) = \frac{4n^2}{(n+1)^2(n+2)}.$$

In these examples, the Fisher information is infinite, but one clearly has monotonicity.

(5) One can call F Pitman regular if

$$\lim_{n\to\infty} n\operatorname{var}(t_n) > 0 \tag{28}$$

and nonregular if the limit in (28) (that always exists) is zero. As mentioned earlier, Ibragimov and Has'minskii [7] showed that under rather mild conditions on F that include the finiteness of the Fisher information I,

$$\lim_{n\to\infty} n\,\mathrm{var}\,(t_n) = 1/I.$$

Under these conditions, Theorem 4 implies monotone convergence of n var (t_n) to its limit. A corollary of Theorem 4 is worth mentioning.

Corollary 4. *Let $(x_1, \ldots, x_{n+m})$, $m+n \geqslant 3$, be a sample from the population $F(x-\theta)$ with finite variance. If t_m is the Pitman estimator of θ from the first m and t_n from the last n observations, then*

$$t_{n+m} = w_1 t_n + w_2 t_m$$

for some w_1, w_2 if and only if F is Gaussian.

Proof. One can easily see that necessarily $w_1 = m/(m+n)$, $w_2 = n/(m+n)$ so that

$$\begin{aligned}\mathrm{var}\,(t_{m+n}) &= \left(\frac{m}{m+n}\right)^2 \mathrm{var}\,(t_m) + \left(\frac{n}{m+n}\right)^2 \mathrm{var}\,(t_n) \\ &\geqslant \left(\frac{m}{m+n}\right) \mathrm{var}\,(t_{m+n}) + \left(\frac{n}{m+n}\right) \mathrm{var}\,(t_{m+n}) = \mathrm{var}\,(t_{m+n}),\end{aligned}$$

the equality sign holding if $(m+n)\,\mathrm{var}\,(t_{m+n}) = m\,\mathrm{var}\,(t_m) = n\,\mathrm{var}\,(t_n)$. □

We can now characterize equality for another special case of Theorem 3.

Corollary 5. *Let t_m be the Pitman estimator from a sample of size m from $F(x-\theta)$. Then one has superadditivity with respect to the sample size,*

$$\frac{1}{\mathrm{var}\,(t_n)} \geqslant \frac{1}{\mathrm{var}\,(t_{n_1})} + \cdots + \frac{1}{\mathrm{var}\,(t_{n_N})},\ n = n_1 + \cdots + n_N, \tag{29}$$

with equality if and only if F is Gaussian.

Proof. Taking $F_1 = \ldots = F_N = F$ in Theorem 3 immediately gives (29). To understand when the equality sign holds in (29), suffice to consider the case of $N = 2$. Set $n_1 = l$, $n_2 = m$, $n = l + m$. The equality sign in

$$\frac{1}{\mathrm{var}\,(t_n)} \geqslant \frac{1}{\mathrm{var}\,(t_l)} + \frac{1}{\mathrm{var}\,(t_m)}$$

holds if and only if

$$t_n = w_1 t_l + w_2 t_m, \text{ with } w_1 = l/n,\ w_2 = m/n.$$

According to Corollary 4, the last relation holds if and only if F is Gaussian. $\square$

Another corollary of interest that looks similar in form to Corollary 2 of Sec. 2 but is of a different nature, follows immediately from combining Theorem 4 and the definition (1).

Corollary 6. *For independent identically distributed* $X_1, X_2, \dots$ *with* $\operatorname{var}(X_i) = \sigma^2 < \infty$ *set*

$$\overline{X}_n = (X_1 + \ldots + X_n)/n.$$

Then for any $n \geqslant 1$,

$$(n+1) \operatorname{var} \mathbf{E}(\overline{X}_{n+1} | X_1 - \overline{X}_{n+1}, \dots, X_{n+1} - \overline{X}_{n+1}) \\ \geqslant n \operatorname{var} \mathbf{E}(\overline{X}_n | X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n).$$

In the regular case when $\lim_{n \to \infty} n \operatorname{var}(t_n) = 1/I$,

$$\lim_{n \to \infty} n \operatorname{var} \mathbf{E}(\overline{X}_n | X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n) = \sigma^2 - 1/I.$$

It would be interesting to study the asymptotic behavior as $n \to \infty$ of the random variable

$$\mathbf{E}(\sqrt{n}\overline{X}_n | X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n).$$

§6. Multivariate extensions

An extension of Theorem 1 to the multivariate case depends on a generalization of the variance drop lemma (Lemma 1) to the case of s-variate vector functions. Using the Cramér-Wold principle, for an arbitrary vector $c \in \mathbb{R}^s$ and vector functions $\psi_{\mathbf{s}} = \psi_{\mathbf{s}}(\mathbf{X}_{\mathbf{s}})$, set

$$\phi_{\mathbf{s}}(\mathbf{X}_{\mathbf{s}}) = c^T \psi_{\mathbf{s}}(\mathbf{X}_{\mathbf{s}}).$$

Thus Lemma 1 implies

$$c^T \operatorname{var}\left\{ \sum_{\mathbf{s}} \psi_{\mathbf{s}} \right\} c \leqslant \binom{N-1}{m-1} \sum_{\mathbf{s}} w_{\mathbf{s}}^2 c^T \operatorname{var}\{\psi_{\mathbf{s}}\} c.$$

This is equivalent to

$$\operatorname{var}\left\{ \sum_{\mathbf{s}} \psi_{\mathbf{s}} \right\} \leqslant \binom{N-1}{m-1} \sum_{\mathbf{s}} w_{\mathbf{s}}^2 \operatorname{var}\{\psi_{\mathbf{s}}\},$$

where var means the covariance matrix; hence Lemma 1 holds in the multivariate case if we interpret the inequality in terms of the Loewner ordering.

In Theorem 1, if $X_1, \dots, X_n$ are independent s-variate random vectors with distribution $F(x-\theta)$, $x, \theta \in \mathbb{R}^s$, all the results and the proof remain true where an inequality $A \geqslant B$ for matrices A, B means, as usual, that the matrix $A - B$ is non-negative definite.

Corollary 5 remains valid in the multivariate case when the above samples come from s-variate populations depending on $\theta \in \mathbb{R}^s$ assuming that the covariance matrices of the involved Pitman estimators are nonsingular. The latter condition is extremely mild. Indeed, if the covariance matrix V of the Pitman estimator τ_n from a sample of size n from an s-variate population $H(x-\theta)$ is singular, then for a nonzero (column) vector $a \in \mathbb{R}^s$

$$\operatorname{var}(a'\tau_n) = a'Va = 0,$$

(prime stands for transposition) meaning that the linear function $a'\theta$ is estimatable with zero variance. This implies that any two distributions in $\mathbb{R}^{ns}$ generated by samples of size n from $F(x-\theta_1)$ and $F(x-\theta_2)$ with $a'\theta_1 \neq a'\theta_2$ are mutually singular and so are the measures in $\mathbb{R}^s$ with distribution functions $F(x-\theta_1)$ and $F(x-\theta_2)$. Since for any θ_1 there exists an arbitrarily close to it θ_2 with $a'\theta_1 \neq a'\theta_2$, singularity of the covariance matrix of the Pitman estimator would imply an extreme irregularity of the family $\{F(x-\theta),\, \theta \in \mathbb{R}^s\}$. In the multivariate case (26) takes the form of

$$V^{-1}(t_n^{(1,\dots,N)}) \geqslant V^{-1}(t_{n_1}^{(1)}) + \cdots + V^{1}(t_{n_N}^{(N)}) \tag{30}$$

where $V(t)$ is the covariance matrix of a random vector t. To prove (30), take matrix-valued weights

$$W_k = \left(V^{-1}(t_{n_1}^{(1)}) + \cdots + V^{-1}(t_{n_N}^{(N)})\right)^{-1} V^{-1}(t_{n_k}^{(k)}), \quad k = 1, \dots, N. \tag{31}$$

Since $W_1 + \cdots + W_N$ is the identity matrix, $W_1 t_{n_1}^{(1)} + \cdots + W_N t_{n_N}^{(N)}$ is an equivariant estimator of θ so that its covariance matrix exceeds that of the Pitman estimator,

$$\begin{aligned} V(t_n^{(1,\dots,N)}) &\leqslant V\left(W_1 t_{n_1}^{(1)} + \cdots + W_N t_{n_N}^{(N)}\right) \\ &= W_1 V(t_{n_1}^{(1)}) W_1' + \cdots + W_N V(t_{n_N}^{(N)}) W_N'. \end{aligned}$$

Substituting the weights (31) into the last inequality gives (30).

If $(x_1, \dots, x_n)$ is a sample from the multivariate population $F(x-\theta)$ (where both x and θ are vectors), the monotonicity of Theorem 4 holds for the covariance matrix V_n of the Pitman estimator, i.e.,

$$nV_n \geqslant (n+1)V_{n+1}.$$

The proof is the same as that of the univariate case, but uses the multivariate version of Lemma 1 discussed at the beginning of this section.

§7. Extensions to polynomial Pitman estimators

Assuming

$$\int x^{2k} dF(x) < \infty \tag{32}$$

for some integer $k \geqslant 1$, the polynomial Pitman estimator $\widehat{t}_n^{(k)}$ of degree k is, by definition, the minimum variance equivariant polynomial estimator (see Kagan [11]). Let $M_k = M_k(x_1 - \overline{x}, \dots, x_n - \overline{x})$ be the space of all polynomials of degree k in the residuals. Also, let $\widehat{\mathbf{E}}(\cdot|M_k)$ be the projection into M_k in the (finite-dimensional) Hilbert space of polynomials in $x_1, \dots, x_n$ of degree k with the standard inner product

$$(q_1, q_2) = \mathbf{E}(q_1 q_2).$$

Then the polynomial Pitman estimator can be represented as

$$\widehat{t}_n^{(k)} = \overline{x} - \widehat{\mathbf{E}}(\overline{x}|M_k). \tag{33}$$

Plainly, it depends only on the first $2k$ moments of F.

To extend our earlier results to the polynomial Pitman estimators $\widehat{t}_n^{(k)}$ under the assumption $\int x^{2k} dF(x) < \infty$, the following properties of the projection operators are useful:

(1) For any index set $\mathbf{s}$,

$$M_k(R_n) = M_k(R_{\mathbf{s}} + R_{\overline{\mathbf{s}}}) \subset M_k(R_{\mathbf{s}}, R_{\overline{\mathbf{s}}})$$

so that for any random variable ξ

$$\operatorname{var}\{\widehat{\mathbf{E}}(\xi|M_k(R_n))\} \leqslant \operatorname{var}\{\widehat{\mathbf{E}}(\xi|M_k(R_{\mathbf{s}}, R_{\overline{\mathbf{s}}}))\}.$$

(2) Let ξ be a random variable such that the pair $(\xi, R_{\mathbf{s}})$ is independent (actually, suffice to assume uncorrelatedness) of $R_{\overline{\mathbf{s}}}$, then

$$\widehat{\mathbf{E}}(\xi|M_k(R_{\mathbf{s}}, R_{\overline{\mathbf{s}}})) = \widehat{\mathbf{E}}(\xi|M_k(R_{\mathbf{s}})).$$

Substituting the conditional expectations in the proof of Theorem 1 by the projection operators $\widehat{\mathbf{E}}(\cdot|M_k)$, the following version of Theorem 1 for polynomial Pitman estimators can be proved.

Theorem $1'$. *If for some integer $k \geqslant 1$, $\int x^{2k} dF_j(x) < \infty$, $j = 1, \dots, N$, the variance of the polynomial Pitman estimators $\widehat{t}_{\mathbf{s},n}^{(k)}$ satisfy the inequality*

$$\operatorname{var}(\widehat{t}_n^{(k)}) \geqslant \frac{1}{\binom{N-1}{m-1}} \sum_{\mathbf{s}} \operatorname{var}(\widehat{t}_{\mathbf{s},n}^{(k)}).$$

Assuming that for some integer $m \geqslant 1$

$$\int x^{2m} dF_k(x) < \infty,\ k = 1, \dots, N,$$

Corollary 5 also easily extends to the polynomial Pitman estimators of degree m.

Similarly, under the condition (32) for some integer $k \geqslant 1$, the Theorem 4 extends to the polynomial Pitman estimator $\widehat{t}_n^{(k)}$ defined in (33). The polynomial Pitman estimator $\widehat{t}_{n,j}^{(k)}$ of degree k from

$$(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$$

is equidistributed with $\widehat{t}_{n-1}^{(k)}$ and thus

$$\operatorname{var}(\widehat{t}_{n,j}^{(k)}) = \operatorname{var}(\widehat{t}_{n-1}^{(k)}).$$

The estimator $(1/n)\sum_1^n \widehat{t}_{n,j}^{(k)}$ is equivariant (for sample $(x_1, \dots, x_n)$) and since $\widehat{t}_n^{(k)}$ is the polynomial Pitman estimator,

$$\operatorname{var}(\widehat{t}_n^{(k)}) \leqslant (1/n^2) \operatorname{var}\left(\sum_1^n \widehat{t}_{n,j}^{(k)}\right). \tag{34}$$

By the $m = N - 1$ special case of the variance drop lemma,

$$\operatorname{var}\left(\sum_1^n \widehat{t}_{n,j}^{(k)}\right) \leqslant (n-1) \sum_1^n \operatorname{var}(\widehat{t}_{n,j}^{(k)}) = n(n-1) \operatorname{var}(\widehat{t}_{n-1}^{(k)}). \tag{35}$$

Combining the last two inequalities gives

$$(n+1) \operatorname{var}(\widehat{t}_{n+1}^{(k)}) \leqslant n \operatorname{var}(\widehat{t}_n^{(k)}), \tag{36}$$

i.e., $n \operatorname{var}(\widehat{t}_n^{(k)})$ decreases with n.

In Kagan *et. al.* [13] it is shown that under only the moment condition (32), $n \operatorname{var}(\widehat{t}_n^{(k)}) \to 1/I^{(k)}$ as $n \to \infty$ where $I^{(k)}$ can be interpreted as the Fisher information on θ contained in the first $2k$ moments of F (see Kagan [12]). For any increasing sequence $k(n)$, one sees that $n \operatorname{var}(\widehat{t}_n^{k(n)})$ decreases with n, and the limit can be equal to $1/I$ under some additional conditions.

Indeed, if the span of all the polynomials in X with distribution function F coincides with $L^2(F)$, the space of all square integrable functions of X, then $I^{(k)} \to I$ as $k \to \infty$.

The above proof of monotonicity is due to the fact that the classes where t_n and $t_n^{(k)}$ are the best are rather large. To illustrate this, consider the following analog of $t_n^{(k)}$:

$$\tau_n^{(k)} = \overline{x} - \widehat{\mathbf{E}}\big(\overline{x} \mid 1, m_2, \dots, m_k\big) = \overline{x} - \sum_{j=0}^{k} a_{j,n} m_j,$$

where $m_j = (1/n)\sum_1^n (x_i - \overline{x})^j$ and $\widehat{\mathbf{E}}(\overline{x} \,|\, 1, m_2, \dots, m_k)$ is the projection of $\overline{x}$ into the space $\operatorname{span}(1, m_2, \dots, m_k)$ (i.e., the best mean square approximation of $\overline{x}$ by linear combinations of the sample central moments of orders up to k). As shown in Kagan *et. al.* [13], if $\int x^{2k} dF(x) < \infty$, the behavior of $\tau_n^{(k)}$ as $n \to \infty$ is the same as of $t_n^{(k)}$:

$$\sqrt{n}(\tau_n^{(k)} - \theta) \overset{\text{dist}}{\to} Z^{(k)}$$

where $Z^{(k)}$ has a Gaussian distribution $N(0, 1/I^{(k)})$ and $n \operatorname{var}(\tau_n^{(k)}) \to 1/I^{(k)}$. However, it does not seem likely that (36) holds for $\tau_n^{(k)}$.

§8. Additive perturbations with a scale parameter

In this section, the setup of a sample $(x_1, \dots, x_n)$ from a population $F_\lambda(x - \theta)$ is considered where

$$F_\lambda(x) = \int F(y)\, dG((x - y)/\lambda).$$

In other words, an observation X with distribution function $F(x - \theta)$ is perturbed by an independent additive noise λY with $\mathbf{P}(Y \leqslant y) = G(y)$.

We study the behavior of the variance $\operatorname{var}(t_{n,\lambda})$, as a function of λ, of the Pitman estimator of θ from a sample of size n from $F_\lambda(x - \theta)$. For the so called self-decomposable Y, it is proved that $\operatorname{var}(t_{n,\lambda})$ behaves "as expected," i.e., monotonically decreases for $\lambda \in (-\infty, 0)$ and increases for $\lambda \in (0, +\infty)$.

They say that a random variable Y is *self-decomposable* if for any $c \in (0, 1)$, Y is equidistributed with $cY + Z_c$, i.e.,

$$Y \cong cY + Z_c, \tag{37}$$

where Z_c is independent of Y. If $f(t)$ is the characteristic function of Y, then (37) is equivalent to

$$f(t) = f(ct)g_c(t)$$

where $g_c(t)$ is a characteristic function. All random variables having stable distributions are self-decomposable. A self-decomposable random variable is necessarily infinitely divisible. In Lukacs [14, Chap. 5] necessary and sufficient conditions are given for self-decomposability in terms of the Lévy spectral function.

Theorem 5. *Let X be an arbitrary random variable with $\mathbf{E}(X^2) < \infty$ and Y a self-decomposable random variable with $\mathbf{E}(Y^2) < \infty$ independent of X. Then the variance $\operatorname{var}(t_{n,\lambda})$ of the Pitman estimator of θ from a sample of size n from $F_\lambda(x-\theta)$, is increasing in λ on $(0,\infty)$ and decreasing on $(-\infty, 0)$.*

Proof. If $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ are independent random variables, the x's with distribution $F(x-\theta)$ and the y's with distribution $G(y)$, then

$$t_{n,\lambda} = \overline{x} + \lambda\overline{y} - \mathbf{E}(\overline{x} + \lambda\overline{y} | x_1 - \overline{x} + \lambda(y_1 - \overline{y}), \ldots, x_n - \overline{x} + \lambda(y_n - \overline{y}))$$

and

$$\begin{aligned}\operatorname{var}(t_{n,\lambda}) = {} & \operatorname{var}(\overline{x} + \lambda\overline{y}) \\ & - \operatorname{var}\Big\{\mathbf{E}(\overline{x} + \lambda\overline{y} | x_1 - \overline{x} + \lambda(y_1 - \overline{y}), \ldots, x_n - \overline{x} + \lambda(y_n - \overline{y}))\Big\}.\end{aligned}$$

If $\lambda_2 > \lambda_1 > 0$, then $\lambda_1 = c\lambda_2$ for some c, $0 < c < 1$.

Due to self-decomposability of y_i, there exist random variables $z_{c,1}, ..., z_{c,n}$ such that

$$y_i - \overline{y} \cong c(y_i - \overline{y}) + (z_{c,i} - \overline{z}_c) \tag{38}$$

and the random variables $x_1, \ldots, x_n, y_1, \ldots, y_n, z_{c,1}, \ldots, z_{c,n}$ are independent.

The σ-algebra

$$\begin{aligned}&\sigma(x_1 - \overline{x} + \lambda_2(y_1 - \overline{y}), \ldots, x_n - \overline{x} + \lambda_2(y_n - \overline{y})) \\ &\quad = \sigma(x_1 - \overline{x} + \lambda_2 c(y_1 - \overline{y}) + \lambda_2(z_{c,1} - \overline{z}_c), \ldots, x_n - \overline{x} \\ &\qquad + \lambda_2 c(y_n - \overline{y}) + \lambda_2 c(y_n - \overline{y}) + \lambda_2(z_{c,n} - \overline{z}_c))\end{aligned}$$

is smaller than the σ-algebra

$$\sigma(x_1 - \overline{x} + \lambda_2 c(y_1 - \overline{y}), \ldots, x_n - \overline{x} + \lambda_2 c(y_n - \overline{y}), z_{c,1} - \overline{z}_c, \ldots, z_{c,n} - \overline{z}_c)$$

and thus

$$\begin{aligned}&\operatorname{var}\{\mathbf{E}(\overline{x}+\lambda_2\overline{y}|x_1-\overline{x}+\lambda_2(y_1-\overline{y}),\dots,x_n-\overline{x}+\lambda_2(y_n-\overline{y}))\}\\&\qquad\leqslant \operatorname{var}\{\mathbf{E}(\overline{x}+\lambda_2\overline{y}|x_1-\overline{x}+\lambda_2c(y_1-\overline{y}),\dots,\\&\qquad\qquad x_n-\overline{x}+\lambda_2c(y_n-\overline{y}),z_{c,1}-\overline{z}_c,\dots,z_{c,n}-\overline{z}_c)\}.\end{aligned}$$

From (38) and Lemma 2 in Sec. 2 one can rewrite the right-hand side of the above inequality

$$\begin{aligned}&\operatorname{var}\{\mathbf{E}(\overline{x}+\lambda_2\overline{y}|x_1-\overline{x}+\lambda_2c(y_1-\overline{y}),\dots,\\&\qquad x_n-\overline{x}+\lambda_2c(y_n-\overline{y}),z_{c,1}-\overline{z}_c,\dots,z_{c,n}-\overline{z}_c)\}\\&=\operatorname{var}\{\mathbf{E}(\overline{x}+\lambda_2c\overline{y}|x_1-\overline{x}+\lambda_2c(y_1-\overline{y}),\dots,x_n-\overline{x}+\lambda_2c(y_n-\overline{y})\}\\&\qquad+\operatorname{var}\{\mathbf{E}(\lambda_2\overline{z}_c|z_{c,1}-\overline{z}_c,\dots,z_{c,n}-\overline{z}_c).\}\end{aligned}\tag{39}$$

Again due to (38)

$$\operatorname{var}(\overline{x}+\lambda_2\overline{y})=\operatorname{var}(\overline{x}+\lambda_2c\overline{y}+\lambda_2\overline{z}_c).$$

Combining this with (39) and recalling that $c\lambda_2=\lambda_1$ leads to

$$\operatorname{var}(t_{n,\lambda_2})\geqslant\operatorname{var}(t_{n,\lambda_1}).$$

The case of $\lambda_1<\lambda_2<0$ is treated similarly. □

Theorem 5 has a counterpart in terms of the Fisher information: *Let* X, Y *be independent random variables. If* Y *is self-decomposable, then* $I(X+\lambda Y)$*, as a function of* λ*, monotonically increases on* $(-\infty,\,0)$ *and decreases on* $(0,\,+\infty)$.

The proof is much simpler than that of Theorem 5. Let $0<\lambda_2=c\lambda_1$ with $0<c<1$. Then $X+\lambda_2Y\cong X+c\lambda_2Y+\lambda_2Z_c$ where X, Y and Z_c are independent and the claim follows from that for independent random variables ξ, η, $I(\xi+\eta)\leqslant I(\xi)$.

§9. Discussion

Few years ago Bulletin of the Institute of Mathematical Statistics published letters [2], and [19] whose authors raised a question of monotonicity in the sample size of risks of standard ("classical") estimators. Natural expectations are that under reasonable conditions the mean square error, say, of the maximum likelihood estimator from a sample of size $n+1$ is less than from a sample of size n.

In this paper a stronger property of the Pitman estimator t_n of a location parameter is proved. Not only $\operatorname{var}(t_n)$ monotonically decreases in n but $\operatorname{var}(t_{n+1}) \leqslant \frac{n}{n+1}\operatorname{var}(t_n)$. However, for another equivariant estimator of a location parameter, that is asymptotically equivalent to t_n and has a "more explicit" form than t_n,

$$\widetilde{t}_n = \overline{x} - \frac{1}{nI}\sum_1^n J(x_i - \overline{x})$$

where J is the Fisher score and I the Fisher information, monotonicity in n of $\operatorname{var}(\widetilde{t}_n)$ is an open question. In a general setup, it is not clear what property of the maximum likelihood estimator is responsible for monotonicity of the risk when monotonicity holds.

In a recent paper [9] was proved monotonicity in the sample size of the length of some confidence intervals.

It seems as a challenge to find out when it is worth to make an extra observation.

References

1. S. Artstein, K. M. Ball, F. Barthe, A. Naor, *Solution of Shannon's problem on the monotonicity of entropy.* — J. Amer. Math. Soc. (electronic) **17**, No. 4 (2004), 975–982.
2. A. DasGupta, *Letter to the editors.* — IMS Bulletin **37**, No. 6 (2008), 16.
3. E. A. Carlen, *Superadditivity of Fisher's information and logarithmic Sobolev inequalities.* — J. Funct. Anal. **101**, No. 1 (1991), 194–211.
4. B. Efron, C. Stein, *The jackknife estimate of variance.* — Ann. Statist. **9**, No. 3 (1981), 586–596.
5. W. Hoeffding, *A class of statistics with asymptotically normal distribution.* — Ann. Math. Statist. **19**, No. 3 (1948), 293–325.
6. J. Hoffmann-Jørgensen, A. M. Kagan, L. D. Pitt, L. A. Shepp, *Strong decomposition of random variables.* — J. Theor. Probab. **20**, No. 2 (2007), 211–220.
7. I. A. Ibragimov, R. Z. Has'minski. *Statistical Estimation: Asymptotic Theory*, Vol. 16. *Applications of Mathematics.* Springer, New York, 1981.
8. A. Kagan, Z. Landsman, *Statistical meaning of Carlen's superadditivity of the Fisher information.* — Statist. Probab. Lett. **32** (1997), 175–179.
9. A. M. Kagan, Ya. Malinovsky, *Monotonicity in the sample size of the length of classical confidence intervals.* — Statist. Probab. Lett. (accepted) (2013).
10. A. Kagan, *An inequality for the Pitman estimators related to the Stam inequality.* — Sankhyā, Ser. A **64** (2002), 281–292.
11. A. M. Kagan. *On the estimation theory of location parameter.* — Sankhyā, Ser. A **28** (1966), 335–352.

12. A. M. Kagan, *Fisher information contained in a finite-dimensional linear space, and a properly formulated version of the method of moments.* — Probl. Peredači Inform. **12**, No. 2 (1976), 20–42.
13. A. M. Kagan, L. B. Klebanov, S. M. Fintušal, *Asymptotic behavior of polynomial Pitman estimators.* — Zap. Nauchn. Semin. LOMI **43** (1974), 30–39.
14. E. Lukacs, *Characteristic Functions*, 2nd ed. Hafner Publishing Co., New York, 1970.
15. M. Madiman, A. R. Barron, A. M. Kagan, T. Yu, *Fundamental limits for distributed estimation: the case of a location parameter.* — Preprint (2009).
16. M. Madiman, A. R. Barron, A. M. Kagan, T. Yu, *A model for pricing data bundles based on minimax risks for estimation of a location parameter.* — In: Proc. IEEE Inform. Theory Workshop. Volos, Greece (June–2009).
17. M. Madiman, A. R. Barron, *Generalized entropy power inequalities and monotonicity properties of information.* — IEEE Trans. Inform. Theory **53**, No. 7 (2007), 2317–2329.
18. J. Shao, *Mathematical Statistics*, 2nd ed. Springer, New York, 2003.
19. N.-Z. Shi, *Letter to the Editors.* — IMS Bulletin **36**, No. 4 (2008), 4.
20. A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon.* — Inform. Control **2** (1959), 101–112.
21. R. Zamir, *A proof of the Fisher information inequality via a data processing argument.* — IEEE Trans. Inform. Theory **4**, No. 4(3) (1998), 1246–1250.

Department of Mathematics,
University of Maryland
College Park, MD 20742
E-mail: `amk@math.umd.edu`,
`yuth@math.umd.edu`

Department of Statistics, Yale University
New Haven, CT 06511
E-mail: `andrew.barron@yale.edu`,
`mokshay.madiman@yale.edu`

Поступило 15 октября 2012 г.