ELECTROSTATIC VIEWS OF STEIN-TYPE ESTIMATION OF LOCATION VECTORS

Toshio Ohnishi***** and Takemi Yanagimoto*

Stein-type estimation of location vectors is discussed with the aid of the theory of electrostatics. We consider a class of estimating functions and assess the superiority of an estimating equation by its mean squared norm. The Coulomb potential function leads to a Pythagorean relationship with respect to this norm. By making full use of the Pythagorean relationship, we improve upon the likelihood estimating function. A further improvement is shown to be feasible under a certain condition which is described. We pursue possible strong relationships between the superiority over the likelihood estimating function and physical quantities appearing in the theory of electrostatics.

Key words and phrases: Coulomb potential function, electrostatics, estimating function, Green's formula, James-Stein estimator, James-Stein positive-part estimator, Pythagorean relationship.

1.   Introduction

A number of works have been devoted to elucidating the reason why the James-Stein estimator (James and Stein, 1961) or its modifiers perform well. The role of the Stein phenomenon is still increasing in the statistical theory due to the recent methodological progress in a statistical model containing a high-dimensional parameter. Since Efron and Morris (1973) pointed out the close relationship between the James-Stein estimator and the empirical Bayes method, justifications of the reason have been proposed from the Bayes and the empirical Bayes viewpoints; reviews are found in Lehmann and Casella (1998, Chapter 4) and Robert (2001, Sections 2.8 and 8.5). Furthermore, decision-theoretic approaches have been employed by many authors, including Brown (1966), Berger (1975), Hudson (1978), Shinozaki (1984), George (1986), Brandwein and Strawderman (1991) and Yanagimoto (1994, 2000).

The aim of this paper is to discuss the Stein-type estimation of location vectors in terms of estimating functions and to give interpretations from the viewpoint of electrostatics, a branch of physics. Although Liang and Waclawiw (1990) pointed out some merits of discussing the Stein phenomenon in terms of estimating functions in place of estimators, it would appear that there has been no other work since. Let a p-dimensional random vector $\boldsymbol{x} \in \mathbb{R}^p$ be distributed according to a location family $f(\boldsymbol{x}-\boldsymbol{\mu})$ where $f(\cdot)$ is a known density function on $\mathbb{R}^p$ and $\boldsymbol{\mu} \in \mathbb{R}^p$. As noted in Godambe and Kale (1991), the likelihood estimating

function

$$(1.1) \qquad \frac{\partial}{\partial \boldsymbol{\mu}} \log f(\boldsymbol{x} - \boldsymbol{\mu}) = -\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})$$

with $\nabla = \left(\partial/\partial x_1, \dots, \partial/\partial x_p\right)^T$ is optimum among unbiased estimating functions with respect to the trace optimality criterion. With respect to a formally extended version of this criterion, we attempt to improve upon the likelihood estimating function. The equality (1.1), characteristic of a location family, is one of the motivations to focus on the location family case. Another motivation is that the density function is defined on $\mathbb{R}^p$ and is invariant under the transformation

$$\boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{x}_0 \ \text{ and } \ \boldsymbol{\mu} \to \boldsymbol{\mu} + \boldsymbol{x}_0,$$

where $\boldsymbol{x}_0$ is an arbitrary point in $\mathbb{R}^p$. This property will facilitate our physical interpretations.

A Stein-type estimator is expected to be obtained from an estimating function of the form

$$(1.2) \qquad -\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \nabla \log u(\boldsymbol{x})$$

by choosing a suitable function $u(\boldsymbol{x})$. When $f(\boldsymbol{x} - \boldsymbol{\mu})$ is the density function of a normal distribution $\mathcal{N}_p(\boldsymbol{\mu}, I_p)$ with I_p being the $p \times p$ identity matrix, the estimator induced from (1.2) has the form $\boldsymbol{x} + \nabla \log u(\boldsymbol{x})$, which was studied by Stein (1981). In the Bayesian context, the estimating function (1.2) is closely related to the posterior mean. See Theorem 3.2 in Lehmann and Cassella (1998, Chapter 4).

If we regard the two functions of $\boldsymbol{x}$, $f(\boldsymbol{x} - \boldsymbol{\mu})$ and $u(\boldsymbol{x})$, as potential functions, then the two vector fields, $-\nabla f(\boldsymbol{x} - \boldsymbol{\mu})$ and $-\nabla u(\boldsymbol{x})$, can be looked upon as forces. We will use the term 'relative force' for such vector fields as $-\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})$ and $-\nabla \log u(\boldsymbol{x})$. It should be noted that the estimating function (1.2) is the difference between the two relative forces induced from $f(\boldsymbol{x} - \boldsymbol{\mu})$ and $u(\boldsymbol{x})$.

Our primary objective is to find a potential function $u_*(\boldsymbol{x})$ satisfying the equality

$$(1.3) \qquad E\left[\left\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) \right\|^2 - \left\| \nabla \log \frac{f(\boldsymbol{x} - \boldsymbol{\mu})}{u_*(\boldsymbol{x})} \right\|^2 - \left\| \nabla \log u_*(\boldsymbol{x}) \right\|^2 \right] = 0.$$

Note that this equality implies that $E\left[\|\nabla \log\{f(\boldsymbol{x} - \boldsymbol{\mu})/u_*(\boldsymbol{x})\}\|^2 \right]$ is less than $E\left[\|\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})\|^2 \right]$ unless $E\left[\|\nabla \log u_*(\boldsymbol{x})\|^2 \right]$ vanishes. If we assess the superiority of an estimating function by its mean squared norm, then the equality (1.3) implies that the estimating function $-\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \nabla \log u_*(\boldsymbol{x})$ dominates the likelihood estimating function. As noted by Stein (1981) in the $\mathcal{N}_p(\boldsymbol{\mu}, I_p)$ case, the estimating function $-\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \nabla \log u_*(\boldsymbol{x})$ has a mild optimum property in the sense that the following function of $a \in \mathbb{R}$

$$E\left[\left\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) - a\nabla \log u_*(\boldsymbol{x}) \right\|^2 \right]$$
$$= (a-1)^2 E\left[\left\| \nabla \log u_*(\boldsymbol{x}) \right\|^2 \right] + E\left[\left\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) \right\|^2 - \left\| \nabla \log u_*(\boldsymbol{x}) \right\|^2 \right]$$

is minimized at $a = 1$.

There are two advantages in our approach. One is to enable us to apply techniques developed in electrostatics to the theory of estimation. Actually, the potential function which we use to derive the Stein-type estimator is the well-known Coulomb potential function, $u_C(\boldsymbol{x} - \boldsymbol{m}) = \|\boldsymbol{x} - \boldsymbol{m}\|^{2-p}$ with $\boldsymbol{m} \in \mathbb{R}^p$. A generalized version of the James-Stein positive-part estimator is obtained by modifying the relative force induced from the Coulomb potential function. In this paper, the Stein identity (Stein, 1981) is replaced by Green's formula, which simplifies mathematical handling of the singular point of the Coulomb potential function. Interpretations from the viewpoints of electrostatics lead us to flexible treatments to the present problem and also provide us with a new insight of the Stein-type estimator.

The other advantage is that our approach is applicable to a wide range of distributions. This is a striking feature of the discussions in terms of estimating functions, as pointed out by Liang and Waclawiw (1990). One example concerns estimation of the location vector of a mutually independent Cauchy distribution. The mean squared error for the maximum likelihood estimator (MLE) is infinite. Instead, the quantity $E\big[\|\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})\|^2\big]$ is finite for the Cauchy distribution since the Fisher information matrix exists. Recall that the Cauchy distribution is becoming familiar in applications such as the financial engineering. See Nagahara (1999) for example.

The organization of this paper is as follows. In Section 2, we will formulate a criterion for assessing the superiority of an estimating function. A justification of the criterion will be given based on the theory of estimating functions. In Section 3, we will derive a Pythagorean relationship playing a fundamental role in subsequent sections. In Section 4, a location family with spherical symmetry will be discussed. We will use the Pythagorean relationship to derive an estimating function superior to the likelihood estimating function. Under a certain condition, we will also obtain another estimating function superior to the derived one above. In Section 5, a location family with mutual independence will be dealt with, and discussions similar to those in Section 4 will be made. In Section 6, we will give electrostatic views to the results in Sections 3 and 4. Some technical lemmas necessary for proving propositions in Sections 4 and 5 will be given in Appendix.

2. Formulation of a criterion

The dimension p is assumed to be an integer equal to or larger than three. Suppose that $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ has the density function $f(\boldsymbol{x} - \boldsymbol{\mu})$ with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ being a location vector. The function $f(\cdot)$ is assumed to be of class $\mathcal{C}^1$. We consider an estimating function of the form

$$(2.1) \qquad\qquad -\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) - \boldsymbol{U}(\boldsymbol{x}),$$

where $\boldsymbol{U}(\boldsymbol{x})$ is a vector field. The estimator $\hat{\boldsymbol{\mu}}$ is given as the parameter value which makes the two vector fields $-\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})$ and $\boldsymbol{U}(\boldsymbol{x})$ balance. If we set

$\boldsymbol{U}(\boldsymbol{x}) = \boldsymbol{0}$ in (2.1), then (2.1) coincides with the likelihood estimating function. As will be seen later, we will put a special emphasis on the case where $\boldsymbol{U}(\boldsymbol{x})$ is given as the relative force of a potential function $u(\boldsymbol{x})$, i.e., $\boldsymbol{U}(\boldsymbol{x}) = -\nabla \log u(\boldsymbol{x})$.

The criterion which we adopt in this paper is that we assess the superiority of the estimating function (2.1) by the following quantity

$$(2.2) \qquad E\Big[\big\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \boldsymbol{U}(\boldsymbol{x}) \big\|^2 \Big],$$

which is expressed also as

$$E\Big[\big\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) - \nabla \log f(\boldsymbol{x} - \hat{\boldsymbol{\mu}}) \big\|^2 \Big].$$

This can be regarded as the risk for $\hat{\boldsymbol{\mu}}$. When $\boldsymbol{U}(\boldsymbol{x}) = -\nabla \log u(\boldsymbol{x})$, the criterion (2.2) is the mean squared norm of the difference between the two relative forces, $-\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})$ and $-\nabla \log u(\boldsymbol{x})$. The estimating function (2.1) is biased in general. Setting $\boldsymbol{b}(\boldsymbol{\mu}) = E\big[-\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) - \boldsymbol{U}(\boldsymbol{x})\big] = E\big[-\boldsymbol{U}(\boldsymbol{x})\big]$, we can decompose (2.2) into the following two terms,

$$\big\| \boldsymbol{b}(\boldsymbol{\mu}) \big\|^2 + E\Big[\big\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \boldsymbol{U}(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{\mu}) \big\|^2 \Big].$$

Therefore, it seems relevant to use (2.2) as a criterion for the estimating function (2.1). Similar criterions for biased estimating functions are found in the existing literature such as Liang and Waclawiw (1990) and Yanagimoto and Yamamoto (1993).

It can be easily verified that the criterion (2.2) for the likelihood estimating function is the trace of the Fisher information matrix. As will be clarified by Theorem 3.2 in the next section, we will be able to employ the unified discussion for the location families having finite Fisher information matrices. This is one of the merits of our formulation. Note that the risk for the MLE happens to be infinite in the traditional formulation where simultaneous estimation is discussed directly in terms of estimators with the squared error as loss.

The criterion (2.2) can be obtained also as a formal extension of the trace optimality criterion presented in Godambe and Kale (1991). To show this, set two matrices A and B as

$$\begin{aligned} A &= E\left[\frac{\partial}{\partial \boldsymbol{\mu}} \Big\{ \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \boldsymbol{U}(\boldsymbol{x}) \Big\}^T \right] \\ &= E\left[-\frac{\partial^2}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T} \log f(\boldsymbol{x} - \boldsymbol{\mu}) \right] \end{aligned}$$

and

$$B = E\left[\Big\{ \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \boldsymbol{U}(\boldsymbol{x}) \Big\} \Big\{ \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \boldsymbol{U}(\boldsymbol{x}) \Big\}^T \right].$$

The trace optimality criterion assesses the superiority of (2.1), if unbiased, by the quantity $\operatorname{tr} A^{-1} B A^{-1}$. We will deal with the two cases where $f(\boldsymbol{x} - \boldsymbol{\mu})$ is

expressed as $f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)$ and as $\prod_{i=1}^{p} f(x_i - \mu_i)$ in Sections 4 and 5, respectively. In these cases the Fisher information matrix A of $f(\boldsymbol{x} - \boldsymbol{\mu})$ has the form $a_0 I_p$ with a_0 being positive and independent of $\boldsymbol{\mu}$. Then $A^{-1}BA^{-1} = a_0^{-2}B$, and a formally extended version of the trace optimality criterion reduces to the criterion (2.2). Note that we take biased estimating functions into consideration.

3. A Pythagorean relationship

First, in this section we will derive a Pythagorean relationship under certain conditions. This relationship will be used as a basic tool of improving upon the likelihood estimating function in the subsequent sections. We will apply Green's formula for the Laplacian $\triangle = \sum_{i=1}^{p} \partial^2/\partial x_i^2$ instead of the Stein identity (Stein, 1981). The Coulomb potential function $u_C(\cdot)$ introduced in Section 1 plays a key role.

First, let us consider a somewhat general situation. Assume that a function of class $\mathcal{C}^1$, $g : \mathbb{R}^p \to \mathbb{R}^+$, satisfies

$$(3.1) \qquad \int_{\mathbb{R}^p} \left\| \nabla \log g(\boldsymbol{x}) \right\|^2 g(\boldsymbol{x})\, d\boldsymbol{x} < \infty$$

and

$$(3.2) \qquad \int_{\mathbb{R}^p} \|\boldsymbol{x} - \boldsymbol{m}\|^{-2} g(\boldsymbol{x})\, d\boldsymbol{x} < \infty,$$

where $\boldsymbol{m}$ is a point in $\mathbb{R}^p$. The function $g(\cdot)$ is not necessarily a density function. The condition (3.2) is equivalent to

$$\int_{\mathbb{R}^p} \left\| \nabla \log u_C(\boldsymbol{x} - \boldsymbol{m}) \right\|^2 g(\boldsymbol{x})\, d\boldsymbol{x} < \infty,$$

and the Schwarz inequality gives

$$\int_{\mathbb{R}^p} \left\| \nabla \log \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \right\|^2 g(\boldsymbol{x})\, d\boldsymbol{x} < \infty.$$

A little calculation leads to the equality

$$\int_{\mathbb{R}^p} \left\| \nabla \log g(\boldsymbol{x}) \right\|^2 g(\boldsymbol{x})\, d\boldsymbol{x} - \int_{\mathbb{R}^p} \left\| \nabla \log \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \right\|^2 g(\boldsymbol{x})\, d\boldsymbol{x}$$

$$- \int_{\mathbb{R}^p} \left\| \nabla \log u_C(\boldsymbol{x} - \boldsymbol{m}) \right\|^2 g(\boldsymbol{x})\, d\boldsymbol{x}$$

$$= 2 \int_{\mathbb{R}^p} \nabla \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \cdot \nabla u_C(\boldsymbol{x} - \boldsymbol{m})\, d\boldsymbol{x}.$$

In order to evaluate the right-hand side of this equality, we apply Green's formula. Note that Green's formula is obtained from the divergence theorem, which was used by Brandwein and Strawderman (1991) in the setting different from ours. In

the following two theorems and their proofs, the dot indicates the scalar product, and the symbols dS and $\boldsymbol{n}$ stand for the surface element and the unit exterior normal vector, respectively.

THEOREM 3.1. *In addition to* (3.1) *and* (3.2), *assume that*

$$(3.3) \qquad \lim_{K \to \infty} \int_{\|\boldsymbol{x}\| = K} g(\boldsymbol{x}) \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} \cdot \boldsymbol{n} dS = 0.$$

Then the following Pythagorean relationship holds :

$$\int_{\mathbb{R}^p} \left\| \nabla \log g(\boldsymbol{x}) \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x} = \int_{\mathbb{R}^p} \left\| \nabla \log \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x}$$
$$+ \int_{\mathbb{R}^p} \left\| \nabla \log u_C(\boldsymbol{x} - \boldsymbol{m}) \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x}.$$

PROOF. Let $B(\boldsymbol{x}_0; r)$ denote the open ball with center $\boldsymbol{x}_0$ and radius r, $S(\boldsymbol{x}_0; r)$ the boundary of $B(\boldsymbol{x}_0; r)$, and $\overline{B(\boldsymbol{x}_0; r)}$ the closure of $B(\boldsymbol{x}_0; r)$. And let K and ε be positive numbers satisfying $K > \|\boldsymbol{m}\| + \varepsilon$ so that $B(\boldsymbol{0}; K) \supset B(\boldsymbol{m}; \varepsilon)$. An application of Green's formula yields the equality

$$(3.4) \qquad \int_V \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \triangle u_C(\boldsymbol{x} - \boldsymbol{m}) \, d\boldsymbol{x} + \int_V \nabla \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \cdot \nabla u_C(\boldsymbol{x} - \boldsymbol{m}) \, d\boldsymbol{x}$$
$$= \int_{S(\boldsymbol{0}; K)} \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \nabla u_C(\boldsymbol{x} - \boldsymbol{m}) \cdot \boldsymbol{n} dS$$
$$- \int_{S(\boldsymbol{m}; \varepsilon)} \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \nabla u_C(\boldsymbol{x} - \boldsymbol{m}) \cdot \boldsymbol{n} dS,$$

where $V = B(\boldsymbol{0}; K) \setminus \overline{B(\boldsymbol{m}; \varepsilon)}$. The former term in the left-hand side in (3.4) is zero since $u_C(\boldsymbol{x} - \boldsymbol{m})$ is harmonic on V. The condition (3.3) implies that the former term in the right-hand side in (3.4) vanishes as $K \to \infty$. In order to complete the proof, it is sufficient to show that the latter term in the right-hand side in (3.4) vanishes as $\varepsilon \to +0$. This is shown as follows:

$$\int_{S(\boldsymbol{m}; \varepsilon)} \frac{g(\boldsymbol{x})}{u_C(\boldsymbol{x} - \boldsymbol{m})} \nabla u_C(\boldsymbol{x} - \boldsymbol{m}) \cdot \boldsymbol{n} dS$$
$$= -(p - 2) \int_{S(\boldsymbol{m}; \varepsilon)} g(\boldsymbol{x}) \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} \cdot \boldsymbol{n} dS$$
$$= -(p - 2) \int_{S(\boldsymbol{m}; \varepsilon)} \frac{g(\boldsymbol{x})}{\|\boldsymbol{x} - \boldsymbol{m}\|} dS$$
$$= -\frac{p - 2}{\varepsilon} \int_{S(\boldsymbol{m}; \varepsilon)} g(\boldsymbol{x}) \, dS$$
$$= O\left(\varepsilon^{p-2}\right). \qquad \square$$

Now, we are to apply this theorem to a density function $f(\boldsymbol{x} - \boldsymbol{\mu})$ such that $f : \mathbb{R}^p \to \mathbb{R}^+$ is of class $\mathcal{C}^1$. Define two vector-valued functions of $\boldsymbol{\mu}$ as

$$(3.5) \qquad \boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x}) = -\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})$$

and

$$(3.6) \qquad \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) = -\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) + \nabla \log u_C(\boldsymbol{x} - \boldsymbol{m})$$
$$= -\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) - \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|}.$$

The following theorem implies that the three vectors $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$, $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ and $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x}) - \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ form on the average a right triangle with $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$ being its hypotenuse.

THEOREM 3.2. *Assume that the Fisher information matrix of $f(\boldsymbol{x} - \boldsymbol{\mu})$ exists. Then the following Pythagorean relationship holds :*

$$(3.7) \quad E\left[\left\| \boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2 - \left\| \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2 - \left\| \boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x}) - \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2 \right] = 0.$$

PROOF. All we need to do is to show that the three conditions (3.1)–(3.3) with $g(\boldsymbol{x}) = f(\boldsymbol{x} - \boldsymbol{\mu})$ are satisfied. The existence of the Fisher information matrix implies that $f(\boldsymbol{x} - \boldsymbol{\mu})$ satisfies (3.1). By transforming to the spherical coordinates around $\boldsymbol{m}$, we see that the contribution of the singular point $\boldsymbol{x} = \boldsymbol{m}$ to the integral (3.2) is zero when $p \geq 3$. Thus $f(\boldsymbol{x} - \boldsymbol{\mu})$ satisfies (3.2). Let K be a positive number larger than $\|\boldsymbol{m}\|$. Note that

$$\left| \int_{\|\boldsymbol{x}\| = K} f(\boldsymbol{x} - \boldsymbol{\mu}) \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} \cdot \boldsymbol{n} dS \right|$$
$$\leq \int_{\|\boldsymbol{x}\| = K} \frac{f(\boldsymbol{x} - \boldsymbol{\mu})}{\|\boldsymbol{x} - \boldsymbol{m}\|} \left| \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|} \cdot \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \right| dS$$
$$\leq \int_{\|\boldsymbol{x}\| = K} \frac{f(\boldsymbol{x} - \boldsymbol{\mu})}{\|\boldsymbol{x} - \boldsymbol{m}\|} dS$$
$$\leq \frac{1}{K - \|\boldsymbol{m}\|} \int_{\|\boldsymbol{x}\| = K} f(\boldsymbol{x} - \boldsymbol{\mu}) dS.$$

Since $f(\boldsymbol{x} - \boldsymbol{\mu})$ is a density function on $\mathbb{R}^p$, we see that

$$\lim_{K \to \infty} \int_{\|\boldsymbol{x}\| = K} f(\boldsymbol{x} - \boldsymbol{\mu}) dS = 0.$$

Thus (3.3) is satisfied. □

This theorem makes it clear that the unified discussion is applicable to all the location families having finite Fisher information matrices, which is one of the

advantages of our formulation over the traditional one. We present an illustrative example.

Example 3.1.   Consider a p-variate spherical t distribution having the density function

$$(3.8) \qquad f(\boldsymbol{x} - \boldsymbol{\mu}) = \frac{\Gamma\left((\phi_0 + p)/2\right)}{(\pi\phi_0)^{p/2}\Gamma\left(\phi_0/2\right)} \left(1 + \frac{\|\boldsymbol{x} - \boldsymbol{\mu}\|^2}{\phi_0}\right)^{-(\phi_0 + p)/2}$$

with ϕ_0 being a known positive constant. The mean squared norm of the likelihood estimating function

$$E\left[\left\|\nabla \log f(\boldsymbol{x} - \boldsymbol{\mu})\right\|^2\right] = \frac{p(p + \phi_0)}{p + \phi_0 + 2}$$

exists for any $\phi_0 > 0$ although the mean squared error for the MLE $\hat{\boldsymbol{\mu}}_M$

$$E\left[\left\|\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu}\right\|^2\right] = \frac{p\phi_0}{\phi_0 - 2}$$

does not exist if $\phi_0 \leq 2$.

Here we note mathematical handling of the case when the observation $\boldsymbol{x}$ equals $\boldsymbol{m}$. As shown by the proof of Theorem 3.2, the contribution of the singular point $\boldsymbol{x} = \boldsymbol{m}$ to the expectation in (3.7) is zero. Therefore we will formally discard the case $\boldsymbol{x} = \boldsymbol{m}$ in the subsequent sections. If the equation with respect to $\boldsymbol{\mu}$, $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) = \boldsymbol{0}$, has a solution for $\boldsymbol{x} \neq \boldsymbol{m}$, then we can regard $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ as an estimating function. Otherwise, we introduce an estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$, a modification of $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$, whose explicit form will be given in Sections 4 and 5.

Electrostatic views of the results in this section will be discussed further in Section 6. In addition, two corollaries to Theorems 3.1 and 3.2 will be obtained there.

4. Spherically symmetric case

In the following two sections, we will discuss the two cases, the spherically symmetric and the mutually independent cases. Further improvements on both $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ and its modification $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ will be shown to be possible under certain conditions. We begin with the former case in this section.

The likelihood estimating function (3.5) is expressed as

$$\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\|\boldsymbol{x} - \boldsymbol{\mu}\|}.$$

We focus on the case where the density function of $\boldsymbol{x}$ is in the family

$$\mathcal{S}_0 = \left\{ f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) \,\middle|\, f'(r) < 0 \text{ on } \mathbb{R}^+ \text{ and } \lim_{r \to +0} f'(r) = 0 \right\}$$

so that the MLE $\hat{\boldsymbol{\mu}}_M$ is obtained as $\boldsymbol{x}$ from the equation $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x}) = \boldsymbol{0}$.

The function $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ in (3.6) reduces to be of the form

$$(4.1) \qquad \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\|\boldsymbol{x} - \boldsymbol{\mu}\|} - \frac{p - 2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|}.$$

Suppose that $\boldsymbol{x} \neq \boldsymbol{m}$. Then a necessary and sufficient condition for the existence of the solution to the equation $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) = \boldsymbol{0}$ is that the function $-f'(r)/f(r)$ is not bounded to the above. In order to simplify later discussions, we introduce two subfamilies of $\mathcal{S}_0$ as

$$\mathcal{S}_1 = \left\{ f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) \in \mathcal{S}_0 \ \middle| \ \frac{-f'(r)}{f(r)} \text{ is not bounded to the above} \right\}$$

and

$$\mathcal{S}_2 = \left\{ f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) \in \mathcal{S}_0 \ \middle| \ \frac{-f'(r)}{f(r)} \text{ attains its maximum in } \mathbb{R}^+ \right\}.$$

When $f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) \in \mathcal{S}_1$ or $f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) \in \mathcal{S}_2$, Theorem 3.2 can be applied to improving upon $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$. If the equation $-f'(r)/f(r) = c$ has solutions, we set the least one as $r_S(c)$ for our later use. Electrostatic views of the estimating functions introduced in Subsections 4.1 and 4.2 will be given in Section 6.

4.1. The case of the subfamily $\mathcal{S}_1$

Applying Theorem 3.2 gives the following proposition.

PROPOSITION 4.1. *The estimating function $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the likelihood estimating function $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$.*

Next, we will show that $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ can be improved upon under the condition

$$(4.2) \qquad \frac{-f'(r)}{rf(r)} \text{ is non-increasing on } \mathbb{R}^+ \text{ and } \lim_{r \to +0} \frac{-f'(r)}{rf(r)} \text{ exists.}$$

In the expression (4.1) of $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$, the quantity $(p-2)/\|\boldsymbol{x} - \boldsymbol{m}\|$ becomes large when $\boldsymbol{x}$ is close to $\boldsymbol{m}$. As suggested in Berger and Bock (1976), it is expected that eliminating the singularity will lead us to better estimation. So let us consider the following estimating function

$$(4.3) \qquad \boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\|\boldsymbol{x} - \boldsymbol{\mu}\|}$$
$$- \min\left\{ h_0 \|\boldsymbol{x} - \boldsymbol{m}\|, \frac{p - 2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|},$$

where $h_0 = \lim_{r \to +0}\{-f'(r)\}/\{rf(r)\}$.

PROPOSITION 4.2. *Under the condition (4.2), the estimating function $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the estimating function $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$.*

PROOF. Let $B = B(\boldsymbol{m}; r_0)$ stand for the open ball with center $\boldsymbol{m}$ and radius $r_0 = \{(p-2)/h_0\}^{1/2}$. Noting that $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ differs from $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ only for $\boldsymbol{x} \in B$, we see that

$$
E\Big[\big\| \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) \big\|^2 \Big] - E\Big[\big\| \boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x}) \big\|^2 \Big]
$$
$$
= \int_B \left\| -\frac{f'(\|\boldsymbol{x}-\boldsymbol{\mu}\|)}{f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)} \frac{\boldsymbol{x}-\boldsymbol{\mu}}{\|\boldsymbol{x}-\boldsymbol{\mu}\|} - \frac{p-2}{\|\boldsymbol{x}-\boldsymbol{m}\|} \frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|} \right\|^2 f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)\, d\boldsymbol{x}
$$
$$
- \int_B \left\| -\frac{f'(\|\boldsymbol{x}-\boldsymbol{\mu}\|)}{f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)} \frac{\boldsymbol{x}-\boldsymbol{\mu}}{\|\boldsymbol{x}-\boldsymbol{\mu}\|} - h_0\,(\boldsymbol{x}-\boldsymbol{m}) \right\|^2 f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)\, d\boldsymbol{x}.
$$

The proof is completed by applying Lemma 2 in Appendix with $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{m}$, $\boldsymbol{a} = \boldsymbol{\mu} - \boldsymbol{m}$, $w_1(r) = -f'(r)/\{rf(r)\}$, $w_2(r) = f(r)$ and $w_3(r) = (p-2)\,r^{-2}$. □

Let us derive expressions of the two estimators $\hat{\boldsymbol{\mu}}_S$ and $\hat{\boldsymbol{\mu}}_S^+$, which are induced from $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ and $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$, respectively. When an estimating function yields multiple estimators, we choose the one closest to the MLE. This choice will be applied also in subsequent sections. Note that choice of an estimator is not essential in our approach. For various approaches for multiple root problems see Small *et al.* (2000). The expressions we obtain are

$$
(4.4) \qquad \hat{\boldsymbol{\mu}}_S = \boldsymbol{x} - r_S\left(\frac{p-2}{\|\boldsymbol{x}-\boldsymbol{m}\|} \right) \frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|}
$$

and

$$
(4.5) \qquad \hat{\boldsymbol{\mu}}_S^+ = \boldsymbol{x} - r_S\left(\min\left\{ h_0\|\boldsymbol{x}-\boldsymbol{m}\|,\ \frac{p-2}{\|\boldsymbol{x}-\boldsymbol{m}\|} \right\} \right) \frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|}.
$$

The condition (4.2) together with the definition of $r_S(\cdot)$ yields that

$$
(4.6) \qquad\qquad\qquad \lim_{c\to+0} \frac{r_S(h_0 c)}{c} = 1.
$$

Therefore, it follows that

$$
\|\hat{\boldsymbol{\mu}}_S^+ - \boldsymbol{m}\| = o\big(\|\boldsymbol{x}-\boldsymbol{m}\|\big), \qquad \text{as } \|\boldsymbol{x}-\boldsymbol{m}\| \to +0.
$$

This favorable property of $\hat{\boldsymbol{\mu}}_S^+$ is consistent with Proposition 4.2, which shows the superiority of $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ over $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$. The following example shows that the estimator $\hat{\boldsymbol{\mu}}_S^+$ is a generalization of the James-Stein positive-part estimator to a location family $f(\|\boldsymbol{x}-\boldsymbol{\mu}\|) \in \mathcal{S}_1$ satisfying the condition (4.2).

Example 4.1. Consider the $\mathcal{N}_p\,(\boldsymbol{\mu},\,I_p)$ case where (4.2) is satisfied. The estimators (4.4) and (4.5) are respectively the James-Stein estimator and the James-Stein positive-part estimator,

$$
\hat{\boldsymbol{\mu}}_S = \boldsymbol{x} - \frac{p-2}{\|\boldsymbol{x}-\boldsymbol{m}\|} \frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|}
$$

and

$$\hat{\boldsymbol{\mu}}_S^+ = \boldsymbol{x} - \min\left\{ \|\boldsymbol{x} - \boldsymbol{m}\|, \ \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|}.$$

4.2. The case of the subfamily $\mathcal{S}_2$

Set $c_1 = \max\{-f'(r)/f(r)\}$ and $r_1 = (p-2)/c_1$. Treatments parallel to those in the previous subsection can be used in this subsection, except for the fact that the equation $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) = \boldsymbol{0}$ does not always have any solution. It does not have any solution if $\|\boldsymbol{x} - \boldsymbol{m}\| < r_1$. By truncating the factor $(p-2)/\|\boldsymbol{x} - \boldsymbol{m}\|$ in $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ at the value c_1, we define the estimating function as

$$(4.7) \quad \tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\|\boldsymbol{x} - \boldsymbol{\mu}\|} - \min\left\{ c_1, \ \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|}.$$

The following proposition is fundamental in proving the superiority of $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ over $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$. In fact, the superiority is obtained as a corollary of the proposition and Theorem 3.2.

PROPOSITION 4.3. *The estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ satisfies the inequality*

$$\left\| \tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2 \leq \left\| \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2$$

for any $\boldsymbol{x} \in \mathbb{R}$ and $\boldsymbol{\mu} \in \mathbb{R}$.

PROOF. It is clear for $\boldsymbol{x} \notin B(\boldsymbol{m}; r_1)$ that $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x}) = \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$. A simple calculation shows for $\boldsymbol{x} \in B(\boldsymbol{m}; r_1)$ that

$$
\begin{aligned}
&\left\| \boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2 - \left\| \tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x}) \right\|^2 \\
&= \left(\frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} - c_1 \right) \left\{ \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} + c_1 + 2 \frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\|\boldsymbol{x} - \boldsymbol{\mu}\|} \cdot \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \\
&\geq \left(\frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} - c_1 \right) \left\{ \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} + c_1 + 2 \frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \right\} \\
&\geq 0.
\end{aligned}
$$
□

COROLLARY 4.1. *The estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the likelihood estimating function $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$.*

We can derive an estimating function superior to $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ under the condition (4.2) in Subsection 4.1.

PROPOSITION 4.4. *Under the condition (4.2), the estimating function*

$$(4.8) \qquad \tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(\|\boldsymbol{x} - \boldsymbol{\mu}\|)}{f(\|\boldsymbol{x} - \boldsymbol{\mu}\|)} \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\|\boldsymbol{x} - \boldsymbol{\mu}\|}$$

$$- \min\left\{ h_0 \|\boldsymbol{x} - \boldsymbol{m}\|, \ c_1, \ \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|}$$

is superior to the estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\boldsymbol{x})$, where $h_0 = \lim_{r\to +0}\{-f'(r)\}/\{rf(r)\}$.

PROOF.   Set $w(r) = \min\{c_1,\,(p-2)r^{-1}\}$ and $B = B(\boldsymbol{m};r_2)$ with r_2 being the unique solution to $h_0 r = w(r)$. It is shown that $\min\{h_0 r,\,c_1,\,(p-2)\,r^{-1}\} = h_0 r$ if $r < r_2$ and that $\min\{h_0 r,\,c_1,\,(p-2)\,r^{-1}\} = w(r)$ otherwise. Noting that $\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu};\boldsymbol{x}) = \tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\boldsymbol{x})$ for $\boldsymbol{x} \notin B$, we have

$$
\begin{aligned}
E\Big[\big\|\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\boldsymbol{x})\big\|^2\Big] &- E\Big[\big\|\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu};\boldsymbol{x})\big\|^2\Big] \\
&= \int_B \left\| -\frac{f'(\|\boldsymbol{x}-\boldsymbol{\mu}\|)}{f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)}\frac{\boldsymbol{x}-\boldsymbol{\mu}}{\|\boldsymbol{x}-\boldsymbol{\mu}\|} - w(\|\boldsymbol{x}-\boldsymbol{m}\|)\frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|}\right\|^2 f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)\,d\boldsymbol{x} \\
&\quad - \int_B \left\| -\frac{f'(\|\boldsymbol{x}-\boldsymbol{\mu}\|)}{f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)}\frac{\boldsymbol{x}-\boldsymbol{\mu}}{\|\boldsymbol{x}-\boldsymbol{\mu}\|} - h_0\,(\boldsymbol{x}-\boldsymbol{m})\right\|^2 f(\|\boldsymbol{x}-\boldsymbol{\mu}\|)\,d\boldsymbol{x}.
\end{aligned}
$$

The proof is completed by applying Lemma 2 in Appendix with $\boldsymbol{y} = \boldsymbol{x}-\boldsymbol{m}$, $\boldsymbol{a} = \boldsymbol{\mu}-\boldsymbol{m}$, $w_1(r) = -f'(r)\,/\,\{rf(r)\}$, $w_2(r) = f(r)$, and $w_3(r) = w(r)/r$.   □

Let $\check{\boldsymbol{\mu}}_S$ and $\check{\boldsymbol{\mu}}_S^+$ denote the two estimators induced from $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\boldsymbol{x})$ and $\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu};\boldsymbol{x})$, respectively. They are expressed as

$$
\check{\boldsymbol{\mu}}_S = \boldsymbol{x} - r_S\left(\min\left\{c_1\,,\,\frac{p-2}{\|\boldsymbol{x}-\boldsymbol{m}\|}\right\}\right)\frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|}
$$

and

$$
\check{\boldsymbol{\mu}}_S^+ = \boldsymbol{x} - r_S\left(\min\left\{h_0\|\boldsymbol{x}-\boldsymbol{m}\|,\,c_1\,,\,\frac{p-2}{\|\boldsymbol{x}-\boldsymbol{m}\|}\right\}\right)\frac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|}.
$$

Again, the equality (4.6) gives that

$$
\|\check{\boldsymbol{\mu}}_S^+ - \boldsymbol{m}\| = o\big(\|\boldsymbol{x}-\boldsymbol{m}\|\big), \qquad \text{as } \|\boldsymbol{x}-\boldsymbol{m}\| \to +0.
$$

Such a desirable property of $\check{\boldsymbol{\mu}}_S^+$ is consistent with Proposition 4.4.

Example 4.2.   Suppose that $\boldsymbol{x}$ has the density function (3.8) of a spherical t distribution. Note that (3.8) satisfies the condition (4.2). Setting $r_1 = 2\phi_0^{1/2}\,(p-2)/(p+\phi_0)$, we obtain an expression of $\check{\boldsymbol{\mu}}_S$ as

$$
\check{\boldsymbol{\mu}}_S = \begin{cases}
\boldsymbol{x} - \sqrt{\phi_0}\,\dfrac{\boldsymbol{x}-\boldsymbol{m}}{\|\boldsymbol{x}-\boldsymbol{m}\|} & \text{if } \boldsymbol{x} \in B(\boldsymbol{m};r_1), \\[2em]
\boldsymbol{x} - \sqrt{\phi_0}\left\{1 - \sqrt{1 - r_1^2\|\boldsymbol{x}-\boldsymbol{m}\|^{-2}}\right\}\dfrac{\boldsymbol{x}-\boldsymbol{m}}{r_1} & \text{otherwise.}
\end{cases}
$$

When $\phi_0 \leq 3p - 8$, we set $r_2 = \phi_0^{1/2}/2$ and obtain

$$
\check{\boldsymbol{\mu}}_S^+ = \begin{cases}
\boldsymbol{x} - 2r_2^2 \left\{ 1 - \sqrt{1 - r_2^{-2} \|\boldsymbol{x} - \boldsymbol{m}\|^2} \right\} \dfrac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} & \text{if } \|\boldsymbol{x} - \boldsymbol{m}\| < r_2, \\[3mm]
\boldsymbol{x} - 2r_2 \dfrac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|} & \text{if } r_2 \leq \|\boldsymbol{x} - \boldsymbol{m}\| < r_1, \\[3mm]
\boldsymbol{x} - \dfrac{2r_2}{r_1} \left\{ 1 - \sqrt{1 - r_1^2 \|\boldsymbol{x} - \boldsymbol{m}\|^{-2}} \right\} (\boldsymbol{x} - \boldsymbol{m}) & \text{otherwise.}
\end{cases}
$$

When $\phi_0 > 3p - 8$, we set $r_2 = \{\phi_0(p - 2)/(p + \phi_0)\}^{1/2}$ and obtain

$$
\check{\boldsymbol{\mu}}_S^+ = \begin{cases}
\boldsymbol{x} - \dfrac{\phi_0}{2} \left\{ 1 - \sqrt{1 - (4/\phi_0) \|\boldsymbol{x} - \boldsymbol{m}\|^2} \right\} \dfrac{\boldsymbol{x} - \boldsymbol{m}}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} & \text{if } \|\boldsymbol{x} - \boldsymbol{m}\| < r_2, \\[3mm]
\boldsymbol{x} - \dfrac{\phi_0}{2} \left\{ 1 - \sqrt{1 - (4r_2^4/\phi_0) \|\boldsymbol{x} - \boldsymbol{m}\|^{-2}} \right\} \dfrac{\boldsymbol{x} - \boldsymbol{m}}{r_2^2} & \text{otherwise.}
\end{cases}
$$

5. Mutually independent case

We proceed with the case where the density function of $\boldsymbol{x}$ is a member of the family

$$
\mathcal{I}_0 = \left\{ \prod_{i=1}^{p} f(x_i - \mu_i) \,\middle|\, \begin{array}{l} \text{the equation } f'(t) = 0 \text{ has a unique solution} \\ \text{and } f(t) \text{ attains its maximum at this point} \end{array} \right\}.
$$

Fortunately, arguments similar to those in the previous section can be employed also in this section. In fact, Theorem 3.2 will be again used to improve upon the likelihood estimating function. A difference lies in the fact that component-wise inequalities are obtained in this section.

The MLE is given as $\hat{\boldsymbol{\mu}}_M = \boldsymbol{x} - t_0 \mathbf{1}$ with t_0 being the unique solution to $f'(t) = 0$ and $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^p$. It is obtained from the likelihood estimating function $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$ in (3.5), whose i-th component has the form

$$
G_{M,i}(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)}.
$$

The i-th component of $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ in (3.6) is of the form

$$
G_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - \frac{p - 2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|}.
$$

Suppose that $\boldsymbol{x} \neq \boldsymbol{m}$. Then a necessary and sufficient condition for the existence of the solution to the equation $G_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) = 0$ is that the range of

$-f'(t)/f(t)$ is $\mathbb{R}$. We will focus on the two subfamilies of $\mathcal{I}_0$,

$$\mathcal{I}_1 = \left\{ \prod_{i=1}^{p} f(x_i - \mu_i) \in \mathcal{I}_0 \; \middle| \; \text{the range of } \frac{-f'(t)}{f(t)} \text{ is } \mathbb{R} \right\}$$

and

$$\mathcal{I}_2 = \left\{ \prod_{i=1}^{p} f(x_i - \mu_i) \in \mathcal{I}_0 \; \middle| \; \text{the range of } \frac{-f'(t)}{f(t)} \text{ is a bounded closed interval} \right\}.$$

In place of $r_S(c)$ in Section 4, we define $t_S(c)$ as the solution to the equation $f'(t)\,/\,f(t) + c = 0$ which is the closest to t_0.

5.1. The case of the subfamily $\mathcal{I}_1$

Theorem 3.2 gives the following proposition.

PROPOSITION 5.1.   *The estimating function $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the likelihood estimating function $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$.*

To improve upon $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$, we assume the following condition

(5.1) $f(t)$ is even, $\dfrac{-f'(t)}{tf(t)}$ is non-increasing on $\mathbb{R}^+$, and $\lim\limits_{t \to 0} \dfrac{-f'(t)}{tf(t)}$ exists,

which is the counterpart of the condition (4.2) in Section 4. This condition implies that $f'(0) = 0$ and therefore that $\hat{\boldsymbol{\mu}}_M = \boldsymbol{x}$.

PROPOSITION 5.2.   *Under the condition (5.1), define $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ as*

$$\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x}) = \left(G_{S,1}^+(\boldsymbol{\mu}; \boldsymbol{x}), \ldots, G_{S,p}^+(\boldsymbol{\mu}; \boldsymbol{x}) \right)^T,$$

$$G_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x}) = -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - \min\left\{ h_0 \|\boldsymbol{x} - \boldsymbol{m}\|, \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|},$$

where $h_0 = \lim_{t \to 0}\{-f'(t)\}\,/\,\{tf(t)\}$. Then the estimating function $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the estimating function $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$. Especially, it holds for each $i \in \{1, \ldots, p\}$ and for any $\boldsymbol{\mu} \in \mathbb{R}$ that

$$E\left[\left\{ G_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right] \leq E\left[\left\{ G_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right].$$

PROOF.   Set $B = B(\boldsymbol{m}; r_0)$ with $r_0 = \{(p-2)/h_0\}^{1/2}$. Since $G_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x})$ differs from $G_{S,i}(\boldsymbol{\mu}; \boldsymbol{x})$ only for $\boldsymbol{x} \in B$, we have

$$E\left[\left\{ G_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right] - E\left[\left\{ G_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right]$$

$$= \int_B \left| -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right|^2 \prod_{j=1}^{p} f(x_j - \mu_j)\, d\boldsymbol{x}$$

$$- \int_B \left| -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - h_0(x_i - m_i) \right|^2 \prod_{j=1}^{p} f(x_j - \mu_j)\, d\boldsymbol{x}.$$

Here note that $\{-f'(t)\} / \{tf(t)\}$ is even since $f(t)$ is even. In order to prove the latter part, we have only to apply Lemma 4 in Appendix with $y_i = x_i - m_i$, $a_i = \mu_i - m_i$, $w_1(s) = -f'(s) / \{sf(s)\}$, $w_2(s) = f(s)$, and $w_3(s, r) = (p-2) r^{-2}$. The former part follows from the latter. $\square$

The latter part of Proposition 5.2 shows that $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ dominates $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ component-wisely. This is a stronger and more general version of the result in Efron and Morris (1973, Section 2). We will obtain a similar proposition in Subsection 5.2.

Let $\hat{\boldsymbol{\mu}}_S$ and $\hat{\boldsymbol{\mu}}_S^+$ be induced from $\boldsymbol{G}_S(\boldsymbol{\mu}; \boldsymbol{x})$ and $\boldsymbol{G}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$, respectively. The i-th components of these estimators are expressed as

$$(5.2) \qquad \hat{\mu}_{S,i} = x_i - t_S \left(\frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right)$$

and

$$(5.3) \qquad \hat{\mu}_{S,i}^+ = x_i - t_S \left(\min\left\{ h_0 \|\boldsymbol{x} - \boldsymbol{m}\|, \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right).$$

The condition (5.1) and the definition of $t_S(\cdot)$ yield the equality

$$(5.4) \qquad \lim_{c \to 0} \frac{t_S(h_0 c)}{c} = 1,$$

which implies that

$$\left| \hat{\mu}_{S,i}^+ - m_i \right| = o\big(|x_i - m_i| \big), \qquad \text{as } \|\boldsymbol{x} - \boldsymbol{m}\| \to +0 \quad (1 \leq i \leq p).$$

Example 5.1. Suppose that $\boldsymbol{x}$ is distributed according to the generalized inverse Gaussian distribution having the density function

$$\prod_{i=1}^p \frac{1}{2\eta_i^{\lambda_0} K_{\lambda_0}(\tau_0)} x_i^{\lambda_0 - 1} \exp\left\{ -\frac{\tau_0}{2} \left(\frac{\eta_i}{x_i} + \frac{x_i}{\eta_i} \right) \right\},$$

where τ_0 and λ_0 are known positive constants and $K_\lambda(\cdot)$ is the modified Bessel function of the third kind. The density function for $y_i = \log x_i$ is given by $f(y_i - \mu_i)$, where $\mu_i = \log \eta_i$ and $f(y) = \exp(\lambda_0 y - \tau_0 \cosh y) / \{2 K_{\lambda_0}(\tau_0)\}$. The transformed density function belongs to $\mathcal{I}_1$. Theorem 3.2 can be applied to obtaining a Pythagorean relationship

$$E\left[\sum_{i=1}^p \left\{ \lambda_0 - \frac{\tau_0}{2} \left(\frac{x_i}{\eta_i} - \frac{\eta_i}{x_i} \right) \right\}^2 \right]$$

$$= E\left[\sum_{i=1}^p \left\{ \lambda_0 - \frac{\tau_0}{2} \left(\frac{x_i}{\eta_i} - \frac{\eta_i}{x_i} \right) + \frac{(p-2)(\log x_i - m_i)}{\sum_{j=1}^p (\log x_j - m_j)^2} \right\}^2 \right]$$

$$+ E\left[\frac{(p-2)^2}{\sum_{i=1}^p (\log x_i - m_i)^2} \right].$$

We obtain an expression of the estimator $\hat{\boldsymbol{\eta}}_S = (\hat{\eta}_{S,1}, \ldots, \hat{\eta}_{S,p})^T$ as

$$\hat{\eta}_{S,i} = \left\{ \sqrt{a_i^2(\boldsymbol{x},\, \boldsymbol{m}) + 1} - a_i(\boldsymbol{x},\, \boldsymbol{m}) \right\} x_i \qquad (1 \le i \le p),$$

where

$$a_i(\boldsymbol{x},\, \boldsymbol{m}) = \frac{1}{\tau_0} \left\{ \lambda_0 + \frac{(p-2)(\log x_i - m_i)}{\sum_{j=1}^p (\log x_j - m_j)^2} \right\}.$$

5.2. The case of the subfamily $\mathcal{I}_2$

Set $c_1 = \max\{-f'(t)\,/\,f(t)\}$ and $d_1 = \min\{-f'(t)\,/\,f(t)\}$. The equation $G_{S,i}(\boldsymbol{\mu};\, \boldsymbol{x}) = 0$ does not have any solution unless $d_1 \le (p-2)(x_i - m_i)/\|\boldsymbol{x} - \boldsymbol{m}\|^2 \le c_1$. In this subsection we will use the notation

$$\mathrm{T}_{[a,b]}\bigl(c(\boldsymbol{x})\bigr) = \begin{cases} a & \text{if } c(\boldsymbol{x}) < a, \\ b & \text{if } c(\boldsymbol{x}) > b, \\ c(\boldsymbol{x}) & \text{otherwise,} \end{cases}$$

where a and b are real numbers such that $a < b$ and $c(\boldsymbol{x})$ is a function on $\mathbb{R}^p$.

Define a new estimating function as

$$\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\, \boldsymbol{x}) = \left(\tilde{G}_{S,1}(\boldsymbol{\mu};\, \boldsymbol{x}),\, \ldots,\, \tilde{G}_{S,p}(\boldsymbol{\mu};\, \boldsymbol{x}) \right)^T,$$

$$\tilde{G}_{S,i}(\boldsymbol{\mu};\, \boldsymbol{x}) = -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - \mathrm{T}_{[d_1,c_1]}\left(\frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right).$$

Then we obtain the following proposition and its corollary.

PROPOSITION 5.3. *The estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\, \boldsymbol{x})$ satisfies the inequality*

$$\left\| \tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\, \boldsymbol{x}) \right\|^2 \le \left\| \boldsymbol{G}_S(\boldsymbol{\mu};\, \boldsymbol{x}) \right\|^2$$

for any $\boldsymbol{x} \in \mathbb{R}$ and $\boldsymbol{\mu} \in \mathbb{R}$. Especially, it holds for each $i \in \{1, \ldots, p\}$ and for any $\boldsymbol{x} \in \mathbb{R}$ and $\boldsymbol{\mu} \in \mathbb{R}$ that

$$\left\{ \tilde{G}_{S,i}(\boldsymbol{\mu};\, \boldsymbol{x}) \right\}^2 \le \left\{ G_{S,i}(\boldsymbol{\mu};\, \boldsymbol{x}) \right\}^2.$$

PROOF. Note that

$$\begin{aligned} G_{S,i}(\boldsymbol{\mu};\, \boldsymbol{x}) &= \left\{ -\frac{f'(x_i - \mu_i)}{f'(x_i - \mu_i)} - c_1 \right\} + \left\{ c_1 - \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \\ &= \left\{ -\frac{f'(x_i - \mu_i)}{f'(x_i - \mu_i)} - d_1 \right\} + \left\{ d_1 - \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\}. \end{aligned}$$

Hence we obtain the latter part, from which the former follows. $\qquad\square$

COROLLARY 5.1. *The estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the likelihood estimating function $\boldsymbol{G}_M(\boldsymbol{\mu}; \boldsymbol{x})$.*

A further improvement can be made on $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ under the condition (5.1) in Subsection 5.1. Since $d_1 = -c_1$, it follows that

$$
\begin{aligned}
\tilde{G}_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) &= -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - \mathrm{T}_{[-c_1, c_1]}\left(\frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right) \\
&= -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - \min\left\{ \frac{c_1 \|\boldsymbol{x} - \boldsymbol{m}\|}{|x_i - m_i|}, \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|}.
\end{aligned}
$$

PROPOSITION 5.4. *Under the condition (5.1), define $\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ as*

$$
\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu}; \boldsymbol{x}) = \left(\tilde{G}_{S,1}^+(\boldsymbol{\mu}; \boldsymbol{x}), \ldots, \tilde{G}_{S,p}^+(\boldsymbol{\mu}; \boldsymbol{x}) \right)^T,
$$

$$
\begin{aligned}
\tilde{G}_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x}) = &-\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} \\
&- \min\left\{ h_0 \|\boldsymbol{x} - \boldsymbol{m}\|, \frac{c_1 \|\boldsymbol{x} - \boldsymbol{m}\|}{|x_i - m_i|}, \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|},
\end{aligned}
$$

where $h_0 = \lim_{t \to 0}\{-f'(t)\} / \{t f(t)\}$. Then the estimating function $\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$ is superior to the estimating function $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$. Especially, it holds for each $i \in \{1, \ldots, p\}$ and for any $\boldsymbol{\mu} \in \mathbb{R}$ that

$$
E\left[\left\{ \tilde{G}_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right] \leq E\left[\left\{ \tilde{G}_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right].
$$

PROOF. Setting $w(s, r) = \min\{c_1 r / s, (p-2)/r\}$ and $B_i = \{\boldsymbol{x} \in \mathbb{R}^p \mid h_0 \|\boldsymbol{x} - \boldsymbol{m}\| < w(|x_i - m_i|, \|\boldsymbol{x} - \boldsymbol{m}\|)\}$, we see that

$$
\begin{aligned}
&E\left[\left\{ \tilde{G}_{S,i}(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right] - E\left[\left\{ \tilde{G}_{S,i}^+(\boldsymbol{\mu}; \boldsymbol{x}) \right\}^2 \right] \\
&= \int_{B_i} \left| -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - w(|x_i - m_i|, \|\boldsymbol{x} - \boldsymbol{m}\|) \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right|^2 \prod_{j=1}^{p} f(x_j - \mu_j)\, d\boldsymbol{x} \\
&\quad - \int_{B_i} \left| -\frac{f'(x_i - \mu_i)}{f(x_i - \mu_i)} - h_0(x_i - m_i) \right|^2 \prod_{j=1}^{p} f(x_j - \mu_j)\, d\boldsymbol{x}.
\end{aligned}
$$

The latter part is obtained by applying Lemma 4 in Appendix with $y_i = x_i - m_i$, $a_i = \mu_i - m_i$, $w_1(s) = -f'(s) / \{s f(s)\}$, $w_2(s) = f(s)$, and $w_3(s, r) = w(s, r)/r$. The former part follows from the latter. $\square$

Let $\check{\boldsymbol{\mu}}_S$ and $\check{\boldsymbol{\mu}}_S^+$ denote the estimators induced from $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu}; \boldsymbol{x})$ and $\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu}; \boldsymbol{x})$, respectively. The i-th components of these estimators are of the form

$$\check{\mu}_{S,i} = x_i - t_S \left(\min \left\{ \frac{c_1 \|\boldsymbol{x} - \boldsymbol{m}\|}{|x_i - m_i|}, \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right)$$

and

$$\check{\mu}_{S,i}^+ = x_i - t_S \left(\min \left\{ h_0 \|\boldsymbol{x} - \boldsymbol{m}\|, \frac{c_1 \|\boldsymbol{x} - \boldsymbol{m}\|}{|x_i - m_i|}, \frac{p-2}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right\} \frac{x_i - m_i}{\|\boldsymbol{x} - \boldsymbol{m}\|} \right).$$

Just as in Subsection 5.1, the equality (5.4) gives an intuitively good property of $\check{\boldsymbol{\mu}}_S^+$:

$$\left| \check{\mu}_{S,i}^+ - m_i \right| = o\big(|x_i - m_i|\big), \qquad \text{as } \|\boldsymbol{x} - \boldsymbol{m}\| \to +0 \quad (1 \le i \le p).$$

Example 5.2.   Consider the case where $\boldsymbol{x}$ is distributed according to a mutually independent t distribution having the density function

$$\prod_{i=1}^p \frac{1}{\sqrt{\phi_0} B(\phi_0/2, \ 1/2)} \left\{ 1 + \frac{(x_i - \mu_i)^2}{\phi_0} \right\}^{-(\phi_0+1)/2},$$

where ϕ_0 is a known positive constant. Note that this density function satisfies the condition (5.1). The density function with small ϕ_0 is useful in the field of the financial engineering (Nagahara, 1999). Especially when $\phi_0 = 1$, it is the density function of a mutually independent Cauchy distribution. The solution t_0 to $f'(t) = 0$ is zero, and $f'(t)/f(t) + c = 0$ has two solutions for $|c| \le (\phi_0+1)/2\phi_0^{1/2}$. The one closer to $t_0 = 0$ is given as

$$t_S(c) = \frac{2\phi_0 c}{\phi_0 + 1 + \sqrt{(\phi_0 + 1)^2 - 4\phi_0 c^2}}.$$

The i-th component of $\check{\boldsymbol{\mu}}_S$ is expressed as

$$\check{\mu}_{S,i} = \begin{cases} x_i - \sqrt{\phi_0} & \text{if } \dfrac{(p-2)(x_i - m_i)}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} > \dfrac{\phi_0 + 1}{2\sqrt{\phi_0}}, \\[2ex] x_i + \sqrt{\phi_0} & \text{if } \dfrac{(p-2)(x_i - m_i)}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} < -\dfrac{\phi_0 + 1}{2\sqrt{\phi_0}}, \\[2ex] x_i - t_S \left(\dfrac{(p-2)(x_i - m_i)}{\|\boldsymbol{x} - \boldsymbol{m}\|^2} \right) & \text{otherwise.} \end{cases}$$

The i-th component of $\check{\boldsymbol{\mu}}_S^+$ is given as

$$
\check{\mu}_{S,i}^{+} =
\begin{cases}
x_i - t_S\left(\dfrac{(\phi_0+1)(x_i-m_i)}{\phi_0}\right) \\
\qquad \text{if } |x_i - m_i| < \dfrac{\sqrt{\phi_0}}{2} \text{ and } \|\boldsymbol{x}-\boldsymbol{m}\|^2 < \dfrac{(p-2)\phi_0}{\phi_0+1}, \\
x_i - \sqrt{\phi_0} \\
\qquad \text{if } x_i - m_i \geq \dfrac{\sqrt{\phi_0}}{2} \text{ and } \dfrac{(p-2)(x_i-m_i)}{\|\boldsymbol{x}-\boldsymbol{m}\|^2} > \dfrac{\phi_0+1}{2\sqrt{\phi_0}}, \\
x_i + \sqrt{\phi_0} \\
\qquad \text{if } x_i - m_i \leq -\dfrac{\sqrt{\phi_0}}{2} \text{ and } \dfrac{(p-2)(x_i-m_i)}{\|\boldsymbol{x}-\boldsymbol{m}\|^2} < -\dfrac{\phi_0+1}{2\sqrt{\phi_0}}, \\
x_i - t_S\left(\dfrac{(p-2)(x_i-m_i)}{\|\boldsymbol{x}-\boldsymbol{m}\|^2}\right) \\
\qquad \text{otherwise.}
\end{cases}
$$

6. Electrostatic views

In this section we are to throw light on the arguments in Sections 3 and 4 from the viewpoint of electrostatics. Electrostatics is the theory of electricity without time-dependency. In this theory, the dimension p is three, and things are described by the Poisson differential equation

(6.1) $$\triangle u(\boldsymbol{x}) = -4\pi\rho(\boldsymbol{x}),$$

where $u(\boldsymbol{x})$ is an electrostatic potential function and $\rho(\boldsymbol{x})$ is the corresponding electric charge density, the amount of electric charge per unit volume. See Purcell (1985, Section 2.10) for details. A naturally extended version of the equation (6.1) to the p-dimensional case is

(6.2) $$\triangle u(\boldsymbol{x}) = -(p-2)\omega_p\rho(\boldsymbol{x})$$

with ω_p being the surface area of the p-dimensional unit ball.

Let $\delta_p(\cdot)$ denote the p-dimensional Dirac δ-function. It should be noted that

$$\triangle u_C(\boldsymbol{x}-\boldsymbol{m}) = -(p-2)\,\omega_p\,\delta_p(\boldsymbol{x}-\boldsymbol{m}).$$

Since $\delta_p(\boldsymbol{x}-\boldsymbol{m})$ represents the electric charge density of a point-charge of magnitude one and located at $\boldsymbol{m}$, the function $u_C(\boldsymbol{x}-\boldsymbol{m})$ is interpreted as the electrostatic potential function produced by that point-charge.

Theorems 3.1 and 3.2 use the electrostatic potential function which one point-charge produces. These are generalized in the following way. Suppose that k positive point-charges of magnitude $\alpha_1, \ldots, \alpha_k$ are located at k different points $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_k$, respectively. It can be easily verified from (6.2) that these point-charges produce the electrostatic potential function $\sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x}-\boldsymbol{m}_i)$. This fact is well known as the *principle of superposition* in electrostatics. This electrostatic potential function is used in the following two corollaries. Note that

these corollaries are closely related to the notion of multiple shrinkage estimation proposed by George (1986).

COROLLARY 6.1.   *In addition to the condition* (3.1), *assume that*

$$(6.3) \qquad \int_{\mathbb{R}^p} \left\| \nabla \log \sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i) \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x} < \infty$$

and that

$$(6.4) \qquad \lim_{K \to \infty} \int_{\|\boldsymbol{x}\| = K} g(\boldsymbol{x}) \left\{ \nabla \log \sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i) \right\} \cdot \boldsymbol{n} dS = 0.$$

Then the following Pythagorean relationship holds :

$$\int_{\mathbb{R}^p} \left\| \nabla \log g(\boldsymbol{x}) \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x} = \int_{\mathbb{R}^p} \left\| \nabla \log \frac{g(\boldsymbol{x})}{\sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i)} \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x}$$
$$+ \int_{\mathbb{R}^p} \left\| \nabla \log \sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i) \right\|^2 g(\boldsymbol{x}) \, d\boldsymbol{x}.$$

PROOF.   The proof is similar to that of Theorem 3.1. Let ε be a positive number less than $\min_{i \neq j} \|\boldsymbol{m}_i - \boldsymbol{m}_j\|/2$. And let K be a positive number larger than $\max_i \|\boldsymbol{m}_i\| + \varepsilon$. We apply Green's formula to the following integral

$$\int_V \nabla \frac{g(\boldsymbol{x})}{\sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i)} \cdot \nabla \sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i) \, d\boldsymbol{x},$$

where $V = B(\boldsymbol{0}; K) \backslash \bigcup_{j=1}^{k} \overline{B(\boldsymbol{m}_j; \varepsilon)}$. The proof is completed by considering the limit $K \to \infty$ and $\varepsilon \to +0$. $\square$

COROLLARY 6.2.   *Assume that the Fisher information matrix of* $f(\boldsymbol{x} - \boldsymbol{\mu})$ *exists. Then the following Pythagorean relationship holds* :

$$E \left[\left\| \nabla \log f(\boldsymbol{x} - \boldsymbol{\mu}) \right\|^2 \right] = E \left[\left\| \nabla \log \frac{f(\boldsymbol{x} - \boldsymbol{\mu})}{\sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i)} \right\|^2 \right]$$
$$+ E \left[\left\| \nabla \log \sum_{i=1}^{k} \alpha_i u_C(\boldsymbol{x} - \boldsymbol{m}_i) \right\|^2 \right].$$

PROOF.   Suffice it to say that the three conditions (3.1), (6.3) and (6.4) with $g(\boldsymbol{x}) = f(\boldsymbol{x} - \boldsymbol{\mu})$ are satisfied. The proof is parallel to that of Theorem 3.2 and is therefore omitted. $\square$

Next, we will give another representation of each of the three estimating functions in Section 4, $\boldsymbol{G}_S^+(\boldsymbol{\mu};\,\boldsymbol{x})$, $\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\,\boldsymbol{x})$ and $\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu};\,\boldsymbol{x})$. Each representation will enable us to interpret each of the estimating functions from the viewpoint of an electrostatic potential function and the corresponding electric charge density. We introduce three electrostatic potential functions as

$$(6.5) \qquad u_C^+(\boldsymbol{x}) = \begin{cases} r_0^{2-p}\exp\{h_0\left(r_0^2 - \|\boldsymbol{x}\|^2\right)/2\} & \text{if } \|\boldsymbol{x}\| < r_0, \\ \|\boldsymbol{x}\|^{2-p} & \text{otherwise,} \end{cases}$$

$$(6.6) \qquad \tilde{u}_C(\boldsymbol{x}) = \begin{cases} r_1^{2-p}\exp\{c_1\left(r_1 - \|\boldsymbol{x}\|\right)\} & \text{if } \|\boldsymbol{x}\| < r_1, \\ \|\boldsymbol{x}\|^{2-p} & \text{otherwise,} \end{cases}$$

and

$$(6.7)$$
$$\tilde{u}_C^+(\boldsymbol{x}) = \begin{cases} r_1^{2-p}\exp\{c_1\left(r_1 - r_2\right)\}\exp\{h_0\left(r_2^2 - \|\boldsymbol{x}\|^2\right)/2\} & \text{if } \|\boldsymbol{x}\| < r_2, \\ r_1^{2-p}\exp\{c_1\left(r_1 - \|\boldsymbol{x}\|\right)\} & \text{if } r_2 \leq \|\boldsymbol{x}\| < r_1, \\ \|\boldsymbol{x}\|^{2-p} & \text{otherwise,} \end{cases}$$

where $h_0 = \lim_{r\to+0}\{-f'(r)/rf(r)\}$, $r_0 = \{(p-2)/h_0\}^{1/2}$, $c_1 = \max\{-f'(r)/f(r)\}$, $r_1 = (p-2)/c_1$ and $r_2 = c_1/h_0$ as defined in Section 4. For simplicity, we assume that $(p-2)h_0 > c_1^2$ in (6.7) so that $r_2 < r_1$. The graphs of $u_C(\boldsymbol{x})$ and $u_C^+(\boldsymbol{x})$ as functions of $r = \|\boldsymbol{x}\|$ are presented in Figure 1, showing that $u_C^+(r)$ is much smaller than $u_C(r)$ for small r, say, $r < 0.3$. Those of $u_C(\boldsymbol{x})$, $\tilde{u}_C(\boldsymbol{x})$ and $\tilde{u}_C^+(\boldsymbol{x})$ as functions of $r = \|\boldsymbol{x}\|$ are drawn in Figure 2. We observe the inequalities among the three potential functions, $\tilde{u}_C^+(r) \leq \tilde{u}_C(r) \leq u_C(r)$. In addition the strict inequalities $\tilde{u}_C^+(r) < \tilde{u}_C(r) < u_C(r)$ hold for small values of r. By using these electrostatic potential functions, we find the following expressions:

$$\boldsymbol{G}_S^+(\boldsymbol{\mu};\,\boldsymbol{x}) = -\nabla\log f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) + \nabla\log u_C^+(\boldsymbol{x} - \boldsymbol{m}),$$
$$\tilde{\boldsymbol{G}}_S(\boldsymbol{\mu};\,\boldsymbol{x}) = -\nabla\log f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) + \nabla\log \tilde{u}_C(\boldsymbol{x} - \boldsymbol{m}),$$
$$\tilde{\boldsymbol{G}}_S^+(\boldsymbol{\mu};\,\boldsymbol{x}) = -\nabla\log f(\|\boldsymbol{x} - \boldsymbol{\mu}\|) + \nabla\log \tilde{u}_C^+(\boldsymbol{x} - \boldsymbol{m}).$$

Recalling the equation (6.2), we obtain the electric charge densities corresponding to the electrostatic potential functions (6.5)–(6.7) as

$$\rho_C^+(\boldsymbol{x}) = \begin{cases} \dfrac{p-2}{\omega_p\,r_0^2}\left(\dfrac{p}{p-2} - \dfrac{\|\boldsymbol{x}\|^2}{r_0^2}\right)u_C^+(\boldsymbol{x}) & \text{if } \|\boldsymbol{x}\| < r_0, \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{\rho}_C(\boldsymbol{x}) = \begin{cases} \dfrac{p-2}{\omega_p\,r_1^2}\left(\dfrac{p-1}{p-2}\dfrac{r_1}{\|\boldsymbol{x}\|} - 1\right)\tilde{u}_C(\boldsymbol{x}) & \text{if } \|\boldsymbol{x}\| < r_1, \\ 0 & \text{otherwise,} \end{cases}$$
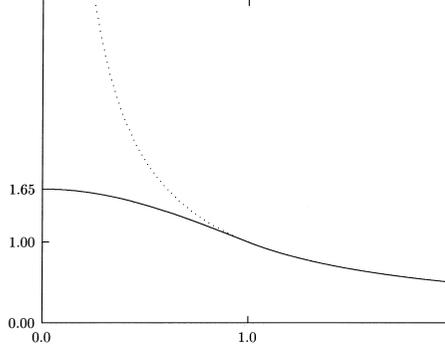
Figure 1. The functions $u_C(r)$ in the dotted line and $u_C^+(r)$ in the solid line for $p = 3$ and $h_0 = 1$.
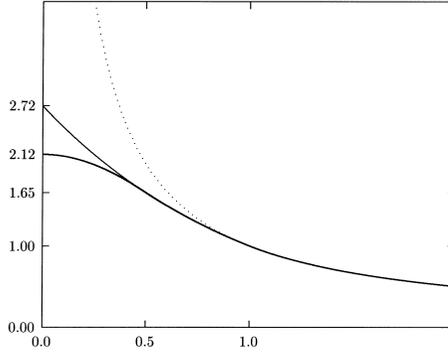
Figure 2. The functions $u_C(r)$ in the dotted line, $\tilde{u}_C(r)$ in the solid thin line and $\tilde{u}_C^+(r)$ in the solid thick line for $p = 3$, $c_1 = 1$ and $h_0 = 2$.

and

$$
\tilde{\rho}_C^+(\boldsymbol{x}) = \begin{cases}
\dfrac{p-2}{\omega_p\, r_1 r_2}\left(\dfrac{p}{p-2} - \dfrac{\|\boldsymbol{x}\|^2}{r_1 r_2}\right)\tilde{u}_C^+(\boldsymbol{x}) & \text{if } \|\boldsymbol{x}\| < r_2, \\[2ex]
\dfrac{p-2}{\omega_p\, r_1^2}\left(\dfrac{p-1}{p-2}\,\dfrac{r_1}{\|\boldsymbol{x}\|} - 1\right)\tilde{u}_C^+(\boldsymbol{x}) & \text{if } r_2 \le \|\boldsymbol{x}\| < r_1, \\[2ex]
0 & \text{otherwise,}
\end{cases}
$$

respectively. Note that the above functions are all non-negative. The behavior of $\rho_C^+(\boldsymbol{x})$ as a function of $r = \|\boldsymbol{x}\|$ is illustrated in Figure 3, indicating that $\rho_C^+(r)$ is decreasing and has a point of discontinuity. Those of $\tilde{\rho}_C(\boldsymbol{x})$ and $\tilde{\rho}_C^+(\boldsymbol{x})$ as functions of $r = \|\boldsymbol{x}\|$ are graphed in Figure 4. We learn that $\tilde{\rho}_C^+(r)$ is bounded to the above while the value of $\tilde{\rho}_C(r)$ becomes large as r approaches zero. We also find that the two functions $\tilde{\rho}_C^+(r)$ and $\tilde{\rho}_C(r)$ cross at some r less than 0.5.

Let us compute the total electric charge in each case. Integration by parts gives that

$$
\int_{\mathbb{R}^p} \rho_C^+(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}^p} \tilde{\rho}_C(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}^p} \tilde{\rho}_C^+(\boldsymbol{x})\, d\boldsymbol{x} = 1.
$$
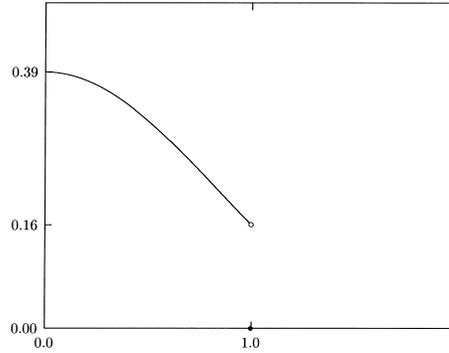
Figure 3. The function $\rho_C^+(r)$ for the same parameter setting as in Figure 1.
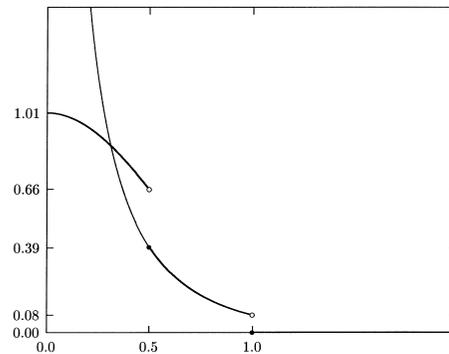
Figure 4. The functions $\tilde{\rho}_C(r)$ in the thin line and $\tilde{\rho}_C^+(r)$ in the thick line for the same parameter setting as in Figure 2.

This shows that in each case the total electric charge is the same as that due to $\delta_p(\boldsymbol{x})$. The electric charge density $\delta_p(\boldsymbol{x})$ is concentrated in the origin. In comparison with $\delta_p(\boldsymbol{x})$, the electric charge density $\rho_C^+(\boldsymbol{x})$ represents that of an electrically charged ball with center $\boldsymbol{0}$ and radius r_0. Similarly, $\tilde{\rho}_C(\boldsymbol{x})$ and $\tilde{\rho}_C^+(\boldsymbol{x})$ are interpreted as those of electrically charged balls with center $\boldsymbol{0}$ and radius r_1. The difference between $\tilde{\rho}_C(\boldsymbol{x})$ and $\tilde{\rho}_C^+(\boldsymbol{x})$ is restricted within the ball with center $\boldsymbol{0}$ and radius r_2.

Appendix

Four technical lemmas are presented here. Lemma 1 is used to prove Lemma 2, which is essential to Propositions 4.2 and 4.4 in Section 4. Lemma 3 gives Lemma 4, which is applied to proving Propositions 5.2 and 5.4 in Section 5.

LEMMA 1.   *Suppose that $w_1(r)$ $(r \geq 0)$ is non-increasing and that $w_2(r)$ $(r \geq 0)$ satisfies the condition that $B = \left\{ \boldsymbol{y} \in \mathbb{R}^p \,|\, w_2(\|\boldsymbol{y}\|) > 0 \right\}$ is not empty. Then,*

it holds for any $\boldsymbol{a} \in \mathbb{R}^p$ that

$$\int_B \boldsymbol{a} \cdot \boldsymbol{y} \, w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \, w_2(\|\boldsymbol{y}\|) \, d\boldsymbol{y} \geq 0,$$

when the left-hand side exists.

PROOF.   Let $\boldsymbol{y}$ be an arbitrary point in B. We have only to consider $\boldsymbol{y} \neq \boldsymbol{0}$ since the integrand takes the value zero for $\boldsymbol{y} = \boldsymbol{0}$. The point $-\boldsymbol{y}$ is also in B because of the spherical symmetry of B. Note that $\|\boldsymbol{y} - \boldsymbol{a}\| \leq \|\boldsymbol{y} + \boldsymbol{a}\|$ if $\boldsymbol{a} \cdot \boldsymbol{y} \geq 0$ and that $\|\boldsymbol{y} - \boldsymbol{a}\| > \|\boldsymbol{y} + \boldsymbol{a}\|$ otherwise. By evaluating the integrand for $\boldsymbol{y}$ and $-\boldsymbol{y}$, we find that

$$\begin{aligned}
\boldsymbol{a} \cdot \boldsymbol{y} \, & w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \, w_2(\|\boldsymbol{y}\|) + \boldsymbol{a} \cdot (-\boldsymbol{y}) \, w_1(\| - \boldsymbol{y} - \boldsymbol{a}\|) \, w_2(\| - \boldsymbol{y}\|) \\
&= \boldsymbol{a} \cdot \boldsymbol{y} \, \{w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) - w_1(\|\boldsymbol{y} + \boldsymbol{a}\|)\} \, w_2(\|\boldsymbol{y}\|) \\
&\geq 0. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square
\end{aligned}$$

LEMMA 2.   *Suppose that $w_1(r)$ $(r \geq 0)$ and $w_2(r)$ $(r \geq 0)$ are both non-increasing and also that $w_2(r)$ is non-negative. Suppose further that $w_3(r)$ $(r \geq 0)$ satisfies the condition that $B = \left\{ \boldsymbol{y} \in \mathbb{R}^p \,|\, w_3(\|\boldsymbol{y}\|) > w_1(0) \right\}$ is not empty. Then, it holds for any $\boldsymbol{a} \in \mathbb{R}^p$ that*

$$\int_B \left\| w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \, (\boldsymbol{y} - \boldsymbol{a}) - w_3(\|\boldsymbol{y}\|) \, \boldsymbol{y} \right\|^2 w_2(\|\boldsymbol{y} - \boldsymbol{a}\|) \, d\boldsymbol{y}$$

$$\geq \int_B \left\| w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \, (\boldsymbol{y} - \boldsymbol{a}) - w_1(0) \, \boldsymbol{y} \right\|^2 w_2(\|\boldsymbol{y} - \boldsymbol{a}\|) \, d\boldsymbol{y},$$

when both sides exist.

PROOF.   Note that

$$\begin{aligned}
\left\| w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \right. & \left. (\boldsymbol{y} - \boldsymbol{a}) - w_3(\|\boldsymbol{y}\|) \, \boldsymbol{y} \right\|^2 - \left\| w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \, (\boldsymbol{y} - \boldsymbol{a}) - w_1(0) \, \boldsymbol{y} \right\|^2 \\
&= \left\{ w_3(\|\boldsymbol{y}\|) + w_1(0) - 2w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \right\} \left\{ w_3(\|\boldsymbol{y}\|) - w_1(0) \right\} \|\boldsymbol{y}\|^2 \\
&\quad + 2\boldsymbol{a} \cdot \boldsymbol{y} \, w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \left\{ w_3(\|\boldsymbol{y}\|) - w_1(0) \right\}.
\end{aligned}$$

It is found for $\boldsymbol{y} \in B$ that

$$\left\{ w_3(\|\boldsymbol{y}\|) + w_1(0) - 2w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \right\} \left\{ w_3(\|\boldsymbol{y}\|) - w_1(0) \right\} \|\boldsymbol{y}\|^2 \, w_2(\|\boldsymbol{y} - \boldsymbol{a}\|) \geq 0.$$

By replacing $w_1(r)$ with $w_1(r) \, w_2(r)$ and $w_2(r)$ with $w_3(r) - w_1(0)$ in Lemma 1, we have

$$\int_B \boldsymbol{a} \cdot \boldsymbol{y} \, w_1(\|\boldsymbol{y} - \boldsymbol{a}\|) \, w_2(\|\boldsymbol{y} - \boldsymbol{a}\|) \left\{ w_3(\|\boldsymbol{y}\|) - w_1(0) \right\} d\boldsymbol{y} \geq 0. \qquad \square$$

LEMMA 3. *Suppose that $w_1(s)$ $(s \geq 0)$ and $w_2(s)$ $(s \geq 0)$ are both non-increasing and also that $w_2(s)$ is non-negative. Suppose further that $w_3(s, r)$ $(s \geq 0, r \geq 0)$ satisfies the condition that $B_i = \{ \boldsymbol{y} \in \mathbb{R}^p \,|\, w_3(|y_i|, \|\boldsymbol{y}\|) > 0 \}$ is not empty. Then, it holds for each $i \in \{1, \ldots, p\}$ and for any $\boldsymbol{a} = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$ that*

$$\int_{B_i} a_i y_i \, w_1(|y_i - a_i|) \, w_3(|y_i|, \|\boldsymbol{y}\|) \prod_{j=1}^{p} w_2(|y_j - a_j|) \, d\boldsymbol{y} \geq 0,$$

when the left-hand side exists.

PROOF. The proof is similar to that of Lemma 1. We have only to evaluate the integrand for a pair of points, $\boldsymbol{y} = (y_1, \ldots, y_i, \ldots, y_p)^T$ and $\tilde{\boldsymbol{y}} = (y_1, \ldots, y_{i-1}, -y_i, y_{i+1}, \ldots, y_p)^T$. $\square$

LEMMA 4. *Suppose that $w_1(s)$ $(s \geq 0)$ and $w_2(s)$ $(s \geq 0)$ are both non-increasing and also that $w_2(s)$ is non-negative. Suppose further that $w_3(s, r)$ $(s \geq 0, r \geq 0)$ satisfies the condition that $B_i = \{ \boldsymbol{y} \in \mathbb{R}^p \,|\, w_3(|y_i|, \|\boldsymbol{y}\|) > w_1(0) \}$ is not empty. Then, it holds for each $i \in \{1, \ldots, p\}$ and for any $\boldsymbol{a} = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$ that*

$$\int_{B_i} \Big\{ w_1(|y_i - a_i|) \, (y_i - a_i) - w_3(|y_i|, \|\boldsymbol{y}\|) \, y_i \Big\}^2 \prod_{j=1}^{p} w_2(|y_j - a_j|) \, d\boldsymbol{y}$$

$$\geq \int_{B_i} \Big\{ w_1(|y_i - a_i|) \, (y_i - a_i) - w_1(0) \, y_i \Big\}^2 \prod_{j=1}^{p} w_2(|y_j - a_j|) \, d\boldsymbol{y},$$

when both sides exist.

PROOF. The proof is similar to that of Lemma 2. We use Lemma 3 instead of Lemma 1. $\square$

Acknowledgements

REFERENCES

Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities, *Ann. Statist.*, **3**, 1318–1328.

Berger, J. O. and Bock, M. E. (1976). Eliminating singularities of Stein-type estimators of location vectors, *J. Roy. Statist. Soc. B*, **38**, 166–170.

Brandwein, A. C. and Strawderman, W. E. (1991). Generalization of James-Stein Estimators under spherical symmetry, *Ann. Statist.*, **19**, 1639–1650.

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters, *Ann. Math. Statist.*, **37**, 1087–1136.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - An empirical Bayes approach, *Jour. Amer. Statist. Assoc.*, **68**, 117–130.

George, E. I. (1986). Minimax multiple shrinkage estimation, *Ann. Statist.*, **14**, 188–205.

Godambe, V. P. and Kale, B. K. (1991). Estimating functions: an overview, *Estimating Functions*, (ed. Godambe, V. P.), 3–20, Clarendon Press, Oxford.

Hudson, H. M. (1978). A natural identity for exponential families with applications in multi-parameter estimation, *Ann. Statist.*, **6**, 473–484.

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, **1**, 361–380.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, Springer-Verlag, New York.

Liang, K.-Y. and Waclawiw, M. A. (1990). Extension of the Stein estimating procedure through the use of estimating functions, *Jour. Amer. Statist. Assoc.*, **85**, 435–440.

Nagahara, Y. (1999). The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters, *Statist. Prob. Lett.*, **43**, 251–264.

Purcell, E. M. (1985). *Electricity and Magnetism*, McGraw-Hill, New York.

Robert, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Motivations to Computational*, Springer-Verlag, New York.

Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss, *Ann. Statist.*, **12**, 322–335.

Small, C. G., Wang, J. and Yang, Z. (2000). Eliminating multiple root problem in Estimation, *Statist. Science*, **15**, 313–341.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution, *Ann. Statist.*, **9**, 1135–1151.

Yanagimoto, T. (1994). The Kullback-Leibler risk of the Stein estimator and the conditional MLE, *Ann. Inst. Statist. Math.*, **46**, 29–41.

Yanagimoto, T. (2000). A pair of estimating equations for a mean vector, *Statist. Prob. Lett.*, **50**, 97–103.

Yanagimoto, T. and Yamamoto, E. (1993). A criterion of sensitivity of an estimating function, *Commun. Statist. -Theory. Meth.*, **22**(2), 451–460.