stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 60/78 OKTOBER
(DEPARTMENT OF MATHEMATICAL STATISTICS)

C.A.J. KLAASSEN

NONUNIFORMITY OF THE CONVERGENCE OF LOCATION ESTIMATORS

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): primary 62E15, secondary 62F10, 62G05

Nonuniformity of the convergence of location estimators*)

by

C.A.J. Klaassen

SUMMARY

The pointwise supremum is considered of the distribution functions of properly normed translation equivariant antisymmetric location estimators based on observations from a distribution with a symmetric density. For positive values of the argument this function converges to the standard normal distribution function as the number of observations increases. Sets of symmetric densities are given over which this convergence - and consequently the convergence of adaptive location estimators - is not uniform.

KEY WORDS & PHRASES: *distribution functions of location estimators, nonuniformity.*

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

We consider the set D of density functions f with respect to Lebesgue measure on $\mathbb{R}$, which are symmetric about zero and absolutely continuous and which have finite Fisher information $I(f) = \int (f'/f)^2 f$. Notice that for all $f \in D$ $I(f)$ is positive.
$X_1, \ldots, X_n$ are independent and identically distributed random variables with density function $f(x-\theta)$ for some unknown $\theta \in \mathbb{R}$ and known $f \in D$.
Let $t_n: \mathbb{R}^n \to \mathbb{R}$ be Borel measurable and $T_n = t_n(X_1, \ldots, X_n)$ be an estimator based on $X_1, \ldots, X_n$ of the location parameter θ. If for all real a and Lebesgue almost all $x_1, \ldots, x_n$ $t_n(x_1+a, \ldots, x_n+a) = t_n(x_1, \ldots, x_n) + a$ then T_n is called translation equivariant. If $t_n(-x_1, \ldots, -x_n) = -t_n(x_1, \ldots, x_n)$ for Lebesgue almost all $x_1, \ldots, x_n$ then we shall call T_n antisymmetric. We denote by $\mathcal{T}_n$ the class of translation equivariant antisymmetric estimators of location T_n.
Let Φ be the standard normal distribution function. For all $f \in D$ there are known estimators $T_n \in \mathcal{T}_n$, $n = 1, 2, \ldots$, for which for all $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$

$$\lim_{n\to\infty} P_{f(.-\theta)}((nI(f))^{\frac{1}{2}}(T_n-\theta) \leq x) = \Phi(x)$$

holds or equivalently for which for all $x > 0$

$$(1.1) \qquad \lim_{n\to\infty} P_f((nI(f))^{\frac{1}{2}} T_n \leq x) = \Phi(x)$$

holds. It can even be shown that for all $f \in D$ and all $x > 0$

$$\lim_{n\to\infty} \sup_{T_n \in \mathcal{T}_n} P_f((nI(f))^{\frac{1}{2}} T_n \le x) = \Phi(x). \tag{1.2}$$

We shall be concerned with the question whether there exist subsets D_0 of D over which the convergence on (1.2) can not be uniform. It turns out that such subsets exist. Intuitively this is clear from the following reasoning. Let the sample size n be fixed. There exist densities f and g belonging to D for which $I(f)/I(g)$ is arbitrarily large, while at the same time the joint densities of a sample $X_1, \ldots, X_n$ under f respectively g are almost indistinguishable. As a consequence of the latter phenomenon it is impossible to estimate the location parameter considerably better under f than under g on the basis of n observations. Because $I(f)$ is much greater than $I(g)$ this implies that for $x > 0$ the supremum in formula (1.2) is much smaller than the same supremum with f replaced by g. From this it is clear that the convergence in (1.2) can not be uniform over D.
A precise formulation of this statement is given in section 2. The assertions of section 2 are proved in sections 3 and 4. Section 5 is an appendix containing a proof of formula (1.2).

2. THE MAIN RESULTS

In this section a theorem and a lemma will be stated which imply the assertions of the preceding section. We define the total variation distance $d_n(f,g)$ of f and g by

$$d_n(f,g) = \tfrac{1}{2} \int \cdots \int_{\mathbb{R}^n} \left| \prod_{i=1}^n f(x_i) - \prod_{i=1}^n g(x_i) \right| dx_1 \ldots dx_n$$

and we denote $d_1(f,g)$ by $d(f,g)$. Furthermore we shall say that a subset D_0 of D has property A iff

> for all $\varepsilon > 0$ and all $\delta > 0$ there exist a density $f \in D_0$ and a density $g \in D$ such that $d(f,g) < \varepsilon$ and $I(g)/I(f) < \delta$.

We may now formulate our theorem.

<u>Theorem 2.1</u>. Let x be a positive real, n a positive integer and D_0 a subset of D. If D_0 has property A, then

$$\inf_{f \in D_0} \sup_{T_n \in \mathcal{T}_n} P_f((nI(f))^{\frac{1}{2}} T_n \le x) = \tfrac{1}{2}. \tag{2.1}$$

This theorem would be meaningless if no subsets of D with property A would exist. However the subsets presented in the following lemma have property A.

Lemma 2.1. *In the following cases formula* (2.1) *holds for all positive reals* x *and all positive integers* n

(2.2) $D_0 = D$

(2.3) $D_0 = \{f \in D | d(f,g) < \eta\}$ *where* $\eta > 0$ *and* $g \in D$

(2.4) $D_0 = \{f \in D | I(f) = 1\}$.

Remark. There exist asymptotically efficient so called adaptive estimators of location $T_n \in \mathcal{T}_n$, n=1, 2, ..., for which (1.1) holds for all $f \in D$ and all $x > 0$. See Beran [1]. Lemma 2.1 implies (with the help of (3.4)) that for all estimators $T_n \in \mathcal{T}_n$ and consequently for all adaptive estimators $T_n \in \mathcal{T}_n$ the following equality holds for all $x > 0$

$$\lim_{n\to\infty} \inf_{f\in D} P_f((nI(f))^{\frac{1}{2}}\, T_n \le x) = \tfrac{1}{2}.$$

3. PROOF OF THEOREM 2.1

In order to prove theorem 2.1 we need the following lemma.

Lemma 3.1. *For each* $f \in D$ *and* $T_n \in \mathcal{T}_n$ *the distribution function of* $(nI(f))^{\frac{1}{2}}\, T_n$ *under* f *is differentiable and has a density which equals at most* $\frac{1}{2}$.

Proof. First we shall prove that the derivative of the distribution function of $(nI(f))^{\frac{1}{2}}\, T_n$ under f exists and that for all reals y

(3.1) $$\frac{d}{dy} P_f((nI(f))^{\frac{1}{2}}\, T_n \le y)$$

$$= \int \cdots\cdots \int_{(nI(f))^{\frac{1}{2}} t_n(x_1,\ldots,x_n)\le y} (nI(f))^{-\frac{1}{2}} \sum_{j=1}^{n} f'(x_j)/f(x_j) \prod_{i=1}^{n} f(x_i)\,dx_1\ldots dx_n.$$

For that purpose we note that for $\theta > 0$

$$\int_{-\infty}^{\infty} |\theta^{-1}(f(x+\theta)-f(x))|dx = \int_{-\infty}^{\infty} \theta^{-1}\left|\int_x^{x+\theta} f'(y)dy\right|dx$$

$$\leq \int_{-\infty}^{\infty}\int_x^{x+\theta} \theta^{-1}|f'(y)|dy\ dx$$

$$= \int_{-\infty}^{\infty}\int_{y-\theta}^{y} \theta^{-1}|f'(y)|dx\ dy$$

$$= \int_{-\infty}^{\infty} |f'(x)|dx$$

Clearly the same is true for $\theta < 0$. So we have for all $\theta \neq 0$ and for $j=1,\dots,n$

$$\int_{\mathbb{R}^n}\cdots\int \left| \prod_{i=1}^{j-1} f(x_i+\theta) \prod_{i=j+1}^{n} f(x_i)\theta^{-1}(f(x_j+\theta)-f(x_j)) \right| dx_1\dots dx_n$$

$$\leq \int_{\mathbb{R}^n}\cdots\int \left| f'(x_j)/f(x_j) \prod_{i=1}^{n} f(x_i) \right| dx_1\dots dx_n.$$

By Vitali's theorem it follows that for $j=1,\dots,n$

$$\lim_{\theta\to 0}\int_{\mathbb{R}^n}\cdots\int \left| \prod_{i=1}^{j-1} f(x_i+\theta) \prod_{i=j+1}^{n} f(x_i)\theta^{-1}(f(x_j+\theta)-f(x_j)) -f'(x_j)/f(x_j) \prod_{i=1}^{n} f(x_i) \right| dx_1\dots dx_n = 0.$$

For each Borel subset B of $\mathbb{R}^n$ this implies

$$\lim_{\theta\to 0}\left| \int_B\cdots\int \{\theta^{-1}(\prod_{i=1}^{n} f(x_i+\theta)- \prod_{i=1}^{n} f(x_i))-\sum_{j=1}^{n} f'(x_j)/f(x_j) \prod_{i=1}^{n} f(x_i)\}\, dx_1\dots dx_n \right| \tag{3.2}$$

$$\leq \lim_{\theta\to 0} \sum_{j=1}^{n} \int_{\mathbb{R}^n}\cdots\int \left| \prod_{i=1}^{j-1} f(x_i+\theta) \prod_{i=j+1}^{n} f(x_i)\theta^{-1}(f(x_j+\theta)-f(x_j)) -f'(x_j)/f(x_j) \prod_{i=1}^{n} f(x_i) \right| dx_1\dots dx_n = 0.$$

Now by the translation equivariance of T_n

$$\theta^{-1}\{P_f((nI(f))^{\frac{1}{2}}\, T_n \leq y+\theta) - P_f((nI(f))^{\frac{1}{2}}\, T_n \leq y)\}$$

$$= \int \cdots\cdots \int_{(nI(f))^{\frac{1}{2}} t_n(x_1,\ldots,x_n)\le y} \{\theta^{-1}(\prod_{i=1}^n f(x_i+(nI(f))^{-\frac{1}{2}}\theta) - \prod_{i=1}^n f(x_i))\}dx_1\ldots dx_n$$

from which (3.1) may be concluded by using (3.2). Because the distribution of $\sum f'(X_j)/f(X_j)$ under f is symmetric it is clear, that for each Borel subset B of $\mathbb{R}^n$

$$\int_B \cdots \int (nI(f))^{-\frac{1}{2}} \sum_{j=1}^n f'(x_j)/f(x_j) \prod_{i=1}^n f(x_i)dx_1\ldots dx_n$$

$$\le \tfrac{1}{2} \int_{\mathbb{R}^n} \cdots \int \left| (nI(f))^{-\frac{1}{2}} \sum_{j=1}^n f'(x_j)/f(x_j) \right| \prod_{i=1}^n f(x_i)dx_1\ldots dx_n$$

$$\le \tfrac{1}{2}\{ \int_{\mathbb{R}^n} \cdots \int (nI(f))^{-1} (\sum_{j=1}^n f'(x_j)/f(x_j))^2 \prod_{i=1}^n f(x_i)dx_1\ldots dx_n\}^{\frac{1}{2}}$$

$$= \tfrac{1}{2}$$

where the second inequality is a consequence of the Cauchy-Schwarz inequality. So we have for all reals y

$$\frac{d}{dy} P_f((nI(f))^{\frac{1}{2}} T_n \le y) \le \tfrac{1}{2}$$

which completes the proof of the lemma. □

Now theorem 2.1 may be proved as follows. Let x, n and D_0 be as in the theorem. Furthermore let ε and δ be positive reals. Since D_0 has property A there exist $f \in D_0$ and $g \in D$ with $d(f,g) < \varepsilon$ and $I(g)/I(f) < \delta$. For these f and g and all $T_n \in \mathcal{T}_n$ we have

$$
\begin{aligned}
P_f((nI(f))^{\frac{1}{2}} T_n \le x) &\le P_g((nI(f))^{\frac{1}{2}} T_n \le x) + d_n(f,g) \\
&\le P_g((nI(g))^{\frac{1}{2}} T_n \le x(I(g)/I(f))^{\frac{1}{2}}) + nd(f,g) \qquad (3.3) \\
&\le P_g((nI(g))^{\frac{1}{2}} T_n \le x\delta^{\frac{1}{2}}) + n\varepsilon.
\end{aligned}
$$

From the translation equivariance and antisymmetry of T_n we obtain for all $h \in D$

$$(3.4) \qquad P_h((nI(h))^{\frac{1}{2}} T_n \le 0) = \tfrac{1}{2}.$$

From (3.3), (3.4) and lemma 3.1 it follows that

$$\tfrac{1}{2} \le P_f((nI(f))^{\frac{1}{2}} T_n \le x) \le \tfrac{1}{2}+\tfrac{1}{2}x\delta^{\frac{1}{2}}+n\varepsilon.$$

Because ε and δ may be chosen arbitrarily small this string of inequalities proves the theorem.

4. PROOF OF LEMMA 2.1.

In this section we shall prove lemma 2.1 by showing that the subsets of D mentioned in this lemma have property A. For case (2.3) and hence case (2.2) this is a consequence of the following lemma.

Lemma 4.1. *For each* $g \in D$ *there exist* $f_\ell \in D$, $\ell=1, 2, \dots,$ *with*

$$\lim_{\ell\to\infty} d(f_\ell,g) = 0 \text{ and } \lim_{\ell\to\infty} I(f_\ell) = \infty.$$

Proof. Let $g \in D$ with distribution function G be fixed and let

$$\Delta = \{\delta \mid 0 < \delta < \tfrac{1}{2},\ 0 < g(G^{-1}(\delta)) < 1\}.$$

For all $\delta \in \Delta$ we define

$$f_\delta(x) = \begin{cases} c_\delta \exp\{b_\delta(x-G^{-1}(\delta))\} & x < G^{-1}(\delta) \\ a_\delta g(x) & \text{for } G^{-1}(\delta) \le x \le -G^{-1}(\delta) \\ c_\delta \exp\{-b_\delta(x+G^{-1}(\delta))\} & -G^{-1}(\delta) < x \end{cases}$$

where $a_\delta = (1-2\delta+2\delta\{g(G^{-1}(\delta))\}^2)^{-1}$, $b_\delta = \{\delta g(G^{-1}(\delta))\}^{-1}$ and $c_\delta = a_\delta g(G^{-1}(\delta))$.
Then $f_\delta \in D$,

$$f'_\delta(x) / f_\delta(x) = \begin{cases} b_\delta & x < G^{-1}(\delta) \\ g'(x)/g(x) & \text{for } G^{-1}(\delta) < x < -G^{-1}(\delta) \\ -b_\delta & -G^{-1}(\delta) < x \end{cases}$$

and

$$I(f_\delta) = a_\delta \int_{G^{-1}(\delta)}^{-G^{-1}(\delta)} (g'(x)/g(x))^2 g(x)\,dx + 2a_\delta \delta^{-1}. \tag{4.1}$$

Furthermore we have

$$d(f_\delta,g) \le \tfrac{1}{2} \int_{G^{-1}(\delta)}^{-G^{-1}(\delta)} |1-a_\delta|\, g(x)\,dx + \int_{-\infty}^{G^{-1}(\delta)} (f_\delta(x)+g(x))\,dx \tag{4.2}$$

$$\leq \tfrac{1}{2}|1-a_\delta|(1-2\delta) + a_\delta \delta\{g(G^{-1}(\delta))\}^2 + \delta.$$

Because there exist $\delta(\ell) \in \Delta$, $\ell=1, 2, \ldots$, with $\lim_{\ell\to\infty} \delta(\ell) = 0$, we may define f_ℓ by $f_{\delta(\ell)}$ for such $\delta(\ell)$, $\ell=1, 2, \ldots$. In view of (4.1) and (4.2) the lemma holds for these f_ℓ, $\ell=1, 2, \ldots$, since $\lim_{\ell\to\infty} a_{\delta(\ell)} = 1$. □

In order to prove that the subset of case (2.4) has property A we proceed as follows. Let f and g belong to D and let $\sigma = (I(f))^{\frac{1}{2}}$. Define $f_\sigma(x) = \sigma^{-1}f(x\sigma^{-1})$ and $g_\sigma(x) = \sigma^{-1}g(x\sigma^{-1})$. Now f_σ and g_σ belong to D and $I(f_\sigma) = 1$ while $d(f_\sigma,g_\sigma) = d(f,g)$ and $I(g_\sigma)/I(f_\sigma) = I(g)/I(f)$. From this observation and from the existence of a subset of D with property A it follows that $\{f|I(f) = 1\}$ has property A.

5. APPENDIX

For the sake of completeness we shall prove here formula (1.2). To this end the following lemma suffices.

Lemma 5.1. *For all* $f \in D$ *and all* $x > 0$

$$\limsup_{n\to\infty} \sup_{T_n\in\mathcal{T}_n} P_f((nI(f))^{\frac{1}{2}}\, T_n \leq x) \leq \Phi(x).$$

Proof. Let $f \in D$, $x > 0$, $\theta_n = (nI(f))^{-\frac{1}{2}}x$, $T_n \in \mathcal{T}_n$ with $T_n = t_n(X_1, \ldots, X_n)$ and $B = \{(x_1, \ldots, x_n) \in \mathbb{R}^n | t_n(x_1, \ldots, x_n) \leq 0\}$. Then we have

$$(5.1)\qquad P_f((nI(f))^{\frac{1}{2}}\, T_n \leq x) = P_{f(.+\theta_n)}((nI(f))^{\frac{1}{2}}\, T_n \leq 0)$$

$$= \int_{\mathbb{R}^n}\!\cdots\!\int 1_B \prod_{i=1}^n f(x_i+\theta_n)dx_1\ldots dx_n$$

and

$$(5.2)\qquad P_f((nI(f))^{\frac{1}{2}}\, T_n \leq 0) = \int_{\mathbb{R}^n}\!\cdots\!\int 1_B \prod_{i=1}^n f(x_i)dx_1\ldots dx_n = \tfrac{1}{2}.$$

By the fundamental lemma of Neyman and Pearson it follows from (5.1) and (5.2) that for all integers n

$$(5.3)\qquad \sup_{T_n\in\mathcal{T}_n} P_f((nI(f))^{\frac{1}{2}}\, T_n \leq x) \leq P_{f(.+\theta_n)}(\textstyle\sum_{i=1}^n \log[f(X_i+\theta_n)/f(X_i)] \geq c_n)$$

where

$$c_n = \sup\{c \,|\, P_f(\sum_{i=1}^n \log[f(X_i+\theta_n)/f(X_i)] \geq c) \geq \tfrac{1}{2}\}.$$

Now from formula (7) of section VI.2.5 of Hájek and Šidák [2] we obtain for all reals y

(5.4) $$\lim_{n\to\infty} P_f(\sum_{i=1}^n \log[f(X_i+\theta_n)/f(X_i)] \leq y) = \Phi((y+\tfrac{1}{2}x^2)x^{-1})$$

and

(5.5) $$\lim_{n\to\infty} P_{f(.+\theta_n)}(\sum_{i=1}^n \log[f(X_i+\theta_n)/f(X_i)] \leq y) = \Phi((y-\tfrac{1}{2}x^2)x^{-1}).$$

From (5.3), (5.4) and (5.5) the lemma follows. □

Acknowledgment. The author wishes to thank Professor W.R. van Zwet for suggesting the problem and for several helpful comments and stimulating discussions during the preparation of this paper.

REFERENCES

[1] Beran, R. (1978). An efficient and robust adaptive estimator of location. Ann. Statist. 6 292 - 313.

[2] Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.