THE ORDER OF A VARIANCE BOUND WHEN ESTIMATING A GENERAL DISCONTINUITY IN DENSITY

BY

THOMAS POLFELDT

TECHNICAL REPORT NO. 63

MARCH 28, 1969

PREPARED UNDER CONTRACT Nonr-225(72)
(NR-042-993)

FOR

OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
THE ORDER OF A VARIANCE BOUND WHEN ESTIMATING A GENERAL DISCONTINUITY IN DENSITY

by

Thomas Polfeldt

TECHNICAL REPORT NO. 63

March 28, 1969

PREPARED UNDER CONTRACT Nonr 225(72)
(NR-042-993) FOR
OFFICE OF NAVAL RESEARCH

This research was carried out with the support of the University of Lund and the Sweden-America Foundation.

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE ORDER OF A VARIANCE BOUND WHEN ESTIMATING A GENERAL DISCONTINUITY IN DENSITY

by

Thomas Polfeldt

1. Introduction and Summary.

In an earlier paper (Polfeldt (1968a); cf. also Polfeldt (1969b)), we showed the following theorem. In a one-sided distribution \( F(x-\theta) \), with \( F(x-\theta) = 0 \) (\( x \leq \theta \)), \( F(x-\theta) > 0 \) (\( x > \theta \)), and with density \( f(x-\theta) \), we estimated the location parameter \( \theta \) by means of \( n \) independent observations. One assumption was that \( f(y) \) varies regularly at \( y = 0 \), with exponent \( c-1 \) (\( c > 0 \)):

\[
\lim_{y \downarrow 0} \frac{f(ky)}{f(y)} = k^{c-1} \quad (\text{all } k > 0).
\]

For any unbiased estimate \( t \) of \( \theta \), we then showed that the exact order of \( \inf_t V_\theta(t) \) is

\[
\begin{align*}
(F^{-1}(1/n))^2 &= n^{-2/c}(Z(1/n))^2 & 0 < c < 2 \\
(G^{-1}(1/n))^2 &= n^{-1} Z_1(1/n) & c = 2 \\
(n^{-1})^2 &= c > 2
\end{align*}
\]  

(1)

where \( G(x) = \int x y^{-2} f(y) dy \), and \( Z \) and \( Z_1 \) vary slowly at zero; \( Z_1(1/n) \) cannot tend to zero as \( n \to \infty \). The results hold as well for left one-sided distributions with \( F(x-\theta) = 1 \) for \( x \geq \theta \).

In this note, we consider estimating the location parameter \( \delta \) of the distribution defined by the density \( f(x-\delta) = f(y) \) using \( n \)
independent observations. We assume that \( f(y) \) has the following property:

\[
\lim_{y \downarrow 0} f(ky)/f(y) = k^{c-1} \quad (\text{all } k > 0)
\]

\[
\lim_{y \uparrow 0} f(ky)/f(y) = k^{d-1} \quad (\text{all } k > 0)
\]

At the unknown location \( \phi \), there is thus (usually) a discontinuity in \( f(x-\phi) \); and the density can tend to zero, infinity or any finite non-zero value as \( x \downarrow \phi \) and \( x \uparrow \phi \). This situation can be considered as the joining of one right and one left one-sided distribution. In these terms, we shall show that the side of \( f(x-\phi) \) that gives the smallest variance bound according to (1) will determine the order of our variance bound for unbiased estimates of \( \phi \). An exception is the case \( c=d=1 \), which is treated in Polfeldt (1968b).

Finally, we indicate how to construct an estimate, the variance of which is believed to be of the same order as the variance bound found here. That order would then be the best attainable.

The tool for our derivation is an inequality proved in Polfeldt (1968a).

**Notation.** \( K \) and \( K' \) are positive, finite constants. If there exist \( K \) and \( K' \) such that

\[
K < a(x)/b(x) < K' \quad (|x-x_0| < \varepsilon)
\]

we write \( a(x) = \Omega(b(x)) \quad (x \to x_0) \).
2. A Lower Variance Bound.

Let \( 0 \leq p \leq 1 \), and denote by \( f(X; \theta) \) the \( n \)-dimensional density of \( X = (x_1, \ldots, x_n) \). Further, \( \mu \) is Lebesgue measure in \( \mathbb{R}^n \), and \( t = t(X) \) is an unbiased estimate of \( \theta \). Finally, define

\[
\begin{align*}
\{+F(y) &= F(y) - F(0) \\
(-F(y) &= F(0) - F(-y)
\end{align*}
\]

\( y > 0 \)

and

\[
\begin{align*}
\{+G(y) &= y^2 \int_y^\infty t^{-2} f(t) dt \\
-G(y) &= y^2 \int_y^\infty t^{-2} f(-t) dt
\end{align*}
\]

\( y > 0 \)

Lemma. (Polfledt (1968a)). For all \( h \) such that \( \theta+h \) is a possible parameter value,

\[
(1-p)V_\theta(t) + pV_{\theta+h}(t) \geq h^2 (Q^{-1}-p),
\]

where

\[
Q = \int \frac{(f(X; \theta+h))^2}{(1-p)f(X; \theta)+pf(X; \theta+h)} \, d\mu
\]

Theorem. If

(i) the set \( \mathcal{X} = \{x | f(x; \theta) > 0 \text{ if } x \neq \theta, \, f(0) \geq 0 \} \) does not depend on \( \theta \)

(ii) there is an \( h_0 \) such that \( \{h | |h| < h_0 \} \subset \mathcal{C} \)

\[
\mathcal{C}[h | \int (f(y-h))^2 f(y) dy \text{ is defined}] \supset \mathcal{C}
\]

\( \mathcal{N} \{h | \theta+h \text{ is a possible parameter value} \} \)
(iii) \[ f(y) = \begin{cases} y^{c-1} + R(y) & y > 0 \\ |y|^{d-1} - R(y) & y < 0 \end{cases} \]

where \( +R(y) \) and \( -R(y) \) vary slowly as \( y^0 \) and \( y^0 \) respectively.

(iv) \( +R(y-h)/+R(y) = 1 + O(|h/y|) \) \( (\lambda|h| < y < \eta, h \to 0) \)

\( -R(y-h)/-R(y) = 1 + O(|h/y|) \) \( (-\eta < y < -\lambda|h|, h \to 0) \)

(v) \( 0 < \lim h^{-2} \int \frac{[f(y-h)/f(y)-1]^2 f(y)dy}{|y| > \eta} < \infty \) \( (h \to 0, \eta > 0 \text{ fixed}) \)

(vi) \( t \) belongs to the class \( T \) of unbiased estimates of \( \phi \)

fulfilling

\[ K' < \frac{V_{\phi}(t)}{V_{\psi}(t)} < K \]

for all \( \phi' \), \( |\phi - \phi'| < \epsilon \), and all \( n \), where \( K' \) and \( K \) may depend on \( \phi \) and \( \epsilon \), remaining positive and finite when \( \epsilon \to 0 \)

then for all \( t \in T \),

\[ V_{\phi}(t) \geq K(\psi^{-1}(1/n))^2, \]

where

\[ \psi(h) = \Omega(-F(h) + F(h) + G(h) + G(h) + h^2). \]

Proof. We shall use the inequality (2). From condition (vi), we obtain for \( t \in T \)

\[ V_{\phi}(t) \geq h^2(Q^{-1} - p)/(1-p+pk). \] \( \quad (4) \)

In our case, the integrand of \( Q \) is (we now assume \( 0 < p < 1 \))
\[
\frac{\Pi(f(x_1 - \theta - h))^2}{(1-p)\Pi f(x_1 - \theta) + p\Pi f(x_1 - \theta - h)} \leq \left\{ \begin{align*}
(1-p)^{-1}\Pi(f(x_1 - \theta - h))^2/f(x_1 - \theta) \\
p^{-1}\Pi f(x_1 - \theta - h)
\end{align*} \right. 
\]

(5)

The integration domain, $R^n$, is divided into

$$A = \{ x | x_1 \notin (\theta \pm \lambda | h) \}$$

(6)

(where $\lambda > 1$), and its complement. Transforming into $y_1 = x_1 - \theta$, we use (5) when integrating over $A$, and (6) for the rest of $R^n$.

Thus $Q = \int_A + \int_{R^n - A}$ and (if $h > 0$)

$$\int_A \leq \frac{1}{1-p} \left[ \int_{-\infty}^{-\lambda h} f(y-h) + \int_{\lambda h}^{\infty} \frac{(f(y-h))^2}{f(y)} \, dy \right]^n$$

Using the identity $f^2(y-h) = 2f(y-h)f(y)-f^2(y)+f(y-h)-f(y))^2$, it is readily seen that the inner integral equals

$$1-2^+P(\lambda' h)-2^-P(\lambda'' h)+^+P(\lambda h)+^+P(\lambda h) + \int_{-\infty}^{-\lambda h} + \int_{\lambda h}^{\infty} \left( \frac{f(y-h)}{f(y)} - 1 \right)^2 f(y) \, dy.$$ 

(where $\lambda' = \lambda - 1$, $\lambda'' = \lambda + 1$). From condition (iv), it follows that

$$\left( \frac{f(y-h)}{f(y)} - 1 \right)^2 \leq \Omega(h \frac{y^2}{y^2} - 2) \quad \text{for} \quad \lambda h < |y| < \eta,$$

then the integrals over ($-\eta$, $-\lambda h$) and ($\lambda h, \eta$) are $\Omega(\frac{\lambda^2}{\eta^2} - \frac{G(\lambda h) - h^2}{\eta^2} - 2G(\eta))$ and $\Omega(\lambda^2 + G(\lambda h) - h^2 - 2G(\eta))$, respectively. From
condition (v), the integral over \(|y| > \eta\) is \(\Omega(h^2)\), and so
\[
\int_A \leq \frac{1}{1-p} (1 + \Omega(-F(h) + F(h) + G(h) + g(h) + h^2))^n
\]
\[
= \frac{1}{1-p} (1 + \psi(h))^n.
\]

Further,
\[
\int_{\mathbb{R}^n-A} \leq p^{-1} \left[ \int_{\mathbb{R}^n} f(x; \psi + h) d\mu - \int_A f(x; \psi + h) d\mu \right]
\]
\[
= p^{-1} \left[ 1 - \int_{-\lambda h}^{\lambda h} \int_{-\lambda h}^{\lambda h} f(y) dy \right]^{n/2}
\]
\[
= p^{-1} (1 - (1 - F(\lambda h) - F(\lambda h))^n).
\]

Now, choose \(h\) such that \(\psi(h) = 1/n\). Then, \(\int_A \leq e/(1-p)\), and
\[
\int_{\mathbb{R}^n-A} \leq p^{-1} (1 - e^{-K\psi}).
\]

When \(\min(c,d) \geq 2\), any \(K\psi > 0\) will do. Thus, for this choice of \(h\), \(Q\) is smaller than a constant depending on \(p\), and \(p\) can be chosen so that \(Q^{-1-p} > K_p > 0\).

From (4) we then obtain
\[
V_{\phi}(t) \geq (\psi^{-1}(1/n))^2 K_p / (1-p+pK),
\]
and the theorem is proved.
3. **Comments on the Theorem.**

The term in $\psi(h)$ that tends to zero most slowly (as $h \to 0$) will decide the order of $\psi(h)$. It is easy to state the leading term when $c \neq d$: it is

$$
-\frac{F(h)}{G(h)} \text{ if } d < c \text{ and } d < 2
$$

$$
-\frac{G(h)}{F(h)} \text{ if } d < c \text{ and } d = 2
$$

$$
\frac{F(h)}{G(h)} \text{ if } c < d \text{ and } c < 2
$$

$$
\frac{G(h)}{F(h)} \text{ if } c < d \text{ and } c = 2
$$

$$
h^2 \text{ if } 2 < \min(c,d).
$$

From known properties of regularly varying functions (cf. Feller (1966), VIII.8-9 and Polfeldt (1969a)), it is seen that $-\frac{G(h)}{F(h)}$ and $-\frac{F(h)}{G(h)}$ are absorbed into $-\frac{F(h)}{F(h)}$ or $h^2$ except when $c = 2$, or $d = 2$. The order of $-\frac{G(h)}{F(h)}$ and $-\frac{F(h)}{G(h)}$, when $d = 2$ or $c = 2$, is easily found from the results of Polfeldt (1969a).

When $c = d = 1$, the inequalities (5) and (6) give too crude a result. In particular, $c = d = 1$ contains as a special case the situation when $\hat{\sigma}$ is not the location of a discontinuity, but where $f(x-\hat{\sigma})$ is differentiable at $x = \hat{\sigma}$. Then, of course, $V(t)$ is expected to be $\Omega(n^{-1})$, while the theorem only gives $V(t) \geq \Omega(n^{-2}(Z(1/n))^2)$, with some slowly varying $Z$. A better variance bound in this situation is given in Polfeldt (1968b).

If $\mathcal{E}$ of condition (i) is of the form $(\hat{\sigma}+a, \hat{\sigma}+b)$, (i) does not hold. However, a modification of the form proposed by Blischke et al. (1966) will give results of the present type. The endpoints of $\mathcal{E}$ will also have to be taken into account, using methods from Polfeldt (1968a).
4. Proposed Estimates of $\phi$.

We intend to base the proposed estimates of $\phi$ on the sample spacings, $x_{(k+1)} - x_{(k)}$ ($x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ is the ordered sample). This is rather natural, e.g., in the case $\min(c,d) < 1$, when $f(x-\phi) \to \infty$ as $x \uparrow \phi$ or $x \downarrow \phi$ or both, and we have a (kind of) mode at $x = \phi$. For easier exposition, assume $^+R \equiv 1$, $^-R \equiv 1$.

Now, if $x_{(k_0)}$ is the first observation to fall to the right of $\phi$ ($k_0$ is then a random variable), we can apply the asymptotic theory for extreme order statistics for a one-sided distribution - see Grenenko (1943), Smirnov (1952). - to obtain the result $E(x_{(k_0+1)} - x_{(k_0)}) = \Omega(n^{-1/c})$.

On the other hand, spacings away from $\phi$ and infinity might generally be expected to have a mean of $\Omega(n^{-1})$. (More generally, we can speak of spacings of $\Omega_p(g(n))$ - cf. Mann and Wald (1943).) Going from left to right on the x-axis, we will expect the following pattern:

Large spacings - spacings of $\Omega(n^{-1})$ - spacings of $\Omega(n^{-1/d})$ - $\phi$ - spacings of $\Omega(n^{-1/c})$ - spacings of $\Omega(n^{-1})$ - large spacings

(7)

Now, consider first the case $\min(c,d) < 1$. We will expect unusually small spacings near $\phi$, and we will base our estimate on the smallest spacings. Grenander (1965) and Venter (1968) have used this idea in different ways to estimate the location of a mode. Both have been led to consider $\kappa$-spacings:

$$x_{(k+\kappa)} - x_{(k)}$$

with $\kappa > 1$; Venter even lets $\kappa$ vary with $n$. We shall not indicate why we also find it necessary to assume $\kappa > 1$, in the following proposed
estimate:

\[ \hat{\theta} = x(k(j-1)) + b \quad \text{if} \quad x(k_j) - x(k(j-1)) < x(k_1) - x(k(i-1)) \]

(all \( i \neq j \))

The quantity \( b \) is chosen so that \( E(\hat{\theta}) = \theta \). We conjecture that with suitable conditions on \( f(y) \), and with \( K \) large enough, (but fixed, finite as \( n \to \infty \)) we will have \( V(\hat{\theta}) = \Omega((-F^{-1}(1/n))^2 + (F^{-1}(1/n))^2) \) — which is the order of the lower variance bound implied by the theorem of this note.

When \( \min(c,d) = 1 \), the estimation problem seems much more complicated. No estimate will be suggested; the reader is referred to the papers of Chernoff and Rubin (1955) and Prakasa Rao (1968) for examples of what may be done.

When \( 1 < \min(c,d) < 2 \), the sample spacings at \( x = \theta \) may be expected to be larger than in the surrounding parts (cf. (7)). We take as our estimate

\[ \hat{\theta} = x(k(j-1)) + b \quad \text{if} \quad x(k_j) - x(k(j-1)) > x(k_1) - x(k(i-1)) \]

for at least \( K \) indices \( i < j \) and at least \( K \) indices \( i > j \), choosing \( K \) so as to practically exclude extreme order statistics.

Again, we believe that \( V(\hat{\theta}) = \Omega((-F^{-1}(1/n))^2 + (F^{-1}(1/n))^2) \).

When \( \min(c,d) = 2 \), problems of the type treated in sections 7 and 8 of Polfeldt (1969b) arise. We refrain from specifying an estimate.
When $\min(c, d) > 2$, we want $V(\hat{\theta}) = \Omega(n^{-1})$, and

$$\hat{\varphi} = \bar{x} + b$$

will be one possibility (if $V(\bar{x}) < \infty$); there are many others.
REFERENCES


Polfeldt, T. (1968b). "Minimum variance order when estimating the location of a discontinuity or a cusp in density." To be published.


The Order of a Variance Bound When Estimating A General Discontinuity in Density

Let the distribution defined by the density $f(x-y) = f(y)$ have a discontinuity at $y$. Suppose that $f(y) = |y|^{-d-1}R(y)$ ($y < 0$), $f(y) = y^{-c+1}R(y)$ ($y > 0$), where $^+R$ and $^-R$ vary slowly as $y \to 0$. Using a variance inequality due to the author, it is shown that an unbiased estimate of $\phi$ based on $n$ independent observations, has a variance not smaller than $K(\psi^{-1}(1/n))^2$, where $\psi(x)$ is of the exact order (as $x \to 0$) $x^{-c+1}R(x) + x^{-d-1}R(x) + x^2 \int_0^x t^{-2}f(t)dt + x^2 \int_0^x t^{-2}f(-t)dt + x^2$. Estimates are suggested that as $n \to \infty$ seem to have variances of the order $(\psi^{-1}(1/n))^2$, (not when $c = d = 1$). This order would then be the best attainable.
Non-regular estimation  
Estimation of a discontinuity  
Estimation of a mode

<table>
<thead>
<tr>
<th>LINK A</th>
<th>LINK B</th>
<th>LINK C</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
</tr>
</tbody>
</table>

**INSTRUCTIONS**

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

9a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

9b. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9c. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9d. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   1. "Qualified requesters may obtain copies of this report from DDC."

   2. "Foreign announcement and dissemination of this report by DDC is not authorized."

   3. "U.S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ."

   4. "U.S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ."

   5. "All distribution of this report is controlled. Qualified DDC users shall request through ."

The report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.