On Local Unbiased Estimation

By D. A. S. Fraser

University of Toronto

[Received February 1963. Revised October 1963]

Summary

A concept of local unbiasedness is derived from an investigation by Fisher (1956, p. 145) of consistency for a neighbourhood of a parameter point $\theta_0$. Connections are made with Cramér–Rao type inequalities and conditions obtained under which the local estimates propagate to provide globally-unbiased estimates.

1. Introduction

As part of a survey of estimation, Fisher (1956) considers concepts of consistency and efficiency for a neighbourhood of a point $\theta_0$ in the parameter space. For his definition of consistency, Fisher uses a discrete sample space and restricts his attention to estimates that are linear in the frequencies. An estimate of this form is said to be a consistent estimate of a parameter if the estimate transforms into the parameter when the relative frequencies on which it is based are replaced by the corresponding probabilities. Fisher examines estimates that are consistent for $\theta$ near to $\theta_0$, and determines the particular estimate that has minimum variance; this optimum estimate has variance equal to the reciprocal of the information at $\theta_0$.

It is seen directly from the definition of consistency that a consistent estimate is also an unbiased estimate. Suppose then that this restriction to estimates linear in the frequencies is relaxed and the larger class of estimates unbiased in the neighbourhood of $\theta_0$ is examined in search of the estimate with minimum variance. This can be accomplished simply, and without the limitation to discrete sample spaces, by utilizing properties of sufficiency and the Cramér–Rao inequality.

2. Local Estimation for a Real Parameter

Let $f(x \mid \theta)$ be a probability or probability density function with a real parameter $\theta$ and suppose that estimation is of interest for $\theta$ near to $\theta_0$. For a sample of $n$ a globally-unbiased estimate $h(x_1, \ldots, x_n)$ satisfies the following conditions at $\theta_0$:

$$E[h(x_1, \ldots, x_n) \mid \theta_0] = \theta_0,$$

(2.1)

$$\frac{d}{d\theta_0} E[h(x_1, \ldots, x_n) \mid \theta] = 1.$$

(2.2)

This suggests the definition: the statistic $h(x_1, \ldots, x_n)$ is locally unbiased for $\theta$ at $\theta_0$ if equations (2.1), (2.2) hold.

In this section the problem of finding the locally-unbiased estimate at $\theta_0$ that has minimum variance will be considered. For this, attention will be restricted to models for which the range of positive density does not depend on $\theta$ at $\theta_0$, and for which differentiation with respect to $\theta$ at $\theta_0$ can be carried through the sign of integration over the sample space.
The definition of a locally-unbiased estimate involves \( f(x \mid \theta_0) \) and \( \partial f(x \mid \theta) / \partial \theta_0 \) and no further properties of the probability density function \( f(x \mid \theta) \). The optimality is in terms of variance at \( \theta_0 \) and it involves only \( f(x \mid \theta_0) \). It would then suffice to consider any model provided it agreed with the given model up to the first derivative at \( \theta_0 \).

The minimal sufficient statistic for a single observation can be given expression by means of the log-likelihood function \( l(\theta \mid x) \) which, in the neighbourhood of \( \theta_0 \), has the form

\[
l(\theta \mid x) = k + \log f(x \mid \theta_0) + (\theta - \theta_0) \frac{\partial}{\partial \theta_0} \log f(x \mid \theta) + \ldots
\]

\[
= k + (\theta - \theta_0) \frac{\partial}{\partial \theta_0} \log f(x \mid \theta) + \ldots.
\]

For \( \theta \) near to \( \theta_0 \) an equivalent sufficient statistic is

\[
s(x) = s(x \mid \theta_0) = \frac{\partial}{\partial \theta_0} \log f(x \mid \theta)
\]

and, for a sample of \( n \), is

\[
S(x) = \sum_{i=1}^{n} s(x_i) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_0} \log f(x_i \mid \theta).
\]

This statistic depends only on \( f(x \mid \theta) \) and \( \partial f(x \mid \theta) / \partial \theta_0 \). It is interesting and perhaps relevant that the locally-sufficient statistic under sampling should involve exactly the same characteristics of \( f(x \mid \theta) \) as does local unbiasedness. The following model of exponential form

\[
f^*(x \mid \theta) = c(\theta) f(x \mid \theta_0) \exp \{(\theta - \theta_0), s(x)\}
\]

is easily seen to agree with the given model at \( \theta_0 \) in the way just described.

The Cramér–Rao inequality can be presented in a form concerned with local unbiased estimation. The ordinary Cramér–Rao type of analysis shows that

\[
E[s(x) \mid \theta_0] = 0,
\]

\[
\frac{d}{d\theta_0} E[s(x) \mid \theta] = \text{var} \{s(x) \mid \theta_0\} = I(\theta_0),
\]

where \( I(\theta) \) is Fisher’s information. And it also shows that an estimate \( h(x_1, \ldots, x_n) \), locally unbiased at \( \theta_0 \), has unit covariance with the locally-sufficient statistic \( S(x) = \Sigma s(x_i) \). The covariance inequality then gives

\[
\text{var} \{h(x_1, \ldots, x_n) \mid \theta_0\} \geq \frac{1}{n I(\theta_0)} = \frac{1}{n I(\theta_0)},
\]

with equality if and only if the variables are linearly related.

The locally-sufficient statistic can be adjusted linearly to produce an unbiased estimate at \( \theta_0 \). The statistic

\[
H(x_1, \ldots, x_n) = \theta_0 + \frac{\Sigma s(x_i)}{I(\theta_0)}
\]
clearly has mean equal to $\theta_0$ at $\theta_0$ and satisfies (2.2); it is thus locally unbiased at $\theta_0$. In addition it is linearly related to $S(x)$ and thus has variance equal to the Cramér–Rao lower bound:
\[
\text{var} \{H(x_1, \ldots, x_n) \mid \theta_0\} = \frac{1}{nI(\theta_0)}.
\]

Thus by the inequality it is the minimum-variance locally-unbiased estimate at $\theta_0$.

For a presentation of the Cramér–Rao inequality there would seem to be several advantages to having it in this local form. Its global relevance would then derive from the fact that a globally-unbiased estimate is necessarily locally unbiased at a parameter value $\theta_0$.

3. LOCAL ESTIMATION FOR A VECTOR PARAMETER

The results on local estimation for a real-valued parameter extend directly to vector-valued parameters. Let $f(x \mid \theta_1, \ldots, \theta_r)$ be a probability or probability density function with parameter $\theta = (\theta_1, \ldots, \theta_r)$ and suppose that the range of positive density does not depend on $\theta$ at $\theta^0$ and that differentiation with respect to $\theta$ at $\theta^0$ can be carried through the sign of integration.

An estimate $h(x)$ is said to be locally unbiased at $\theta^0$ if
\[
E \{h(x) \mid \theta^0\} = \theta^0, \quad \frac{\partial}{\partial \theta^0} E \{h(x) \mid \theta\} = I,
\]
where $I$ is the identity matrix. A globally-unbiased estimate is necessarily locally unbiased at $\theta^0$.

The locally-sufficient statistic is $S(x) = \{S_1(x), \ldots, S_r(x)\}$, where
\[
S_i(x) = \frac{\partial}{\partial \theta_i^0} \log f(x \mid \theta),
\]
and at $\theta^0$ it has mean equal to zero and covariance with $h(x)$ equal to the identity matrix. The generalized covariance inequality then gives
\[
\text{var} \{h(x) \mid \theta^0\} \geq I^{-1}(\theta^0),
\]
where $I(\theta^0)$ is the information matrix at $\theta^0$ and the symbol $\succ$ designates the natural partial ordering of positive definite matrices ($A \succ B$ if and only if $A - B$ is positive semi-definite), and where equality occurs if and only if $h$ and $S$ are linearly related.

The locally-sufficient statistic can be adjusted linearly to produce an unbiased estimate,
\[
H(x) = \theta^0 + I^{-1}(\theta_0) S(x);
\]
this estimate is the minimum-covariance-matrix unbiased estimate at $\theta^0$.

The notion of local unbiasedness can be extended in a variety of ways and provides a basis for giving local expression and local meaning to generalized Cramér–Rao and Bhattacharyya inequalities. A particular form of this extended local unbiasedness at $\theta_0$ would supplement or replace equations (2.1) and (2.2) or their vector analogues by equations such as
\[
\frac{d^2}{d\theta_0^2} E \{h(x_1, \ldots, x_n) \mid \theta\} = 0
\]
or
\[
[\theta_1, \theta_0] E \{h(x_1, \ldots, x_n) \mid \theta\} = 1,
\]
where \([\theta_1, \theta_0]g(\theta) = g(\theta_1) - g(\theta_0)\); in other words, local unbiasedness would be given extended meaning in terms of higher derivatives or differences set equal to zero at \(\theta_0\) and possibly by replacing the first derivative condition by a first difference condition. Then corresponding operators such as

\[
\frac{1}{f(x|\theta_0)} d^2 \quad \text{and} \quad \frac{1}{f(x|\theta_0)} [\theta_1, \theta_0],
\]

when applied to \(f(x|\theta)\), would produce statistics which collectively would be locally sufficient in the particular extended sense. And the generalized covariance inequality would produce a lower bound for the covariance matrix of the extended locally-unbiased estimates. Also, an estimate of minimum covariance matrix (at the lower bound) could be constructed from the extended locally-sufficient statistic.

This form of analysis of extended unbiasedness provides a means for interpreting the succession of bounds presented by Barankin (1949).

4. Propagation of Estimates

The locally-unbiased estimate having minimum variance is based on the locally-sufficient statistic and it therefore preserves all the information available locally concerning the parameter, information being in the sense used by R. A. Fisher. In this section conditions will be examined briefly under which these local information-preserving estimates can be pieced together to form a single overall estimate.

First, by examining the derivation of the estimates in the preceding sections, it is seen that each estimate is essentially unique.

In the univariate case the estimate for a neighbourhood of \(\theta_0\) has the form

\[
t(x_1, \ldots, x_n) = \theta_0 + \frac{\sum s(x_i)}{nI(\theta_0)} \frac{1}{nI(\theta_0)} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_0} \log f(x_i|\theta).
\]

If these estimates piece together to form a single overall estimate having locally the properties of the component estimates then the above expression must be independent of \(\theta_0\). Thus,

\[
\frac{\partial}{\partial \theta_0} \sum_{i=1}^{n} \log f(x_i|\theta) = -n\theta I(\theta) + nI(\theta) t(x_1, \ldots, x_n).
\]

Integration over a connected region of \(\theta\) values then produces

\[
\Pi f(x_i|\theta) = k(x_1, \ldots, x_n) \exp \left\{ t(x_1, \ldots, x_n) \int nI(\theta) d\theta - \int n\theta I(\theta) d\theta \right\};
\]

this has exponential form. Then by noting the factorization of the left side it follows that the density function \(f(x|\theta)\) itself must be of exponential form on each connected region of \(\theta\) values; this is the condition for the estimates to piece together into a single overall estimate.

The multi-parameter case can be handled in a similar manner.

In the preceding two sections the results show that locally on the parameter space adequate information-preserving estimates are available. In this Section the presence of an exponential family is seen to be the necessary and sufficient condition under which these local estimates can be subsumed in a single overall estimate. The presence
of an exponential distribution would seem then to be of advantage more for simplicity of expression than for more fundamental aspects of estimation.

Without an exponential distribution with its single overall estimate, the following programme for estimation seems to have many attractive characteristics. Partition the parameter space into many small neighbourhoods. For each neighbourhood calculate the locally-unbiased estimate or leave available the information for calculating such an estimate. As a single terminal estimate, report the local estimate that falls within the parameter neighbourhood for which it was designed. The reported estimate is easily seen to be the solution of the likelihood equation

$$\Sigma s(x_i | \theta_0) = 0,$$

the maximum-likelihood estimate. Then, if further data become available, new estimates for each neighbourhood on the parameter space can be calculated; and these are seen to have the form of an average of component estimates using the component information values as weights. Again a terminal estimate can be reported — the local estimate which falls in the neighbourhood for which it was designed. All of this can be arranged in a tableau in which a horizontal row for a neighbourhood $\theta_0 \pm \delta$ would contain values for

$$\Sigma s(x_i | \theta_0), \quad nI(\theta_0), \quad \theta_0 + n^{-1}I^{-1}(\theta_0) \Sigma s(x_i | \theta_0),$$

for each body of data concerning the parameter $\theta$. It is interesting in this estimation programme based on unbiasedness and minimum variance that the terminal estimate at each stage should turn out to be the maximum-likelihood estimate.

5. COHERENT INFERENCE

In the preceding Sections some aspects of inference have been examined locally with respect to the parameter space. For a probability function $f(x | \theta)$ at $\theta_0$, the investigation involved only $f(x | \theta_0)$ and $s(x | \theta_0)$ and could thus be interpreted in terms of an approximating distribution,

$$f^*(x | \theta) = c(\theta)f(x | \theta_0)\exp\{(\theta - \theta_0)s(x)\},$$

of exponential form. The local estimate became a global estimate only when this local approximating distribution was the statistical model globally. And in this case the locally-sufficient statistic was globally sufficient.

Methods of inference seem to have greater strength and meaning in the case of translation-invariance and, in particular, the fiducial method is available with a frequency interpretation; see Fraser (1961). It is relevant then to enquire under what conditions the approximating distribution just presented is of translation-invariant form. The answer is available in Lindley (1958); the distribution must be normal with $\theta$ as location parameter or be gamma with $\theta$ as scale parameter. For this case, the function $s(x)$ is determined by $f(x | \theta_0)$ or conversely, and the approximating distribution thus no longer has the generality of two functions of $x$ but is restricted to a single function of $x$.

In some related work (Fraser, 1964) the concept of conditional sufficiency has been examined locally with respect to the parameter space. Local estimation of translation-invariant form can then be developed and is available for the full generality of approximation using two functions of $x$. And interestingly, the special case here in which local estimation is of translation-invariant form is equivalently analysed from the conditional point of view.
REFERENCES


