HIGHER ORDER LOCAL UNBIASEDNESS WITH COMPUTER ALGEBRA

M. Ennis
Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

D.A.S. Fraser
Department of Mathematics & Statistics
York University
North York, Canada, M3J 1P3

Key Words and Phrases: Unbiased estimation; local unbiasedness; minimum variance; computer algebra; approximate unbiasedness

ABSTRACT

Unbiased estimation methods that are in some way local on the parameter space have been initiated by Barankin (1948) and Fraser (1964). The later methods involving local unbiasedness are extended to higher (derivative or difference) order at a parameter value ($\theta_0$) and provide a basis for a computer algebra implementation concerning the improvement of unbiased estimates.

1. INTRODUCTION

Unbiased estimation has a substantial place in contemporary statistical inference, although recent directions emphasize distributional methods; for a recent review, see Fraser (1991).

Most theory of unbiased estimation requires global unbiasedness and targets on uniform minimum variance. Barankin (1949) retained global unbiasedness and targeted on minimum variance at a parameter value of interest. Fraser (1964) following patterns in Fisher (1956) defined first order (derivative) unbiasedness at a parameter value $\theta_0$ and explored minimum variance at $\theta_0$ for such estimates. A byproduct of this was an interpretation of the Cramér-Rao inequality with the minimum of assumptions needed for its basic derivation (Fraser, 1976).
In this paper we define $p^{th}$ order local $(\theta_0)$ unbiasedness and describe the procedure for calculating the corresponding minimum variance estimate. For other than very simple problems, computer algebra (for example, Maple, Mathematica or Maxcyma) would be needed, but within such a context we would have a simple procedure for developing progressively improved estimates.

The theory is presented in Section 2, some aspects of implementation discussed in Section 3 and an example given in Section 4.

2. LOCAL UNBIASEDNESS OF ORDER $p$

An estimate $t(y)$ for a scalar $\theta$ was defined (Fraser, 1964) to be locally unbiased at $\theta_0$ if its mean value coincided with $\theta$ to the first order at $\theta = \theta_0$:

\[ E(t(y); \theta_0) = \theta_0 \]
\[ \frac{d}{d\theta} E(t(y); \theta)_{\theta_0} = 1 \]  \hspace{1cm} (1)

We generalize as follows:

Definition 1: An estimate $t(y)$ is locally unbiased of order $p$ for a scalar $\beta(\theta)$ at $\theta_0$ if:

\[ E(t(y); \theta_0) = \beta(\theta_0) \]  \hspace{1cm} (2)
\[ \beta_i E(t(y); \theta) = \beta_i, \quad i = 1, \ldots, p \]  \hspace{1cm} (3)

where $\beta_i = \beta_i(\theta)$ and

$\beta_i$ is the $i$th derivative at $\theta_0$

\[ \beta_i h(\theta) = \frac{d^i}{d\theta^i} h(\theta) |_{\theta_0} \]  \hspace{1cm} (4)

or the $i$th divided difference with respect to some sequence $\theta_0, \theta_1, \ldots, \theta_p$ typically local to $\theta_0$.

With most inference theory we commonly use the scores

\[ S_i(y) = \beta_i \ln f(y; \theta) \]  \hspace{1cm} (5)

For the theory here the related estimation scores are more appropriate:

\[ U_i(y) = \frac{1}{f(y; \theta)} \beta_i f(y; \theta) \]  \hspace{1cm} (6)

With the differentiation definition, the usual regularity conditions are needed for differentiation under the sample space integration sign. With the divided difference definition, less regularity is
needed, specifically differencing through the sign of integration is immediate; for some background for this case, see Chapman and Robbins (1951) and Fraser and Guttman (1952).

The following lemma holds under the usual regularity conditions for the differentiation and differencing cases.

**Lemma 1:** If \( t(y) \) is locally unbiased of order \( p \) for \( \beta(\theta) \) at \( \theta_0 \), then
\[
\text{cov} \{ t(y), U(y); \theta_0 \} = \beta
\]  
(7)

where \( U = (U_1(y), \ldots, U_p(y))^\prime \) and \( \beta = (\beta_1, \ldots, \beta_p)^\prime \).

**Proof:** The result follows from
\[
\partial_i E \{ t(y); \theta \} = E \{ t(y) U_i(y); \theta \}
\]
\[E \{ U_i(y); \theta_0 \} = 0 \quad i = 1, \ldots, p.
\]  
(8)

The locally unbiased estimates of order \( p \) for \( \beta(\theta) \) at \( \theta_0 \) satisfy a Cramér-Rao lower bound:

**Lemma 2:** The variance \( \sigma_t^2 \) of an estimate \( t(y) \) that is locally unbiased of order \( p \) for \( \beta(\theta) \) at \( \theta_0 \) satisfies
\[
\sigma_t^2 \geq \beta^\prime \Sigma^{-1} \beta.
\]  
(9)

where \( \Sigma \) is the variance matrix of \( U \) at \( \theta = \theta_0 \). Proof: The variance matrix for \( (t(y), U_1(y), \ldots, U_p(y)) \) has the following form:
\[
\begin{bmatrix}
\sigma_t^2 & \beta_1 & \ldots & \beta_p \\
\beta_1 & \sigma_{11} & \ldots & \sigma_{1p} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_p & \sigma_{p1} & \ldots & \sigma_{pp}
\end{bmatrix} =
\begin{bmatrix}
\sigma_t^2 & \beta^\prime \\
\beta & \Sigma
\end{bmatrix}
\]  
(10)

where \( \sigma_{ij} = \text{cov} \{ U_i(y), U_j(y); \theta_0 \} \). The linear regression residual of \( t \) regressed on \( U_1, \ldots, U_p \) is \( t - \beta^\prime \Sigma^{-1} U \) and has variance
\[
\sigma_t^2 - \beta^\prime \Sigma^{-1} \beta.
\]  
(11)

In particular, if \( t(y) \) is globally unbiased for \( \beta(\theta) \), then it is locally unbiased of order \( p \) at \( \theta_0 \) and satisfies (9) where the left and right sides are calculated for any \( \theta_0 \) value of interest; these would then be the ordinary Cramér-Rao and Bhattacharyya bounds. We feel however that the considerations in terms of \( p^\text{th} \) order unbiasedness focuses on the essentials of the inequality for the bounds on variance. The following results support this view.

As a method of constructing \( p^\text{th} \) order locally unbiased estimates we consider a linear combination of the functions \( U_0, U_1, \ldots, U_p \) treating these estimation score functions as basic building blocks for estimation:
where \( c = (c_1, \ldots, c_p) \) gives coefficients for the corresponding coordinates of \( U \). The following lemma shows that the coefficients are unique.

**Lemma 3:** If \( t(y) \) given by (12) is locally unbiased of order \( p \) for \( \beta(\theta) \) at \( \theta_0 \), then

\[
t(y) = \beta(\theta_0) + \beta' \Sigma^{-1} U(y). \tag{13}
\]

**Proof:** For local unbiasedness, \( t(y) \) must have the correct covariance (Lemma 2) with \( U \):

\[
\beta' = \text{cov}(c'U, U') = c'\Sigma; \tag{14}
\]

thus \( c = \Sigma^{-1} \beta \). Also \( t(y) \) must have mean zero at \( \theta_0 \); thus \( c_0 = \beta_0 \). The constructed estimator then satisfies Definition 1, and in fact has the minimum variance defined by (11).

The uniqueness of the linearly constructed locally unbiased estimate then shows that (13) is minimum variance locally unbiased of order \( p \). Note that there is an implicit assumption in our development: that the variance matrix \( \Sigma \) is nonsingular.

3. IMPLEMENTATION

A computationally easier case arises if the parameter \( \beta(\theta) \) to be estimated is in fact the parameter \( \theta \). Then \( \beta(\theta_0) = \theta_0 \) and \( \beta = (1, 0, \ldots, 0) \). It follows that the minimum variance local estimate is

\[
t(y) = \theta_0 + \Sigma^{11} U_1 + \Sigma^{12} U_{(2)} = \theta_0 + \Sigma^{11} \tilde{U}_1 \tag{15}
\]

where \( U_{(2)}(y) = (U_2(y), \ldots, U_p(y)) \), with corresponding partition

\[
\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \tag{16}
\]

and

\[
\tilde{U} = U_1 + (\Sigma^{11})^{-1} \Sigma^{12} U_{(2)} = U_1 - \Sigma_{12} \Sigma_{22}^{-1} U_{(2)} \tag{17}
\]

which is the regression residual. The minimum variance is then \( \Sigma^{11} \).

For most standard problems the ordinary scores \( S_1(y), \ldots, S_p(y) \) will be more readily available as they will have the additivity that goes with the log density or likelihood. We also know
that standard programming in computer algebra will produce \( U_1, \ldots, U_p \) as functions of
\( S_1, \ldots, S_p \). The covariance matrix \( \Sigma \) for \( U_1, \ldots, U_p \) will require higher moments of the \( S \)'s, but if
these are available then routine computer calculations will give the \( \Sigma \) matrix for the estimation
scores.

Thus, for a statistical problem with an available likelihood function and programs for calculat-
ing variances and higher moments of the standard scores, computer algebra can give the \( U \)
scores and the unbiased estimates (13), (14) for successive values of \( p \). We can thus generate pro-
gressively 'broader band' minimum variance unbiased estimates.

4. EXAMPLE

We present a simple transparent example to illustrate the results in Section 3. For a realistic
example, however, we would expect to resort to computer algebra for the calculations.

Consider a variable \( y \) on the range \((-1, +1)\) with density
\[
    f(y; \theta) = \frac{1}{2} \left( 1 + \theta p_1(y) + \frac{\theta^2}{2} p_2(y) + \frac{\theta^3}{6} p_3(y) \right),
\]
where the \( p_i(y) \) are the Legendre polynomials (Abramowitz and Stegun, 1965):
\[
p_1(y) = y, \quad p_2(y) = \frac{1}{2} (3y^2 - 1), \quad p_3(y) = \frac{1}{2} (5y^3 - 3y)
\]
and \( \theta \) lies in a neighbourhood of \( \theta_0 = 0 \) ensuring that \( f(y; \theta) \) is a density. As interest parameter
consider \( \beta = e^{\theta} \), in the neighbourhood of \( \theta_0 = 0 \).

From Sections 2 and 3 with \( p = 3 \) we obtain \( \beta' = (1, 1, 1) \) and
\[
    U_0 = 1, \quad U_i = p_i(y).
\]
At \( \theta = 0 \), the variance matrix \( \Sigma \) for \( U \) is available from data in Abramowitz and Stegun (1965):
\[
    \Sigma = \begin{bmatrix}
        \frac{1}{3} & 0 & 0 \\
        0 & \frac{1}{5} & 0 \\
        0 & 0 & \frac{1}{7}
    \end{bmatrix}
\]
The orthogonality of the polynomials at \( \theta = 0 \) clearly will simplify computations.
The estimates (13) for \( p = 1, 2, 3 \) are then easily calculated:

\[
\begin{align*}
  t_1(y) &= 1 + 1 \cdot \begin{bmatrix} \frac{1}{3} \end{bmatrix}^{-1} p_1(y) = 1 + 3y, \\
  t_2(y) &= 1 + (1,1) \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{5} \end{bmatrix}^{-1} \begin{bmatrix} p_1(y) \\ p_2(y) \end{bmatrix} \\
  &= 1 + 3y + \frac{5}{2} (3y^2 - 1), \\
  t_3(y) &= 1 + (1,1,1) \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{7} \end{bmatrix}^{-1} \begin{bmatrix} p_1(y) \\ p_2(y) \\ p_3(y) \end{bmatrix} \\
  &= 1 + 3y + \frac{5}{2} (3y^2 - 1) + \frac{7}{2} (5y^3 - 3y).
\end{align*}
\]

The estimates have progressively higher order unbiased near \( \theta = 0 \) but at the price of increasing variance: 3, 3+5, 3+5+7.

BIBLIOGRAPHY


Received August 1991; Revised July 1992