Bias In The Maximum Likelihood Estimation Of Parameters Of Nonlinear Regression Models

B. Mahaboob, B.Venkateswaru, C. Narayana, M. Sivaiah

Abstract: Nonlinear model building has become an important tool in Predictive Analysis and Forecasting Theory. MLE is a phenomenon in which one can obtain unknown coefficients of a distribution by optimizing a likelihood function. Maximum likelihood estimate is the vector in parameter space which optimizes the likelihood function. This research article throws a light on the BIAS in the MLE of unknown coefficients of statistical models which are not linear. In addition to this a test for the linearity of regression has been proposed. If the ML function possesses derivatives one can apply first derivative test to obtain optimum values. But in some situations the equations of first degree of ML function are to be solved in explicit manner. For example in linear statistical model OLS estimator optimize the ML function. In vast number of cases advanced numerical techniques should be implemented in order to get ML function. As the application of ML technique is both flexible and intuitive this technique has become an indispensable tool in statistical inference.

Index Terms: Bias, MLE(Maximum Likelihood Estimation), Multivariate Normal distribution, Variance-Covariance Matrix, Residual Vector, Linearity of Regression, Rank correlation coefficient.

1. INTRODUCTION

A Mathematician or Statistician suggests a model $M(\beta)$ for describing some hidden relationships where $\beta$'s are the unknown coefficients which tell the characteristics of such a connection. The method of estimation of parameters is a procedure by which the Mathematician or Statistician traces the best parameter choice $\hat{\beta}$ for observed entries. Usually $\beta$ is used to encoding the best features of the model $M$. An estimate of parameter $\beta$ say $\hat{\beta}$ accepts the validation of the model (i.e., verifying if the derived model fits the given entries) and forecasting future outcomes. Most renowned techniques in the estimation of parameters are MLE,, NLLSE,,MOM and MINIMAX Estimation. Gauss M. Corriero et al.[1] in their paper derived most general formulae of biases of order two in ML estimates of the regression, dispersion and precision parameters in nonlinear statistical models with t- distributed errors. In 2003, Mukesh Sharma et al.[2] in their paper, presented the maximum likelihood method for regression analyses of censored data below detection limit for nonlinear models and there proposed ML method has been translated into an equivalent least squares method. Man Chen et al. in 2016[3], in their paper proposed analytic bias -corrected MLEs to reduce the biases of the MLEs of two parameter Weibull distributions. R D Cook et al. [4], in their paper, investigated the biases of the residuals and maximum likelihood parameter estimates from normal nonlinear regression models. If the ML function possesses derivatives on can apply first derivative test to obtain optimum values. David E. Giles et.al [2013] depicted the bias of order two of the ML estimators of parameters of Lomax distribution (aka Pareto II) which is most useful in vast number of fields for finite sample sizes. Besides they shown that the bias is positive.

2 BIAS IN THE MLE OF PARAMETERS OF NONLINEAR REGRESSION MODELS

A set of $l$ nonlinear regression models can be specified by

$$Z_{ab} = g_{a}(Y_{a},\psi) + \varepsilon_{ab}$$

Where $a = 1, 2, \ldots, l$ and $b = 1, 2, \ldots, m$. Here $\varepsilon_{ab}$ stand for random errors following MVND satisfy

$$E(\varepsilon_{ab}) = 0, E(\varepsilon_{ab}, \varepsilon_{cv}) = 0, \psi = (\psi_{1}, \psi_{2}, \ldots, \psi_{q})_{q \times 1}$$

vectors of unknown parameters. The forms of the $l$ models $g_{a}(Y_{a},\psi)$ are known. $X_{b} = \{X_{pb}\}$ is a $v \times 1$ vector of the entries observed for the $v$ (assume) free quantities in the both experiment $Z_{b} = \{Z_{ab}\}$ is a $l \times 1$ vector of the entries observed for the $l$ responses in the both investigations unless $b = v$ when $E(\varepsilon_{ab}, \varepsilon_{cv}) = \sigma_{ap}$ and hence the vectors of entries in the both investigations has the Var-Covar-Matrix

$$\sum_{b} = \left(\sigma_{ap}^{(p)}\right)$$
In order to evaluate the estimates of the biases in the MLE of the parameters of nonlinear regression models, the following are defined
\[
e \equiv (e_1', e_2', \ldots, e_m') \quad \text{and} \quad Z = (z_1', z_2', \ldots, z_m')' \quad \text{and} \quad g(\psi) = \{g_1(Y, \psi), \ldots, g_m(Y, \psi)\}
\]
Where \( e \) is the \( ml \times 1 \) vector of \( ml \) residuals. \( z \) is the \( ml \times 1 \) vector observations and \( g(\psi) \) are \( mk \) forecasts of \( \psi \). By the maximum likelihood or the generalized least squares estimation one can minimize the expression
\[
(z - g(\psi))' \left( \sum^{-1} - g(\psi) \right) \quad \text{with respect to} \quad \psi
\]
Where is the Var-Covar-Matrix for all \( ml \) entries and is block diagonal matrix as every one of the \( m \) investigations are taken to be statistically independent. First order condition gives
\[
G(\hat{\psi})' = \sum^{-1} [z - g(\hat{\psi})] = 0
\]
(3)
Here \( G(\psi) \) represents \( ml \times q \) matrix of first order derivatives of \( g(\psi) \), the models for every response in every investigation w.r.t every one of the \( q \)-parameters, and those differential coefficients are found out at \( \psi \). Let \( \Delta = \hat{\psi} - \psi \)
represent the discrepancy in the parameter estimate \( \psi \) following (2.4). Now vector residual is expressed as
\[
E = Z - g(\hat{\psi}) = e - G(\psi)\Delta - \frac{1}{2} H\Delta
\]
(4)
Here the models are taken to be sufficiently denoted by Taylor power series expansion in exponents of \( \Delta \) shortened after terms of order 2 and \( G \) is the \( ml \times q \) matrix
\[
H = \{J_{1}\Delta, \ldots, J_{m}\Delta\}
\]
(5)
Where each \( J_{m} \) is a \( q \times q \) matrix of order two differential coefficients of the \( a_{ik} \) response model w.r.t its \( q \) unknown constants at the constraints of the both investigations obtained at \( \psi \). Further up to this order of correction
\[
G(\hat{\psi}) = G(\psi) + H
\]
(6)
Suppose that \( \Delta = B \in + N \)
(7)
Here \( N \) stands for second order polynomial of the random error variables and
\[
N = \{e'_1, e'_2, \ldots, e'_q\} \quad \text{and} \quad B = q \times ml \quad \text{and} \quad C's \quad \text{are} \quad ml \times ml \quad \text{matrices and} \quad N = q \times 1 \quad \text{vector. It remains to find the matrices} \quad A \quad \text{and} \quad B \quad \text{. This supposed structure has the essential principle that when random error variables have null variance the expectation of the discrepancy i.e., the bias vanishes, from (2.8) the bias is obtained by}
\]
\[
E(\Delta) = E(N)
\]
(8)
Put (4) and (6) in (2) to get
\[
(G + H)\sum^{-1} (e - G\Delta - \frac{1}{2} H\Delta) = 0
\]
(9)
Where \( G = G(\psi) \). Now (7) into (9) and equalizing coefficients to 0 and hence this expression is identically 0 to order 2 in the random error variables \( e \). For the term linear in \( e \), on can obtain
\[
G'\sum^{-1} (e - GB e) = 0
\]
(10)
\[
B = UG^{-1}\sum^{-1} G^{-1}
\]
(11)
Here \( U = (G'\sum^{-1} G)^{-1} \)
(12)
For the second order term in \( e \) one can get
\[
G'\sum^{-1} (-GN - \frac{1}{2} KB e) + K'\sum^{-1} (-GB e) = 0
\]
(13)
Where \( K \) is the \( ml \times q \) and is given by
\[
K = \{J_{1}B, \ldots, J_{m}B\}
\]
(14)
By taking expectations in (13) one can get
\[
G'\sum^{-1} (-GE(N) - \frac{1}{2} W) = 0
\]
(15)
Where \( W \) is an \( ml \times 1 \) vector which is given by
\[
W = \{tr(J_{1}U), tr(J_{2}U), \ldots, tr(J_{m}U)\}
\]
(16)
From the bias in the MLE of properties of nonlinear models is given by
\[
E(\Delta) = -\frac{1}{2} UG'\sum^{-1} W
\]
(17)
Where \( U \) and \( W \) are given by (12) and (16). In practice the parameter estimates \( \hat{\psi} \) must be used instead of unknown parameters \( \psi \). Put \( G_{ml\times q} = \{G_{1}^1, \ldots, \ldots, G_{m}^1\} \). Where each \( G^1 \) and \( I \times q \) matrix of first order derivatives of each of the \( K \) responses with respect to each of the \( q \)-parameters, evaluate at the conditions of the both experiment and at parameter values \( \psi \). Now the bias can be written as
It is essential to identify that for linear models all the J’s vanish and the biases also vanish. Also even if just one response is nonlinear in the parameters then in general all the parameter estimates are biased.

The term \( U = \left( \sum_{i=1}^{m} G_i \sum_{j=1}^{m} G_j \right)^{-1} \) is in fact the VAR-COVAR Matrix of the estimates of parameters and determinant is the “Generalized variance” of the estimates of parameters.

### 3 A TEST FOR LINEARITY OF REGRESSION

Choose a model specified by

\[ Z_j = \gamma + \delta g(Y_j) + \epsilon_j, \quad j = 1, 2, \ldots, m \]

where \( \gamma, \delta \) unknown parameters are \( Y_j \)'s are fixed constants.

Assumptions:

\( B_1 \): The errors \( \epsilon_j \)'s are continuous, \( B_2 \): the errors \( \epsilon_j \)'s are independently, symmetrically and continuously distributed around the same median.

The testing hypothesis are \( H_0 : g(Y) = Y \quad \text{if} \quad H_1 : g \) is a nonlinear convex or concave function. Rename the \((Y, Z)\) pairs such that \( Y_i < Y_2 < \ldots < Y_m \)

Assume that \( m = zp \) and if \( m \) is odd delete the middle pair \((Y_{m+1}(0.5), Z_{m+1}(0.5))\)

Evaluate the \( p \) sample slopes as

\[ \hat{\delta}_i = \frac{Z_p - Z_i}{Y_p - Y_i} \] (18)

The test statistic is given by

\[ D = \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \phi_{jk}, \text{where} \quad \phi_{jk} = \begin{cases} 1, & \text{if} \quad \hat{\delta}_j < \hat{\delta}_k \\ -1, & \text{if} \quad \hat{\delta}_j > \hat{\delta}_k \end{cases} \]

For significantly large values of \( D \) accept \( H_0 \). This is equivalent either to subject \( H_0 \) or accept \( H_0 \) for this significantly large or small values of Kendall’s rank coefficient

\[ \rho = \frac{2D}{p(p-1)} \]

and this is calculated for the sequences \( (1, 2, \ldots, p) \) and \( (\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_p) \) if \( H_0 \) is true the successive slopes \( \hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_p \) cannot show an appreciable trend where it \( H_1 \) is true the slopes should tend to increase or decrease. One can point out that (i) Under either \( B_0 \) or \( B_2 \) the \( \hat{\delta}_j \)'s are independent and (ii) Under \( H_0 \) and either \( B_0 \) or \( B_2 \)

\[ P\left( \hat{\gamma}_j > \hat{\gamma}_k \right) = P\left( \hat{\gamma}_j < \hat{\gamma}_k \right) = \frac{1}{2} \text{ for all } j \neq k \]

### 4 CONCLUSIONS

In the above research work the BIAS in the MLE of parameters of nonlinear statistical models has been extensively discussed. Besides a test for linearity of regression has been proposed. In the context of future research one can propose a test for the specification of errors in statistical models which are nonlinear and can estimate a mixed general Cobb-Douglas type function with multiplicative and additive errors. Moreover the above propositions can be executed in estimating Seemingly Unrelated Nonlinear Regression Equations Model and Gompertz Growth type function considering additive error term.

### REFERENCES


