Computation of an exact confidence set for a maximum point of a univariate polynomial function in a given interval

June 20, 2016

Abstract

Construction of a confidence set for a maximum point of a function is an important statistical problem. Wan et al. (2015) provided an exact $1 - \alpha$ confidence set for a maximum point of a univariate polynomial function in a given interval. In this paper, we give an efficient computational method for computing the confidence set of Wan et al. (2015). We demonstrate with two examples that the new method is substantially more efficient than the proposals by Wan et al. (2015). Matlab programs have been written which make the implementation of the new method straightforward.

Keywords: Confidence set; Numerical quadrature; P-value; Statistical inference; Parametric regression; Semi-parametric regression.

1 Introduction

Determination of the maximum point of a function is an important problem due to its wide applications; see e.g. Wan et al. (2015, 2016) and the references therein.

Consider the function

$$l(x, \theta_0, \theta) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p = \theta_0 + f(x, \theta), \quad (1)$$

where $\theta = (\theta_1, \cdots, \theta_p)^T$ and $f(x, \theta) = \theta_1 x + \cdots + \theta_p x^p$ contains all the information on the maximum points of $l(x, \theta_0, \theta)$. The interest is in the maximum points of $l(x, \theta_0, \theta)$ in a given interval of $x \in [a, b]$. It is clear that the maximum points of $l(x, \theta_0, \theta)$ do not depend on $\theta_0$ and hence are the same as those of $f(x, \theta)$. If the value of $\theta$ is known, then this is a simple calculus problem. The difficulty lies in that the value of $\theta$ is unknown and only an estimator $\hat{\theta}$ of $\theta$ with a certain distribution, specified in (2) and (3) below, is available. Hence the maximum points of $l(x, \theta_0, \theta)$ can only be estimated based on $\hat{\theta}$. We assume that the estimator $\hat{\theta}$ of $\theta$ has the
normal distribution

\[ \hat{\theta} \sim N(\theta, \sigma^2 \Sigma), \]  

(2)

where \( \Sigma \) is a known positive definite covariance matrix, and \( \hat{\sigma}^2 \) is an available estimator of the unknown error variance \( \sigma^2 \) with distribution \( \hat{\sigma}^2 \sim \sigma^2 \chi^2_v/v \) independent of \( \hat{\theta} \), where \( v \) is the degrees of freedom (df) of the chi-square distribution. In the special case that \( \sigma^2 \) is a known constant, then \( \hat{\sigma}^2 = \sigma^2 \) for \( v = \infty \) and we can assume without loss of generality that

\[ \hat{\theta} \sim N(\theta, \Sigma). \]  

(3)

The distributional assumption (2) follows naturally from the standard \( p \)th order univariate polynomial regression model: \( y = \theta_0 + f(x, \theta) + e \), where \( e \) is the usual random error with distribution \( N(0, \sigma^2) \). Based on \( n \) observations \( (y_i, x_i), \ i = 1, \cdots, n \), \( \Sigma \) results from deleting the first row and the first column of \( (X^T X)^{-1} \), where \( X \) is the usual \( n \times (p + 1) \) design matrix. The distributional assumption (3) holds asymptotically for many parametric and semi-parametric models (cf. Wan et al., 2015). A maximum point of \( f(x, \theta) \) in a covariate interval of \( x \in [a, b] \) may represent the dose in the range \( [a, b] \) that maximizes the response, or the amount of fertilizer that produces the highest yield, etc.

Let \( k \) be a global maximum point of \( f(x, \theta) \) in \( x \in [a, b] \). Wan et al. (2015), denoted as WLHB henceforth, provides an exact \( 1 - \alpha \) level confidence set for \( k \) by inverting a family of exact \( 1 - \alpha \) level acceptance sets for testing \( H_0 : k = k_0 \) for each \( k_0 \in [a, b] \). This method computes one critical constant \( c(k_0) \) for each \( k_0 \in [a, b] \), using numerical quadrature and search for \( p = 2 \) and using statistical simulation for \( p \geq 3 \).

In this paper, we first show that the acceptance probability used in the computation of a critical point can be computed using \( p - 1 \) dimensional numerical quadrature. Hence, when \( p = 4 \) for example, the acceptance probability can be computed using 3-dimensional numerical quadrature, which is usually much faster than statistical simulation to achieve the same computational accuracy. As polynomial regression models of order higher than 4 are rarely used in applications (cf. Liu et al., 2014), the computation method developed in this paper for computing the exact confidence set of WLHB (2015) is of considerable practical relevance.

We further show that, for testing \( H_0 : k = k_0 \), we only need to compute the p-value of the test, which requires the computation of one acceptance probability only. This requires substantially less computation time than the computation of the critical constant \( c(k_0) \), which involves repeated computation of the acceptance
probability for several candidate values of $c(k_0)$, whose corresponding acceptance probability is equal to $1 - \alpha$.

The paper is organized as follows. Section 2 gives a brief review of the confidence set of WLHB (2015). Section 3 considers how the acceptance probability can be computed efficiently at least for $p \leq 4$ using numerical quadrature. Section 4 then shows how to determine whether a $k_0 \in [a, b]$ belongs to the confidence set by computing just one acceptance probability. Section 5 illustrates the new computational method with two examples to demonstrate the substantial saving of computation time and the improved accuracy over the computation methods given in WLHB (2015). Finally, section 6 contains discussions and conclusions.

2 WLHB method

Assume that $k$ is a maximum point of $f(x; \theta)$ in $x \in [a, b]$. Wan et al. (2015) uses the following $1 - \alpha$ level acceptance set for testing $H_0: k = k_0$ for each $k_0 \in [a, b]$: 

$$A(k_0) = \{Y : f(k_0, \hat{\theta}) - f(x; \hat{\theta}) \geq -c(k_0) \hat{v}(k_0, x), \forall x \in [a, b]\backslash k_0\},$$

where

$$\hat{v}(k_0, x) = \left(\hat{\sigma}/\sigma\right) \sqrt{\text{var}(f(k_0; \hat{\theta}) - f(x; \hat{\theta}))}$$

$$= \hat{\sigma}|k_0 - x| \sqrt{g(k_0, x, p)^T \Sigma g(k_0, x, p)}$$

with

$$g(k_0, x, p) = \frac{(k_0 - x, k_0^2 - x^2, \cdots, k_0^p - x^p)^T}{k_0 - x},$$

and $c(k_0)$ is the critical value chosen so that the acceptance probability is equal to $1 - \alpha$, that is,

$$\inf_{H_0} P\{Y \in A(k_0)\} = \inf_{H_0} P\{Y : f(k_0, \hat{\theta}) - f(x_0, \hat{\theta}) \geq -c(k_0) \hat{v}(k_0, x), \forall x \in [a, b]\backslash k_0\}$$

$$= \inf_{H_0} P\{Y : \sup_{x \in [a, b]\backslash k_0} \frac{(k_0 - x)}{|k_0 - x|} \hat{\sigma} \sqrt{g(k_0, x, p)^T \Sigma g(k_0, x, p)} \leq c(k_0)\} = 1 - \alpha. \quad (4)$$

The $1 - \alpha$ level confidence set of Wan et al. (2015, 2016) is then given by

$$C_E(Y) = \{k \in [a, b] : Y \in A(k)\}.$$
The key of the WLHB method for the construction of $C_{E}(Y)$ is the computation of the critical constant $c(k_{0})$ for each $k_{0} \in [a, b]$. A simulation-based method to obtain $c(k_{0})$ is given for a general $p \geq 2$, and for the special case $p = 2$, a numerical method is provided involving one-dimensional numerical quadrature.

3 Computation of the acceptance probability

Let $P$ denote the unique positive definite square-root matrix of $\Sigma$ and so $P^2 = \Sigma$. Then we have $N = P^{-1}(\hat{\theta} - \theta)/\sigma \sim N(0, I_p)$. Furthermore, since $\hat{\sigma}/\sigma$ and $\hat{\theta}$ are independent random variables and $\hat{\sigma}/\sigma \sim \sqrt{\chi^2_{\nu}}/\nu$, we have that $T = P^{-1}(\hat{\theta} - \theta)/\hat{\sigma} = N/(\hat{\sigma}/\sigma)$ is a standard $p$-dimensional $t$ random vector with $\nu$ degrees of freedom, and $||T|| \sim \sqrt{pF_{p, \nu}}$, where $F_{p, \nu}$ denotes an $F$ distributed random variable with $p$ and $\nu$ degrees of freedom. Define the polar coordinates $(R_T, \psi_{T1}, \cdots, \psi_{Tp-1})^{T1}$ of the vector $T = (T_0, T_1, \cdots, T_{p-1})^{T}$ by

\[
\begin{align*}
T_0 &= R_T \cos \psi_{T1} \\
T_1 &= R_T \sin \psi_{T1} \cos \psi_{T2} \\
T_2 &= R_T \sin \psi_{T1} \sin \psi_{T2} \cos \psi_{T3} \\
& \quad \cdots \quad \cdots \\
T_{p-2} &= R_T \sin \psi_{T1} \sin \psi_{T2} \cdots \sin \psi_{Tp-2} \cos \psi_{Tp-1} \\
T_{p-1} &= R_T \sin \psi_{T1} \sin \psi_{T2} \cdots \sin \psi_{Tp-2} \sin \psi_{Tp-1},
\end{align*}
\]

where

\[
\begin{align*}
0 &\leq \psi_{T1} \leq \pi \\
& \quad \cdots \quad \cdots \\
0 &\leq \psi_{Tp-2} \leq \pi \\
0 &\leq \psi_{Tp-1} < 2\pi \\
R_T &\geq 0.
\end{align*}
\]

The Jacobian of the transformation is

\[
|J| = R_{T}^{p-1} \sin^{p-2} \psi_{T1} \sin^{p-3} \psi_{T2} \cdots \sin \psi_{Tp-2}.
\]

It is well known (cf. Liu et al. 2014) that $||T|| = R_T$ and $(\psi_{T1}, \cdots, \psi_{Tp-1})^{T}$ are independent, and the joint density function of $(\psi_{T1}, \cdots, \psi_{Tp-1})^{T}$ is

\[
h(\psi_{T1}, \cdots, \psi_{Tp-1}) = \frac{1}{2} \pi^{-p/2} \Gamma(p/2) \sin^{p-2} \psi_{T1} \sin^{p-3} \psi_{T2} \cdots \sin \psi_{Tp-2},
\]

where $\Gamma(\cdot)$ denotes the gamma function.

\footnote{suggest using subscripts, like $(R_T, \psi_{T1}, \cdots, \psi_{Tp-1})^{T}$; here and elsewhere}
Now the acceptance probability in (4) can be expressed as

\[
\inf_{H_0} P\{ \frac{Y}{\sup_{x \in [a,b] \setminus k_0}} - \left( k_0 - x \right) \left[ \mathbf{P}g(k_0, x, p) \right]^T \mathbf{P}^{-1} \frac{\mathbf{P}^{-1}(\tilde{\theta} - \theta)/\hat{\sigma}}{\mathbf{P}g(k_0, x, p)} \leq c(k_0) \}
\]

\[
= P\{ Y : \sup_{x \in [a,b] \setminus k_0} - \left( k_0 - x \right) \left[ \mathbf{P}g(k_0, x, p) \right]^T \mathbf{T} \left\| \mathbf{P}g(k_0, x, p) \right\| \leq c(k_0) \}
\]

\[
= P\{ - \inf_{x \in [a,b]} \frac{\left[ \mathbf{P}g(k_0, x, p) \right]^T \mathbf{T}}{\left\| \mathbf{P}g(k_0, x, p) \right\|} \leq c(k_0) \quad \text{and} \quad \sup_{x \in [a,b]} \frac{\left[ \mathbf{P}g(k_0, x, p) \right]^T \mathbf{T}}{\left\| \mathbf{P}g(k_0, x, p) \right\|} \leq c(k_0) \}
\]

\[
= P\{ \left\| \mathbf{T} \| S_{k_0} \leq c(k_0) \}
\]

\[
P\{ \left\| \mathbf{T} \| \leq c(k_0) / S_{k_0} \}, \quad (5)
\]

where \( \mathbf{T} = \mathbf{P}^{-1}(\tilde{\theta} - \theta)/\hat{\sigma} \) and \( S_{k_0} = \max(Q_{k_0}, R_{k_0}) \) with

\[
Q_{k_0} = - \inf_{x \in [a,b]} \frac{\left[ \mathbf{P}g(k_0, x, p) \right]^T \mathbf{u}}{\left\| \mathbf{P}g(k_0, x, p) \right\|}, \quad R_{k_0} = \sup_{x \in [a,b]} \frac{\left[ \mathbf{P}g(k_0, x, p) \right]^T \mathbf{u}}{\left\| \mathbf{P}g(k_0, x, p) \right\|}
\]

and \( \mathbf{u} = \mathbf{T}/\|\mathbf{T}\| \). Note that \( Q_{k_0} \) and \( R_{k_0} \) depend on \( \mathbf{T} \) only through \( (\psi_{T1}, \cdots, \psi_{Tp})^T \). Since \( \|\mathbf{T}\| \) and \( (\psi_{T1}, \cdots, \psi_{Tp-1})^T \) are independent and \( \|\mathbf{T}\| \sim \sqrt{pF_{p,\nu}} \), the expression in (5) can be written as

\[
\int_{\psi_{T1}=0}^{\pi} \cdots \int_{\psi_{Tp-2}=0}^{\pi} \int_{\psi_{Tp-1}=0}^{2\pi} h(\psi_{T1}, \cdots, \psi_{Tp-2}, \psi_{Tp-1})
\]

\[
\times F_{p,\nu}(c_{k_0}^2/(pS_{k_0}^2(\psi_{T1}, \cdots, \psi_{Tp-2}, \psi_{Tp-1}))) d\psi_{T1} \cdots d\psi_{Tp-2} d\psi_{Tp-1}, \quad (6)
\]

where \( F_{p,\nu}(\cdot) \) denotes the cumulative distribution function of an \( F \) random variable.

For a given \( (\psi_{T1}, \cdots, \psi_{Tp-1}) \), the method given in WLHB (2015) can be used to accurately and quickly compute \( Q_{k_0}, R_{k_0} \) and so \( S_{k_0} \). The function \( F_{p,\nu}(\cdot) \) can be computed efficiently by using the Matlab built-in function \( fcdf \). Hence the integral in expression (6) can be computed quickly at least for small values of \( p \). For example, expression (6) involves one dimensional integration for \( p = 2 \) and three dimensional integration for \( p = 4 \).

WLHB (2015) provided an alternative expression for the acceptance probability when \( p = 2 \), which also involves one-dimensional integration. As the derivation of that expression uses the geometry of the acceptance region, the method is difficult to be generalized. On the other hand, the expression (6) is derived using only algebra. From the expression (6), one can compute the acceptance probability for each given value of \( c(k_0) \). A numerical searching method, such as the bisection method, can then be used to find the critical constant \( c(k_0) \) so that the acceptance probability is equal to \( 1 - \alpha \). This is used in Wan et al. (2015) for \( p = 2 \) only to find \( c(k_0) \), which is then used to determine whether each \( k_0 \) belongs to the confidence set \( C_E(Y) \).
Note, however, that in the process of finding \( c(k_0) \), the acceptance probability in (6) needs to be computed several times for different candidate values of \( c(k_0) \). In the next section, it is shown that the acceptance probability in (6) needs to be computed only once for one particular value of \( c(k_0) \) in order to determine whether \( k_0 \) belongs to the confidence set \( C_E(Y) \).

### 4 P-value Method

From the definition of the acceptance set \( A(k_0) \) and the confidence set \( C_E(Y) \) in Section 2, a given point \( k_0 \in [a, b] \) belongs to \( C_E(Y) \) if and only if

\[
P\left\{ \sup_{x \in [a, b] \setminus k_0} \frac{(k_0 - x)}{|k_0 - x|} \left[ \frac{P g(k_0, x, p)^T T}{\sqrt{g(k_0, x, p)^T \Sigma g(k_0, x, p)}} \right] \leq c^* \right\} \leq 1 - \alpha, \tag{7}
\]

where

\[
c^* = \sup_{x \in [a, b] \setminus k_0} \frac{(k_0 - x)}{|k_0 - x|} \frac{g(k_0, x, p)^T \hat{\theta} / \hat{\sigma}}{\sqrt{g(k_0, x, p)^T \Sigma g(k_0, x, p)}}
\]

with \( \hat{\theta} \) and \( \hat{\sigma} \) being the estimates of \( \theta \) and \( \sigma \) based on the observed data. For the given \( \hat{\theta} \) and \( \hat{\sigma} \), \( c^* \) can be computed accurately and quickly by using the method given in WLHB (2015). Hence we need to compute the acceptance probability in (7) only once and compare it with \( 1 - \alpha \) in order to determine whether \( k_0 \) is in \( C_E(Y) \). This is much faster than the computation of \( c(k_0) \), which requires the computation of several acceptance probabilities.

In essence, this method uses the p-value of the test based on the acceptance set \( A(k_0) \) for testing \( H_0 : k = k_0 \) to determine whether \( H_0 \) is accepted or rejected. We therefore call this the p-value method.

### 5 Examples

In this section two data examples, one from WLHB (2015) and the other from Liu et al. (2014), are used to demonstrate that the p-value method is substantially more efficient than the WLHB method. All the computations were done on an ordinary windows laptop (Windows edition: Windows Vista™ Home Basic, Processor: Intel(R) Core(TM)2 Duo CPU T7100 @ 1.80GHz, Installed memory: 2534 MB, System type: 32-bit Operating System).

WLHB showed that \( c_0 = \sqrt{pF(1 - \alpha; p, v)} \) is a conservative critical value which can be used to quickly compute the conservative confidence set \( C_0(Y) \). Moreover, the exact \( 1 - \alpha \) level confidence set \( C_E(Y) \) can be computed efficiently by checking
only the points $k_0$ in $C_0(Y)$ to see whether they belong to $C_E(Y)$. This is used in our programs for the computation of the two examples.

The first example involves a fourth order polynomial regression model of the transformed perinatal mortality rate (PMR), $Y = \log(-\log(PMR))$, and the birth weight (BW) ($x$); see WLHB (2015, Example 1) for more details. Based on the data, we have $\hat{\theta} = (4.18, 1.80, 0.42, -0.04)^T$, 

$$
\Sigma = \begin{pmatrix}
175.673 & -116.729 & 31.771 & -3.039 \\
* & 78.591 & -21.628 & 2.087 \\
* & * & 6.010 & -0.585 \\
* & * & * & 0.057
\end{pmatrix},
$$

$\hat{\sigma} = 0.0567$ and $v = 30$. It is of interest to identify the BW level in the observed range $[a, b] = [0.85, 4.25]$ that maximizes the response $Y$ (i.e., minimizes the PMR). We compare the methods from WLHB with the proposed method in this paper by constructing a $1 - \alpha = 95\%$ level confidence set for this optimal BW level. We replace the interval $[a, b]$ by the grid of points that are $d = 0.01$ distance apart, and check each $k_0$ in the grid to see whether it belongs to the confidence set.

We first construct the conservative confidence set $C_0(Y)$ by using $c_0 = \sqrt{4F(0.95; 4, 30)} = 3.280$, which is given by $C_0(Y) = [3.72, 4.25]$ and takes 0.4194 seconds to compute. Next, we check whether each grid point $k_0$ in $C_0(Y)$ belongs to the exact confidence set $C_E(Y)$. The WLHB method computes each $c(k_0)$ based on $n = 100,000$ simulations; this takes about 22.0123 seconds for each $k_0$, and the accuracy of the acceptance probability is about $1 - \alpha \pm 0.002$ (WLHB, 2015, pp. 564); The exact confidence set is $C_E(Y) = [3.75, 4.21]$, taking 3015.6904 seconds all together.

The new method computes one acceptance probability for each grid point $k_0$ in $C_0(Y)$, using 3-dimensional numerical quadrature. Setting the accuracy of numerical quadrature at $1 - \alpha \pm 0.001$, the computation of one acceptance probability takes about 0.8570 seconds. The exact confidence set is $C_E = [3.76, 4.20]$, taking 117.8277 seconds all together. It is clear from this example that the computation time of the new method is only about 3.91% of that used by the WLHB method, while the new method achieves a better accuracy than the WLHB method.

The second example involves a third order polynomial regression model for modelling the mean dose response; see Liu et al. (2014, Example 2) for more details. Based on the data, we have $\hat{\theta} = (0.0953, -5.186 \times 10^{-4}, 7.5 \times 10^{-7})^T$ and 

$$
\Sigma = \begin{pmatrix}
8.6247 \times 10^{-5} & -6.3460 \times 10^{-7} & 1.0699 \times 10^{-9} \\
* & 5.0830 \times 10^{-9} & -8.8834 \times 10^{-12} \\
* & * & 1.5787 \times 10^{-14}
\end{pmatrix}.
$$
It is of interest to identify the dose level in the observed range \([a, b] = [0, 400]\) that maximizes the mean response. We use the methods of WLHB and this paper to construct a \(1 - \alpha = 95\%\) level confidence set for this optimal dose level. We replace the interval \([a, b]\) by the grid of points that are \(d = 1\) distance apart, and check each \(k_0\) in the grid whether it belongs to the confidence set.

The conservative confidence set \(C_0(Y)\) with \(c_0 = \sqrt{3F(0.95; 3, 38)} = 2.9249\) is \(C_0(Y) = [109, 166]\), taking 0.2736 seconds. Next, we check whether each grid point \(k_0\) in \(C_0(Y)\) belongs to the exact confidence set \(C_E(Y)\). The WLHB method computes each \(c(k_0)\) based on \(n = 100,000\) simulations; this takes about 54.6896 seconds for each \(k_0\), and the accuracy of the acceptance probability is about \(1 - \alpha \pm 0.002\); The exact confidence set is \(C_E(Y) = [111, 156]\), taking 3172 seconds all together. The new method computes one acceptance probability for each grid point \(k_0\) in \(C_0(Y)\), using one 2-dimensional numerical quadrature. Setting the accuracy of numerical quadrature at \(1 - \alpha \pm 0.001\), the computation of one acceptance probability takes about 0.0975 seconds. The exact confidence set is \(C_E(Y) = [112, 155]\), taking 7.48 seconds all together. It is clear from this example that the computation time of the new method is only about 0.23% of that used by the WLHB method, while the new method achieves a better accuracy than the WLHB method.

For quadratic polynomial regression, WLHB (2015) already used 1-dimensional quadrature to compute the acceptance probability, employing the Matlab built-in function \texttt{fzero} to search for the critical constant \(c(k_0)\) for each given \(k_0\). As \texttt{fzero} combines efficiently the bisection, secant and inverse quadratic interpolation methods, it often takes no more than four computations of the acceptance probabilities to find the \(c(k_0)\). Also, the computation of one acceptance probability, involving one 1-dimensional numerical quadrature, often takes a fraction of a second, e.g. 0.001 second, and the total computation time of \(C_E(Y)\) is just a few seconds, e.g. about 18 second for the example given in WLHB (2015), when the computational error tolerance is set at \(10^{-4}\). Hence the computational saving of the new method of this paper over the WLHB (2015) method is not of practical importance for quadratic polynomial regression.

From the examples above and several other data sets we have tried, the p-value method requires substantially less time to compute the confidence set \(C_E(Y)\) than the WLHB method for cubic and quartic polynomial regressions at least.
6 Conclusions

In this article, the new p-value method is given for computing the confidence set $C_E(Y)$ of WLHB (2015).

The p-value method hinges on the efficient computation of an acceptance probability, and this is accomplished by using numerical quadrature for at least $p \leq 4$, involving no more than a 3-dimensional integral, which is much faster than using statistical simulation to achieve a better accuracy. It is computationally efficient also because it needs to compute one acceptance probability only, while finding the critical constant $c(k_0)$ requires the computation of several acceptance probabilities, when judging whether $k_0$ belongs to $C_E(Y)$.

As the polynomial regression of order higher than four is rarely used in applications, the new method is sufficient for most real problems. Matlab programs for implementing the new method for $p \leq 4$ have been written and are available from the authors, which make the new method easily applicable.

References

[1] Wan, F., Liu, W., Bretz, F., and Han, Y., 2015. An exact confidence set for a maximum point of a univariate polynomial function in a given interval. Technometrics. 57, 559-565.

[2] Wan, F., Liu, W., Bretz, F., and Han, Y., 2016. Confidence sets for optimal factor levels of a response surface. Biometrics, to appear.

[3] Liu, W., Ah-Kine, P., Zhou, S., 2015. Simultaneous confidence bands for linear regression with covariates constrained in intervals. Scand. J. Stat. 39(3): 543-553.

[4] Liu, W., Zhou, S., and Bretz, F., 2014. Exact simultaneous confidence bands for quadratic and cubic polynomial regression with applications in dose response study. Aust. N. Z. J. Stat. 55(4), 421-434.