Influence Analysis of the Common Mean Problem

Myung Geun Kim¹,a

¹Department of Mathematics Education, Seowon University

Abstract

Two influence diagnostic methods for the common mean model are proposed. First, an investigation of the influence of observations according to minor perturbations of the common mean model is made by adapting the local influence method which is based on the likelihood displacement. It is well known that the maximum likelihood estimates are in general sensitive to influential observations. Case-deletions can be a candidate for detecting influential observations. However, the maximum likelihood estimators are iteratively computed and therefore case-deletions involve an enormous amount of computations. An approximation by Newton’s method to the maximum likelihood estimator obtained after a single observation was deleted can reduce much of computational burden, which will be treated in this work. A numerical example is given for illustration and it shows that the proposed diagnostic methods can be useful tools.

Keywords: Case deletions, common mean, local influence, Newton method.

1. Introduction

The common mean problem has a long history of research since the pioneering work by Graybill and Deal (1959). It has real-life applications in wide areas, for example in clinical trials (Kelleher, 1996), in soil engineering (Zacks, 1966, 1970), in a balanced incomplete block design (Montgomery, 1997; Pal et al., 2007), and in meta analysis (Chang and Pal, 2008). Some detailed references about point estimation of the common mean can be found in Kubokawa (1991) and Ma et al. (2011). Recently Pal et al. (2007) studied the maximum likelihood estimation of the parameters of the common mean model. For estimating the common mean with unknown variances, they showed that the maximum likelihood estimator has better overall performance than the Graybill-Deal estimator.

In this work, an investigation of the influence of observations according to minor perturbations of the common mean model is considered. In Section 2, it is performed by adapting the local influence method introduced by Cook (1986) which is based on the likelihood displacement. It is well known that the maximum likelihood estimates are in general sensitive to influential observations. Thus we need some diagnostic methods of detecting observations that have large influence on the estimates. To this end case-deletions can be a candidate. As will be seen later, the maximum likelihood estimators are iteratively computed and therefore case-deletions involve an enormous amount of computations. An approximation by Newton’s method to the maximum likelihood estimator obtained after a single observation was deleted can reduce much of computational burden, which will be treated in Section 3. A numerical example is given for illustration in Section 4.

¹Professor, Department of Mathematics Education, Seowon University, Cheongju 361-742, Korea.
E-mail: mgkim@seowon.ac.kr
2. Local Influence

2.1. Maximum likelihood estimation

We consider the common mean model characterized by two normal distributions with common mean, \( N(\mu, \sigma_1^2) \) and \( N(\mu, \sigma_2^2) \). A random sample \( \{x_i, \ldots, x_{im}\} \) is drawn from \( N(\mu, \sigma_i^2) \) for each \( i = 1, 2 \) and both samples are assumed to be independent. The joint log-likelihood function of \( \mu, \sigma_1^2, \sigma_2^2 \) based on both samples can be written as

\[
L(\mu, \sigma_1^2, \sigma_2^2) = -\sum_{i=1}^{n} \frac{n_i}{2} \left[ \log(2\pi) + \log(\sigma_i^2) + \frac{s_i^2}{\sigma_i^2} + \frac{1}{\sigma_i^2} (\bar{x}_i - \mu)^2 \right],
\]

where

\[
\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad \text{and} \quad s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2.
\]

Let \( \hat{\mu}, \hat{\sigma_1^2}, \hat{\sigma_2^2} \) be the maximum likelihood estimators (MLEs) of \( \mu, \sigma_1^2, \sigma_2^2 \), respectively and then they are found by iteratively solving the following likelihood equations

\[
\begin{align*}
\hat{\mu} &= \frac{\sum_{i=1}^{2} n_i \bar{x}_i / \sum_{i=1}^{2} n_i \hat{\sigma}_i^2,} \\
\hat{\sigma}_1^2 &= s_1^2 + (\bar{x}_1 - \hat{\mu})^2, \quad (i = 1, 2).
\end{align*}
\]

Since \( \bar{x}_i - \hat{\mu} = n_2 \hat{\sigma}_2^2 (\bar{x}_1 - \bar{x}_2)/(n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2) \), an alternative form for \( \hat{\sigma}_1^2 \) becomes

\[
\hat{\sigma}_1^2 = s_1^2 + \left[ \frac{n_2 \hat{\sigma}_2^2}{n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2} \right] (\bar{x}_1 - \bar{x}_2)^2.
\]

Exchanging the role of the subscripts, we can have an equivalent form for \( \hat{\sigma}_2^2 \). More details about the maximum likelihood estimation can be found in Pal et al. (2007).

For later use, let \( \hat{L} = \partial^2 L(\mu, \sigma_1^2, \sigma_2^2) / \partial \theta \partial \theta' \) evaluated at \( \theta = \hat{\theta} \), where \( \theta = (\mu, \sigma_1^2, \sigma_2^2)' \). Then \( \hat{L} \) is a 3 \times 3 matrix whose components are computed as

\[
\begin{align*}
\frac{\partial^2 L(\mu, \sigma_1^2, \sigma_2^2)}{\partial \mu^2} &= -\frac{n_1}{\sigma_1^2} - \frac{n_2}{\sigma_2^2}, \\
\frac{\partial^2 L(\mu, \sigma_1^2, \sigma_2^2)}{\partial \mu \partial (\sigma_1^2)} &= -\frac{n_1}{\sigma_1^2} (\bar{x}_1 - \hat{\mu}), \\
\frac{\partial^2 L(\mu, \sigma_1^2, \sigma_2^2)}{\partial \mu \partial (\sigma_2^2)} &= -\frac{n_2}{\sigma_2^2} (\bar{x}_2 - \hat{\mu}), \\
\frac{\partial^2 L(\mu, \sigma_1^2, \sigma_2^2)}{\partial (\sigma_1^2)^2} &= -\frac{n_1}{2} \left[ \frac{1}{\sigma_1^4} + \frac{2s_1^2}{\sigma_1^6} + \frac{2 (\bar{x}_1 - \hat{\mu})^2}{\sigma_1^6} \right] = -\frac{n_1}{2\sigma_1^4}, \quad (i = 1, 2), \\
\frac{\partial^2 L(\mu, \sigma_1^2, \sigma_2^2)}{\partial (\sigma_1^2) \partial (\sigma_2^2)} &= 0,
\end{align*}
\]

where all of the above partial derivatives are evaluated at \( \theta = \hat{\theta} \).
2.2. Perturbed model

Let $n = n_1 + n_2$. A perturbation vector $w = (w_1, \ldots, w_n)^T$ in the $n$-dimensional Euclidean space characterizes the perturbed model, and it can be expressed as $w = 1_α + αd$, where $1_α$ is an $n \times 1$ vector whose elements are all equal to one, the vector $d$ of unit length represents the direction towards which the perturbation $w$ moves, and the scalar $α$ denotes the magnitude of the perturbation $w$ along the direction $d$. We consider a perturbed model in which

\[
\begin{align*}
  x_{1j} &\sim N(μ, \frac{σ_1^2}{w_j}), \quad (j = 1, \ldots, n_1), \\
  x_{2k} &\sim N(μ, \frac{σ_2^2}{w_{n_1+k}}), \quad (k = 1, \ldots, n_2).
\end{align*}
\]

Under this perturbation scheme, the perturbed joint log-likelihood function of $μ, σ_1^2, σ_2^2$ based on both samples becomes

\[
L(μ, σ_1^2, σ_2^2 | w) = -\frac{n}{2} \log(2π) + \frac{1}{2} \sum_{j=1}^{n_1} \log(w_j) - \frac{n_1}{2} \log(σ_1^2) - \frac{1}{2} \sum_{j=1}^{n_1} \frac{w_j(x_{1j} - μ)^2}{σ_1^2} \\
- \frac{n_2}{2} \log(σ_2^2) - \frac{1}{2} \sum_{k=1}^{n_2} \frac{w_{n_1+k}(x_{2k} - μ)^2}{σ_2^2}.
\]

Let $\hat{μ}(w), \hat{σ}_1^2(w), \hat{σ}_2^2(w)$ be the MLEs of $μ, σ_1^2, σ_2^2$ under the above perturbed model, respectively, and then it is easily seen that they are also iteratively computed by the following equations

\[
\begin{align*}
  \hat{μ}(w) &= \frac{\sum_{j=1}^{n_1} w_j x_{1j} / \hat{σ}_1^2(w) + \sum_{k=1}^{n_2} w_{n_1+k} x_{2k} / \hat{σ}_2^2(w)}{\sum_{j=1}^{n_1} w_j / \hat{σ}_1^2(w) + \sum_{k=1}^{n_2} w_{n_1+k} / \hat{σ}_2^2(w)}, \\
  \hat{σ}_1^2(w) &= \frac{1}{n_1} \sum_{j=1}^{n_1} w_j (x_{1j} - \hat{μ}(w))^2, \\
  \hat{σ}_2^2(w) &= \frac{1}{n_2} \sum_{k=1}^{n_2} w_{n_1+k} (x_{2k} - \hat{μ}(w))^2.
\end{align*}
\]

Since $\hat{σ}_1^2 = s_1^2 + (\bar{x}_1 - \hat{μ})^2 = (1/n_1) \sum_{j=1}^{n_1} (x_{1j} - \hat{μ})^2$ in the unperturbed model, the form of the $\hat{σ}_1^2(w)$ in the perturbed model can easily be guessed just as given in the above.

Let $Δ = \frac{∂^2 L(μ, σ_1^2, σ_2^2 | w)}{∂ μ∂ w_j}$ evaluated at $θ = \hat{θ}$ and $w = 1_α$. A closed form of $Δ$ is not available but its components are easily computed as

\[
\begin{align*}
  \frac{∂^2 L(μ, σ_1^2, σ_2^2 | w)}{∂ μ∂ w_j} &= \begin{cases} 
    \frac{x_{1j} - \hat{μ}}{\hat{σ}_1^2}, & (1 \leq j \leq n_1), \\
    \frac{x_{2k} - \hat{μ}}{\hat{σ}_2^2}, & (j = n_1 + k, 1 \leq k \leq n_2), \\
  \end{cases} \\
  \frac{∂^2 L(μ, σ_1^2, σ_2^2 | w)}{∂ (σ_1^2)∂ w_j} &= \begin{cases} 
    \frac{1}{2} (x_{1j} - \hat{μ})^2 / \hat{σ}_1^4, & (1 \leq j \leq n_1), \\
    0, & (n_1 + 1 \leq j \leq n),
  \end{cases}
\]
3. Case Deletion Diagnostics

Refer to O’Neill (2006) for more details about diagnostics. The largest absolute eigenvalue of \( \mathbf{2} \mathbf{\mathbf{\dot{F}}} \) which is the portion of the surface cut out by the plane spanned by the vectors \( \mathbf{1} \) would be influential. The largest absolute eigenvalue of \( \mathbf{2} \mathbf{\mathbf{\dot{F}}} \) is the maximum curvature of the curve which is the portion of the surface cut out by the plane spanned by the vectors \( \mathbf{1} \) and \( \mathbf{l_{\text{max}}} \), where \( \mathbf{l_{\text{max}}} \) is the \((n+1) \times 1\) vector with its \((n+1)\)st element equal to one and the others being zero. Let \( \mathbf{LD}(w) = 2 \left[ L(\hat{\mu}, \hat{\sigma}^2_1, \hat{\sigma}^2_2) - L(\hat{\mu}(w), \hat{\sigma}^2_1(w), \hat{\sigma}^2_2(w)) \right] \).

The surface of interest is formed by the vector \((w, \mathbf{LD}(w))\) as \( w \) varies over a certain space. The eigenvector \( \mathbf{l_{\text{max}}} \) corresponding to the largest absolute eigenvalue of \( \mathbf{2} \mathbf{\mathbf{\dot{F}}} \) provides information about influential observations. Observations associated with relatively large absolute component of \( \mathbf{l_{\text{max}}} \) would be influential. The largest absolute eigenvalue of \( \mathbf{2} \mathbf{\mathbf{\dot{F}}} \) is the maximum curvature of the curve which is the portion of the surface cut out by the plane spanned by the vectors \( \mathbf{1} \) and \( \mathbf{l_{\text{max}}} \). Let \( \mathbf{l_{\text{max}}} \) is the \((n+1) \times 1\) vector with its \((n+1)\)st element equal to one and the others being zero. Refer to O’Neill (2006) for more details about differential geometry.

2.3. Local influence procedure

In this subsection, we will briefly describe the local influence introduced by Cook (1986) which can be adapted to the common mean model with computational results obtained in the previous subsections. Let \( \mathbf{F} = \mathbf{\Delta^T L^{-1} \Delta} \). Then we can compute \( \mathbf{F} \), numerically. In the case of the common mean model, the likelihood displacement can be defined by

\[
LD(w) = 2 \left[ L(\hat{\mu}_1, \hat{\sigma}^2_1, \hat{\sigma}^2_2) - L(\hat{\mu}(w), \hat{\sigma}^2_1(w), \hat{\sigma}^2_2(w)) \right].
\]

The surface of interest is formed by the vector \((w, LD(w))\) as \( w \) varies over a certain space. The eigenvector \( l_{max} \) corresponding to the largest absolute eigenvalue of \( 2F \) provides information about influential observations. Observations associated with relatively large absolute component of \( l_{max} \) would be influential. The largest absolute eigenvalue of \( 2F \) is the maximum curvature of the curve which is the portion of the surface cut out by the plane spanned by the vectors \( 1 \) and \( l_{max} \). The influence of deleting the \( r^{-}\)th observation of the second sample can be treated, which is therefore not reported here. Let \( L(\hat{\mu}, \hat{\sigma}^2_1, \hat{\sigma}^2_2) \) be the joint log-likelihood based on the reduced sample with the \( r^{-}\)th observation of the first sample deleted. Similarly, the case of deleting the \( r^{-}\)th observation of the second sample can be treated, which is therefore not reported here. Let \( L_r = \partial L(\hat{\mu}_1, \hat{\sigma}^2_1, \hat{\sigma}^2_2) / \partial \theta \) evaluated at \( \theta = \hat{\theta} \) and \( L_{(\hat{\mu}_1, \hat{\sigma}^2_1, \hat{\sigma}^2_2)} \) be the joint log-likelihood based on the reduced sample with the \( r^{-}\)th observation of the second sample deleted.

The influence of deleting the \( r^{-}\)th observation on the MLE \( \hat{\theta} \) can be investigated by the quantity \( \hat{\theta} - \hat{\theta}_r \). If \( \hat{\theta}_r \) resides far away from \( \hat{\theta} \), then the \( r^{-}\)th observation is regarded as influential one. As in \( \hat{\theta} \), an explicit form of \( \hat{\theta}_r \) is not available and for each \( r \) we should take steps of iteratively estimating the parameters, which incurs a heavy computational burden. To avoid this computational difficulty, we can use an approximation to \( \hat{\theta}_r \), which can be obtained by the Newton’s method as

\[
\hat{\theta}_r = \hat{\theta} - L(\hat{\mu}_1, \hat{\sigma}^2_1, \hat{\sigma}^2_2)
\]

For more details about the Newton’s method, refer to Chap. 5 of Cook and Weisberg (1982) or Chap. 10 of Kennedy and Gentle (1980). An approximation to \( \hat{\theta}_r \) given in the above depends on \( \hat{\theta} \) for all \( r \). Hence there is no need to do an estimation process again and this approximation reduces computational burden. An application of Newton method to case deletion diagnostics in other fields of statistics, for example in nonlinear structural equation model can be found in Lee and Lu (2003).

Based on the reduced sample with the \( r^{-}\)th observation deleted from the \( i^{-}\)th sample, the sample mean and the sample variance are

\[
\bar{x}_{(i-r)} = \frac{1}{n_i - 1} \sum_{j \neq r} x_{ij} = \frac{1}{n_i - 1} x_i - \frac{1}{n_i - 1} x_{ir},
\]

\[
\hat{x}^2_{(i-r)} = \frac{1}{n_i - 1} \sum_{j \neq r} (x_{ij} - \bar{x}_{(i-r)})^2 = \frac{n_i}{n_i - 1} \hat{x}^2_i - \frac{n_i}{(n_i - 1)^2} (x_{ir} - \bar{x}_i)^2.
\]
The joint log-likelihood function of \( \mu, \sigma_1^2, \sigma_2^2 \) based on the reduced sample is

\[
L_{(-r)} = -\frac{n_1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_1^2) - \frac{n_1}{2} \frac{(\bar{x}_{(-r)} - \mu)^2}{\sigma_1^2} - \frac{n_2}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_2^2) - \frac{n_2}{2} \frac{(\bar{x}_2 - \mu)^2}{\sigma_2^2}.
\]

For computing \( L_{(-r)} \), we need the following computational results:

\[
\frac{\partial L_{(-r)}}{\partial \mu} = n_1 \frac{1}{\sigma_1^2} (\bar{x}_{(-r)} - \mu) + n_2 \frac{1}{\sigma_2^2} (\bar{x}_2 - \mu),
\]

\[
\frac{\partial L_{(-r)}}{\partial (\sigma_1^2)} = \frac{n_1 - 1}{2} \frac{1}{\sigma_1^4} \left[ \frac{s_{(-r)}^2}{\sigma_1^4} - \frac{(\bar{x}_{(-r)} - \mu)^2}{\sigma_1^4} \right],
\]

\[
\frac{\partial L_{(-r)}}{\partial (\sigma_2^2)} = \frac{n_2}{2} \frac{1}{\sigma_2^4} \left[ \frac{s_2^2}{\sigma_2^4} - \frac{(\bar{x}_2 - \mu)^2}{\sigma_2^4} \right] = 0.
\]

Also, the elements of \( L_{(-r)} \) are computed as

\[
\frac{\partial^2 L_{(-r)}}{\partial \mu^2} = -\frac{n_1}{\sigma_1^4}, \quad \frac{\partial^2 L_{(-r)}}{\partial \sigma_1^4} = -\frac{n_2}{\sigma_2^4},
\]

\[
\frac{\partial^2 L_{(-r)}}{\partial \mu \partial (\sigma_1^2)} = \frac{n_1 - 1}{\sigma_1^6} (\bar{x}_{(-r)} - \mu),
\]

\[
\frac{\partial^2 L_{(-r)}}{\partial \mu \partial (\sigma_2^2)} = \frac{n_2}{\sigma_2^6} (\bar{x}_2 - \mu),
\]

\[
\frac{\partial^2 L_{(-r)}}{\partial (\sigma_1^2)^2} = \frac{n_1 - 1}{2} \frac{1}{\sigma_1^6} \left[ \frac{2s_{(-r)}^2}{\sigma_1^6} + \frac{(\bar{x}_{(-r)} - \mu)^2}{\sigma_1^6} \right],
\]

\[
\frac{\partial^2 L_{(-r)}}{\partial (\sigma_2^2)^2} = \frac{n_2}{\sigma_2^6} \left[ \frac{2s_2^2}{\sigma_2^6} + \frac{(\bar{x}_2 - \mu)^2}{\sigma_2^6} \right] = -\frac{n_2}{2\sigma_2^4},
\]

\[
\frac{\partial^2 L_{(-r)}}{\partial (\sigma_1^2) \partial (\sigma_2^2)} = 0.
\]

Note that all of the above partial derivatives are evaluated at \( \hat{\theta} \).

4. A Numerical Example

In this section we will illustrate the local influence method and an approximation to case-deletions with a data set taken from Problem 14.10 of Neter et al. (1990) in which data on productivity improvements for three levels of firms were collected. Here we will investigate the first two levels: Low level for the first group and Moderate level for the second group. The observations for the first group are labeled as 1 to 9 in order and those for the second group as 10 to 21.
The maximum likelihood estimates based on the full sample are $\hat{\mu} = 7.84$, $\hat{\sigma}_1^2 = 1.52$ and $\hat{\sigma}_2^2 = 0.61$. The largest absolute eigenvalue of $2\hat{F}$ is 2.46. An index plot of $l_{\text{max}}$ is included in Figure 1(a). It shows that observations 10, 20, 12 are potentially influential and observation 4 is a little bit influential.

Single case deletions are performed. Index plots for the values $\hat{\mu} - \hat{\mu}_{(-r)}$, $\hat{\sigma}_1^2 - \hat{\sigma}_{1(-r)}^2$, $\hat{\sigma}_2^2 - \hat{\sigma}_{2(-r)}^2$ are displayed in Figures 1(b) to (d), respectively. We can see that observation 10 has the largest influence on all of the estimates $\hat{\mu}$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$. Observation 4 also influences the estimation of $\hat{\sigma}_1^2$. We can see that the role of observation 10 in estimating $\hat{\sigma}_1^2$ is opposite to that of observation 4 because deletion of observation 10 increases the estimate $\hat{\sigma}_1^2$ while deletion of observation 4 decreases it. The influence of observation 12 is remarkable in estimating $\hat{\sigma}_2^2$ compared with the other parameters. Observations 10 and 12 have a similar role in estimating $\hat{\sigma}_2^2$ in that deleting each of them reduces the estimate $\hat{\sigma}_2^2$. Observation 20 has a little influence on $\hat{\mu}$. This example shows that case deletions yield some useful information about the behavior of observations in estimating the parameters.

References


Received April 2, 2013; Revised April 25, 2013; Accepted April 25, 2013