

リッジ回帰における尤度距離による影響力評価

竹内 秀一

Assessment of Influence Based on Likelihood Distance in Ridge Regression

Hidekazu TAKEUCHI

Some influence measures based on the case deletion procedure are proposed in linear regression. Several of them are applied to ridge regression. An influence measure based on the log-likelihood is taken up in this paper. It is called the likelihood distance. The definition of the likelihood distance in ridge regression is given by the same way as in linear regression. Then the likelihood distance in ridge regression is reduced to an appropriate expression for the sake of assessment of the influence with the case deletion procedure. Some characteristics of the likelihood distance are investigated in contrast to Cook's distance both in linear regression and in ridge regression. The likelihood distance in ridge regression has a peculiar characteristic on the change of the ridge parameter. Furthermore a numerical example for a real data set shows that the likelihood distance in ridge regression is effective for the assessment of the influence of the observation(s).

1 はじめに

線形回帰分析における観測値の影響力評価を、Hoerl and Kennard [6] によって提案されたリッジ回帰 (ridge regression) へ応用することを考える。リッジ回帰における影響力評価 (竹内 [12] を参照) のための診断統計量 (influence measure) として、ノルム化診断統計量である Cook の距離 (Cook's distance) や行列式型診断統計量である一般化分散比 (covariance ratio) などを取り上げてきた。本研究では、これらに加えて、回帰係数の対数尤度規準に基づく診断統計量である尤度距離 (likelihood distance) をリッジ回帰へ応用することを検討する。また、これまで明確に示されていなかった尤度距離と Cook の距離との関係についても言及する。なお、観測値の影響力評価については、個々に観測値を除去する方法 (以

下、観測値除去法)に基づく場合、つまり single-case のみを考える。複数個の観測値を同時に除去する方法に基づく場合、つまり multiple-case については、リッジ回帰においては複雑なベクトルあるいは行列表現になるので取り上げない。

尤度距離を線形回帰に適用したときの性質については、竹内 [13] において数値実験により検証されている。本研究においては、尤度距離をリッジ回帰に適用した場合の性質について、実際のデータ解析に基づいた検討をする。この結果から、リッジ回帰の特性であるリッジパラメータの変化(増加させること)に対して、観測値により尤度距離が減少してから増加する場合と、増加のみを続ける場合があることがわかる。前者の場合は、リッジ回帰を適用することにより観測値の影響力を小さくする(縮小する)ことが可能であるが、後者の場合は、観測値の影響力がより大きくなるのでリッジ回帰の適用に注意を要することが特徴として挙げられる。

また、尤度距離と Cook の距離との関係はこれまであまり研究されてこなかったが、本研究においてある一つの関係式の導出を試みる。これにより通常の線形回帰とリッジ回帰の両方の場合について同様な関数関係を示すことができる。この関数関係により、観測値の影響力評価においては尤度距離と Cook の距離は同じような傾向を示すことがわかる。さらに、リッジ回帰においては、リッジパラメータの変化に対しても、尤度距離と Cook の距離は似通った傾向があることを、実際のデータ解析に基づいて示す。

本論文の構成は以下のとおりである。第2節では基本的な各種の統計量の定義を与える。第3節において、リッジ回帰における尤度距離を定義し、さらに Cook の距離との関係式を導く。第4節では、実際のデータに基づく数値例を示す。第5節は全体のまとめと今後の課題である。

2 定義

本節では、線形回帰モデルにおけるリッジ推定量を導入し、診断統計量として Cook の距離の定義をする。また、個々の観測値の影響力を評価するために、観測値除去法に基づく基本的な統計量を定義する。

2.1 線形回帰およびリッジ回帰

ここでは、線形回帰モデルとして、

$$y = X\beta + \varepsilon$$

を考える。このとき、 y は $n \times 1$ の目的変数ベクトル、 X は $n \times p$ のフルランクの説明変数行列、 β は $p \times 1$ の回帰係数ベクトル、そして ε は $n \times 1$ の誤差ベクトルであり、正規分布 $N(0, \sigma^2 I_n)$ に従うものとする。ただし、 I_n は n 次の単位行列を表す。また、 β の最小2乗推定

量は $\hat{\beta} = (X'X)^{-1}X'y$ として得られ、 σ^2 の不偏推定量は $\hat{\sigma}^2 = e'e/(n-p)$ となる。ただし、「 $'$ 」は行列あるいはベクトルの転置を表し、 e は残差ベクトルであり、 $e = y - X\hat{\beta} = (I_n - H)y$ である。このとき、 H は説明変数行列 X から構成されるハット行列 (hat matrix) $H = X(X'X)^{-1}X'$ であり、その第 i 対角成分 h_{ii} がてこ比である。ただし、 $1/n \leq h_{ii} < 1$ とする。さらに、残差ベクトル e の第 i 成分 e_i を標準化した $t_i = e_i/(\hat{\sigma}\sqrt{1-h_{ii}})$ を標準化残差 (内的スチューデント化残差) と呼び、 t_i の定義式において、 $\hat{\sigma}$ の代わりに $\hat{\sigma}_{(i)}$ を用いた $t_i^* = e_i/\{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}\}$ をスチューデント化残差 (外的スチューデント化残差) と呼ぶ。ここで、添字の (\cdot) は n 個の観測値の中から除去される観測値の番号を表す。

つぎに、 t_i を列ベクトルとして並べた標準化残差ベクトルを

$$t = \frac{1}{\hat{\sigma}}[\text{diag}(I_n - H)]^{-\frac{1}{2}}e$$

とする。ただし、 $\text{diag}(A)$ は正方行列 A の対角成分のみを取り出し、非対角成分をすべて 0 にした行列を表す。同様にスチューデント化残差 t_i^* については

$$t^* = D^{-\frac{1}{2}}[\text{diag}(I_n - H)]^{-\frac{1}{2}}e$$

とする。ただし、

$$D = \begin{pmatrix} \hat{\sigma}_{(1)}^2 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_{(2)}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}_{(n)}^2 \end{pmatrix}$$

である。

ところで、 $\hat{\sigma}^2$ および $\hat{\sigma}_{(i)}^2$ の関係式は、

$$\hat{\sigma}_{(i)}^2 = \frac{n-p-t_i^2}{n-p-1} \hat{\sigma}^2$$

であり、また、 t_i および t_i^* の関係式は、Weisberg [15] などから

$$t_i = t_i^* \sqrt{\frac{n-p}{n-p-1+t_i^{*2}}} \quad \text{または} \quad t_i^* = t_i \sqrt{\frac{n-p-1}{n-p-t_i^2}} \quad (2.1)$$

である。

線形回帰の一つの代替的方法としてリッジ回帰がある (Groβ [5] などを参照)。リッジ回帰においては、回帰係数ベクトル β の推定量 (以下、リッジ推定量) を、リッジパラメータ $k (\geq 0)$ を使って、

$$\hat{\beta}^R \equiv (X'X + kI_p)^{-1}X'y$$

と定義する。すると、最小 2 乗推定量の場合と同じく、残差ベクトル e^R は、 $e^R = y - X\hat{\beta}^R = (I_n - H^R)y$ となる。このとき、ハット行列 H^R は $H^R = X(X'X + kI_p)^{-1}X'$ であり、その第 i 対角成分 h_{ii}^R がリッジ回帰におけるてこ比である。ただし、 $0 < h_{ii}^R < 1$ である。

さらに、第 i 番目の観測値を除去したときの回帰係数ベクトル β の最小 2 乗推定量は $\hat{\beta}_{(i)}$

リッジ回帰における尤度距離による影響力評価

$= (X'_{(i)}X_{(i)})^{-1}X'_{(i)}y_{(i)}$ と定義されるので、これと同様に、第 i 番目の観測値を除去したときのリッジ推定量を

$$\hat{\beta}_{(i)}^R = (X'_{(i)}X_{(i)} + kI_p)^{-1}X'_{(i)}y_{(i)}$$

とする。このとき、観測値除去に関して、リッジパラメータ k は一定（独立）であると仮定する。この理由は、 $\hat{\beta}_{(i)}^R$ における逆行列の部分の計算が大変困難になり複雑になるからである。 k についても $k_{(i)}$ とする場合については、Takeuchi [10] において一つの方法が試みられている。

以上から、特に、 $k=0$ とすれば、 $\hat{\beta}^R = \hat{\beta}$ 、 $\hat{\beta}_{(i)}^R = \hat{\beta}_{(i)}$ 、あるいは $H^R = H$ など、リッジ回帰の統計量と通常の線形回帰の統計量が一致することがわかる。

実際のデータ解析においては、この他にもリッジ推定量に基づく診断統計量を検討する場合に生じる固有の問題について考慮する必要がある。リッジ推定量は、説明変数を中心化（センターリング）したり、尺度化（スケーリング）したりすることにより導出されることがよくある。たとえば、Belsley, Kuh and Welsch [1] や Draper and van Nostrand [4] などでもこうしたことは以前から議論されている。一般に、説明変数の尺度を変えたりする、あるいは k （の推定量）が観測値除去に依存する、といったことは無関係に影響力の大きさは変化する。けれども、ここでは、一般的なリッジ回帰において議論されているように、 $X'X$ を相関行列に変換することは必ずしも仮定しない。もし、線形回帰モデルに定数項が含まれている場合に相関行列へ変換するときには、 p および I_p をそれぞれ $p+1$ および I_{p+1}^* に置き換えて考える（あるいは、 $p+1$ をあらためて p と定義し直す）。ただし、 I_{p+1}^* は、Walker and Birch [14] で行われている手法と同様に、第 1 対角成分が 0 である以外は単位行列 I_p と同じ行列を表すものとして扱う。

2.2 リッジ回帰における Cook の距離

この節では、通常の線形回帰における観測値の影響力評価を行うための Cook の距離を拡張して、リッジ回帰における観測値の影響力評価を行うための診断統計量としての Cook の距離を示す。第 3 節において、尤度距離を Cook の距離に基づいて表すことを考えるので、ここであらかじめ定義しておく。

通常の線形回帰における観測値の影響力評価を行うための Cook の距離 CD_i は、Takeuchi [9] から

$$CD_i \equiv \frac{(\hat{\beta} - \hat{\beta}_{(i)})'X'X(\hat{\beta} - \hat{\beta}_{(i)})}{p\hat{\sigma}^2} = \left(\frac{1}{\sqrt{p}} \sqrt{\pi_i} t_i \right)^2 = c_i^2 \quad (2.2)$$

と表現できる。ただし、 c_i は

$$c = \frac{1}{\sqrt{p}} \Pi^{\frac{1}{2}} t$$

の第 i 成分であり, π_i は

$$\Pi = \text{diag}(H)[\text{diag}(I_n - H)]^{-1}$$

の第 i 対角成分であり, $\pi_i = h_{ii}/(1 - h_{ii})$ となる。

リッジ回帰における観測値の影響力評価を行うための Cook の距離 CD_i^R を, (2.2) 式を拡張することにより導入すると, Takeuchi [11] から

$$\begin{aligned} CD_i^R &\equiv \frac{(\hat{\beta}^R - \hat{\beta}_{(i)}^R)'(X'X + kI_p)(X'X)^{-1}(X'X + kI_p)(\hat{\beta}^R - \hat{\beta}_{(i)}^R)}{p\hat{\sigma}^2} \\ &= \left(\frac{1}{\sqrt{p}} \sqrt{\frac{h_{ii}}{1 - h_{ii}}} \cdot \frac{e_i^R}{\hat{\sigma}\sqrt{1 - h_{ii}}} \right)^2 = \{s_i(c_i + d_i)\}^2 \end{aligned} \quad (2.3)$$

となる。ただし, e_i^R はリッジ回帰における残差ベクトル e^R の第 i 成分であり, d_i は

$$d = \frac{1}{\sqrt{p}} \Pi^{\frac{1}{2}} u$$

の第 i 成分である。このとき,

$$u = \frac{1}{\hat{\sigma}} [\text{diag}(I_n - H)]^{-\frac{1}{2}} (e^R - e)$$

である。また, s_i は

$$S = \text{diag}(I_n - H)[\text{diag}(I_n - H^R)]^{-1}$$

の第 i 対角成分である。(2.3) 式の第三表現は, Cook の距離に関して 3 つの効果を測る部分に分けているので, 分離表現と呼ばれる。

(2.3) 式で与えられるリッジ回帰における Cook の距離 CD_i^R の導入においては, $\hat{\beta}^R - \hat{\beta}_{(i)}^R$ の中に挟まれる行列の選び方によっていくつかの定義式が考えられるが, ここでは, $\hat{\beta}^R$ の分散共分散行列 $\text{Var}(\hat{\beta}^R) = \sigma^2(X'X + kI_p)^{-1}(X'X)(X'X + kI_p)^{-1}$ の逆行列を選ぶものとする (たとえば, Walker and Birch [14] を参照)。

3 リッジ回帰における尤度距離

本節では, リッジ回帰における尤度距離を定義し, 観測値の影響力評価における基本的な統計量に基づく計算式を導出する。また, 導出された尤度距離と Cook の距離との関係式についても検討する。

3.1 リッジ回帰における尤度距離の導出

Cook and Weisberg [3] から, 通常の線形回帰における尤度距離 LD_i は, 以下のように定義される。

$$LD_i \equiv 2[L(\hat{\beta}) - L(\hat{\beta}_{(i)})] \quad (3.1)$$

ただし,

リッジ回帰における尤度距離による影響力評価

$L(\hat{\beta})$: β が $\hat{\beta}$ のときの対数尤度 および $L(\hat{\beta}_{(i)})$: β が $\hat{\beta}_{(i)}$ のときの対数尤度である。詳しく言えば、尤度距離 LD_i とは、未知回帰係数ベクトル β が $\hat{\beta}$ のとき、つまり、すべての観測値を利用したときの対数尤度と、同様に β が $\hat{\beta}_{(i)}$ のとき、つまり、第 i 番目の観測値を除去したときの対数尤度の差を 2 倍した診断統計量である。この差の大きさから、その除去された第 i 番目の観測値の影響力を測定するのである。したがって、この差が大きい、つまり LD_i の値が大きい第 i 番目の観測値を影響力が大きいものとして評価するのである。

また、Chatterjee and Hadi [2] や竹内 [13] で示されているように、(3.1) 式をてこ比 h_{ii} とスチューデント化残差 t_i^* を使って表すと、以下のように表現することができる。

$$LD_i = n \log \left(\frac{n}{n-1} \cdot \frac{n-p-1}{n-p-1+t_i^{*2}} \right) + \frac{n-1}{n-p-1} \cdot \frac{t_i^{*2}}{1-h_{ii}} - 1 \quad (3.2)$$

この (3.2) 式については、Chatterjee and Hadi [2] などでは、(2.1) 式の関係から標準化残差 t_i を使って

$$LD_i = n \log \left(\frac{n}{n-1} \cdot \frac{n-p-t_i^2}{n-p} \right) + \frac{n-1}{n-p-t_i^2} \cdot \frac{t_i^2}{1-h_{ii}} - 1 \quad (3.3)$$

とも表現している。

(3.1) 式の尤度距離の考え方をリッジ回帰の場合に適用して、以下のように定義する。

$$LD_i^R \equiv 2[L(\hat{\beta}^R) - L(\hat{\beta}_{(i)}^R)] \quad (3.4)$$

ただし、通常の線形回帰の場合である (3.1) 式と同じく、

$L(\hat{\beta}^R)$: β が $\hat{\beta}^R$ のときの対数尤度 および $L(\hat{\beta}_{(i)}^R)$: β が $\hat{\beta}_{(i)}^R$ のときの対数尤度とする (具体的な対数尤度については付録 1 を参照)。(3.4) 式の定義に基づいて、リッジ回帰における尤度距離 LD_i^R を導入すると、付録 1 から

$$LD_i^R = n \log T_i + \frac{n}{n-p} \cdot \frac{S_i}{T_i} \left[\frac{h_{ii}^{R2}}{1-h_{ii}^R} (t_i + u_i)^2 + 2(t_i + u_i) u_i^* \right] + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{u' \text{diag}(\mathbf{I}_n - \mathbf{H}) u}{n-p} \right] \quad (3.5)$$

となる。ただし、

$$T_i = \frac{n}{n-1} \cdot \frac{n-p-t_i^2}{n-p} = \frac{n}{n-1} \cdot \frac{n-p-1}{n-p-1+t_i^{*2}}$$

であり、 u_i^* は

$$u_i^* = \frac{1}{\sigma} [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}} (e^{R2} - e^R) \quad (3.6)$$

の第 i 成分であり、それに h_{ii}^{R2} は $(\mathbf{H}^R)^2$ の第 i 対角成分である。ここで、(3.6) 式において、 $e^{R2} = \{\mathbf{I}_n - (\mathbf{H}^R)^2\} y$ である。

特に、リッジパラメータ k が $k=0$ のとき、リッジ回帰における尤度距離 LD_i^R は、

$$LD_i^R \Big|_{k=0} = LD_i = n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \cdot \frac{h_{ii}}{1-h_{ii}} t_i^2 + n \left(\frac{1}{T_i} - 1 \right) \quad (3.7)$$

として得られ、付録2から(3.3)式の尤度距離 LD_i と一致することがわかる。

3.2 尤度距離と Cook の距離との関係

(3.7) 式の尤度距離 LD_i は、(2.2) 式の Cook の距離 CD_i を利用すれば、以下のように表現することもできる。

$$LD_i = n \log T_i + \frac{1}{T_i} \cdot \frac{np}{n-p} CD_i + n \left(\frac{1}{T_i} - 1 \right) \quad (3.8)$$

なお、Chatterjee and Hadi [2] では、分散の推定について条件を付けた場合の尤度距離を提案し、それが Cook の距離と実質的に同一であることを示している。

(3.8) 式と同様にして、リッジ回帰における尤度距離 LD_i^R も (2.3) 式の Cook の距離 CD_i^R を利用して、以下のように表現することができる。

$$LD_i^R = n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \left[\frac{h_{ii}^R}{h_{ii}} p CD_i^R + 2s_i(t_i + u_i) u_i^* \right] + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{u' \text{diag}(\mathbf{I}_n - \mathbf{H}) u}{n-p} \right] \quad (3.9)$$

さらに、(2.3) 式で与えられるリッジ回帰における Cook の距離 CD_i^R を拡張して、(3.9) 式を変形すると

$$LD_i^R = n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} CD_i^{R*} + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{u' \text{diag}(\mathbf{I}_n - \mathbf{H}) u}{n-p} \right] \quad (3.10)$$

と表現することもできる。ただし、(3.10) 式において

$$CD_i^{R*} \equiv \frac{\pi_i^*}{\pi_i} \left[\left\{ s_i(c_i + d_i) + \frac{d_i^*}{\pi_i^*} \right\}^2 - \left(\frac{d_i^*}{\pi_i^*} \right)^2 \right] \quad (3.11)$$

と定義する。このとき、(3.11) 式において、 d_i^* は

$$d_i^* = \frac{1}{\sqrt{p}} \Pi^{\frac{1}{2}} u^*$$

の第 i 成分であり、 π_i^* は

$$\Pi^* = \text{diag}\{(H^R)^2\} [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-1}$$

の第 i 対角成分であり、 $\pi_i^* = h_{ii}^R / (1 - h_{ii})$ となる。特に、 $k = 0$ のとき、

$$CD_i^{R*} \Big|_{k=0} = CD_i^R \Big|_{k=0} = CD_i$$

となる。

したがって、通常の線形回帰においてもリッジ回帰においても、尤度距離は Cook の距離の関数であるとみなすこともできる。

4 数値例

(3.9) 式あるいは (3.10) 式で与えられるリッジ回帰における尤度距離 LD_i^k の性質を調べるために、実際のデータに基づく数値計算を行う。また、リッジ回帰における代表的な診断統計量である (2.3) 式で与えられる Cook の距離 CD_i^k との比較検討もする。

数値計算をするためのデータとしては、Takeuchi [11] において例示されている「Artificial Data」(データ数は $n=20$ であり、説明変数の数は定数項を含め $p=4+1=5$) を利用する。数値例を示す前に、リッジ推定量に対してよく利用されるリッジパラメータの候補の中から、以下の2つの k を取り上げる。1つは、Hoerl, Kennard and Baldwin [7] によって提案されている $k_1 = p\hat{\sigma}^2 / \hat{\beta}'\hat{\beta}$ であり、もう1つは Lawless and Wang [8] によって提案されている $k_2 = p\hat{\sigma}^2 / \hat{\beta}'X'X\hat{\beta}$ である。リッジパラメータは様々な観点を基にして数多く導かれているが、尤度距離の基本的な性質(挙動)を例示するために、代表的なこの2つの場合を取り上げる。

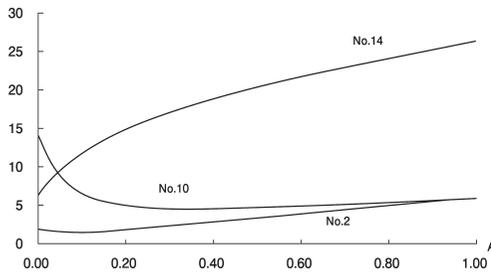
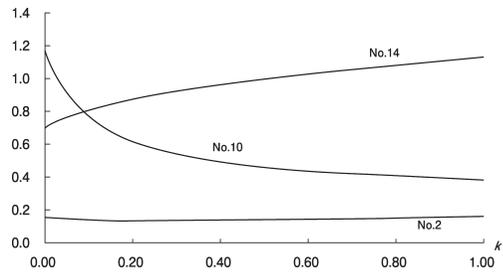
リッジ回帰における尤度距離 LD_i^k の計算結果は表 4.1 のとおりである。通常の線形回帰の場合の尤度距離 LD_i は LD_i^k において $k=0$ の場合の計算結果に相当する。竹内 [13] において提案されている LD_i の一次近似に基づく打切り点を計算すると、

$$LD_i^{(1)*} > 4n \cdot \frac{4n + p(n-p)}{(n-p)^2(n-p+4)} = 4 \times 20 \times \frac{4 \times 20 + 5 \times (20-5)}{(20-5)^2 \times (20-5+4)} \cong 2.901$$

となる。つまり、 LD_i の打切り点を 2.901 とみなす。これを使うと、観測値 No. 10 および No. 14 が影響力の大きい観測値であると判断できる。 LD_i^k については、リッジパラメータの変化を考慮する必要があるので、打切り点を導くことが容易ではない。よって、 LD_i の打切

表 4.1 尤度距離の計算結果

観測値 No.	LD_i ($k=0$)	LD_i^k		観測値 No.	LD_i ($k=0$)	LD_i^k	
		$k=k_2$ =0.073	$k=k_1$ =0.109			$k=k_2$ =0.073	$k=k_1$ =0.109
1	0.940	-0.701	-1.052	11	0.247	-0.102	-0.138
2	1.842	1.428	1.466	12	0.047	-0.044	-0.074
3	0.220	-0.115	-0.210	13	0.079	0.063	0.030
4	0.223	-0.164	-0.296	14	6.251	10.543	12.007
5	0.069	0.277	0.370	15	0.028	0.098	0.113
6	0.619	0.845	0.901	16	0.273	0.619	0.730
7	0.433	-0.326	-0.417	17	0.246	0.832	0.926
8	0.032	0.168	0.205	18	0.281	0.249	0.233
9	0.029	0.045	0.096	19	0.092	0.096	0.048
10	13.537	7.443	6.256	20	0.156	-0.365	-0.440

図 4.1 尤度距離 LD_i^k のグラフ図 4.2 Cook の距離 CD_i^k のグラフ

り点を考慮しながら、 LD_i^k の値が相対的に大きい観測値を影響力の大きいものとして判断することにする。すると、 k_1 および k_2 の両方の場合において、観測値 No. 10 および No. 14 の LD_i^k の値が LD_i の打ち切り点よりも大きくなる。また、その打ち切り点よりも小さい値であるが観測値 No. 2 の LD_i^k の値も他の 17 個の観測値の影響力に比べれば大きな値であると考えられる。したがって、これらの結果から影響力の大きい観測値としては、観測値 No. 2、No. 10 それに No. 14 の 3 個が挙げられる。その他の 17 個の観測値については、この 3 個の観測値に比べて影響力がかなり小さいものとみなせる。

さらに、影響力の大きい観測値であると判断される 3 個の観測値 No. 2、No. 10 それに No. 14 については、 k の値を 0 から 1 まで変化 (増加) させて、尤度距離 LD_i^k の挙動を細かく調べる。その結果が図 4.1 である。参考までにリッジ回帰における Cook の距離 CD_i^k についても、同じ 3 個の観測値について挙動を調べたのが図 4.2 である。両図とも、 k の値を 1 より大きくすることも考えられるが、通常のデータ解析では $0 \leq k \leq 1$ に設定することが多いので、 $k > 1$ の場合に広げることは現実的には想定しにくい。

図 4.1 から観測値 No. 2 および No. 10 については、リッジパラメータ k の値が 0 から少し増えたときに尤度距離 LD_i^k の値が最小になり、その後は徐々に増加する傾向が見られる。また、 k の値が 1 に近づくにつれて観測値 No. 2 と No. 10 の LD_i^k の値が接近し、最後には逆転する (観測値 No. 2 の LD_i^k の値が LD_i の打ち切り点を超える場合があることもわかる)。これに対して、観測値 No. 14 は k の値が増加すると尤度距離 LD_i^k の値も単調に増加する傾向が見られる。通常の線形回帰における尤度距離 LD_i の値から観測値の影響力評価を行うと、観測値 No. 10 が最も影響力が大きいと判断されるが、リッジ回帰を適用した場合には、その尤度距離 LD_i^k から観測値 No. 14 の影響力の変化が最も大きく、なおかつ影響力の大きさ自体も k の値が大きくなるに従い増加する傾向にある。通常は、最小 2 乗推定量を縮小したものの 1 つがリッジ推定量であるので、ある k については、最小 2 乗推定量 $\hat{\beta}$ よりもリッジ推定量 $\hat{\beta}^R$ の方が影響力の大きさが小さくなることが期待される。けれども、観測値 No. 14 については任意の k ($0 \leq k \leq 1$) について影響力がより大きくなり、リッジ回帰の適用が芳しくない結果になったのである。その意味では、単に観測値の影響力の大きさを比較検討

リッジ回帰における尤度距離による影響力評価

するだけでなく、リッジ回帰の適用についての是非についても検討することが重要であると考えられる。よって、図 4.1 のように、リッジパラメータ k について、尤度距離 LD_i^k の変化を調べる必要があるといえる。

なお、参考までに図 4.2 の Cook の距離 CD_i^k の挙動を調べると、尤度距離 LD_i^k の値ほどの変化は見られないが、増加や減少等の変化については、よく似た傾向を示している。リッジパラメータ k の増加に伴い、観測値 No.2 は減少してから増加に転じるが、観測値 No.10 は $k=1$ までの範囲では単調減少している。これに対して、観測値 No.14 については、 CD_i^k の値が単調に増加する傾向を示している。観測値 No.10 を除き、 LD_i^k と CD_i^k における挙動についての差異はあまり見受けられない。観測値 No.10 の場合のような差異は、導出された (3.9) 式の関係式から、 LD_i^k には CD_i^k 以外にもリッジパラメータ k に関わる統計量として、 s_i や u_i などを含むので、挙動としては CD_i^k と異なる微小変動の部分が少なからずあるものと考えられる。

5 まとめと今後の課題

本論文では、リッジ回帰における尤度距離の定義を与えて、観測値の影響力評価における基本的な統計量に基づく計算式を導出した。その結果、Cook の距離との関連性 (関数関係) を見出すことができた。また、実際のデータを基にして、リッジパラメータの変化に対する尤度距離 LD_i^k の挙動を調べることにより、観測値により概ね 2 種類の傾向があることがわかった。この傾向は Cook の距離 CD_i^k の挙動とよく似ているが、 LD_i^k の方がより特徴がはっきりしているものと考えられる。

今後の課題としては、複数個の観測値の影響力評価を検討することと、リッジパラメータに対する尤度距離 LD_i^k の変化についてさらに検討を加える必要があるものと考えている。実際のデータに対する個別研究も重要であるが、理論的に挙動の特徴を調べるためには、尤度距離 LD_i^k のパラメータである k についての微分を検討することも重要である。この微分については計算結果を導出することはできているが、増加や減少などの理論的な特性を解明する段階までには至っていないので、今回の論文では取り上げていない。

付録 1： (3.5) 式の導出

リッジ回帰における対数尤度の定義である (3.4) 式から、対数尤度 $L(\hat{\beta}^k)$ および $L(\hat{\beta}_{(i)}^k)$ を求める。このとき、 σ^2 の推定量としては、最小 2 乗推定量である $\hat{\sigma}^2 = e'e/n$ を利用する。通常の線形回帰における尤度距離の導出過程を示した竹内 [13] の付録 1 の方法と同様にし、対数尤度は

$$\begin{aligned}
L(\hat{\beta}^R) &= \log f(y; \hat{\beta}^R, \hat{\sigma}^2) \\
&= \log \left[\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \right)^n \exp \left\{ -\frac{(y - X\hat{\beta}^R)'(y - X\hat{\beta}^R)}{2\hat{\sigma}^2} \right\} \right] \\
&= -\frac{n}{2} \log 2\pi\hat{\sigma}^2 - \frac{e^{R'}e^R}{2\hat{\sigma}^2}
\end{aligned}$$

および

$$\begin{aligned}
L(\hat{\beta}_{(i)}^R) &= \log f(y; \hat{\beta}_{(i)}^R, \hat{\sigma}_{(i)}^2) \\
&= \log \left[\left(\frac{1}{\sqrt{2\pi\hat{\sigma}_{(i)}^2}} \right)^n \exp \left\{ -\frac{(y - X\hat{\beta}_{(i)}^R)'(y - X\hat{\beta}_{(i)}^R)}{2\hat{\sigma}_{(i)}^2} \right\} \right] \\
&= -\frac{n}{2} \log 2\pi\hat{\sigma}_{(i)}^2 - \frac{(y - X\hat{\beta}_{(i)}^R)'(y - X\hat{\beta}_{(i)}^R)}{2\hat{\sigma}_{(i)}^2}
\end{aligned}$$

となる。ただし、

$$\begin{aligned}
&(y - X\hat{\beta}_{(i)}^R)'(y - X\hat{\beta}_{(i)}^R) \\
&= \{y - X\hat{\beta}^R + X(\hat{\beta}^R - \hat{\beta}_{(i)}^R)\}'\{y - X\hat{\beta}^R + X(\hat{\beta}^R - \hat{\beta}_{(i)}^R)\} \\
&= (y - X\hat{\beta}^R)'(y - X\hat{\beta}^R) + 2(\hat{\beta}^R - \hat{\beta}_{(i)}^R)'X'(y - X\hat{\beta}^R) + (\hat{\beta}^R - \hat{\beta}_{(i)}^R)'X'X(\hat{\beta}^R - \hat{\beta}_{(i)}^R) \\
&= e^{R'}e^R + 2(\hat{\beta}^R - \hat{\beta}_{(i)}^R)'X'e^R + \left(\frac{e_i^R}{1 - h_{ii}^R} \right)^2 x_i (X'X + kI_p)^{-1} X'X (X'X + kI_p)^{-1} x_i' \\
&= e^{R'}e^R + 2 \frac{e_i^R e_i^{R2} - (e_i^{R2})^2}{1 - h_{ii}^R} + \left(\frac{e_i^R}{1 - h_{ii}^R} \right)^2 h_{ii}^{R2} \\
&= e^{R'}e^R + \frac{2e_i^R e_i^{R2}}{1 - h_{ii}^R} + \left(\frac{e_i^R}{1 - h_{ii}^R} \right)^2 \{h_{ii}^{R2} - 2(1 - h_{ii}^R)\}
\end{aligned}$$

である。このとき、

$$\begin{aligned}
2(\hat{\beta}^R - \hat{\beta}_{(i)}^R)'X'e^R &= 2 \frac{e_i^R}{1 - h_{ii}^R} x_i (X'X + kI_p)^{-1} X' (I_n - H^R) y \\
&= 2 \frac{e_i^R}{1 - h_{ii}^R} \{1_i' X \hat{\beta}^R - 1_i' (H^R)^2 y\} \\
&= 2 \frac{e_i^R}{1 - h_{ii}^R} \{-(y_i - \hat{y}_i^R) + y_i - 1_i' (H^R)^2 y\} \\
&= 2 \frac{e_i^R}{1 - h_{ii}^R} [-e_i^R + 1_i' \{I_n - (H^R)^2\} y] = 2 \frac{e_i^R}{1 - h_{ii}^R} (-e_i^R + e_i^{R2}) \\
&= 2 \frac{e_i^R e_i^{R2} - (e_i^R)^2}{1 - h_{ii}^R}
\end{aligned}$$

であり、Takeuchi [11] の Appendix A から

$$\hat{\beta}^R - \hat{\beta}_{(i)}^R = (X'X + kI_p)^{-1} x_i' \frac{e_i^R}{1 - h_{ii}^R}$$

の関係を利用している。ここで、 1_i は第 i 成分が 1 で他の成分が 0 の列ベクトルであり、 \hat{y}_i^R は $\hat{y}^R = X\hat{\beta}^R$ の第 i 成分であり、 $\hat{y}_i^R = 1_i' \hat{y}^R$ と表すことができ、同様に y_i は y の第 i 成

リッジ回帰における尤度距離による影響力評価

分であり, $y_i = 1_i'y$ と表すこともできる。また, x_i は X の第 i 行であり, $x_i = 1_i'X$ である。さらに, $h_{ii}^R = x_i(X'X + kI_p)^{-1}X'X(X'X + kI_p)^{-1}x_i'$ となることは, $(H^R)^2 = X(X'X + kI_p)^{-1}X'X(X'X + kI_p)^{-1}X'$ の第 i 対角成分であることから自明である。

したがって, 以上のことから,

$$\begin{aligned} LD_i^R &= 2[L(\hat{\beta}^R) - L(\hat{\beta}_{(i)}^R)] \\ &= n \log \frac{\tilde{\sigma}_{(i)}^2}{\tilde{\sigma}^2} + \frac{\tilde{\sigma}^2}{\tilde{\sigma}_{(i)}^2} \cdot \frac{1 - h_{ii}}{1 - h_{ii}^R} \left[\frac{2e_i^R e_i^{R2}}{\tilde{\sigma}^2(1 - h_{ii})} + \frac{(e_i^R)^2}{\tilde{\sigma}^2(1 - h_{ii})} \cdot \frac{h_{ii}^{R2} - 2(1 - h_{ii}^R)}{1 - h_{ii}^R} \right] \\ &\quad + \frac{1}{\tilde{\sigma}_{(i)}^2} \left(1 - \frac{\tilde{\sigma}_{(i)}^2}{\tilde{\sigma}^2} \right) e^{R'} e^R \end{aligned}$$

となる。ここで, 第一項および第三項において

$$\frac{\tilde{\sigma}_{(i)}^2}{\tilde{\sigma}^2} = \frac{n(n-p-1)}{(n-1)(n-p)} \cdot \frac{\tilde{\sigma}_{(i)}^2}{\tilde{\sigma}^2} = \frac{n}{n-1} \cdot \frac{n-p-t_i^2}{n-p} = T_i$$

である。また, 第二項において

$$\frac{\tilde{\sigma}^2}{\tilde{\sigma}_{(i)}^2} = \frac{\tilde{\sigma}^2}{\tilde{\sigma}_{(i)}^2} \cdot \frac{\tilde{\sigma}^2}{\tilde{\sigma}^2} = \frac{1}{T_i} \cdot \frac{n}{n-p}$$

と変形することもできる。

つぎに, 第三項において,

$$\begin{aligned} e^{R'} e^R &= \{e + (e^R - e)\}' \{e + (e^R - e)\} \\ &= e'e \left[1 + 2 \frac{e'(e^R - e)}{e'e} + \frac{(e^R - e)'(e^R - e)}{e'e} \right] \\ &= n\tilde{\sigma}^2 \left[1 + \frac{u' \text{diag}(I_n - H) u}{n-p} \right] \end{aligned}$$

である。このとき, $e'(e^R - e) = 0$ (直交性) を利用した。

第二項においても, $e_i^R = e_i + (e_i^R - e_i)$ および $e_i^{R2} = e_i + (e_i^R - e_i) + (e_i^{R2} - e_i^R)$ を利用して,

$$\begin{aligned} &\frac{2e_i^R e_i^{R2}}{\tilde{\sigma}^2(1 - h_{ii})} + \frac{(e_i^R)^2}{\tilde{\sigma}^2(1 - h_{ii})} \cdot \frac{h_{ii}^{R2} - 2(1 - h_{ii}^R)}{1 - h_{ii}^R} \\ &= 2(t_i + u_i)(t_i + u_i + u_i^*) + (t_i + u_i)^2 \frac{h_{ii}^{R2} - 2(1 - h_{ii}^R)}{1 - h_{ii}^R} \\ &= 2(t_i + u_i)u_i^* + (t_i + u_i)^2 \frac{2(1 - h_{ii}^R) + h_{ii}^{R2} - 2(1 - h_{ii}^R)}{1 - h_{ii}^R} \\ &= 2(t_i + u_i)u_i^* + (t_i + u_i)^2 \frac{h_{ii}^{R2}}{1 - h_{ii}^R} \end{aligned}$$

である。

以上から, 最終的に (3.5) 式として,

$$\begin{aligned} LD_i^R &= n \log T_i + \frac{n}{n-p} \cdot \frac{s_i}{T_i} \left[\frac{h_{ii}^{R2}}{1 - h_{ii}^R} (t_i + u_i)^2 + 2(t_i + u_i)u_i^* \right] \\ &\quad + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{u' \text{diag}(I_n - H) u}{n-p} \right] \end{aligned}$$

を導くことができる。

付録 2： (3.3) 式と (3.7) 式の同一性

(3.3) 式を式変形すると

$$\begin{aligned}
 LD_i &= n \log \left(\frac{n}{n-1} \cdot \frac{n-p-t_i^2}{n-p} \right) + \frac{n-1}{n-p-t_i^2} \cdot \frac{t_i^2}{1-h_{ii}} - 1 \\
 &= n \log T_i + \frac{n}{n-p} \left(\frac{n-p}{n} \cdot \frac{n-1}{n-p-t_i^2} \right) t_i^2 \left(1 + \frac{h_{ii}}{1-h_{ii}} \right) - 1 \\
 &= n \log T_i + \frac{n}{n-p} \cdot \frac{1}{T_i} \cdot t_i^2 \cdot \frac{h_{ii}}{1-h_{ii}} + \frac{n}{n-p} \cdot \frac{t_i^2}{T_i} - 1 \\
 &= n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \cdot \frac{h_{ii}}{1-h_{ii}} t_i^2 + \frac{nt_i^2 - (n-p)}{n-p-t_i^2} \\
 &= n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \cdot \frac{h_{ii}}{1-h_{ii}} t_i^2 + \frac{(n-1)(n-p) - n(n-p-t_i^2)}{n-p-t_i^2} \\
 &= n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \cdot \frac{h_{ii}}{1-h_{ii}} t_i^2 + n \left(\frac{n-1}{n} \cdot \frac{n-p}{n-p-t_i^2} - 1 \right) \\
 &= n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \cdot \frac{h_{ii}}{1-h_{ii}} t_i^2 + n \left(\frac{1}{T_i} - 1 \right)
 \end{aligned}$$

となり, (3.7) 式と一致することがわかる。

参考文献

- [1] Belsley, D. A., Kuh, E. and Welsch, R. E. (1980), *Regression Diagnostics ; Identifying Influential Data and Sources of Collinearity*, New York : Wiley.
- [2] Chatterjee, S. and Hadi, A. S. (1988), *Sensitivity Analysis in Linear Regression*, New York : Wiley.
- [3] Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York : Chapman and Hall.
- [4] Draper, N. R. and van Nostrand, R. C. (1979), Ridge regression and James-Stein estimation : Review and comments, *Technometrics*, **21**, 451-466.
- [5] Groß, J. (2003), *Linear Regression*, Berlin : Springer.
- [6] Hoerl, A. E. and Kennard, R. W. (1970), Ridge regression : biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55-67.
- [7] Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. (1975), Ridge regression : some simulations, *Communications in Statistics—Theory and Methods*, **4**, 105-123.
- [8] Lawless, J. F. and Wang, P. (1976), A simulation study of ridge and other regression estimators, *Communications in Statistics—Theory and Methods*, **5**, 307-323.
- [9] Takeuchi, H. (1991), Detecting influential observations by using a new expression of

- Cook's distance, *Communications in Statistics—Theory and Methods*, **20**, 261-274.
- [10] Takeuchi, H. (1993), Influence analysis with an extension of covariance ratio in ridge regression, 静岡県立大学経営情報学部報「経営と情報」, **5**, 9-17.
- [11] Takeuchi, H. (1994), Sensitivity analysis with an extension of Cook's distance in ridge regression, *Journal of the Japan Statistical Society*, **24**, 221-236.
- [12] 竹内秀一 (1997), リッジ回帰における診断統計量の代替表現, 東京経大会誌 **203**号, 5-18。
- [13] 竹内秀一 (2005), 線形回帰分析における尤度距離による影響力評価, 人文自然科学論集, **119**号, 19-30。
- [14] Walker, E. and Birch, J. B. (1988), Influence measures in ridge regression, *Technometrics*, **30**, 221-227 (Correction **30**, 469-470).
- [15] Weisberg, S. (2005), *Applied Linear Regression*, 3rd edition, New York : Wiley.