福山平成大学経営学部紀要
第 17 号（2021），＊－＊＊頁

対数尤度関数の視覚化プログラム

福井　正康[*1]

[*1] 福山平成大学経営学部経営学科

要旨：最尤法における対数尤度関数はどのような形状をして、データ数の増加と共にどのように変化して行くのか、著者は College Analysis の教育用ツールとして、簡単に対数尤度関数のグラフが描けるプログラムを作成した。それを用いて観察した結果、対数尤度関数は必ずしも山のような形状をしていないこと、パラメータ間の相関がない場合、標準誤差の範囲を横軸に取ると、データ数によらず、ほぼ一定の形状をしていることが確認できた。そのため、固定された範囲では、データ数が増えると次第に細くなって行くことになる。

キーワード：最尤法、対数尤度関数、College Analysis

1.　はじめに

モデルを用いたパラメータの推定問題では、1990 年ごろまで、最小 2 乗法が用いられることが多かったが、それ以降は最尤法が多く用いられるようになった。これにはコンピュータの発達が大きく寄与し、最近ではパソコンを使って簡単に最尤法の計算ができるようになった。特に共分散構造分析では、最尤法の複雑なパラメータの推定問題も扱えるようになった。しかし、最尤法は初心者にとって直感的に分かり易い手法ではない。そしてその分かりにくさの原因の一つとして考えられるのが、尤度関数のイメージを掴みにくいということである。尤度関数は多くの実測値とパラメータから作られる関数であるが、利用者にはそれが、実測値が与えられた下でのパラメータの関数、と感じられているであろうか。実際最尤法を使っている人でも、尤度関数や対数尤度関数をパラメータの関数として表示してみた人は少ないだろう。

そこで著者は少しでもイメージをつかめるように、対数尤度関数をできるだけ簡単に視覚化するツールを作ってみようと考えた。もちろん分析で、ソフトを使ってパラメータを推定するだけなら、このようなことを考える必要はない。このプログラムの意味は、「対数尤度関数は山のような形をしているんだろうか」、「データ数が多くなるとパラメータの推定の幅は縮まるけれど、対数尤度関数の形や幅はどうなるんだろう」などと余計なことを考える利用者にこそ理解されると思う。

対数尤度関数を表示する際の困難は、この関数が多くの座標値で与えられたパラメータ関数の合計である点である。例えば、1 つの関数の形が $\log f(x;a)$ で与えられる場合を考える。これをパラメータ a の関数と考えると、計算には、まず数式の中に x の値を（文

字列として）代入し、その後 a の値を少しずつ変化させて代入してゆくことで、関数形は描画点の個数回の計算によって容易に描かれる。しかし、対数尤度関数では以下のように関数形が $\log f(x_\lambda;a)$ の λ についての和になっている。

$$\log L = \sum_{\lambda=1}^{N} \log f(x_\lambda;a)$$

このため、最初に x_λ の値を（文字列として）入れると、関数式が長くなり過ぎて計算が困難になる。これを解決するためには、まず a の値を決めて、関数形 $\log f(x_\lambda;a)$ の中に（文字列として）代入し、関数形を定める。その後 1 つずつ x_λ の値を代入して計算、それを足すことで 1 つの a に対する対数尤度関数の値を定める。これは各 a に対する関数値を求めるために通常の関数計算を N 回行うことになり、データ数が多い場合には計算時間がかかる。

次に、対数尤度関数は対数をとった密度関数を N 回足したものであるので、データ数 N に依存する。そのため、我々は対数尤度関数をそのまま表示する場合と、データ数で対数尤度関数を割った場合を考えてみた。元々我々は対数尤度関数の形を議論する場合、データ 1 つ当たりの形が有効であると考えていた。データの数によって対数尤度関数の大きさが大きく左右されるのは問題があると理由もなしに考えたからである。しかしこれは大きな誤りであった。データ数の増加は対数尤度関数の高さを高くする。これが対数尤度関数の尖り方に影響を与え、パラメータ推定の標準誤差に $1/\sqrt{N}$ で影響を与える。このため対数尤度関数の幅を議論するときには、平均化されたものでなく、本来の対数尤度関数を利用すべきであった。

最後にグラフを描く際に問題になるのがいかに簡単にグラフ化するかである。このグラフ化の目的は初学者の最尤法の理解のためのものであるから、表示の手順が困難では意味がない。そのため、分析をいくつかに絞り、利用するデータ数を簡単に選べて、グラフの表示範囲もできれば自動化したい。グラフの頂点はデータ数により変化する。そのため、頂点を画面上で固定し、描画範囲を調整することによってグラフの形状を細かく見ることができるようにした。

ここではまず最尤法による正規分布のパラメータ推定を説明し、回帰分析とロジスティック回帰分析の最尤法による解法について説明する。次に最尤法の数値計算法について簡単に説明し、計算の過程で求められる推定パラメータの標準誤差について言及する。その後一般論から対数尤度関数の形状を議論し、プログラムの利用法の節で、プログラムを使って一般論の検証と最初に述べた尤度関数についての疑問に答えて行く。また補遺で、2 値ロジスティック回帰分析で実際に計算を進めるための計算式をまとめておく。

2.　正規分布のパラメータ推定

最尤推定法の簡単な例を考えてみる。我々は平均 μ と分散 σ^2 を推測する場合に以下のような推定量を用いる。

$$\mu \to \bar{x} = \frac{1}{N}\sum_{\lambda=1}^{N} x_\lambda \quad , \quad \sigma^2 \to s^2 = \frac{1}{N}\sum_{\lambda=1}^{N}(x_\lambda - \bar{x})^2 \tag{1}$$

この妥当性を、正規分布に基づく最尤法を使って考えてみる。

まず、変数 x_λ が正規分布 $N(\mu,\sigma^2)$ に従うとするとその確率密度関数は以下となる。

$$f(x_\lambda;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-(x_\lambda-\mu)^2/2\sigma^2\right]$$

これを使って尤度関数 L と対数尤度関数 $\log L$ を作ると以下となる。

$$L = \prod_{\lambda=1}^{N} f(x_\lambda;\mu,\sigma) = \frac{1}{(2\pi)^{N/2}\sigma^N}\prod_{\lambda=1}^{N}\exp\left[-(x_\lambda-\mu)^2/2\sigma^2\right] \tag{2}$$

$$\begin{aligned}\log L &= \sum_{\lambda=1}^{N}\log f(x_\lambda;\mu,\sigma) \\ &= -\frac{1}{2\sigma^2}\sum_{\lambda=1}^{N}(x_\lambda-\mu)^2 - N\log\sigma - \frac{N}{2}\log(2\pi)\end{aligned} \tag{3}$$

これから、パラメータ μ,σ を推定してみよう。尤度関数は確率密度関数の掛け算であるから、頂点が最も確からしい点であると考えられる。これは対数尤度関数でも同じ位置になる。そこで、対数尤度関数を微分してその頂点の位置を求めることにする。

$$\partial\log L/\partial\mu = \frac{1}{\sigma^2}\sum_{\lambda=1}^{N}(x_\lambda-\mu) = 0 \qquad \hat{\mu} = \frac{1}{N}\sum_{\lambda=1}^{N}x_\lambda$$

$$\partial\log L/\partial\sigma = \frac{1}{\sigma^3}\sum_{\lambda=1}^{N}(x_\lambda-\mu)^2 - \frac{N}{\sigma} = 0 \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{\lambda=1}^{N}(x_\lambda-\mu)^2$$

ここでは(1)で与えた推定量の通りの結果が得られている。

次に、情報行列 $\mathbf{J}$ をまとめておく。

$$\mathbf{J} = -\begin{pmatrix}\partial^2\log L/\partial\mu^2 & \partial^2\log L/\partial\mu\partial\sigma \\ \partial^2\log L/\partial\mu\partial\sigma & \partial^2\log L/\partial\sigma^2\end{pmatrix} \to \begin{pmatrix} N/\hat{\sigma}^2 & 0 \\ 0 & 2N/\hat{\sigma}^2\end{pmatrix}$$

$$\partial^2\log L/\partial\mu^2 = -\frac{N}{\sigma^2} \to -\frac{N}{\hat{\sigma}^2}$$

$$\partial^2\log L/\partial\mu\partial\sigma = -\frac{2}{\sigma^3}\sum_{\lambda=1}^{N}(x_\lambda-\mu) = -\frac{2N}{\sigma^3}(\bar{x}-\mu) \to 0$$

$$\partial^2\log L/\partial\sigma^2 = -\frac{3}{\sigma^4}\sum_{\lambda=1}^{N}(x_\lambda-\mu)^2 + \frac{N}{\sigma^2} \to -\frac{2N}{\hat{\sigma}^2}$$

情報行列の逆行列 $\mathbf{J}^{-1}$ は、パラメータの共分散行列として知られている。それによると推定値の標準誤差は、データ数 N が増加するにつれ、$1/\sqrt{N}$ で小さくなって行く。この場合、推定値と標準誤差は解析的に求めることが可能であるが、プログラムではニュートン・ラフソン法を用いて求めている。

3. 回帰分析におけるパラメータ推定

この節では回帰分析について最尤法を使って考えてみる。回帰分析は目的変数 y_λ を説明変数 x_λ で以下のように予測する分析である。

$$y_\lambda = a x_\lambda + b + u_\lambda \qquad (\lambda = 1, 2, \cdots, N) \tag{1}$$

ここに u_λ は予測の誤差である。 a, b はパラメータであるが、最尤法を使ってどのように求められるのだろうか。

最初に誤差 u_λ の分布を、平均 0、分散 σ^2 の正規分布とする。この場合、推定値や標準誤差が解析的に求められる。

平均 0 で、分散 σ^2 の正規分布の密度関数は以下のように表される。

$$\begin{aligned} f(u_\lambda;\sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-u_\lambda^2 \big/ 2\sigma^2\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-(y_\lambda - a x_\lambda - b)^2 \big/ 2\sigma^2\right] \end{aligned}$$

これを使って尤度関数を作ってみる。

$$L = \prod_{\lambda=1}^{N} f(u_\lambda;\sigma) = \frac{1}{(2\pi)^{N/2}\sigma^N}\prod_{\lambda=1}^{N}\exp\left[-(y_\lambda - a x_\lambda - b)^2 \big/ 2\sigma^2\right]$$

この尤度関数を使うと対数尤度関数は以下のようになる。

$$\begin{aligned} \log L &= \sum_{\lambda=1}^{N}\log f(u_\lambda;\sigma) \\ &= -\frac{1}{2\sigma^2}\sum_{\lambda=1}^{N}(y_\lambda - a x_\lambda - b)^2 - N\log\sigma - \frac{N}{2}\log(2\pi) \end{aligned}$$

この対数尤度関数をパラメータ a, b, σ で微分する。

$$\frac{\partial}{\partial a}\log L = \frac{1}{\sigma^2}\sum_{\lambda=1}^{N} x_\lambda (y_\lambda - a x_\lambda - b) = 0$$

$$\frac{\partial}{\partial b}\log L = \frac{1}{\sigma^2}\sum_{\lambda=1}^{N} (y_\lambda - a x_\lambda - b) = 0$$

$$\frac{\partial}{\partial \sigma}\log L = \frac{1}{\sigma^3}\sum_{\lambda=1}^{N} (y_\lambda - a x_\lambda - b)^2 - \frac{N}{\sigma} = 0$$

上の 2 つの式は一般に利用される最小 2 乗法の制約式と同じで、以下のような同じ結果を与える。

$$\hat{a} = \sum_{\lambda=1}^{N}(y_\lambda - \overline{y})(x_\lambda - \overline{x}) \Big/ \sum_{\lambda=1}^{N}(x_\lambda - \overline{x})^2 = s_{xy} \big/ s_x^2$$

$$\hat{b} = \overline{y} - \hat{a}\overline{x}$$

また、最後の式から σ の推定値が求まる。

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{\lambda=1}^{N}(y_\lambda - \hat{a}x_\lambda - \hat{b})^2 = \frac{1}{N}\sum_{\lambda=1}^{N}\hat{u}_\lambda^2$$

情報行列 $\mathbf{J}$ をまとめておく。

$$\mathbf{J} = -\begin{pmatrix} \partial^2 \log L/\partial a^2 & \partial^2 \log L/\partial a \partial b & \partial^2 \log L/\partial a \partial \sigma \\ \partial^2 \log L/\partial a \partial b & \partial^2 \log L/\partial b^2 & \partial^2 \log L/\partial b \partial \sigma \\ \partial^2 \log L/\partial a \partial \sigma & \partial^2 \log L/\partial b \partial \sigma & \partial^2 \log L/\partial \sigma^2 \end{pmatrix}$$

$$\to \frac{N}{\hat{\sigma}^2}\begin{pmatrix} s_x^2 + \bar{x}^2 & \bar{x} & 0 \\ \bar{x} & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\partial^2 \log L/\partial a^2 = -\frac{1}{\sigma^2}\sum_{\lambda=1}^{N} x_\lambda^2 = -\frac{N(s_x^2 + \bar{x}^2)}{\sigma^2} \to -\frac{N(s_x^2 + \bar{x}^2)}{\hat{\sigma}^2}$$

$$\partial^2 \log L/\partial a \partial b = -\frac{1}{\sigma^2}\sum_{\lambda=1}^{N} x_\lambda = -\frac{N\bar{x}}{\sigma^2} \to -\frac{N\bar{x}}{\hat{\sigma}^2}$$

$$\partial^2 \log L/\partial a \partial \sigma = -\frac{2}{\sigma^3}\sum_{\lambda=1}^{N} x_\lambda (y_\lambda - a x_\lambda - b) \to 0$$

$$\partial^2 \log L/\partial b^2 = -\frac{N}{\sigma^2} \to -\frac{N}{\hat{\sigma}^2}$$

$$\partial^2 \log L/\partial b \partial \sigma = -\frac{2}{\sigma^3}\sum_{\lambda=1}^{N} (y_\lambda - a x_\lambda - b) \to 0$$

$$\partial^2 \log L/\partial \sigma^2 = -\frac{3}{\sigma^4}\sum_{\lambda=1}^{N} (y_\lambda - a x_\lambda - b)^2 + \frac{N}{\sigma^2} \to -\frac{2N}{\hat{\sigma}^2}$$

ここでも、パラメータの標準誤差はデータ数 N が増加するにつれ、$1/\sqrt{N}$ で小さくなって行くことが分かる。

4. ロジスティック回帰分析のパラメータ推定

よく使われる 2 値ロジスティック回帰分析とは、事象の出現確率を説明変数の 1 次式のロジスティック関数で推測するモデルである。説明変数が 1 つの 2 値ロジスティック回帰分析の事象の出現確率 p_λ は以下で与えられる。

$$p_\lambda = \frac{e^{z_\lambda}}{1 + e^{z_\lambda}} \quad , \qquad z_\lambda = a x_\lambda + b \tag{1}$$

また、事象の出現と非出現を $y_\lambda = \{1, 0\}$ で表すとその尤度関数は (1) 式の p_λ を用いて以下で与えられる。これはベルヌーイ分布の確率関数を掛けたものである。

$$L=\prod_{\lambda=1}^{N} p_\lambda^{y_\lambda}(1-p_\lambda)^{1-y_\lambda}$$

ここにロジスティック関数の形状は図 1 で与えられ、変域が $(-\infty,\infty)$ で、値域が $(0,1)$ であり、確率の値域と一致している。

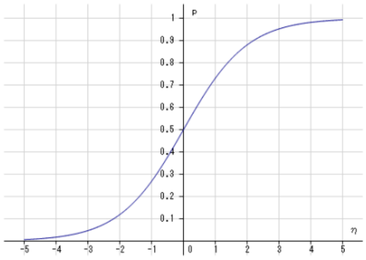


図 1 ロジスティック関数

対数尤度関数は以下で与えられる。

$$\log L=\sum_{\lambda=1}^{N} y_\lambda \log p_\lambda+\sum_{\lambda=1}^{N}(1-y_\lambda)\log(1-p_\lambda)$$

この式に (1) 式を代入した式は補遺に譲るが、これを微分してパラメータ a,b を求める計算は式の上では不可能で、次節で述べる数値計算を必要とする。

5. 最尤法の数値計算

対数尤度関数作成後、ニュートン・ラフソン法を用いてパラメータの推定を行うが、その計算を簡単に示しておく。但しパラメータは２つと仮定し、以下のような行列表現を用いる。

$$\mathbf{b}=\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

まず、対数尤度をパラメータで微分してスコアベクトル $\mathbf{U}$ と情報行列 $\mathbf{J}$ を求める。

$$\mathbf{U}=\begin{pmatrix} \partial\log L/\partial b_1 \\ \partial\log L/\partial b_2 \end{pmatrix},\quad \mathbf{J}=-\begin{pmatrix} \partial^2\log L/\partial b_1^2 & \partial^2\log L/\partial b_1\partial b_2 \\ \partial^2\log L/\partial b_2\partial b_1 & \partial^2\log L/\partial b_2^2 \end{pmatrix}$$

これらを用いてニュートン・ラフソン法でパラメータの推定を行う。ニュートン・ラフソン法は以下のような計算ステップを繰り返し、パラメータの推定値を求める。

$$\mathbf{b}^{(m+1)}=\mathbf{b}^{(m)}+(\mathbf{J}^{(m)})^{-1}\mathbf{U}^{(m)}$$

ここで右肩の (m) や $(m+1)$ は、ニュートン・ラフソン法の計算のステップを表す。ニュートン・ラフソン法では、最初にパラメータの推定値に近い値を何らかの方法で与えて $\mathbf{b}^{(0)}$ とし、それを用いて $\mathbf{U}^{(0)}$, $\mathbf{J}^{(0)}$ を計算し、上の繰り返し計算を実行する。繰り返し計算は、$\mathbf{b}^{(m)}$ の値が落ち着いたら終了する。この情報行列 $\mathbf{J}$ の逆行列 $\mathbf{J}^{-1}$ の対角成分はパラメータの推定値の分散を与えることが知られている。

6.　標準誤差の幾何的意味

ここでは対数尤度関数の形状の特徴を一般的に見てみよう。2 変数関数の形状は推定値（極値点）$(\hat{a},\hat{b})$ の近傍で以下のように与えられる。

$$G=\log L\simeq\frac{1}{2}(a-\hat{a})^2G_{aa}+(a-\hat{a})(b-\hat{b})G_{ab}+\frac{1}{2}(b-\hat{b})^2G_{bb}+c_0$$

ここに、c_0 は対数尤度関数の頂点の高さである。また、

$$G_{aa}=\partial^2G\big/\partial a^2\Big|_{a=\hat{a},b=\hat{b}},\quad G_{bb}=\partial^2G\big/\partial b^2\Big|_{a=\hat{a},b=\hat{b}},\quad G_{ab}=\partial^2G\big/\partial a\partial b\Big|_{a=\hat{a},b=\hat{b}}$$

これを利用すると、情報行列 $\mathbf{J}$ とその逆行列 $\mathbf{J}^{-1}$ は以下のように求められる。

$$\mathbf{J}=-\begin{pmatrix}G_{aa} & G_{ab}\\ G_{ab} & G_{bb}\end{pmatrix},\qquad \mathbf{J}^{-1}=\frac{1}{|\mathbf{J}|}\begin{pmatrix}-G_{bb} & G_{ab}\\ G_{ab} & -G_{aa}\end{pmatrix}$$

情報行列の行列式 $|\mathbf{J}|$ は以下となる。

$G_{ab}=0$ のとき、$|\mathbf{J}|=G_{aa}G_{bb}$

$G_{ab}\neq 0$ のとき、$|\mathbf{J}|=G_{aa}G_{bb}-G_{ab}^2=(1-\rho^2)G_{aa}G_{bb}$

ここに、$\rho=G_{ab}\big/\sqrt{G_{aa}G_{bb}}$

情報行列の逆行列はパラメータの分散、共分散行列になるから、ρ はパラメータの相関係数である。

ここで、パラメータの標準誤差を $\hat{\sigma}_a,\hat{\sigma}_b$ とすると、

$$\hat{\sigma}_a^2=-G_{bb}\big/|\mathbf{J}|=-1\big/[(1-\rho^2)G_{aa}]$$

$$\hat{\sigma}_b^2=-G_{aa}\big/|\mathbf{J}|=-1\big/[(1-\rho^2)G_{bb}]$$

これらより、

$$G_{aa}=-1\big/(1-\rho^2)\,\hat{\sigma}_a^2$$

$$G_{bb}=-1\big/(1-\rho^2)\,\hat{\sigma}_b^2$$

$$G_{ab}=\rho\big/(1-\rho^2)\hat{\sigma}_a\hat{\sigma}_b$$

以上のことから、対数尤度関数 $\log L$ は以下となる。

$$\log L=-\frac{(a-\hat{a})^2}{2(1-\rho^2)\hat{\sigma}_a^2}+\frac{\rho(a-\hat{a})(b-\hat{b})}{(1-\rho^2)\hat{\sigma}_a\hat{\sigma}_b}-\frac{(b-\hat{b})^2}{2(1-\rho^2)\hat{\sigma}_b^2}+c_0$$

次に、特別な点での対数尤度関数の値について見てみる。上の結果から、

$a=\hat{a}\pm 1\big/\sqrt{-G_{aa}}=\hat{a}\pm\sqrt{1-\rho^2}\,\hat{\sigma}_a,\ b=\hat{b}$ のとき、　$\log L\simeq -1/2+c_0$

$a=\hat{a},\ b=\hat{b}\pm 1\big/\sqrt{-G_{bb}}=\hat{b}\pm\sqrt{1-\rho^2}\,\hat{\sigma}_b$ のとき、　$\log L\simeq -1/2+c_0$

$a=\hat{a}+\sqrt{1-\rho^2}\,\hat{\sigma}_a t,\ b=\hat{b}\pm\sqrt{1-\rho^2}\,\hat{\sigma}_b t$ のとき、　$\log L\simeq -(1\mp\rho)t^2+c_0$

上の範囲は広い可能性があるので、正確には範囲を $\alpha\ (0<\alpha\ll 1)$ 倍にした場合、対数尤度関数は（上の値×α^2+c_0）になるということであるが、以後、標準誤差は十分小さいとして上の式をそのまま使うことにする。

次に、対数尤度関数の頂点での軸に沿った曲率を考えてみよう。今、中心が $(\hat{a},\hat{b})$ でパラメータ a の軸に沿った半径 R_a の半円形の関数を考える。

$$f(a)=\sqrt{R_a^2-(a-\hat{a})^2}$$

この２回微分は、

$$\frac{d^2}{da^2}f(a)=-\frac{1}{\sqrt{R_a^2-(a-\hat{a})^2}}-\frac{(a-\hat{a})^2}{[R_a^2-(a-\hat{a})^2]^{3/2}}$$

これより $a=\hat{a}$ のところでは、

$$\left.\frac{d^2}{da^2}f(a)\right|_{a=\hat{a}}=-\frac{1}{R_a}$$

これと $G_{aa}=-1\big/(1-\rho^2)\hat{\sigma}_a^2$ より、頂点での曲率を合わせると、

$$R_a=-1\big/G_{aa}=(1-\rho^2)\hat{\sigma}_a^2$$

同様にして、

$$R_b=-1\big/G_{bb}=(1-\rho^2)\hat{\sigma}_b^2$$

これより、パラメータ間の相関がない場合、対数尤度関数の頂点での軸に沿った曲率半径は標準誤差の 2 乗になる。ここで少し例を考えてみよう。

正規分布のパラメータ推定

正規分布のパラメータ推定では、2 章で述べたように、平均と分散の推定値は以下のように与えられる。

$$\hat{\mu}=\frac{1}{N}\sum_{\lambda=1}^{N}x_\lambda=\bar{x}\,,\qquad \hat{\sigma}^2=\frac{1}{N}\sum_{\lambda=1}^{N}(x_\lambda-\mu)^2=s^2$$

また、情報行列 $\mathbf{J}$ とその逆行列 $\mathbf{J}^{-1}$ は以下で与えられる。

$$\mathbf{J}=\begin{pmatrix} N/s^2 & 0 \\ 0 & 2N/s^2 \end{pmatrix},\qquad \mathbf{J}^{-1}=\begin{pmatrix} s^2/N & 0 \\ 0 & s^2/2N \end{pmatrix}$$

対数尤度関数は、特別の点で以下の値をとる。

$\mu=\bar{x}\pm s\big/\sqrt{N}\,,\ \sigma=s$ のとき、 $\log L\simeq -1/2+c_0$

$\mu=\bar{x},\ \sigma=s\pm s\big/\sqrt{2N}$ のとき、 $\log L\simeq -1/2+c_0$

$\mu=\bar{x}+ts\big/\sqrt{N}\,,\ \sigma=s\pm ts\big/\sqrt{2N}$ のとき、 $\log L\simeq -t^2+c_0$

回帰分析におけるパラメータ推定

回帰分析のパラメータ推定では、3 章で述べたように、パラメータの推定値は以下のように与えられる。

$$\hat{a} = \sum_{\lambda=1}^{N}(y_\lambda - \bar{y})(x_\lambda - \bar{x}) \Big/ \sum_{\lambda=1}^{N}(x_\lambda - \bar{x})^2 = s_{xy} \Big/ s_x^2$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{\lambda=1}^{N}(y_\lambda - \hat{a}x_\lambda - \hat{b})^2 = \frac{1}{N}\sum_{\lambda=1}^{N}\hat{u}_\lambda^2$$

情報行列 $\mathbf{J}$ とその逆行列 $\mathbf{J}^{-1}$ は以下で与えられる。

$$\mathbf{J} = \frac{N}{\hat{\sigma}^2}\begin{pmatrix} s_x^2 + \bar{x}^2 & \bar{x} & 0 \\ \bar{x} & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \qquad \mathbf{J}^{-1} = \frac{\sigma^2}{Ns_x^2}\begin{pmatrix} 1 & -\bar{x} & 0 \\ -\bar{x} & s_x^2 + \bar{x}^2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$

ここでパラメータ a, b, σ の標準誤差と a, b の相関係数は以下となるが、a, b と σ との相関係数は 0 である。

$$\hat{\sigma}_a = \frac{\hat{\sigma}}{\sqrt{N}s_x}, \quad \hat{\sigma}_b = \frac{\hat{\sigma}\sqrt{s_x^2 + \bar{x}^2}}{\sqrt{N}s_x},$$

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{\lambda=1}^{N}(y_\lambda - \hat{a}x_\lambda - \hat{b})^2} = \sqrt{\frac{1}{N}\sum_{\lambda=1}^{N}\hat{u}_\lambda^2}$$

$$\rho = -\bar{x} \Big/ \sqrt{s_x^2 + \bar{x}^2}$$

対数尤度関数は、特別の点で以下の値をとる。

$a = \hat{a} \pm \sqrt{1-\rho^2}\hat{\sigma}_a,\ b = \hat{b},\ \sigma = \hat{\sigma}$ のとき、 $\log L \simeq -1/2 + c_0$

$a = \hat{a},\ b = \hat{b} \pm \sqrt{1-\rho^2}\hat{\sigma}_b,\ \sigma = \hat{\sigma}$ のとき、 $\log L \simeq -1/2 + c_0$

$a = \hat{a} + \sqrt{1-\rho^2}\hat{\sigma}_a t,\ b = \hat{b} \pm \sqrt{1-\rho^2}\hat{\sigma}_b t,\ \sigma = \hat{\sigma}$ のとき、

$$\log L \simeq -(1 \mp \rho)t^2 + c_0$$

7. プログラムの利用法

メニュー［分析－基本統計－ユーティリティー対数尤度関数グラフ］を選択すると図 1 の分析実行画面が表示される。

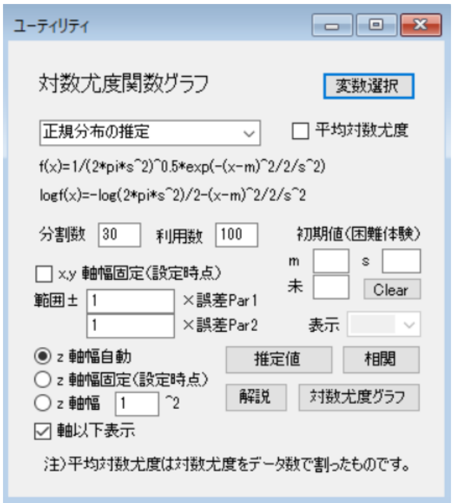


図 1　分析実行画面

このプログラムは実用の計算を行うものではなく、最尤法の対数尤度関数について視覚的に体験するものである。そのため、利用できる分析は、「正規分布の推定」、「回帰分析」及び、「0/1 ロジスティック回帰分析」の 3 つの推定問題である。さらに、回帰分析は説明変数が 1 つ、0/1 ロジスティック回帰分析は目的変数が 0/1 のデータで、説明変数が 1 つという制約が付く。実用的な問題は、多変量解析の中の「重回帰分析」や「2 値ロジスティック回帰分析」で解決できる。これは真に教育的なプログラムである。

まず始めに正規分布のパラメータ推定の問題を考える。図 2 に示すデータ（尤度関数グラフ 1.txt）の変数「y」を利用し、最尤法を用いて平均値と標準偏差を推定する問題を考える。データの数は 500 あるが、まず先頭から 100 個使って最尤法の結果を求める。

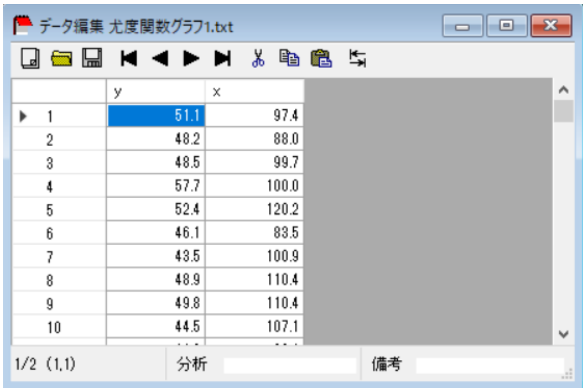


図 2　正規分布のパラメータの推定データ

分析実行メニューの「変数選択」で「y」を選択し、「利用数」をデフォルトの 100 に設定する。「推定値」ボタンをクリックすると、図 3 に示す推定結果を得る。

パラメータの最尤推定

	推定値	標準誤差	2.5%下限	2.5%上限
m(平均)	49.5610	0.5323	48.5178	50.6042
s(標準偏差)	5.3227	0.3764	4.5850	6.0604

図 3　平均と標準偏差の推定結果

ここでは、結果が分かっているので、実際の計算では、その結果の近くをニュートン・ラフソン法の初期値にしている。真にニュートン・ラフソン法を体験する希望があれば、「初期値（困難体験）」の部分に初期値を与えて実行してみることをお勧めする。正しい結果を得るのにかなり苦労すると思う。特に最尤法による回帰分析のパラメータ推定などは至難である。我々は同じ結果が出るものは最も結果が出しやすい手法を使うことを基本にし、最尤法でどうしても難しい場合は、MCMC 乱数発生などで、尤度関数が最も大きくなる点の近傍を求めて初期値にしている。これは紙上の計算通りにはいかない数値計算の難しいところである。

データを 100 個使ったこの問題の対数尤度関数を描いてみよう。x 軸に平均値のパラメータ、y 軸に標準偏差のパラメータを取り、対数尤度関数の表示領域として図 3 で与えた結果を使い、「推定値±1×標準誤差」の範囲をとる。幅が標準誤差の何倍かは、分析実行メニューの「範囲」のところで設定できる。ここではデフォルトの 1 としている。図 4 に「対数尤度グラフ」ボタンをクリックした結果を示す。これは推定値を中心とした素直な山型をしている。

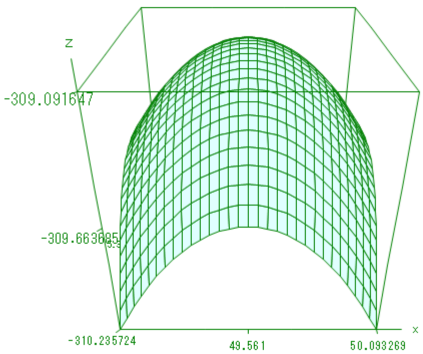


図 4　データ数 100 個の対数尤度関数グラフ

次に、データ数を減らした場合や増やした場合、この形がどのように変わるか見てみよう。データ数 50 個、200 個の対数尤度関数を描いてみる。z 軸の幅を不変にして、結果を図 5 にまとめて示す。

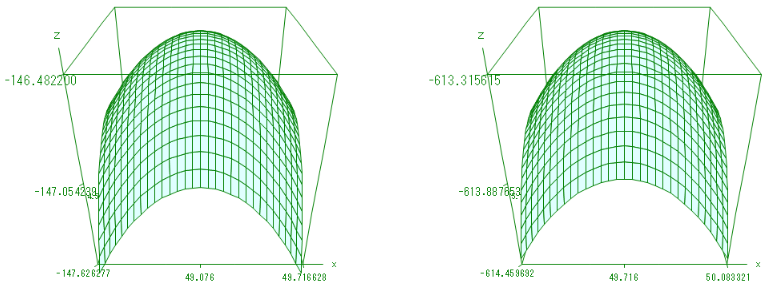


図 5　対数尤度関数（左 50 個，右 200 個）

これを見ると、z 軸の幅は変わっていないが、横軸の幅は、推定値の標準誤差の 2 倍に設定してあるので、標準誤差の値に応じて変わっている。

ここで、実際の標準誤差の値を求めておこう。結果を表 1 に示す。

表 1　データによる標準誤差の変化

	標準誤差（50 個）	標準誤差（100 個）	標準誤差（200 個）
m(平均)	0.6406	0.5323	0.3673
s(標準偏差)	0.4530	0.3764	0.2597

これを見ると、パラメータ推定の標準誤差はデータが増えるに従って表 1 のように小さくなって行く（$\propto 1/\sqrt{N}$）。そのため、対数尤度関数のグラフは、データ数が増えるに従って細くなって行く。

これをはっきりと見るため、軸の幅を図 4 で与えた z 軸と横軸の幅に固定してグラフを描いてみる。結果を図 6 にまとめて示す。

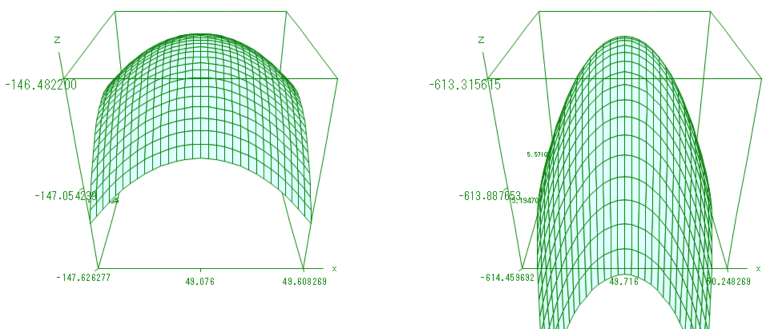


図 6　対数尤度関数（左 50 個，右 200 個）

これより対数尤度関数の曲がり方（尖り方）はデータ数によることがはっきりと分かる。

次に描画の範囲を拡げて、推定値±5×標準誤差の領域をとる。データ数 100 で描かれる対数尤度関数は図 7 の通りである。

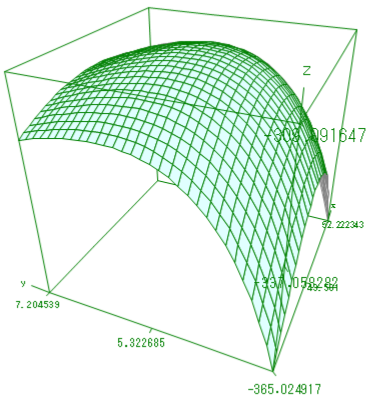


図 7　領域を拡げた対数尤度関数（100 個）

推定値±1×標準誤差の範囲では平均と分散が同じような形状を取っていたが、範囲を拡げると違いがはっきり見えてくる。対数尤度関数は、標準偏差について、推定値より大きな領域と小さな領域で非対称である。

ここで、6 章で調べた結果を再現してみる。近似が成り立つように、描画範囲を小さくして標準誤差の 0.1 倍にし、「z 軸幅」を 0.1^2 に固定する。データを 50 個と 200 個とし、対数尤度関数を描画する。結果を図 8 にまとめて示す。

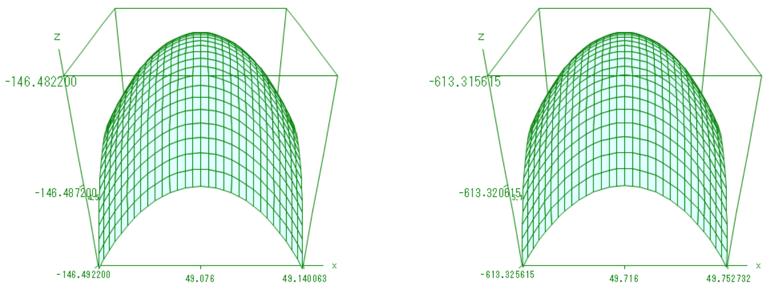


図 8　対数尤度関数（左 50 個，右 200 個）

これらのグラフはデータ数によらず、4 隅からの高さが 0.01 になっている。領域が±標準誤差の範囲だと、関数形が 2 次曲線から少しずれて、高さは 1 と少し異なる。

次に、回帰分析のパラメータ推定について調べてみる。まず分析の種類を「回帰分析」に設定する。図 2 のデータをそのまま用いて、目的変数を「y」、説明変数を「x」として最尤法で推定する。パラメータは回帰係数 2 つと誤差の標準偏差 1 つで 3 つである。パラメ

ータの推定値は数式から計算できるので、簡単のため、その結果を初期値としている。初期値を「初期値（困難体験）」で別に指定する場合は、なかなかよい結果が得られないことを覚悟する必要がある。推定結果を図 9 に示す。

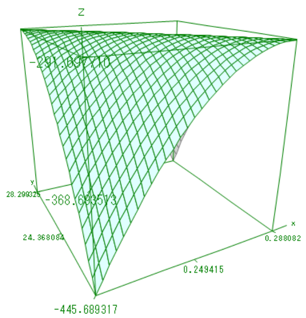
パラメータの最尤推定

	推定値	標準誤差	2.5%下限	2.5%上限
a	0.2494	0.0387	0.1736	0.3252
b	24.3681	3.9312	16.6629	32.0733
s	4.4729	0.3163	3.8530	5.0928

図 9　回帰のパラメータと誤差標準偏差の推定結果（100 個）

対数尤度関数はパラメータの中から 2 組を選んで表示する。まず、分析実行メニューの「表示パラメータ」で回帰係数の２つ「a,b」を選択する。もう 1 つのパラメータは推定値をとるものとして計算する。表示範囲を推定値±1×標準誤差として、データの「利用数」を 100 個で対数尤度関数を表示すると図 10 のようになる。

対数尤度関数は山のような形状かという問いに対して、これは壁のような形状である。この壁はこのサイズではどこが推定点か見分けがつかない（実際は中央がそうであるが）。そのため、z 軸幅を固定して、x,y 軸の幅を推定値±5×標準誤差のように拡げてみる。見やすくするため「分割数」を 50 に増やし、「軸以下表示」のチェックを外した結果を図 11 に示す。



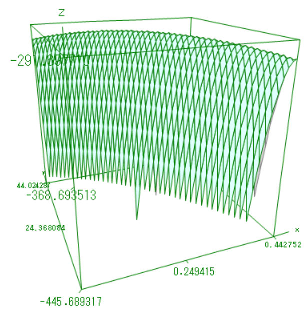


図 10　回帰係数の対数尤度関数　　　図 11　同対数尤度関数（描画領域 5 倍）

図 11 を見ると、稜線はわずかにカーブしており、中央が最も高い。回帰係数の対数尤度関数は壁のような形状であるが、高い部分は確かに存在する。

この結果は、6 章からパラメータの相関係数によっていると考えられるので、この相関係数 ρ と、相関がある場合の描画範囲に掛ける係数 $\sqrt{1-\rho^2}$ を求めてみる。実行画面の「相関」ボタンをクリックすると、図 12 のような結果が表示される。

パラメータの相関係数

	a	b	s
a	1.0000	-0.9935	0.0000
b	-0.9935	1.0000	0.0000
s	0.0000	0.0000	1.0000
描画範囲	aとbに掛ける	0.1138	

図 12　相関係数と描画範囲に掛ける係数

確かに相関係数は -1 に近い値となっている。そこで、「範囲」を±0.1×0.1138×標準誤差、「z 軸幅」を 0.1^2 として、データ数 100 の対数尤度関数を描いてみる。結果を図 13 に示す。

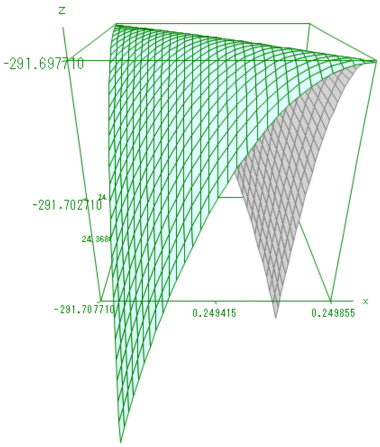


図 13　相関のあるパラメータ間の対数尤度関数

これは 6 章で述べた結果とよく一致している。

　最後に表示パラメータを「a,s」に設定し、「範囲」を推定値±1×標準誤差として対数尤度関数を表示すると図 14 のようになる。

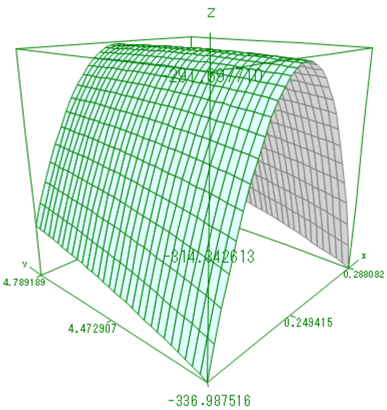


図 14　回帰係数と誤差標準偏差の対数尤度関数

これは回帰係数の対数尤度関数と同じような壁の構造であるが、方向が誤差標準偏差の方向に連なる壁である。この形状は表示範囲と関係があると思われるので、「範囲」をパラメータ a は±0.1×0.1138×標準誤差、パラメータ s は±0.1×標準誤差、z 軸幅を 0.1^2 に固定して対数尤度関数を表示する。結果を図 15 に示す。

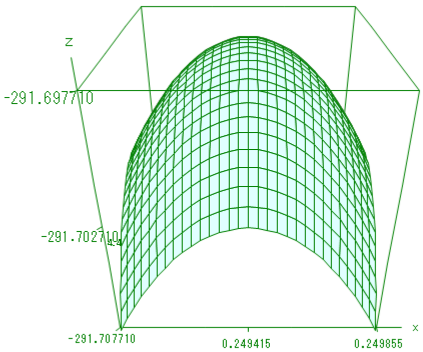


図 15 前節のように設定した回帰係数と誤差標準偏差の対数尤度関数

これを見ると、パラメータ間の相関が 0 の形状で、6 章で述べた結果と一致している。ロジスティック回帰分析の対数尤度関数については、回帰分析と同様であるので省略する。

8. おわりに

最尤法における対数尤度関数は、それがどのような形をしているのか、正確に想像できる利用者はあまり多くないと思われる。その理由はパソコンで最尤推定値が簡単に表示され、気に留める必要がないこと、またその関数が多くの実測値を含んだ形で作られており、実際に描画することが容易でないこと、などが考えられる。

我々はこの多少厄介な関数について、実際に目で見られるツールを開発し、それを用いて代表的な分析の対数尤度関数を観察した。その結果と簡単な理論的考察から、対数尤度関数の形状に関して見落としがちな以下の性質を確認した。

1. 対数尤度関数は必ずしも山のような形状をしていない。パラメータ間に相関がある場合は壁のような構造にもなる。
2. 縦軸を固定し、標準誤差の範囲を横軸に取ると、対数尤度関数は、データ数によらずほぼ一定の形状をしている。

特に 2.から、縦軸と横軸を固定した場合、対数尤度関数はデータ数が増えると次第に細くなっていく。これは推定値の標準誤差が小さくなることと同値である。

これらの性質は理論的に考えるともっともな性質であるが、通常あまり気にしていない

性質である。たまには自分が行っている分析の基本的な部分について、思いを巡らせてもよいのではなかろうか。

謝辞

このプログラムは淵上由衣花氏の卒業論文用に作成したものである。通常ではあまり考えないようなものであるが、議論を進めることで最尤法の理解、特にデータ数と対数尤度関数の形状との関係が明白になった。心より感謝します。

参考文献

［１］ 福井正康､淵上由衣花､尤度関数の視覚化､日本教育情報学会第 36 回年会論文集､(2020) 332-333, （札幌学院大学, 2020/8/22-23）

［２］ 福井正康，College Analysis 総合マニュアル －多変量解析２－、https://www.heisei-u.ac.jp/ba/fukui/gmanual/gmanual03_2.pdf

補遺　0/1 ロジスティック回帰分析の推定

ここでは目的変数が 0/1、説明変数が 1 つのロジスティック回帰分析の最尤推定の方法をまとめておく。

確率関数： $p^y(1-p)^{1-y} \quad y=\{0,1\}$　ベルヌーイ分布

尤度関数： $L=\prod_{\lambda=1}^{N} p_\lambda^{y_\lambda}(1-p_\lambda)^{1-y_\lambda}$

対数尤度： $\log L=\sum_{\lambda=1}^{N} y_\lambda \log p_\lambda + \sum_{\lambda=1}^{N}(1-y_\lambda)\log(1-p_\lambda)$

ここで、

$$p_\lambda=\frac{e^{z_\lambda}}{1+e^{z_\lambda}}, \quad z_\lambda = ax_\lambda + b, \quad \text{さらに、} z_\lambda=\log\frac{p_\lambda}{1-p_\lambda}, \quad e^{z_\lambda}=\frac{p_\lambda}{1-p_\lambda}$$

スコアベクトル $\mathbf{U}$ と情報行列 $\mathbf{J}$ （$\mathbf{b}=\boldsymbol{\beta}+\mathbf{J}^{-1}\mathbf{U}\sim N(\boldsymbol{\beta},\mathbf{J}^{-1})$）

$$\boldsymbol{\beta}=\begin{pmatrix} a \\ b \end{pmatrix}, \quad \mathbf{U}=\begin{pmatrix} \partial\log L/\partial a \\ \partial\log L/\partial b \end{pmatrix}, \quad \mathbf{J}=-\begin{pmatrix} \partial^2\log L/\partial a^2 & \partial^2\log L/\partial a\partial b \\ \partial^2\log L/\partial a\partial b & \partial^2\log L/\partial b^2 \end{pmatrix}$$

最初に以下を求め、

$$\frac{\partial p_\lambda}{\partial a}=\frac{\partial z_\lambda}{\partial a}\frac{\partial p_\lambda}{\partial z_\lambda}=x_\lambda\frac{e^{z_\lambda}(1+e^{z_\lambda})-e^{z_\lambda}e^{z_\lambda}}{(1+e^{z_\lambda})^2}=\frac{x_\lambda e^{z_\lambda}}{(1+e^{z_\lambda})^2}=x_\lambda p_\lambda(1-p_\lambda)$$

$$\frac{\partial p_\lambda}{\partial b} = \frac{\partial z_\lambda}{\partial b}\frac{\partial p_\lambda}{\partial z_\lambda} = \frac{e^{z_\lambda}(1+e^{z_\lambda}) - e^{z_\lambda}e^{z_\lambda}}{(1+e^{z_\lambda})^2} = \frac{e^{z_\lambda}}{(1+e^{z_\lambda})^2} = p_\lambda(1-p_\lambda)$$

この関係を用いて、対数尤度関数の微分を得る。

$$\frac{\partial \log L}{\partial a} = \sum_{\lambda=1}^{N}\frac{y_\lambda}{p_\lambda}\frac{\partial p_\lambda}{\partial a} - \sum_{\lambda=1}^{N}\frac{1-y_\lambda}{1-p_\lambda}\frac{\partial p_\lambda}{\partial a} = \sum_{\lambda=1}^{N}\frac{y_\lambda - p_\lambda}{p_\lambda(1-p_\lambda)}\frac{\partial p_\lambda}{\partial a} = \sum_{\lambda=1}^{N}(y_\lambda - p_\lambda)x_\lambda ,$$

$$\frac{\partial \log L}{\partial b} = \sum_{\lambda=1}^{N}\frac{y_\lambda - p_\lambda}{p_\lambda(1-p_\lambda)}\frac{\partial p_\lambda}{\partial b} = \sum_{\lambda=1}^{N}(y_\lambda - p_\lambda),$$

$$\frac{\partial^2 \log L}{\partial a^2} = -\sum_{\lambda=1}^{N}\frac{\partial p_\lambda}{\partial a}x_\lambda = -\sum_{\lambda=1}^{N}x_\lambda^2 p_\lambda(1-p_\lambda)$$

$$\frac{\partial^2 \log L}{\partial a \partial b} = -\sum_{\lambda=1}^{N}\frac{\partial p_\lambda}{\partial b}x_\lambda = -\sum_{\lambda=1}^{N}x_\lambda p_\lambda(1-p_\lambda)$$

$$\frac{\partial^2 \log L}{\partial b^2} = -\sum_{\lambda=1}^{N}\frac{\partial p_\lambda}{\partial b} = -\sum_{\lambda=1}^{N}p_\lambda(1-p_\lambda)$$

複数の説明変数を持つ一般的なモデルについては、総合マニュアルの多変量解析２の２値ロジスティック回帰分析の章[2]に詳しい。