最小自乗法と最尤法における係数の標準誤差の「直観的」な理解

2018年11月7日

森平　爽一郎
（慶應義塾大学名誉教授）

線形回帰分析における傾きbの推定値の標準誤差は

$$y_t = \alpha + \beta x_t + e_t$$

母集団におけるyとxの関係。αとβの具体的な値は未知であり、それらを推定したい。

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{e}_t$$

母集団からの標本yとxを得て。αとβの値を推定したい。それが

最小自乗法:　誤差の自乗和を最小にする係数　α、β、誤差項eの標準誤差σを推定する。

$$\hat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{T}(x_t - \overline{x})^2} = \frac{Cov(x,y)}{Var(y)}$$

係数の標準誤差の「意味」

$$\sigma_{\hat{\beta}}^2 \equiv Var(\hat{\beta}) = \frac{\sigma_{\hat{e}}^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2} \cdot \quad where \ \sigma_{\hat{e}}^2 \equiv \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}\left(y_t - \left(\hat{\alpha} + \hat{\beta} x_t\right)\right)^2$$

係数の標準誤差は何を意味しているのか？ 「t値=2以上あり」？

分母の意味は $\sum_{t=1}^{T}(x_t - \bar{x})^2$

分子の意味は $\sigma_{\hat{e}}^2 \equiv \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}\left(y_t - (\hat{\alpha} + \hat{\beta} x_t)\right)^2$

ー最尤法とは何か？ー

最尤法による係数の標準誤差の計算は
何を意味しているのか？

コイン投げ実験を例にとる

1. いま、3回コインを投げた。
2. 1回目に表（H),2回目にも表（H),3回目に裏（T)が出た。
3. 知りたいこと：まったく他に情報が無かったときに、このコインが表（H)が出る確率（q）は幾らだろうか？

表＝Head

注：　もし、事前に、このコインがイカサマで無いというある程度の確信があったときには、前にあげた3つの方法に加え、4）ベイズ推定方法がある。これに関しては、新しい倒産確率方法として、時間があれば、後で触れる。

コイン投げ実験

実験番号 (i)	1	2	3
イベント	表 （H)	表 （H)	裏 （T)
数値で表す:y_i	y_1=1	y_2=1	y_3=0

y_iは確率変数であることに注意。ベルヌイ分布

y_i=1　もしi番目のコイン投げで表（H)が出たら

y_i=0　もしi番目のコイン投げで裏（T)が出たら

最尤法1　尤度とは何か？

1. 尤度: Likelihood ⇒ 確率≒確からしさ
2. 表の出る確率qが与えられたときに、コインの表が3回出る確率
3. コインの表（H)がX回出る確からしさ
   – 例1　1枚のコインを3回なげて、「表(H)、表(H)、裏(T)」がでた
   – 例2　コインを3個投げて、2個が表、1個が裏がでた。

$$\Pr\left(HHT\,|q\right)$$

尤度 L: Likelihood の計算

1. 第1回目のコイン投げで表がでる確率：$\Pr(y_1=1)=\Pr(H)=q$　同じコインを投げていることに注意
2. 第2回目のコイン投げで表がでる確率：$\Pr(y_2=1)=\Pr(H)=q$
3. 第3回目のコイン投げで裏がでる確率：$\Pr(y_3=0)=1-\Pr(H)=1-q$
4. 3回のコイン投げで、{表、表、裏}={H,H,T}が出る「同時」確率：$\Pr(y_1=1)\times\Pr(y_2=1)\times\Pr(y_3=0)=\Pr(H)\Pr(H)\Pr(T)=q\times q\times(1-q)$

重要な仮定：コイン投げ実験は互いに「独立」である。確率の「積」の公式が利用できる。

3回のコイン投げで {表、表、裏}={H,H,T}がでる「同時」確率：尤度関数

$$
\begin{aligned}
L &= \Pr\left(H\cdot H\cdot T\,|q\right)=\Pr\left(y_1=1,y_2=1,y_3=0\,|q\right)\\
&= \Pr\left(Y|q\right)\quad where\ Y=\left(y_1=1,y_2=1,y_3=0\right)\\
&= \Pr\left(y_1=1\right)\times\Pr\left(y_2=1\right)\times\Pr\left(y_3=0\right)\\
&= q\times q\times\left(1-q\right)\\
&= q^2\left(1-q\right)^1
\end{aligned}
$$

コインの表が出る確率qがわかったとして、、実験の結果、情報（データ）Yの生じる確率、確からしさをしめしている。

Pr(HHT)と言う結果（同時確率）を最大にするような、このコイン投げの構造、すなわち、表が出る確率qはいくらか？　⇒　尤度関数を最大にするqをもとめる 。

重要：　実験（Y=(y1,y2,y3)）から得られた事実を尊重しよう。そうした事実の背後にある未知の確率qを推定する。

より一般的に表現すると

尤度

$$
y_i=\begin{cases}1 & \text{もしi番目コイン投げで「表(H)」がでたら。}\\ 0 & \text{もし「裏（T)]が出たら}\end{cases}
$$

$$
\begin{aligned}
L &= \Pr\left(H\cdot H\cdot T\,|q\right)\\
&= \Pr\left(y_1=1\right)\times\Pr\left(y_2=1\right)\times\Pr\left(y_3=0\right)\\
&= q\times q\times\left(1-q\right)\\
&= \left[q^{d_1}\left(1-q\right)^{1-y_1}\right]\times\left[q^{d_2}\left(1-q\right)^{1-y_2}\right]\times\left[q^{d_3}\left(1-q\right)^{1-y_3}\right]\\
&= q\times q\times\left(1-q\right)\\
&= q^2\left(1-q\right)^1
\end{aligned}
$$

ここで　$y_1=1$　$y_2=1$　$y_3=0$

尤度L(q)を最大にする！確率qの推定値は？

ハットが付いていることに注意。qの値の推定値の意味。

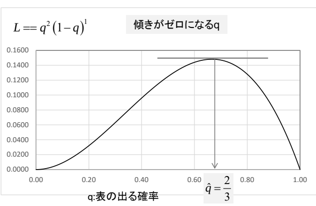
$$0\equiv\frac{\partial L}{\partial q}=\frac{\partial q^2\left(1-q\right)^1}{\partial q}$$ 積の微分公式をつかう

$$0=2\hat{q}\left(1-\hat{q}\right)^1+\hat{q}^2\left(-1\right)$$

$$0=2\hat{q}-2\hat{q}^2-\hat{q}^2=2\hat{q}-3\hat{q}^2$$

$$
\begin{aligned}
&0\equiv 2\hat{q}-3\hat{q}^2\\
&\Rightarrow\ 2\hat{q}=3\hat{q}^2\\
&\Rightarrow\ 2=3\hat{q}\\
&\Rightarrow\ \hat{q}=\frac{2}{3}
\end{aligned}
$$

モーメント法、最少二乗法と同じ結果！



「対数」尤度関数lnL(q)の最大化

同時確率＝尤度の両辺の自然対数　ln()=$\log_e$ ()をとる。何故、対数をとるのか？

$$\ln L=\ln\left(q^2\left(1-q\right)^1\right)=2\ln q+1\ln\left(1-q\right)$$

1. 対数をとることにより、積が和になった。和の微分は難しくない。
2. 対数変換は、単調増加変換。尤度を最大にするπも、対数尤度を最大にするqも、同じqの値になっている。山（尤度関数と対数尤度関数）の位置は変わらず、上に並行移動した

$$0\equiv\frac{\partial\ln L}{\partial q}=\frac{\partial\left[2\ln q+1\ln\left(1-q\right)\right]}{\partial q}\qquad 0=2\left(\frac{1}{\hat{q}}\right)+1\left(\frac{-1}{1-\hat{q}}\right)$$

思い起こそう:　対数関数　y=ln(X)のxに関する微分は？

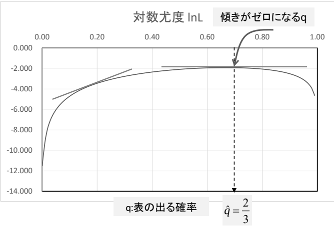
対数尤度関数:lnL(q)の最大化

$$0 \equiv \frac{\partial \ln L}{\partial q} = \frac{\partial\left[2\ln q + 1\ln(1-q)\right]}{\partial q} \quad 0 = 2\left(\frac{1}{\hat{q}}\right) + 1\left(\frac{-1}{1-\hat{q}}\right)$$

Ln(x)のxに関する微分は1/xである。

$$0 = 2\left(\frac{1}{\hat{q}}\right) + 1\left(\frac{-1}{1-\hat{q}}\right)$$
$$\Rightarrow \left(\frac{2}{\hat{q}}\right) = \left(\frac{1}{1-\hat{q}}\right)$$
$$\Rightarrow 2(1-\hat{q}) = \hat{q}$$
$$\Rightarrow 2 = 3\hat{q}$$
$$\Rightarrow \hat{q} = \frac{2}{3} = 0.66666\cdots$$

尤度L(q)を最大にした時と同じ結果！



q:表の出る確率 $\hat{q} = \frac{2}{3}$

尤度関数の数値計算解法

	尤度	対数尤度
確率π	π^2×π	2lnπ+ln(1−π)
0.0	0.000	−9.220
0.1	0.009	−4.711
0.2	0.032	−3.442
0.3	0.063	−2.765
0.4	0.096	−2.343
0.5	0.125	−2.079
0.6	0.144	−1.938
0.7	0.147	−1.917
0.8	0.128	−2.056
0.9	0.081	−2.513
1.0	0.010	−4.625

q=0.666・・・で山の頂上＝対数尤度が最大になっている

$q^2(1-q)^1$　　$2\ln q + 1\ln(1-q)$

「表の出る確率」を確率変数とみなす

1. わずかN=3回の試行（実験）から得たqの推定値は信頼できるか？
2. Nを増やす必要があるのでは？ N回の試行を行ったときのqの散らばりの尺度である分散を計算する。
3. 証明なしに、分散は次のようにして計算する。

$$I(q) \equiv E\left[-\frac{\partial^2 \ln L}{\partial q^2}\right]$$

対数尤度の二階微分、「傾き」の傾きを計算している。

「表の出る確率」の標本情報（Sample Information)と呼ぶ。

$$Var(q) = I(q)^{-1} = E\left[-\frac{\partial^2 \ln L}{\partial q^2}\right]^{-1}$$

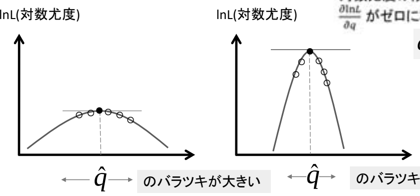
表の出る確率qの推定値の分散（散らばり）は標本情報の逆数（逆行列で計算できる

期待値計算は、通常大変なので、$Var(q)_{|q=\hat{q}} = I(\hat{q})^{-1} = E\left[-\frac{\partial^2 \ln L}{\partial \hat{q}^2}\right]^{-1}$　$q = \hat{q}$ で評価した時の、標本情報の逆数

重要、この式の直感的な理解をえる。

標本情報の意味：直感的な理解

対数尤度の傾き（接戦） $\frac{\partial \ln L}{\partial q}$ がゼロになる点は $q = \hat{q}$



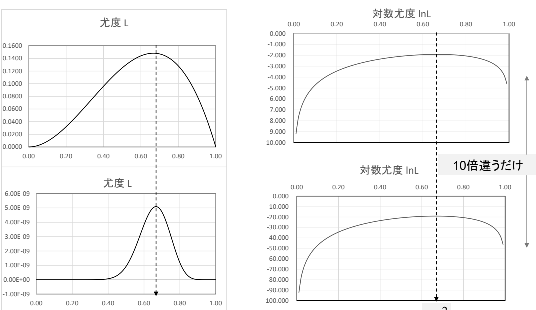
$\hat{q}$ のバラツキが大きい　$\hat{q}$ のバラツキが小さい

表の出る確率（q）

対数尤度が一番大きくなる点（●）は、その近くの対数尤度（○）とは差がない。⇒ 情報量が少ない。

対数尤度が一番大きくなる点（●）は、その近くの点（○）とは対数尤度の大きさに差がある。⇒ 情報量が多い。言い換えれば推定された $\hat{q}$ の「散らばり（分散）」は小さいはず。

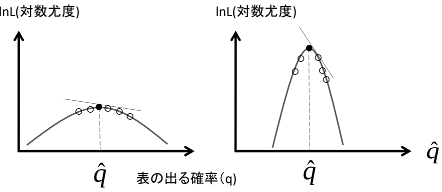
$L = q^2(1-q)^1$ → 対数変換 → $\ln L = 2\ln q + 1\ln(1-q)$



$L = q^{20}(1-q)^{10}$　$\hat{q} = \frac{2}{3}$　$\hat{q} = \frac{2}{3}$

$\ln L = 20\ln q + 10\ln(1-q)$

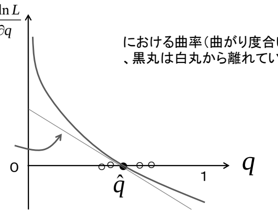
標本情報の意味



前ページのようなことを数式でどのように表したら良いのか？
黒丸●の「近辺」における対数尤度の「傾き」の「傾き」。
つまり「曲率（加速度）」の大きさ、つまり対数尤度の二次微分の大きさでそのことを表すことができる

対数尤度の1次微分と二次微分

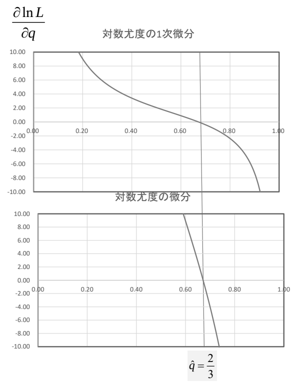
縦軸は対数尤度の1次微分 $\frac{\partial \ln L}{\partial q}$



における曲率(曲がり度合い)が大きいほど、黒丸は白丸から離れているといえる

は対数尤度の1次微分がゼロになる点

$$0 \equiv \frac{\partial \ln L}{\partial q} \Rightarrow 0 \equiv \frac{1}{\hat{q}}\sum_{i=1}^{N} y_i + \frac{-1}{1-\hat{q}}\sum_{i=1}^{N}(1-y_i) \Rightarrow \hat{q} = \frac{1}{N}\sum_{i=1}^{N} y_i = \bar{y}$$



$$\frac{\partial \ln L}{\partial q} = 2\left(\frac{1}{\hat{q}}\right) + 1\left(\frac{-1}{1-\hat{q}}\right)$$

$$\frac{\partial \ln L}{\partial q} = 20\left(\frac{1}{\hat{q}}\right) + 10\left(\frac{-1}{1-\hat{q}}\right)$$

qの標本情報を計算する

標本情報
スコア(Score)は $\frac{\partial \ln L}{\partial q} \Rightarrow \frac{1}{\hat{q}}\sum_{i=1}^{N} y_i + \frac{-1}{1-\hat{q}}\sum_{i=1}^{N}(1-y_i)$ であったので、標本情報は更にこの結果をqで偏微分するとにより、

$$\frac{\partial^2 \ln L}{\partial \hat{q}^2} = \frac{-1}{\hat{q}^2}\sum_{i=1}^{N} y_i + \frac{1}{(1-\hat{q})^2}\sum_{i=1}^{N}(1-y_i)$$

$\hat{q} = \frac{1}{N}\sum_{i=1}^{N} y_i$ であることを思い出そう

$$= \frac{-1}{\hat{q}^2}\sum_{i=1}^{N} y_i + \frac{1}{(1-\hat{q})^2}\left[N - \sum_{i=1}^{N} y_i\right] = \frac{-N\hat{q}}{\hat{q}^2} + \frac{1}{(1-\hat{q})^2}[N - N\hat{q}]$$

$$= \frac{-N}{\hat{q}} + \frac{N(1-\hat{q})}{(1-\hat{q})^2} = \frac{-N}{\hat{q}} + \frac{N}{(1-\hat{q})} = \frac{N}{\hat{q}(1-\hat{q})}$$

$Var(\hat{q}) = I(\hat{q})^{-1} = E\left[-\frac{\partial^2 \ln L}{\partial \hat{q}^2}\right]^{-1}$ から $Var(\hat{q}) = I(\hat{q})^{-1} = \frac{\hat{q}(1-\hat{q})}{N}$ ベルヌイ分布の分散得た。

数値解法によるqの推定

k=1,2,・・・K回の繰り返し計算による推定

K=1回目の計算では、$\hat{q}(0)$ が初期値として与えられたとして、次の計算を行う。

$$\hat{q}(1) = \hat{q}(0) + \left[-\frac{\partial^2 \ln L}{\partial \hat{q}(0)^2}\right]^{-1} \frac{\partial \ln L}{\partial \hat{q}(0)}$$

Newton and Raphson法

もし$\frac{\partial^2 \ln L}{\partial \hat{q}(0)^2}$の値がゼロに近い値をとるとその逆数を計算することができない。左辺の$\hat{q}(1)$は「非常に大きな値」を取ることになる。

また、右辺第2項の二次偏微分の計算そのものが困難であることもある。

そうした場合BHHH法と呼ばれる方法を用いることがある。これは

数値解法によるqの推定:BHHH法

$$-\frac{\partial^2 \ln L_i}{\partial \hat{q}(0)^2}$$ を $$\left(\frac{\partial \ln L_i}{\partial \hat{q}(0)}\right)^2$$ で置き換える

ここで $\ln L = \sum_{i=1}^{N} \ln L_i$ であるので $-\frac{\partial \ln L}{\partial \hat{q}(0)} = -\frac{\partial \sum \ln L_i}{\partial \hat{q}(0)} = \sum_{i=1}^{N} \frac{\partial \ln L_i}{\partial \hat{q}(0)}$

従って

$$\hat{q}(1) = \hat{q}(0) + \left[-\frac{\partial^2 \ln L}{\partial \hat{q}(0)^2}\right]^{-1} \frac{\partial \ln L}{\partial \hat{q}(0)}$$

$$= \hat{q}(0) + \left[\left(\sum_{i=1}^{N} \frac{\partial \ln L_i}{\partial \hat{q}(0)}\right)^2\right]^{-1} \frac{\partial \ln L}{\partial \hat{q}(0)}$$