第6章
線型回帰モデルのベイズ推定

6.1　正規線型モデル
6.2　回帰分析

松本良太

概要

6.1　正規線型モデル

- 因果関係
- 正規線型モデル

6.2　回帰分析

- 標本回帰方程式
- 2つのアプローチ

因果関係

例：y0(℃)の水m(g)にQカロリーの熱を加え、
y(℃)になったとする。

$$m\left(y - y_0\right) = Q$$

$$y = y_0 + \left(\frac{1}{m}\right)Q$$

$$\downarrow \qquad \downarrow$$

$$y = \beta_1 + \beta_2 x \quad (Q = x)$$

このときxとyはxを原因、yを結果とする因果関係を表す。

正規線型モデル(1)

例：1年間の1人当たりの投資をx(ドル)、所得をy(ドル)とする。(詳しいデータについては表6.1参照)

もし、所得yと投資xの間に、

所得＝β1＋β2×投資

に近い関係があるなら、

$$y_i = \beta_1 + \beta_2 x_i + e_i \ (i = 1,2,\dots,n) \qquad (6.1.5)$$

の等式を仮定してもよい。ただし、

$$E(e_i) = 0,\ V(e_i) = \sigma^2$$

正規線型モデル(2)

e_i・・・誤差項

誤差は平均的には0であり、分散は誤差の大きさを表す。(誤差分散という)

さらに、各e_iは、正規分布 $N\left(0,\sigma^2\right)$ に従い、かつ各誤差項は独立であるとする。

ここまで仮定したとき、(6.1.5）式を

<u>正規線型モデル</u>という。

正規線型モデル(3)

β2･･･投資乗数(投資が1ドル増えたとき、所得がいくつ増えるか)

経済理論的には常に正だが、ここでは以下のように定義する。

$$-\infty < \beta_2 < \infty \quad \left(-\infty < \beta_1 < \infty\right)$$

また、 投資＝γ1＋γ2×所得　のような定式化もできる。

↓

不確実性が生じる。

このような状況では、ベイズ統計学のようなアプローチが効果を発揮する。

標本回帰方程式(1)

$$L = \sum_{i=1}^{n} \left\{ y_i - (\beta_1 + \beta_2 x_i)^2 \right\}$$

を最小にするβ1、β2を求める。

Lを微分して $\dfrac{\partial L}{\partial \beta_1} = 0, \quad \dfrac{\partial L}{\partial \beta_2} = 0$

とおき、β1、β2をとくと、それぞれ

$$\hat{\beta}_2 = \frac{\sum \left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right)}{\sum \left(x_i - \bar{x} \right)^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

標本回帰方程式(2)

$\bar{x} = 45.35, \ \bar{y} = 482.95$

$\sum \left(x_i - \bar{x} \right)^2 = 5711, \sum \left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right) = 17418.5$

から、 $\hat{\beta}_1 = 345, \ \hat{\beta}_2 = 3.05$

つまり、投資をy、所得をxとすると、

y=345+3.05x

これを***<u>標本回帰方程式</u>***という。

<u>直線としてみ</u>ると、この傾き(標本回帰係数という)は

3.05であり、負の値からはかけ離れている。

2つのアプローチ(1)

1. 標本回帰係数 β_2 に対して $P(\beta_2 < 0)$ (通常)
2. 母回帰係数 $\hat{\beta}_2$ に対して $P\left(\hat{\beta}_2 < 0\right)$ (ベイズ)

結論から言うと、1の方法では結果に到達できない。

まず、1の場合、$\hat{\beta}_2$ の標本分布は、正規分布に従う。

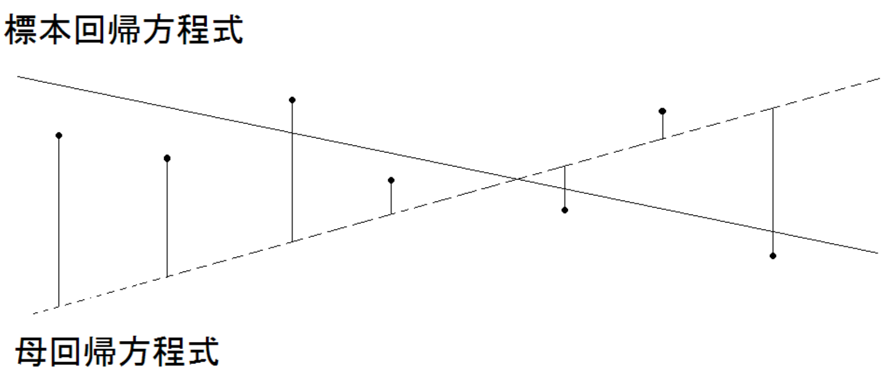
$$N\left(\beta_{2,}\ \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right)$$

2つのアプローチ(2)

$\hat{\beta}_2$ は、元の母集団の真の値 β_2 を中心に分布し、

そのばらつきは元になった標本の誤差分散 σ^2 になる。

一般的に誤差の大きさは限定されていないので、下図のように傾きの正負が変わってしまうこともありうる。



誤差が小さければ真の値を実現できるが、そもそも誤差の値がわからないとどうなるのかわからない。

2つのアプローチ(3)

誤差分散は標本から残差によって

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2}{n-2}, \quad \left(\hat{y}_i \equiv \hat{\beta}_1 + \hat{\beta}_2 x_i\right)$$

と推定できる。これを用いて、$\hat{\beta}_2$ の分散を

$$\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

で推定する。

2つのアプローチ(4)

標準化した変数は、自由度n-2のt分布に従う。

$$t = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\dfrac{\hat{\sigma}_2}{\sum (x_i - \bar{x})^2}}}$$

しかし、この式には仮定しなければいけない β_2 の値が含まれていて、循環論法に陥ってしまうので、この方法では結果に到達できない。