独立成分分析ゼミ

7　推定関数と推定効率

発表日 : 2007/11/7

発表者 : 阿部 竜之介

7章で紹介していること

- 推定関数の定義
- 推定方程式
- 効率の良い推定関数
- 推定誤差解析の基本方針
- 標準推定関数

　推定関数を用いて、未知関数を含むモデルから W を求める方法の紹介

推定関数

推定関数

- 信号 s の未知の確率密度関数 r を知らずに W をうまく推定するのに用いる関数
- 次の性質を満たす関数 $F(x,W)$ のこと

どのような r 、W に対しても

$$E_{w,r}\left[F(x,W')\right]\begin{cases} =0, & W'=W\text{のとき} \\ \neq 0, & W'\neq W\text{のとき} \end{cases}$$

$E_{w,r}$ は x の確率分布が r と W とで決まることを意味している

推定方程式(1/2)

推定方程式

■推定関数を用いた方程式

観測データ $x_1, x_2, \cdots, x_T$ があったとすると以下のような式になる。

$$\sum_{t=1}^{T} F(x_t, W) = 0$$

T が十分大きければ上式は F の期待値に近づく(大数の法則)のでこれを解いて、真の W が求まる。

推定方程式(2/2)

もし r がわかっているなら、F としては

$$F(x,W)=\frac{\partial}{\partial W}\log p(x,W)$$

を使えばよい。これは対数尤度の微分でスコア関数ともいう。

ただ、一般のセミパラメトリックモデル(確率分布が未知関数を含んでいるようなモデル)の場合、間違った r でスコア関数を作ると0にならない問題点がある。

効率の良い推定関数（1/3）

「効率の良い推定関数の族は、スコア関数の情報空間への射影で張られる」という一般定理があるので、とりあえずこれだけを考える。

$F(x,W)$ は行列なので、その要素を $F_{ij}(x,W)$ とする。ここで $\boldsymbol{R}(W)$ を行列 F に働く可逆な線形演算子とする。R は W に依存してもよい。すると

$$\boldsymbol{F}(x,W) = \boldsymbol{R}(W)F(x,W)$$

も推定関数である。

効率の良い推定関数(2/3)

$$\sum_{t=1}^{T} F(x_t, W) = 0$$

$$\sum_{t=1}^{T} \boldsymbol{R}(W) F(x_t, W) = \boldsymbol{R}(W)\{\sum_{t=1}^{T} F(x_t, W)\} = 0$$

２つの推定方程式の答は全く同じで、F と $\boldsymbol{R}F$ は実質的に同じもの、同値だといっていい。

　$\boldsymbol{R}$ は行列から行列への可逆な写像なら何でもよい。$\boldsymbol{R}$ は一般にはインデックスを4つ持った R_{ijkl} という量で、

$$\boldsymbol{F}_{ij}(x, W) = \sum_{k,l} R_{ijkl}(W) F_{kl}(x, W)$$

と書ける。

効率の良い推定関数(3/3)

$\boldsymbol{R}$ をうまく選べば、α, β を定数として

$$\boldsymbol{F}_{ij}(x, w) = I - \alpha\varphi(y)y^T + \beta y\varphi(y)^T$$

というような推定関数も考えられる。これが「$\varphi(y_i)y_j$ で張られる」という意味になる。

推定関数はたくさんある。それならば、どれが良いのか？具体的にはどのφ を選ぶのが良いのか？

↓

信号の真の分布が分からないので、φ は山勘で選んだり、データを見ながら学習で選ぶ方法がある。

■選んだφ によりどの位の誤差になるかという解析が必要

推定誤差の解析（1/4）

推定方程式の解をW_t とする。すると、真の値W に対して推定誤差は $\Delta W_t = W_t - W$　となる。ここで相対誤差$\Delta X_t = W_t W^{-1}$ を用いることにする。

誤差の解析をするには、推定方程式の解も$W_t = W + \Delta X_t W$ と置いて、

$$0 = \sum_{i=1}^{t} F(x_i, W + \Delta X_t W)$$

$$= \sum F(x_i, W) + \sum \frac{\partial F(x_i, W)}{\partial W} \Delta X_t W$$

と書ける。ここで、第1項は期待値0の項をt個足したものである。

推定誤差の解析（2/4）

また、

$$\frac{\partial F}{\partial X} = \frac{\partial F}{\partial W} W^T$$

として

$$\boldsymbol{K} = E\left[\frac{\partial F(x,w)}{\partial X}\right]$$

とおこう。2つのインデックス(a,b) などをひとまとめにして

$A=(a,b), B=(c,d)$ などと書くと、$\boldsymbol{K}$ の成分は

$$\boldsymbol{K}_{AB} = E\left[\frac{\partial F_A}{\partial F_B}\right]$$

のように書ける。これを拡大行列と見ればよい。

推定誤差の解析(3/4)

大数の法則によって、

$$\frac{1}{t}\sum_{i=1}^{t} F(x_i, W) = \boldsymbol{K}$$

と $\boldsymbol{F}$ の期待値に収束する。よって

$$\Delta X_t = -\frac{1}{\sqrt{t}}\boldsymbol{K}^{-1}\left\{\frac{1}{\sqrt{t}}\sum \boldsymbol{F}\right\}$$

と書ける。$\boldsymbol{F}$ の和の方は、中心極限定理によって多次元正規分布に従う。その平均は0で、分散は $\boldsymbol{G} = (G_{ab,cd}) = (G_{AB})$,

$$G_{ab,cd} = E\left[F_{ab}(x, W) F_{cd}(x, W)\right]$$

推定誤差の解析(4/4)

誤差 ΔX_t は正規分布に従い、その分散は

$$E[\Delta X_t \Delta X_t] = \frac{1}{t}\boldsymbol{K}^{-1}\boldsymbol{G}\boldsymbol{K}^{-1}$$

となる。あとは $\boldsymbol{K}$ や $\boldsymbol{G}$ を具体的に計算すればよい。

標準推定関数(1/2)

推定関数F が与えられたときに、同値の推定関数

$$F^* = \boldsymbol{K}^{-1}(W)F(x,W)$$

を作ってみる($\boldsymbol{R}(W) = \boldsymbol{K}^{-1}(W)$ とおいた)。

F^* の微分で作った $\boldsymbol{K}^*$は、

$$\boldsymbol{K}^* = E\left[\frac{\partial F^*}{\partial X}\right] = K^{-1}K = I$$

で単位行列になっている。これを使えば、バッチ式の推定誤差は

$$E\left[\Delta X_t \Delta X_t\right] = \frac{1}{t}\left[F^* F^*\right]$$

と簡単に計算できる。

このF^* のことを F から作られた標準推定関数という。

標準推定関数(2/2)

標準関数を用いた学習は、推定方程式をニュートン法で解くことに相当し、真の W はいつでも安定になる。それに、2次収束で収束がきわめて速い。

具体的な F^* (答のみ)

$$F^*_{aa} = \frac{1}{n_a + 1}\left\{\varphi_a(y_a)y_b - 1\right\}$$

$$F^*_{ab} = C_{ab}\left\{k_b \sigma_a^2 \varphi(y_a)y_b - \varphi(y_b)y_a\right\}$$

$$n_a = E\left[s_a^2 \varphi'_a(s_a)\right]$$

$$k_a = E\left[\varphi'(s_a)\right]$$

$$\sigma_a^2 = E\left[(s_a)^2\right]$$

$$E\left[s_a \varphi_a(s_a)\right] = 1$$