質的変数モデル (2)

別所俊一郎

2006 年 7 月 7 日

Today's attraction

- 最尤法（maximum likelihood estimation）とは
- 非線形最小 2 乗法とは
- その他の質的変数モデル

Probit, logit の推定と推測

- Ch.6 でも非線形な関係の推定について扱った
  - 説明変数について非線形：2 次式などの多項式，交差項，対数変換
  - ただし，推定される係数については線形
- Probit, Logit の推定は係数についても非線形
  - 推定される係数が標準正規分布の関数の「なか」にある
  - OLS 推定を使うことはできない
  - 非線形最小 2 乗推定?
  - 最尤推定?

非線形最小 2 乗法

- OLS は予測誤差の 2 乗和を最小化するような係数を選ぶ

$$\min_{\{b_0,b_1,\cdots,b_k\}} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_{1i} + \cdots + b_k X_{ki})]^2$$

- この考え方を非線形関数にそのまま応用すると，一般に非線形関数 f に対して

$$\min_{\{b_0,b_1,\cdots,b_k\}} \sum_{i=1}^{n} [Y_i - f(b_0 + b_1 X_{1i} + \cdots + b_k X_{ki})]^2$$

  最小化問題の解になるような $b_0, b_1, \cdots, b_k$ を見つけてくればよい

- ただし，この最小化問題には一般には解の公式が存在しない
  - 数値計算（収束計算）で求めるほかない
  - たいていの計量経済ソフトウェアには組み込まれている
  - 最小化すべき式と，求めたい係数さえ指定すればよい

非線形最小 2 乗法

- 推定量の性質
  - 一致性を持つ：サンプルサイズが大きくなると，真の値の周りに分布する確率が 1 に近づく
  - 漸近的に正規分布に従う：サンプルサイズが大きくなると，推定量の分布が正規分布で近似される
  - ただし，より分散の小さい推定量が存在する：有効な（efficient）な推定量ではない
- Probit, Logit の推定には通常は最尤法を用いる

最尤推定

- **尤度関数 Likelihood function**：
  - 未知の（推定したい）係数の関数として書かれたデータの同時確率分布
- **最尤推定量 Maximum Likelihood Estimator**：
  - 尤度関数を最大化するような係数の値
  - 実際に観測されるデータとなるような確率を最大化する
  - この意味で，観測されたデータを生成するのに「最も尤もらしい」係数を選んでいる
  - 分布の形状が既知で，パラメタのみ未知の状況
  - 尤度関数を最大化する係数の値は，対数変換された尤度関数をも最大化するから，実際には対数尤度関数を最大化するものとして計算されることが多い

最尤推定の例

ベルヌーイ試行の例

- 確率変数 Y_1, Y_2 を考える．確率 p のベルヌーイ試行の実現値であり，2 つの変数は互いに独立に同じ分布に従う (i.i.d.) とする
- 確率変数 Y_1, Y_2 が従う確率分布の形状は既知（ベルヌーイ分布）だが，そのパラメタ（p）は未知
- パラメタ p が決まれば，手許にある値が実現する確率を計算することができる．当然，その確率はパラメタ p の値によって異なる．このとき，手許にある値を実現させる確率を最大にするような p を選ぶというのが最尤推定.
- 具体的には，i.i.d. だから，

$$
\begin{aligned}
\Pr(Y_1 = y_1, Y_2 = y_2) &= \Pr(Y_1 = y_1)\Pr(Y_2 = y_2) \\
&= \left(p^{y_1}(1-p)^{1-y_1}\right)\left(p^{y_2}(1-p)^{1-y_2}\right)
\end{aligned}
$$

最尤推定の例

- 手許にある状況が実現する確率は，パラメタ p と実現値 (y_1, y_2) の関数となるが，とくにパラメタ p の関数と見たものを尤度関数と呼ぶ．すなわち，

$$\text{尤度} = L(p; y_1, y_2) = \left(p^{y_1}(1-p)^{1-y_1}\right)\left(p^{y_2}(1-p)^{1-y_2}\right)$$

p の最尤推定量とは，この尤度 $L(p; y_1, y_2)$ を最大化する p

- ベルヌーイ試行の場合には，最尤推定量は単純平均となる
- ただし，最尤推定量のかたちが公式として簡単にもとまることはほとんどなく，最大化問題を数値計算で解くことになる
- 対数変換した尤度 $\ln L(p; y_1, y_2)$ の最大化問題のほうが簡単になることが多い

プロビット推定の場合

説明変数の条件付きで，被説明変数が 1 あるいは 0 の値を取る確率は

$$
\begin{aligned}
\Pr(Y_i = 1|X_{1i}, \cdots, X_{ki}) &= \Phi(b_0 + b_1 X_{1i} + \cdots + b_k X_{ki}) \\
\Pr(Y_i = 0|X_{1i}, \cdots, X_{ki}) &= 1 - \Phi(b_0 + b_1 X_{1i} + \cdots + b_k X_{ki})
\end{aligned}
$$

まとめて書くと，

$$
\Pr(Y_i = y_i|X_{1i}, \cdots, X_{ki}) = \Phi(\mathbf{b}, \mathbf{X_i})^{y_i}(1 - \Phi(\mathbf{b}, \mathbf{X_i}))^{1-y_i}
$$

だから，$Y_1 = y_1, Y_2 = y_2, \cdots, Y_n = y_n$ が同時に起きる確率は，i.i.d. を仮定すると

$$
\begin{aligned}
L(\mathbf{b}|\mathbf{X}) &= \Phi(\mathbf{b}, \mathbf{X_1})^{y_1}(1 - \Phi(\mathbf{b}, \mathbf{X_1}))^{1-y_1} \times \Phi(\mathbf{b}, \mathbf{X_2})^{y_2}(1 - \Phi(\mathbf{b}, \mathbf{X_2}))^{1-y_2} \\
&\quad \times \cdots \times \Phi(\mathbf{b}, \mathbf{X_n})^{y_n}(1 - \Phi(\mathbf{b}, \mathbf{X_n}))^{1-y_n}
\end{aligned}
$$

対数変換して整理すると

$$
\ln L(\mathbf{b}|\mathbf{X}) = \sum_{i=1}^{n} y_i \ln(\Phi(\mathbf{b}, \mathbf{X_n})) + \sum_{i=1}^{n}(1 - y_i)\ln(1 - \Phi(\mathbf{b}, \mathbf{X_n}))
$$

この対数尤度関数を最大化する $\mathbf{b}$ が ML 推定値

最尤推定量の性質

定式化が正しければ

- 一致性を持つ
- 漸近的に正規分布に従う
- 仮説検定（t 検定，F 検定），信頼区間の形成もこれまでと同様
- ただし，計量ソフトウェアによっては χ^2 統計量を出力するものも．
- 行っている検定は同値だが $\chi^2 = qF$（q は制約の数）の関係

適合度の指標

- Fraction correctly predicted
  - 予測が「当たっている」比率
  - 当てはめ値を各観測値ごとにもとめ，予測値が 0.5 以上なら 1，以下なら 0 を当てはめる．この予測値と現実の 0,1 のうち，あたっているものの比率.
  - 考え方は簡単だが，予測の質を考慮しにくい：当てはめ値が 0.51 でも 0.93 でも 1 は 1.
- Pseudo-R^2
  - 尤度は説明変数の数が増えると大きくなる
  - 説明変数を追加したことでどれほど尤度が大きくなっているか，の指標

$$\text{pseudo-}R^2 = 1 - \frac{\text{最大化された対数尤度}}{\text{説明変数がないときの最大化された対数尤度}}$$

質的変数モデルのいろいろ

適当に仮定を置くと尤度関数を構成できる

- **Censored regression model**
  - 被説明変数の値がそのまま観測されないようなデータ
  - 例：留保賃金率より低い賃金率が提示された場合，労働者の賃金率は観測できない
  - ただし，説明変数のほうは観測できる
  - 被説明変数の真の値が観測できない状況を表現する必要
- **Truncated regression model**
  - 被説明変数の値によっては観測できないデータ
  - 例：何も買わなかった買い手のデータは観測できない
  - 説明変数も観測できない．**Sample selection model** とも
  - Censored regression model と合わせて **Tobit model** とも
  - サンプル選択の状況を表現する

質的変数モデルのいろいろ

- **順序プロビット**
  - 被説明変数が質的変数で，順序がある場合
  - 例：学歴
  - 標準正規分布の分布関数を用いて，各状況になりうる確率を求めて尤度関数を構成
- **多項選択モデル　multinomial logit/probit**
  - 被説明変数が質的変数で，順序がない場合
  - 例：交通手段の選択，治療法の選択…
  - 各状況になりうる確率を求めて尤度関数を構成

質的変数モデルのいろいろ

- **Count data**
  - 「回数」が被説明変数になるようなデータ
  - 例：食事の回数，通院回数……
  - 回数は負の値を取らないので，OLS では不適切
  - ポワソン回帰（Poisson regression）か，負の二項回帰（negative binomial regression）が代表的
- **生存期間分析　survival analysis**
  - 「期間」が被説明変数になるような回帰
  - 例：薬剤投与後の生存日数，倒産までの年数，切り替えまでの年数…
  - 期間は負の値を取らないので，OLS では不適切
  - Cox 回帰，ワイブル分布を用いた回帰など