

Parametric Model

労働経済学 (補論)

川田恵介

Table of contents

1	Parametric Model	1
1.1	Motivation	2
1.2	Population model	2
1.3	Parametric Model	2
1.4	例: 男女別労働供給	2
1.5	例: 賃金モデル	3
1.6	Maximum likelihood	3
1.7	間違ったモデル	3
1.8	Estimand	3
1.9	Estimation	4
1.10	Example: Bootstrap	4
1.11	Example: モデルの正しさを前提	4
1.12	まとめ	5
1.13	まとめ: OLS VS 最尤法	5
	Reference	5

1 Parametric Model

- データ生成の”簡単な”確率モデルを仮定し、そのパラメタを推定する
 - (Parametric) 最尤法/ベイズ法
 - 複雑なモデルも推定できるが、
- 批判: 推定結果が、“モデルの正しさ”に強く依存する
 - 最尤法については、モデルの正しさに依存しない解釈が可能

1.1 Motivation

- 労働経済学において、 Y が離散の応用も多い
 - 例: 労働供給関数の推定: $Y = 1$: 働いている、 $Y = 0$: 働いていない
- OLS でも推定できる: $Y \sim X$ を推定すると、母集団における供給関数 $E[Y|X]$ (“就業率”) の線形近似モデル $g(X)$ を推定する
 - しばしば $g(X)$ の値は、 $[0, 1]$ を”はみ出す”
 - 代替的に Probit/Logit を推定する

1.2 Population model

- 議論を単純化するために、ランダムサンプリングを仮定
- 手元のデータのある事例 $i \{Y_i, X_i\}$ は、何らかの確率モデルに従って”実現する”
 - 単純化のために、条件付き確率に限定
 - * 属性 x をもつ事例について、 $Y_i = y$ となる事例を観察する確率 $f_{Y|X}(y|x)$
 - = 母集団における $Y = y$ の割合

1.3 Parametric Model

- Population model を近似するモデル $= g_{Y|X}(y|x; \theta)$
 - $\theta(\in \Theta) = \text{Parameter}$ ($\Theta = \text{Parameter space}$)
- $g_{Y|X}(y|x, \theta) = f(y|x)$ を満たす θ が (Θ 内に) 存在するのであれば、“モデルは正しい”と呼ぶ

1.4 例: 男女別労働供給

- 働いている: $y = 1$ 、いない: $y = 0$
- 女性: $x = 1$ 、男性: $x = 0$
- $g_{Y|X}(y|x; \theta)$
 - $= \theta_{Women} : y = 1|x = 1$
 - $= 1 - \theta_{Women} : y = 0|x = 1$
 - $= \theta_{Men} : y = 1|x = 0$

$$- = 1 - \theta_{Men} : y = 0 | x = 0$$

- 必ず正しいモデル

1.5 例: 賃金モデル

- $y =$ 賃金
- 女性: $x = 1$ 、男性: $x = 0$
- $g_{Y|X}(y|x; \theta)$

$$- = Normal(\theta_{Mean, Women}, \theta_{SD, Women}) : x = 1$$

$$- = Normal(\theta_{Mean, Men}, \theta_{SD, Men}) : x = 0$$

- 各個人の賃金が正規分布に従うと仮定 = おそらく正しくないモデル

1.6 Maximum likelihood

- θ を推定する Algorithm
- 手元のデータが実現する確率 (尤度 Likelihood) が最大になるようなモデルを探す

—

$$\max_{\theta \in \Theta} g_{Y|X}(Y_1|X_1; \theta) \times g_{Y|X}(Y_2|X_2; \theta) \times \dots$$

- “モデルが正しい” のであれば、推定された $g_{Y|X}(y|x; \theta)$ は、 $f_{Y|X}(y|x)$ の優れた推定結果

1.7 間違ったモデル

- $f_{Y|X}(y|x) = g_{Y|X}(y|x; \theta)$ を満たす θ が存在しない場合は?
 - OLS と同様に、最善の近似モデルを推定していると見做せる
- KL divergence:

—

$$KL(\theta) = \int f_{Y|X}(Y|X) \log \left(\frac{f_{Y|X}(Y|x)}{g_{Y|X}(Y|x; \theta)} \right) dY$$

1.8 Estimand

- θ^* ただし

$$\theta = \arg \min_{\theta} KL(\theta)$$

- Recap: OLS では、 β^* ただし

$$\beta^* = \arg \min_{\beta} E[(Y - \beta X)^2]$$

- モデルと母集団の”乖離”を図る物差しが異なる

1.9 Estimation

- 最尤法による推定結果 θ は、 θ^* の一致推定量
 - モデルの正しさに依存しない信頼区間も、ブートストラップなどによって、計算できる
 - * OLS と同様に古典的な推定方法は、モデルの正しさに依存

1.10 Example: Bootstrap

```
car::Boot(
  glm(
    Women ~ education,
    Data,
    family = "binomial"
  ),
  method = "case",
  R = 1000
) |>
car::Confint()
```

Bootstrap bca confidence intervals

	Estimate	2.5 %	97.5 %
(Intercept)	-0.185475889	-1.12207349	0.69667553
education	0.001559809	-0.06477347	0.07132626

1.11 Example: モデルの正しさを前提

```
glm(
  Women ~ education,
  Data,
  family = "binomial"
) |>
confint()
```

	2.5 %	97.5 %
(Intercept)	-1.05426431	0.67989067
education	-0.06368983	0.06689599

1.12 まとめ

- 最尤推定 + 間違ったモデル、であったとしても、OLS と同じような解釈が可能
 - Population Model の母集団における近似モデルを推定
 - * 詳細は Aronow and Miller (2019)などを参照
- ベイズ法は?
 - Section 14 in Buja et al. (2019)などを参照

1.13 まとめ: OLS VS 最尤法

- Y が離散のケースではしばしば logit/probit モデルの最尤法推定が推奨されるが。。。
 - 労働経済学の実践においては、どちらも近似モデル
 - どちらかに明確な優位性があるわけではない
 - * 発展議論と、Chen, Martin, and Wooldridge (2023)などを参照

Reference

- Aronow, Peter M, and Benjamin T Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge University Press.
- Buja, Andreas, Lawrence Brown, Richard Berk, Edward George, Emil Pitkin, Mikhail Traskin, Kai Zhang, and Linda Zhao. 2019. “Models as Approximations i.” *Statistical Science* 34 (4): 523–44.
- Chen, Kaicheng, Robert S Martin, and Jeffrey M Wooldridge. 2023. “Another Look at the Linear Probability Model and Nonlinear Index Models.” *arXiv Preprint arXiv:2308.15338*.