

Maximum Likelihood for Approximation Model

労働経済学 1

川田恵介

1 最尤法 (Maximum Likelihood)

1.1 動機

- OLS よりも”複雑なモデルを推定できる方法”として紹介されがち
 - ▶ 有名な応用は、"Binary outcome (Y が 0 または 1) についてのモデル (Logit/Probit) を最尤法で推定"
- 本質的な違いは、平均値ではなく、**分布の近似モデル**を推定する手法
 - ▶ Generative model を推定する手法とみなすこともできる
- 代替的な推定方法: ベイズ法/Generative Adversarial Network

1.2 Takeaway

- 研究目標と推定対象に応じた、適切な使い分けが必要
- OLS との使い分けについて、大きな誤解がある
 - ▶ “ Y が Binary ならば、OLS は使っては行けない”など
 - 有力な反論 (Angrist & Pischke, 2009)

1.3 Statistical model

- 変数の分布を表すモデル (wiki)
 - ▶ Parametric model: 有限個のパラメタからなるモデル
 - ▶ Non-parametric model: “無限個のパラメタ”からなるモデル
- 教科書的な最尤法/ベイズは、Parametric な統計モデルを推定する方法

1.4 例. Logit 型労働供給モデル

- 労働供給の意思決定 (誰が働いているか) は、労働経済学の古典的な関心
- $Y =$ 就業 (=1) /非就業 (=0)
 - ▶ $Y = 1$ である割合(密度関数)は、 $f(Y = 1 | X)$

- $$f(Y = 1 | X) \simeq \underbrace{g(X)}_{\text{モデル}}$$

$$= \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_L X_L)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_L X_L)}$$

1.5 例. 古典的線型モデル

- ・ ミンサー型賃金”分布”モデル

- ・ $Y = \text{賃金}$

- ▶ Y の密度関数は、 $f(Y | X)$

- $$f(Y | X) \simeq g(X) = \beta_0 + \beta_1 + \dots + \underbrace{u}_{\text{Normal}}$$

- ▶ $u \sim N(0, \beta_\sigma)$

- ・ X が同じであれば、賃金分布は必ず正規分布に従うことを仮定

1.6 生成モデルとしての解釈

- ・ パラメトリックモデルを用いれば、“PC 上で”データを生成できる
 - ▶ 事例を生成するモデルとして解釈できる

2 データ上の計算

2.1 最尤法のアイデア

- ・ 研究者が事前に設定した分布のモデルを、極力データに当てはまるように推定する
 - ▶ KL divergence を最小化する

2.2 実例. 正規分布モデル

- $$f(\text{education}) \simeq \beta_0 + \underbrace{u}_{\text{Normal}}$$

を推定

- ▶ Y が正規分布 $N(\beta_0, \sigma^2)$ に従うモデルを推定

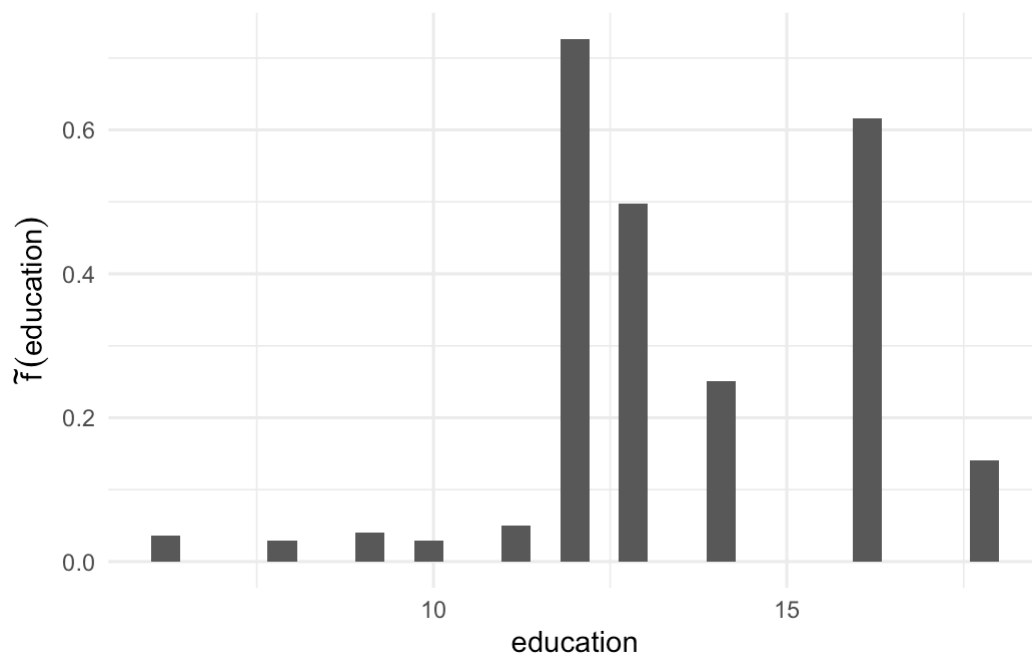
- パラメタ (平均 β_0 と分散 σ^2) が決まれば、education のモデル上の分布を計算 (生成)できる

- ・ データ上の分布 $\tilde{f}(\text{education})$ に最も適合するように、パラメタを選ぶ

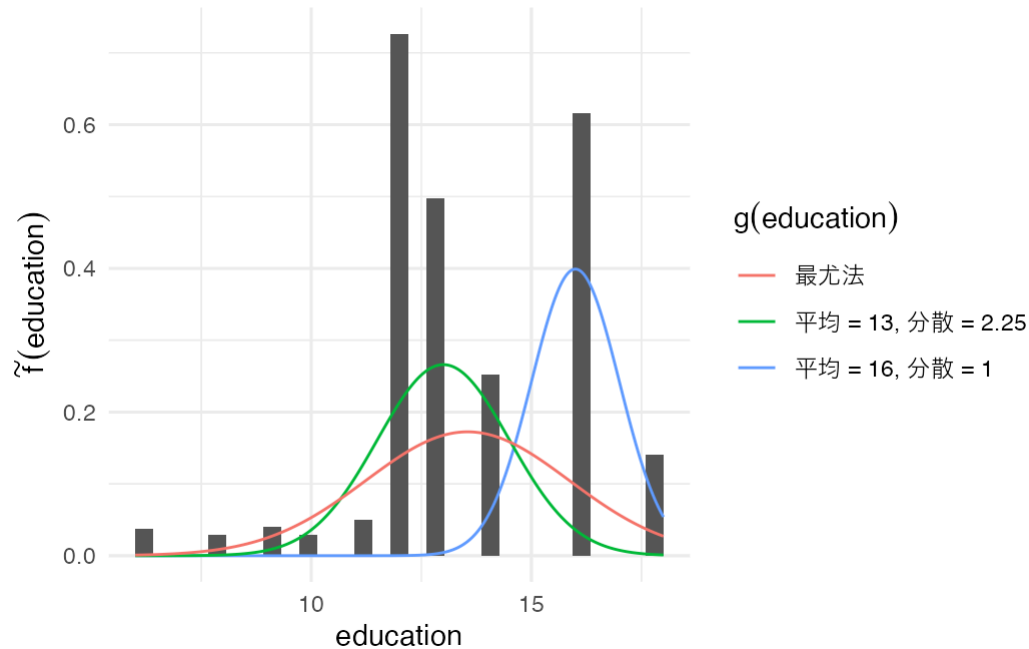
2.3 実例. データ上の分布

education	N	f
6	45	0.015
8	35	0.012
9	49	0.017
10	35	0.012
11	61	0.021
12	887	0.301
13	607	0.206
14	307	0.104
16	752	0.255
18	172	0.058

2.4 実例. データ上の分布



2.5 実例. さまざまなモデル



2.6 KL divergence

- KL divergence を最小化するようにパラメタを推定する
 - ▶ KL divergence = データ上の分布 $\tilde{f}(Y)$ とモデルの分布 $g(Y)$ の乖離度 =

$$\left[\log(\tilde{f}(Y)) - \log(g(Y)) \right] \times \tilde{f}(Y)$$

のすべてのYについての総和

2.7 実例. データ上の分布

education	N	Share	Green	KL_Green	Red	KL_Red
18	172	0.058	0.033	0.564	0.027	0.765
16	752	0.255	0.081	1.147	0.098	0.956
14	307	0.104	0.126	-0.192	0.169	-0.486
13	607	0.206	0.133	0.438	0.168	0.204
12	887	0.301	0.126	0.871	0.138	0.780
11	61	0.021	0.106	-1.619	0.094	-1.499
10	35	0.012	0.081	-1.910	0.053	-1.485
9	49	0.017	0.055	-1.174	0.025	-0.386

education	N	Share	Green	KL_Green	Red	KL_Red
8	35	0.012	0.033	-1.012	0.010	0.182
6	45	0.015	0.009	0.511	0.001	2.708

2.8 実例. データ上の分布

- KL divergence
 - Red model: 0.501
 - Green model: 0.576

2.9 最尤法の別解釈

- KL divergence の最小化 = 以下の総和の最大化と同じ結果

$$\log(g(Y)) \times \tilde{f}(Y)$$

- 対数尤度の最大化

2.10 拡張

- より複雑な分布のモデルも、同じ理屈で推定できる
 - 経済学の応用論文では、 Y の条件付き分布のモデルとして、書かれることが多い
- 例: 古典的線型モデル

$$g(Y | X) = \beta_0 + \dots + \beta_L X_L + \underbrace{u}_{Normal(0, \sigma^2)}$$

- パラメタ: $\beta_0, \dots, \beta_L, \sigma^2$

2.11 拡張

- 以下の KL divergence を最小化するように推定

$$\underbrace{\left[\log(\tilde{f}(Y | X)) - \log(g(Y | X)) \right] \times \tilde{f}(Y | X) \times \tilde{f}(X)}_{X \text{ についての乖離}} \text{ の総和}$$

3 推定対象

3.1 推定対象

- OLS と同様に、母分布の母集団上での近似モデルを推定対象と解釈できる
 - 研究者が設定するモデルが正しい場合、母分布そのものを推定していると解釈できる

3.2 推定対象

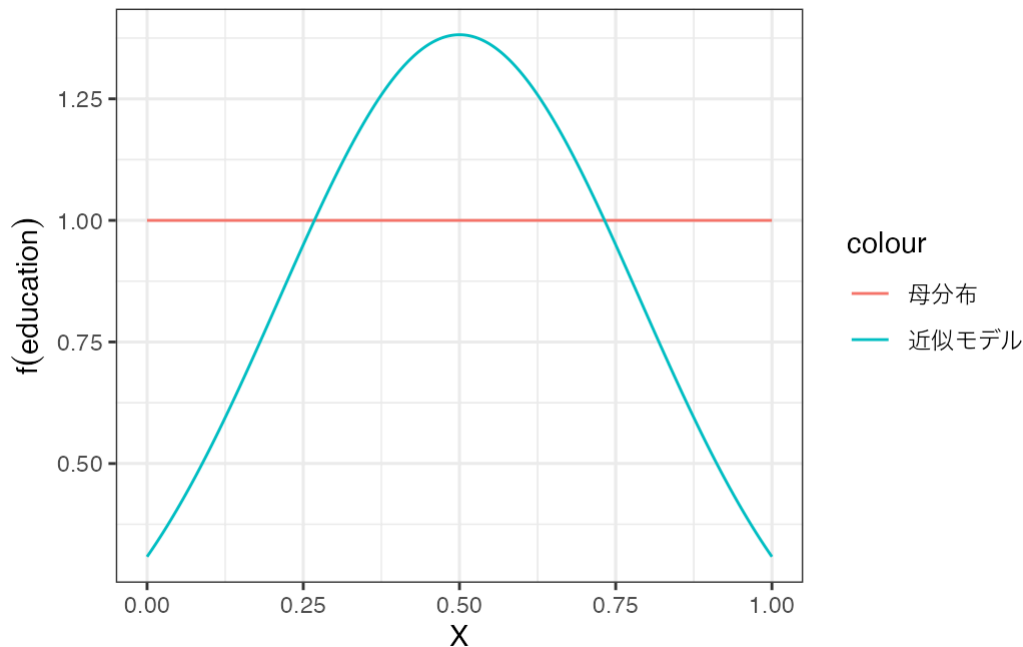
- あるモデル $g(Y)$ を母集団上で最尤法で推定すると、以下を最小化するモデルが計算される

$$[\log(f(Y)) - \log(g(Y))] \times f(Y)$$

のすべての Y についての総和

- 真の母分布と母集団上でのモデル $g(Y)$ を最小化する
 - 近似モデルが推定対象

3.3 例



3.4 推定対象

- もしパラメタを適切に選べば、 $f(Y) = g(Y)$ が達成できるのであれば、推定対象は母分布
 - Misspecification がない状況
 - 多くの入門書が想定

3.5 まとめ

- 母分布の近似モデルを推定する手法として解釈できる
 - 特殊ケース (Misspecification がない) のみ、母分布を推定していると解釈できる
- 後述: OLS と同様に、Misspecification を前提とした信頼区間計算も可能
 - 詳細は、Aronow & Miller (2019) の 5 章、Hansen (2022) の 10.16-19 章

4 実践: 最尤法 VS OLS

4.1 どちらを用いるべきか?

- 一般に研究課題と推定目標に応じて、決める必要がある
 - ▶ 推定値の性質改善を目指して、最尤法を採用するケースもあるが、労働経済学ではあまり有効ではないケースが多い

4.2 最尤法が比較優位

- 分布のモデルを推定したいのであれば、OLS を用いることはできない
 - ▶ 生成モデル、(経済理論などを用いた)構造モデル全体を推定したいなど
- 平均値の非線形モデル (β についての足し算ではない)も、OLS での推定は困難

4.3 OLS が比較優位

- Y の平均値について、線型モデルを推定したいのであれば、OLS に比較優位
 - ▶ Y が Binary (例: 就業状態)であったとしても、OLS は利用できる
 - 予測値が負になり得るが、そもそも近似モデルを推定していると解釈するのであれば、致命的な欠陥ではない (Angrist & Pischke, 2009)
 - ▶ Y の分布の特定の性質に関心があったとしても、活用できる
 - 例: 月給が 20 万円以下の労働者の割合について、モデルを作りたい $Y = Wage \geq 20$ であれば 1、それ以外であれば 0

4.4 OLS が比較優位

- 「バランス後の比較」を行う手法としては、OLS に大きな優位性

- 後述

4.5 推定の問題

- もし Misspecification がない統計モデル を前提にできるのであれば、最尤法の方が優れた推定手法
 - ▶ 私見: 労働経済学の実践においては、Misspecification を前提とすべきであり、今日では最尤法を選択する積極的な理由にならない

5 補論: Parametric VS Semiparametric

5.1 Semiparametric model (wiki)

- 限られた数の推定対象 (Parameter of interest) と”その他の部分”からなる、統計モデルの定式化
- 例:

$$f(Wage) = \beta_0 + \underbrace{u}_{\text{その他の部分}}$$

- ▶ 推定対象は、母分布を用いて、明確に定義する

$$\beta_0 = E[Wage]$$

5.2 Semiparametric model

- その他の部分 u の分布は、有限個のパラメタによる定式化は行わない
- ▶ Parameter of interest の定義から、必然的に満たすべき性質を導出できる

$$Wage = Y - E[Wage]$$

- ▶ 推定のために必要な仮定もある
- 一般に Parametric model に比べて、Misspecification を犯す可能性が低い

5.3 Semiparametric model の推定

- 一定の仮定のもとで、Parameter of interest は推論可能
- 例: ランダムサンプリングデータ (u の分布が独立・無相関) であれば、
 - ▶ 事例数が無限大であれば、データ上の平均賃金 $= \beta_0$
 - ▶ 事例数が大きければ、データ上の平均賃金の分布は、 β_0 を平均とする正規分布で近似可能
 - 信頼区間の近似計算が可能
- その他の部分 u の分布について、信頼できる推定を行っていないことに注意

5.4 Reference

Bibliography

- Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.
- Aronow, P. M., & Miller, B. T. (2019). Foundations of agnostic statistics. Cambridge University Press.
- Hansen, B. (2022). Probability and statistics for economists. Princeton University Press.