2023/06/08

藪友良

第 12 章の答え

練習問題 1

(a) 限界効果は、年齢の係数 0.032 である。つまり、1 年歳をとると約 3%だけ結婚確率は上がる。

(b) 結婚確率は、30 歳と 45 歳において、それぞれ次のようになる。

$$\hat{P}\{Y_i = 1\} = -0.5 + 0.032 \times 30 = 0.46$$

$$\hat{P}\{Y_i = 1\} = -0.5 + 0.032 \times 45 = 0.94$$

(c) 年齢X_iが十分に大きい、もしくは十分に小さいと、確率$P\{Y_i = 1\}$は 1 より大きい、もしくは 0 より小さくなる。たとえば、年齢が 50 歳なら、次のように確率は 100%を超える。

$$P\{\widehat{Y_i = 1}\} = -0.5 + 0.032 \times 50 = 1.1$$

練習問題 2

最終学歴が大卒なら、$X_{2i} = 1$、$X_{3i} = 0$となる。よって、大卒の 30 歳と 45 歳の結婚確率は、それぞれ次のようになる。

$$\hat{P}\{Y_i = 1\} = F(-3.5 + 0.1 \times 30 + 0.3 \times 1 + 0.6 \times 0) = 0.42$$

$$\hat{P}\{Y_i = 1\} = F(-3.5 + 0.1 \times 45 + 0.3 \times 1 + 0.6 \times 0) = 0.90$$

Excel では、「=NORM.S.DIST(-3.5+0.1*30+0.3*1+0.6*0,TRUE）」と入力すれば、確率を 0.42 と計算できる。

練習問題 3

モデルのパラメータ(α^*、β^*、c、σ^2)を用いると、次のようになる(導出は 12 章補足参照)。

$$\alpha = \frac{\alpha^* - c}{\sigma} \qquad \beta = \frac{\beta^*}{\sigma}$$

練習問題 4

$\hat{\alpha}_{ML}$、$\hat{\beta}_{ML}$は最尤推定量としよう。このとき、どのような$\tilde{\alpha}$、$\tilde{\beta}$に対しても、

$$P(\tilde{\alpha}、\tilde{\beta}) \leq P(\hat{\alpha}_{ML}、\hat{\beta}_{ML})$$

が成立する。ただし、$P(\tilde{\alpha}、\tilde{\beta})$は、パラメータが$\tilde{\alpha}$、$\tilde{\beta}$としたときの尤度(同時確率)とする。対数は単調増加する関数であるため[1]、どのような$\tilde{\alpha}$、$\tilde{\beta}$に対しても、次式が成立する。

$$\ln\left(P(\tilde{\alpha}、\tilde{\beta})\right) \leq \ln\left(P(\hat{\alpha}_{ML}、\hat{\beta}_{ML})\right)$$

以上から、尤度を最大にする最尤推定量($\hat{\alpha}_{ML}$、$\hat{\beta}_{ML}$)は、対数尤度を最大にする推定量となる。

練習問題 5

説明変数を増やすと当てはまりは改善するため、次の式が成立する。

$$P_0^{max} \leq P^{max} \leq 1$$

上式の対数をとると、次の関係式が得られる(1 の対数は 0 となることに注意)。

$$\ln\left(P_0^{max}\right) \leq \ln\left(P^{max}\right) \leq 0$$

さらに、両辺を$\ln\left(P_0^{max}\right)$で割ると、次の関係式が得らえる($\ln\left(P_0^{max}\right)$は負の値であるため、不等号は逆になることに注意)。

$$1 \geq \frac{\ln(P^{max})}{\ln(P_0^{max})} \geq 0$$

この関係式から、疑似決定係数は 0 以上 1 以下となる。

$$1 - \frac{\ln(P^{max})}{\ln(P_0^{max})}$$

練習問題 6

$Y_i = y_i$となるときの確率密度関数$f\{Y_i = y_i\}$は、次のようになる。

$$\begin{aligned} f\{Y_i = y_i\} &= f\{\alpha + \beta X_i + u_i = y_i\} \\ &= f\{u_i = y_i - \alpha - \beta X_i\} \end{aligned}$$

1 対数は単調増加関数であるため、$P_1 \leq P_2$であれば$\ln(P_1) \leq \ln(P_2)$となり、逆に、$\ln(P_1) \leq \ln(P_2)$であれば$P_1 \leq P_2$となる。

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \alpha - \beta X_i)^2}{2\sigma^2}}$$

$$= (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} e^{-\frac{(y_i - \alpha - \beta X_i)^2}{2\sigma^2}}$$

誤差項u_iは互いに独立であるから、Y_iも互いに独立となる。よって、尤度は

$$L = f\{Y_1 = y_1, \dots, Y_n = y_n\} = f\{Y_1 = y_1\} \times \dots \times f\{Y_n = y_n\}$$

$$= (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} e^{-\frac{(y_1 - \alpha - \beta X_1)^2}{2\sigma^2}} \times \dots \dots \times (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} e^{-\frac{(y_n - \alpha - \beta X_n)^2}{2\sigma^2}}$$

$$= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2}$$

となる。尤度関数は積の形になっており、最大化問題を解くのが困難である。このため、対数尤度関数を用いて、最大化問題を解く。

尤度の両辺について対数をとると、対数尤度が得られる。

$$ln\, f\{Y_1 = y_1, \dots, Y_n = y_n\} = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2$$

ここで、αとβは、右辺第 3 項だけにあるので、(第 3 項に－1 が掛けられていることに注意すると）対数尤度関数の最大化は、残差 2 乗和$\sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2$の最小化と同じである。つまり、残差 2 乗和を最小化する OLS 推定量が、対数尤度関数を最大化する最尤推定量となる。従って最尤推定量は、次式となる。

$$\hat{\alpha}_{ML} = \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$\hat{\beta}_{ML} = \hat{\beta} = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

次に、σ^2の最尤推定量を求めよう。これは対数尤度を最大にするσ^2であるため、対数尤度をσ^2で偏微分してから 0 と置くことで求める。偏微分して 0 とおくと、次の式となる[2]。

[2] 計算は少し難しいため、丁寧に記述しよう。σ^2で偏微分すると、

$$\frac{\partial ln\, f\{Y_1 = y_1, \dots, Y_n = y_n\}}{\partial \sigma^2} = -\frac{n}{2} \left\{ \frac{\partial\, ln(\sigma^2)}{\partial \sigma^2} \right\} - \frac{1}{2} \left\{ \frac{\partial 1/\sigma^2}{\partial \sigma^2} \right\} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2$$

となる。ここで、対数の微分の公式から、

$$\frac{\partial\, ln(\sigma^2)}{\partial \sigma^2} = \frac{1}{\sigma^2}$$

となる(サポートウェブサイトの補足資料「ネイピア数と自然対数の微分」を参照)。また、$1/\sigma^2 = (\sigma^2)^{-1}$であるから、

$$\frac{\partial (1/\sigma^2)}{\partial \sigma^2} = \frac{\partial (\sigma^2)^{-1}}{\partial \sigma^2} = -(\sigma^2)^{-2} = -\frac{1}{\sigma^4}$$

となる。以上から、先の式は次のとおりとなる。

$$\frac{\partial ln\, f\{Y_1 = y_1, \dots, Y_n = y_n\}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2$$

$$\frac{\partial ln\, f\,\{Y_1 = y_1, \ldots, Y_n = y_n\}}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}_{ML}^2} + \frac{1}{2\hat{\sigma}_{ML}^4}\sum_{i=1}^{n}(y_i - \hat{\alpha}_{ML} - \hat{\beta}_{ML}\, X_i)^2 = 0$$

この式の解は最尤推定量のため、下添字に ML を付けた。上式の両辺に$2\hat{\sigma}_{ML}^2$を掛けてから整理すると、最尤推定量が得られる[3]。

$$\hat{\sigma}_{ML}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha}_{ML} - \hat{\beta}_{ML}\, X_i)^2$$

練習問題 7

補足証明で確認したとおり、

$$P\{Y_i = 1\} = P\{u_i < \alpha + \beta X_i\}$$

であるから、$\beta = 0$とすると、

$$P\{Y_i = 1\} = P\{u_i < \alpha\}$$

となる。ここで、$F(\alpha)$は定数であるため、$F(\alpha) = p$と表記しよう。確率の合計は 1 であるため、Y_iが 1 となる確率はpであれば、0 となる確率は$1 - p$となる。まとめると、次のようになる。

$$P\{Y_i = y_i\} = p^{y_i}(1-p)^{1-y_i}$$

このとき、尤度は、

$$\begin{aligned} P\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n\} &= P\{Y_1 = y_1\}P\{Y_2 = y_2\} \ldots P\{Y_n = y_n\} \\ &= p^{y_1}(1-p)^{1-y_1}p^{y_2}(1-p)^{1-y_2} \ldots p^{y_n}(1-p)^{1-y_n} \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i} \end{aligned}$$

となる。これをpに関して微分して 0 と置くことで、

$$\sum y_i\, \hat{p}_{ML}^{\sum y_i - 1}(1 - \hat{p}_{ML})^{n-\sum y_i} - (n - \sum y_i)\, \hat{p}_{ML}^{\sum y_i}(1 - \hat{p}_{ML})^{n-\sum y_i - 1} = 0$$

となる。さらに両辺を$\hat{p}_{ML}^{\sum y_i - 1}(1 - \hat{p}_{ML})^{n-\sum y_i - 1}$で割ると、

$$\sum y_i\,(1 - \hat{p}_{ML}) - (n - \sum y_i)\hat{p}_{ML} = 0$$

となる。これを整理すると、

$$\sum y_i - n\hat{p}_{ML} = 0$$

となるから、$\hat{p}_{ML}$について解くと、

[3] なお、最尤推定量$\hat{\sigma}_{ML}^2$は、残差 2 乗和をnで割った値である。3 章補足で確認したとおり、σ^2の不偏推定量は、残差 2 乗和を$n-2$で割った値であるため、最尤推定量$\hat{\sigma}_{ML}^2$は不偏推定量ではない。サンプルサイズが大きいと、nで割っても、$n-2$で割ってもほぼ同じであるため、最尤推定量$\hat{\sigma}_{ML}^2$は一致推定量となる。まとめると、最尤推定量は一致推定量だが、不偏推定量ではないといえる。

$$\hat{p}_{ML} = \frac{\sum y_i}{n} = \bar{y}$$

となる。したがって、尤度関数に$\hat{p}_{ML} = \bar{y}$を代入することで、最大尤度P_0^{max}が次のようなる。

$$P_0^{max} = P\{Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n\} = \bar{y}^{\sum y_i}(1 - \bar{y})^{n - \sum y_i}$$

練習問題 8

(a) 投票率P_iは質的データではなく、連続的な値をとる量的データである。しかし、投票率は割合を表しているため、0 以上 1 以下となる必要がある。線形モデルを考えると、投票率P_iの予測値は 0 を下回る、また、1 を上回る可能性がある。これが線形モデルの問題である。

(b) 投票率P_iは 0 から 1 までの値をとるため、$P_i/(1 - P_i)$は 0 から＋∞までの値をとる。さらに対数をとると、$\ln(P_i/(1 - P_i))$は－∞から＋∞の値をとる。したがって、ロジット変換した推定式、つまり、

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \alpha + \beta X_i + u_i$$

では、投票率P_iが 0 以上 1 以下という制約を考慮したうえで推定がされている。この点を確認してみよう。上式の対数をとると、

$$\frac{P_i}{1 - P_i} = e^{\alpha + \beta X_i + u_i}$$

であり、これをP_iについて解くと、

$$P_i = \frac{1}{1 + e^{\alpha + \beta X_i + u_i}}$$

となる。ここで、$e^{\alpha + \beta X_i + u_i}$は 0 以上である。$e^{\alpha + \beta X_i + u_i}$が 0 なら$P_i = 1$となり、$e^{\alpha + \beta X_i + u_i}$が大きくなると、$P_i$は 0 に近づく。

(c) $P_i = 0$もしくは 1 になると、$\ln(P_i/(1 - P_i))$は定義できない。ただし、実際の投票率が 0%や 100%という状況はありえないため、問題とはならない。

練習問題 9

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
use marriage_data.dta

**12.1.1 節の推定結果
reg partner age if(male==1 & age<=50)   /*50 歳以下の男性に限定した分析*/
display _b[_cons]+_b[age]*25
probit partner age if(male==1 & age<=50)   /*50 歳以下の女性に限定した分析*/
display normal(_b[_cons]+_b[age]*25)

*12.1.3 節の推定結果
reg partner age undergrad grad if(male==1 & age<=50), r
display e(r2_a)
probit partner age undergrad grad if(male==1 & age<=50), r
margins, dydx(age undergrad grad)
reg partner age undergrad grad if(male==0 & age<=50),r
display e(r2_a)
probit partner age undergrad grad if(male==0 & age<=50), r
margins, dydx(age undergrad grad)

*練習問題の推定結果(50 歳以下の男性)
probit partner age undergrad grad if(male==1 & age<=50), r
gen p_grad = normal(_b[_cons]+_b[age]*age+_b[undergrad]*0+_b[grad]*1)
gen p_no = normal(_b[_cons]+_b[age]*age+_b[undergrad]*0+_b[grad]*0)
scatter   p_grad   p_no   age
```

① 12.1.1 節において、プロビット推定は、

probit partner age if(male==1 & age<=50)

とすればよい。ここでは、男性で 50 歳以下のデータだけを用いている。また、

25 歳の予測値は、次のようにすればよい。ここで、noromal は標準正規分布の累積密度関数である。

```
display normal(_b[_cons]+_b[age]*25)
```

② 12.1.3 節にといて、プロビット推定は、

```
probit partner age undergrad grad if(male==1 & age<=50), r
margins, dydx(age undergrad grad)
```

とすればよい。ここで、margin, dydx は平均限界効果を計算するためのコマンドである。

練習問題 10

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。