

情報工学実験 2 数理計画法 第二回

学生番号 4617043 神保光洋

2018 年 12 月 7 日

1 要旨

最尤推定を行う場合の応用のされ方に付いて学習・演習する。

2 目的

統計モデルを用いた分析に欠かせない、パラメタの推定の方法について、その基本的な技術を取得することを目的とする。

3 理論

3.1 ベイズの定理

ベイズの定理は以下の式で与えられる。

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

ここで、 $P(A|B)$ は B が起ったときに A が起こる確率を表す条件付き確率である。ここで、 B が起こった時に A が起こる確率 $P(A|B)$ は B が起こっている事象の割合である。つまり以下のように表すことができる。

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

また、これらの関係は A, B を入れ替えても同様に成立するため、以下を得る

$$P(B|A) = \frac{P(A, B)}{P(A)}$$

これらの 2 つより、ベイズの定理が導出される。

3.2 最尤推定

このベイズの定理を使用することで、ある確率変数の値から別の確率変数の値を推定することが可能である。

ある大学の男女別学部別の在籍者数が表 1 のようであったとする。

この時、「この大学に所属する女子学生の学部」を推定することを考える。この時、「女子学生のそれぞれの学部」に所属している確率」が最も高い学部」に所属していると考えることが自然である。

表 1 ある大学の学部別、男女別在籍者数

学部	男子学生数	女子学生数
理学部第一部	2230	611
理学部第二部	1271	390
薬学部	492	547
工学部	1771	417
工学部第二部	701	148
理工学部	4289	870
基礎工学部	1028	372
経営学部	950	441
学部合計	12732	3796

つまり、以下のような数式を考えることが自然である。

$$(\hat{\text{学部}}) = \arg \max_{\text{学部}} P(\text{学部} | \text{女子学生})$$

この考え方は、じご分布最大化 (Maximum a Posteriori:MAP) 推定と呼ばれる。それぞれの学部について計算を行うと以下ようになる。

$$\begin{aligned} P(\text{理学部第一} | \text{女子学生}) &= \frac{611}{3796}, \\ P(\text{理学部第二} | \text{女子学生}) &= \frac{390}{3796}, \\ &\vdots \\ P(\text{経営学部} | \text{女子学生}) &= \frac{441}{3796} \end{aligned}$$

また、式 (1) はベイズの定理を用いると以下のようにもかける。

$$\begin{aligned} (\hat{\text{学部}}) &= \arg \max_{\text{学部}} P(\text{学部} | \text{女子学生}) \\ (\hat{\text{学部}}) &= \arg \max_{\text{学部}} \frac{P(\text{学部})}{P(\text{女子学生})} P(\text{女子学生} | \text{学部}) \end{aligned}$$

これらをそれぞれの学部について計算すると以下ようになる。

$$\frac{P(\text{理学部第一})}{P(\text{女子学生})} P(\text{女子学生} | \text{理学部第一}) = \frac{\frac{2230+611}{12732+3796}}{\frac{3796}{12732+3796}} \times \frac{611}{2230 + 611}$$

$$\begin{aligned}\frac{P(\text{理学部第二})}{P(\text{女子学生})}P(\text{女子学生} \mid \text{理学部第二}) &= \frac{\frac{1271+390}{12732+3796}}{\frac{3796}{12732+3796}} \times \frac{390}{1271+390} \\ &\vdots \\ \frac{P(\text{経営学部})}{P(\text{女子学生})}P(\text{女子学生} \mid \text{経営学部}) &= \frac{\frac{950+441}{12732+3796}}{\frac{3796}{12732+3796}} \times \frac{441}{950+441}\end{aligned}$$

これらの式は計算の結果が全く変わらないことがわかる。

ここで上式をよく見れば、 $P(\text{女子学生})$ は学部に関係のない定数となっているため、以下のように考えても推定される学部は変化しない。

$$\frac{P(\text{学部})}{P(\text{女子学生})}P(\text{女子学生} \mid \text{学部}) \propto P(\text{学部})P(\text{女子学生} \mid \text{学部})$$

であり、よって、

$$\begin{aligned}\hat{(\text{学部})} &= \arg \max_{\text{学部}} P(\text{学部} \mid \text{女子学生}) \\ &= \arg \max_{\text{学部}} P(\text{学部})P(\text{女子学生} \mid \text{学部})\end{aligned}$$

このように、確率ではないが、確率に比例するスコア：尤度を用いて推定を行うこともできる。この尤度が最大のものを推定値として採用する推定法が最尤推定である。

4 課題 1

表 2 ある大学の学部別、男女別在籍者数の推定結果

学部	男子学生数	女子学生数	頻度による推定 $P(\text{学部} \mid \text{女子})$	ベイズの定理に変換 $\frac{P(\text{学生})}{P(\text{女子学生})}P(\text{女子学生} \mid \text{学部})$	最優推定 $P(\text{学部})P(\text{女子学生} \mid \text{学部})$
理学部第一部	2230	611	0.160	0.160	0.036
理学部第二部	1271	390	0.102	0.102	0.023
薬学部	492	547	0.14	0.144	0.033
工学部	1771	417	0.109	0.109	0.025
工学部第二部	701	148	0.038	0.038	0.008
理工学部	4289	870	0.229	0.229	0.052
基礎工学部	1028	372	0.097	0.097	0.022
経営学部	950	441	0.116	0.116	0.026
学部合計	12732	3796	1	1	0.229

4.1 課題 1-1

この大学にある女子学生がいた場合、その学部を推定せよ。またなぜそう考えるのか思考の過程も記述せよ。

表 2 とより頻度による推定値が最も高いのは理学部第一部であるが、上記した理論より尤度が最も高いのは理学部であることより、大学にある女子学生がいた場合その学部は理学部であると推定できる。

4.2 課題 1-2

この大学にある経営学部学生がいた場合、その性別を推定せよ。なたなぜそう考えられるのか思考の課程も記述せよ。

表より

$$\begin{aligned}P(\text{男子学生} \mid \text{経営学部生}) &= \frac{\frac{950}{12732+3796}}{\frac{12732}{12732+3796}} \\&= 0.0746 \\P(\text{女子学生} \mid \text{経営学部生}) &= \frac{\frac{441}{12732+3796}}{\frac{3796}{12732+3796}} \\&= 0.116\end{aligned}$$

であるので女子学生であると推定される。

5 まとめ

最尤推定を行う場合の応用のされ方に付いて学ぶことができた。

6 参考文献

参考文献

[1] 東京大学教養学部統計学教室 統計学入門 (基礎統計学)・東大出版会

7 付録