レポート提出票

科目名:	情報工学実験2
実験テーマ:	実験テーマ5 教育システム設計
実施日:	2020年 11月 2日
学籍番号:	4619055
氏名:	辰川力駆
共同実験者:	

1　要旨

ベイズの定理は既に習っているが、最尤推定を行う場合の応用のされ方について、演習を通して学習する。今回は、ある大学の男女別や学部別の在籍者数から推定する実験を行う。

2　目的

統計モデルを用いた分析は、例えば商品の推薦や迷惑メールの削除機能など、身近な機能を支える基本的な技術となっている。本実験では、このような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、基本的な技術を習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。

$$P(A|B) = \frac{P(A)}{P(B)}P(B|A) \tag{1}$$

ここで、$P(A)$ と $P(B)$ はそれぞれ事象 A と B が起こる確率、$P(A|B)$ は事象 B が起こったときに事象 A が起こる確率を表す条件付き確率である。

B が起こったときに A が起こる確率 $P(A|B)$ は B が起こった事象中で A も同時に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。

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \tag{2}$$

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

$$P(B|A) = \frac{P(A \wedge B)}{P(A)} \tag{3}$$

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。例えば、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。

表 1: ある大学の学部別、男女別在籍者数

学部	学生数	
	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

ここで「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ると考えることが自然である。つまり、以下のような数式を考えることが自然である。

$$\hat{(\text{学部})} = \arg\max_{\text{学部}} P(\text{学部} \mid \text{女子学生}) \tag{4}$$

(この時 $\arg\max_x$ は x を変数とみなし、後に続く式が最も大きくなる x の値を返すことを意味する。) この考え方は、事後分布最大化 (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のようになる。

$$\begin{aligned}
P(\text{理学部第一部} \mid \text{女子学生}) &= \frac{611}{3796}, \\
P(\text{理学部第二部} \mid \text{女子学生}) &= \frac{390}{3796}, \\
&\vdots \quad \vdots \\
P(\text{経営学部} \mid \text{女子学生}) &= \frac{441}{3796}
\end{aligned}$$

また、式 4 はベイズの定理を用いると以下のようにもかける。

$$\begin{aligned}
\hat{(\text{学部})} &= \arg\max_{\text{学部}} P(\text{学部} \mid \text{女子学生}) \\
&= \arg\max_{\text{学部}} \frac{P(\text{学部})}{P(\text{女子学生})} P(\text{女子学生} \mid \text{学部})
\end{aligned} \tag{5}$$

これらをそれぞれの学部について計算すると以下のようになる。

$$
\begin{aligned}
\frac{P(\text{理学部第一部})}{P(\text{女子学生})}P(\text{女子学生}\,|\,\text{理学部第一部}) &= \frac{\frac{2230+611}{12732+3796}}{\frac{3796}{12732+3796}} \times \frac{611}{2230+611}, \\
\frac{P(\text{理学部第二部})}{P(\text{女子学生})}P(\text{女子学生}\,|\,\text{理学部第二部}) &= \frac{\frac{1271+390}{12732+3796}}{\frac{3796}{12732+3796}} \times \frac{390}{1271+390}, \\
&\vdots \quad \vdots \\
\frac{P(\text{経営学部})}{P(\text{女子学生})}P(\text{女子学生}\,|\,\text{経営学部}) &= \frac{\frac{950+441}{12732+3796}}{\frac{3796}{12732+3796}} \times \frac{441}{950+441}
\end{aligned}
$$

こち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、式 5 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。

$$
\frac{P(\text{学部})}{P(\text{女子学生})}P(\text{女子学生}\,|\,\text{学部}) \propto P(\text{学部})P(\text{女子学生}\,|\,\text{学部}) \tag{6}
$$

したがって、式 4 は以下のようになる。

$$
\begin{aligned}
(\hat{\text{学部}}) &= \mathop{\arg\max}_{\text{学部}} P(\text{学部}\,|\,\text{女子学生}) \\
&= \mathop{\arg\max}_{\text{学部}} P(\text{学部})P(\text{女子学生}\,|\,\text{学部}) \qquad (7)
\end{aligned}
$$

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 (Maximum likelihood estimation:MLE) である。

4 課題

4.1 課題 1-1

この大学にある女子学生がいた場合、その学部を推定する。

表 2: 各学部ごとの推定

学部	頻度による推定	尤度推定
	P(学部 | 女子学生)	P(学部)P(女子学生 | 学部)
理学部第一部	0.161	0.037
理学部第二部	0.103	0.024
薬学部	0.144	0.033
工学部	0.110	0.025
工学部第二部	0.039	0.009
理工学部	0.229	0.053
基礎工学部	0.098	0.023
経営学部	0.116	0.027

ある女子学生の学部を推定するために、まずは女子学生のそれぞれの学部に所属している確率を求めた。結果は表 2 にまとめた。縦の合計が 1 になっていないのは、小数第 4 位を四捨五入しているからである。また、尤度推定も行った。計算式は表 2 の通りである。

したがって表 2 より、確率が最も大きい学部は理工学部である。よって、この大学にある女子学生がいた場合理工学部である可能性が高いと言える。

今回は、尤度推定と頻度による推定の両方が行えたが、これらの違いは使っている情報が違うことである。尤度推定は、例えば P(学部) と P(女子学生 | 学部) しか分かっていない時に有効である。

4.2 課題 1-2

この大学にある経営学部学生がいた場合、その性別を推定する。

4.3 課題 1-3

公開された問題を 4 題以上ランダムに選び解答する。

問 3、1 である。

問 30、T を中心に対称

問

4.4 課題 1-4

4.5 課題 1-5

5 まとめ

参考文献