

Problem Set 2. ベイズの法則と最尤法

PUBLISHED

June 4, 2024

提出期限

- 2024年6月19日（水）17:00

提出形式

- LMSにアップロードされている"problemset2_base"を使用して、(1)レポートと(2)コードをフォルダー構造にしたがって格納し、提出すること
- 提出の際には、**フォルダー名を"problemset2_学生番号"に変更して提出する**
 - 例: 学籍番号が"9999999"の学生の場合は以下のフォルダーを提出する
 - "problemset2_9999999"
- レポートはpdf形式で作成し、03_reportに"report2_学籍番号.pdf"という名前で保存すること
- 作成したグラフはレポートに加えること。(コードを実行しても図は自動的に保存されない)

フォルダ構造

- problemset2_base
 - 01_data
 - population.csv
 - population.xlsx
 - 02_code
 - pareto.r
 - 03_report

課題

本課題は、ベイズの定理と最尤法についての理解を深めることを目的としている。課題2を解くために必要なコードは、全て配布したフォルダ内に含まれている。しかし、いくつかの関数内のコードがスクランブルされているため、そのままでは実行できない。この点に留意して課題に取り組むこと。また、問題1全てと問題(2.1)は記述式回答であるため、コードを実行する必要はない。

1. 囚人と主観的確率

(a) 3つの扉

ある1人の囚人の前にはA,B,Cの3つの扉があり、1つだけが正解の扉で自由を手に入れられます。他の2つの扉は不正解で後ろには罰が待っています。囚人が1つの扉を選択した後、看守が残りの2つの扉のうち、不正解の扉を1つ開けます。ここで囚人は最初に選んだ扉のままでよいし、残っている扉に変更してもよいと告げられます。この時囚人は喜んで扉を選びなおし、選択しました。

問い) ベイズの定理を用いて、この囚人の選択は正しいか示しなさい。

(b) 3人の囚人

囚人A,B,Cがいる。3人とも処刑されることになっていたが、王は1人だけ恩赦(釈放)することにした。誰を釈放するかは決まっているが、まだ囚人たちには知らせていない。囚人Aが、結果を知っている看守に対して、「BとCのうち、どちらかは必ず処刑されるのだから、処刑される1人の名前を教えてください」と頼み、看守は「囚人Bが処刑される」と教えたところ、囚人Aは笑みを浮かべた。理由を聞くとAは「最初、私が助かる確率は $\frac{1}{3}$ だったが、いまや助かるのは自分とCだけなので、助かる確率は $\frac{1}{2}$ になった」と語っている。

問い) 看守の返事を聞いた囚人Aは本当に喜んでいいのか、ベイズの定理を用いて示しなさい。

(c) 比較

(a)と(b)を比較して、気づいたことをまとめなさい。

2. パレート分布と最尤法

この課題では、パレート分布の性質を最尤法を用いて理解することを目的としている。都市の人口規模はパレート分布に従っていることが知られている(Rosen and Resnick, 1980; Levy,2009)。以下の問題では、日本の人口上位200市区町村データ (`01_data/population.csv`) を用いて、パレート分布と都市の人口規模について考える。

以下は、パレート分布の(累積)分布関数である。

$$F(x) = \begin{cases} 1 - (\frac{\beta}{x})^\alpha & x \geq \beta, \\ 0 & x < \beta \end{cases}$$

以下の問題に答えなさい。

(a) パレート分布の尤度関数

パレート分布の尤度関数(対数尤度関数)を求め、解析的に α を求めなさい (記述式)

- ヒント：最尤推定の流れ
 - 分布関数を微分して、密度関数を求める。
 - 密度関数を掛け合わせ、尤度関数を求める。
 - 対数尤度関数に変換し、尤度が最大となる α を探す。
- ヒント： $\ln \beta$ は β の単調増加関数であり、 $\beta \leq x$ が全ての x について成り立っている時の、 β の最大値を考える。

(b) 解析解を用いた推定

(a)の結果と `population.csv` に含まれる人口データを用いて、 α と β を計算しなさい。

(c) グリッド探索を用いた推定

次に、数値シミュレーションを用いて α を求めることを考えます。(b)で求めた α の時に、尤度が最大となることを、「グリッド探索」を用いて検証します。`simulate_mle`関数に含まれる、`l = eval(parse(text =`

`paste0(a, b, c, d, e)))` のa~eの順序を正しく並べ替え、尤度関数を作成しなさい。また、コードはスクランブルされているので、順序を正しく並べ替えなさい。なお、 β は(a)で推定した値を用いなさい。

- ヒント：講義動画「最尤法と数値的最適化」"4. 最適化アルゴリズム"を参照

(d) 勾配法を用いた推定

次に、rの最適化関数 (`optim()` 関数) を用いて、 α を求めます。(c)と同様にa~eを並べかえ、正しい尤度関数を作成してください。また、複数の初期値(`alpha_0`)を試して、収束するまでの計算時間を比較しなさい。

- ヒント：`optim()`関数は、最小値を求めるように設定されています。
- 参考サイト：<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/optim>

(e) パレート分布の仮定の検証(Adv.)

ここまでは、人口がパレート分布に従っていることを所与として考えましたが、次の問題ではこの仮定が正しいのかを検証します。検証方法は自身で考え、コードを作成してください。

`generate_pareto_simulation()` は一例です。推定した α と β のパラメータ値を持つパレート分布に従う乱数を、逆関数法を用いて200個生成します。その後、シミュレーションで生成した値と、実際の値の累積分布関数を比較しています。(ここでは、各種検定量を用いた統計的有意性は考えていません。)

`generate_pareto_simulation()` の中身もスクランブルしているため、この関数内のコードを正しく並び替え、検証しても構いません。

取り組むにあたって気をつけてほしいこと

- コードを調べるときは、英語のリソース(Stack Overflowなど)を探すことを強く勧める
- プログラミングの課題と一緒に取り組む仲間も、大切なリソースである。まず自力で取り組み、その後、お互いに助言をし合うことを勧める

References

Rosen, K., & Resnick, M. (1980). The size distribution of cities: An examination of the Pareto Law and primacy. *Journal of Urban Economics*, 8, 165–186.

Levy, M. (2009). Gibrats's Law for (all) cities: Comment. *American Economic Review*, 99, 1672–1675.

参考資料

R

[私たちのR](#)

[R for Data Science \(2e\)](#)

cheat sheet

[Posit Cheatsheets](#)

レポート関連

[Quarto](#)^[^2]

[Overleaf](#)