

摂動を用いた歪ロジスティック分布の生成とその推定

2017SS038 黒澤周作

指導教員：阿部俊弘

1 はじめに

正規分布やロジスティック分布は数学的には性質の良い対称分布であるが、実際の現象は対称であるとは限らず、非対称であるほうが自然である。本研究では、Azzalini による非対称化の手法を用いて、ロジスティック分布を非対称化する手法を与え、その数学的性質を考える。最後に 3 つの実データに対してパラメータ推定の例を与える。

2 ロジスティック分布について

ロジスティック分布は対称な連続型確率分布の一つで、累積分布関数がロジスティック関数であるものである。正規分布同様に対称なシグモイド型の分布関数であり、釣鐘型の確率密度関数であるが、ロジスティック分布の方が裾が重く、密度関数は平均から離れても下がりにくい。また、このような性質から、外れ値のあるようなデータにも適用される。

ロジスティック分布の分布関数 $F_L(x)$ は

$$F_L(x) = \frac{1}{1 + e^{-x}}, \quad -\infty < x < \infty$$

である。このとき確率密度関数 $f_L(x)$ は

$$f_L(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \quad (1)$$

である。このような分布を $X \sim SL(0, 1)$ とする。期待値は 0、分散は $\pi^2/3$ である。

3 摂動を用いたロジスティック分布の非対称化と性質

3.1 Azzalini の歪正規分布

正規分布の非対称化として、Azzalini により提案された歪正規分布がある (Azzalini, 1985)。これは 3 パラメータで表現され、そのうちの 1 つのパラメータを 0 とするとき正規分布となることから、分布を平均と分散の 2 パラメータで表現する正規分布の拡張であるといえる。このとき、標準正規分布に対応する歪正規分布 $SN(0, 1, \lambda)$ の確率密度関数は

$$f(x) = 2\phi(x)\Phi(\lambda x)$$

で与えられる。ここで、 $-\infty < \lambda < \infty$ であり、 ϕ と Φ は標準正規分布の密度関数と分布関数である：

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad \Phi(x) = \int_{-\infty}^x \phi(t) dt.$$

この分布の特徴として、

- 特別な場合に正規分布を含む

- 数学的に扱いやすい
- 幅広い歪度と尖度を持っている

があげられる。本研究では、この考え方をロジスティック分布に適用した歪対称モデルについて統計的性質を考える。

4 ロジスティック分布の非対称化

[1] の Lemma 1 の公式を用いて、ロジスティック分布 (1) の非対称をする：

$$\begin{aligned} f_S(x) &= 2F_L(\lambda x)f_L(x) \\ &= \frac{2}{1 + \exp(-\lambda x)} \frac{\exp(-x)}{[1 + \exp(-x)]^2}. \end{aligned}$$

ここで、位置パラメータ μ と尺度パラメータ σ は通常の手法 $x \mapsto \frac{x - \mu}{\sigma}$ により導入することができるが、本研究では、 $\frac{\lambda}{\sigma} \mapsto \lambda$ とリパラメータ化する。このとき、 f_{SL} を

$$\begin{aligned} f_{SL}(x) &= 2F_L(\lambda(x - \mu))f_L\left(\frac{x - \mu}{\sigma}\right) \\ &= \frac{2}{1 + \exp[-\lambda(x - \mu)]} \frac{1}{\sigma} \frac{\exp\left(-\frac{x - \mu}{\sigma}\right)}{\left[1 + \exp\left(-\frac{x - \mu}{\sigma}\right)\right]^2} \end{aligned}$$

と置くと $2/(1 + \exp(\lambda x)) - 1$ が奇関数なので、 f_{SL} は密度関数になっていることを確認できる。本研究ではこの分布を $SL(\mu, \sigma, \lambda)$ と表記する。

5 歪ロジスティック分布の最尤推定

5.1 歪ロジスティック分布の尤度関数

x_1, \dots, x_n がパラメータ (μ, σ, λ) の歪ロジスティック分布 $SL(\mu, \sigma, \lambda)$ に従うとき、対数尤度関数は

$$\begin{aligned} \ell(\xi) &= n \log 2 - \sum_{i=1}^n \log[1 + \exp[-\lambda(x_i - \mu)]] - n \log \sigma \\ &\quad - \sum_{i=1}^n \frac{x_i - \mu}{\sigma} - 2 \sum_{i=1}^n \log \left[1 + \exp\left(-\frac{x_i - \mu}{\sigma}\right) \right] \end{aligned}$$

により与えられる。ここで、 $\xi = (\mu, \sigma, \lambda)^\top$ とし、上記の対数尤度関数を各パラメータに対して微分すると、スコア関数を得られる。

5.2 SL 分布のフィッシャー情報行列

本節では SL 分布のフィッシャー情報を与える。まず、フィッシャー情報の対角要素は

$$\begin{aligned}\frac{\partial^2 \ell(\xi)}{\partial \mu^2} &= - \sum_{i=1}^n \frac{\lambda \exp[-\lambda(x_i - \mu)]}{(1 + \exp[-\lambda(x_i - \mu)])^2}, \\ \frac{\partial^2 \ell(\xi)}{\partial \sigma^2} &= \frac{n}{\sigma^2} + \frac{4}{\sigma^3} \sum_{i=1}^n \frac{1}{\exp\left(\frac{x_i - \mu}{\sigma}\right) + 1} \\ &\quad - \frac{2}{\sigma^3} \sum_{i=1}^n (x_i - \mu) + \frac{2}{\sigma^2} \sum_{i=1}^n \frac{(x_i - \mu) \exp\left(\frac{x_i - \mu}{\sigma}\right)}{\left[1 + \exp\left(\frac{x_i - \mu}{\sigma}\right)\right]^2}, \\ \frac{\partial^2 \ell(\xi)}{\partial \lambda^2} &= \sum_{i=1}^n \frac{(x_i - \mu)^2 \exp[-\lambda(x_i - \mu)]}{(1 + \exp[-\lambda(x_i - \mu)])^2}\end{aligned}$$

となり、非対角要素は

$$\begin{aligned}\frac{\partial^2 \ell(\xi)}{\partial \mu \partial \sigma} &= - \frac{n}{\sigma^2} + \frac{2}{\sigma^2} \sum_{i=1}^n \frac{1}{\exp\left(\frac{x_i - \mu}{\sigma}\right) + 1} \\ &\quad + \frac{2}{\sigma^3} \sum_{i=1}^n \frac{(x_i - \mu) \exp\left(\frac{x_i - \mu}{\sigma}\right)}{\left[1 + \exp\left(\frac{x_i - \mu}{\sigma}\right)\right]^2}, \\ \frac{\partial^2 \ell(\xi)}{\partial \mu \partial \lambda} &= \sum_{i=1}^n \frac{\exp[-\lambda(x_i - \mu)]}{1 + \exp[-\lambda(x_i - \mu)]} \\ &\quad - \sum_{i=1}^n \frac{(\lambda(x_i - \mu)) \exp[-\lambda(x_i - \mu)]}{(1 + \exp[-\lambda(x_i - \mu)])^2}, \\ \frac{\partial^2 \ell(\xi)}{\partial \sigma \partial \lambda} &= 0\end{aligned}$$

となる。上記のフィッシャー情報において、 $\frac{\partial \ell(\xi)}{\partial \sigma \partial \lambda} = 0$ なので、 σ と λ のパラメータは直交していることが分かる。

また、 $\frac{\partial^2 \ell(\xi)}{\partial \lambda^2}$ において、Azzalini の歪正規分布については、歪パラメータの 2 階微分は $\lambda = 0$ で 0 となるが、この SL 分布ではこのような問題は起きない。

5.3 Newton-Raphson 法

スコアを計算する。スコア関数は

$$S(\mathbf{x}, \xi) = \frac{\partial \ell(\xi)}{\partial \xi} = \left(\frac{\partial \ell(\xi)}{\partial \mu}, \frac{\partial \ell(\xi)}{\partial \sigma}, \frac{\partial \ell(\xi)}{\partial \lambda} \right)^\top$$

となる。ここで、 $\mathbf{x} = (x_1, \dots, x_n)^\top$ である。これより、観測 Fisher 情報行列 $J(\xi)$ を用いて Newton-Raphson 法を適用する。kth step のパラメータを $\xi^{(k)}$ とすると、(k + 1)th step のパラメータは

$$\xi^{(k+1)} = \xi^{(k)} + J(\xi^{(k)})^{-1} S(\mathbf{x}, \xi^{(k)})$$

で更新される。

6 実データへの適用例

実データとして、(1)Fiberdata[3]、(2) 南山大学内の放射線量データ、(3) アスリートの肩甲骨データ (SSF)[2] に対してパラメータ推定を行った。

DATA	MLE			MLL	AIC	BIC
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$			
(1)	2.33	0.55	6.76	-56.79	119.59	126.02
	3.02	0.35	0	-59.33	122.66	126.94
(2)	0.12	0.02	25.05	296.60	-587.20	-578.08
	0.12	0.02	0	295.46	-586.92	-580.84
(3)	29.52	28.62	1.51	-994.82	1995.64	1903.50
	65.17	17.98	0	-986.85	1977.70	1984.32

その結果、Fiberdata とアスリートの肩甲骨データ (SSF) に対しては歪ロジスティック分布が良いモデルとなり、南山大学内の放射線量データに対しては、ロジスティック分布の方が良いモデルとなった。

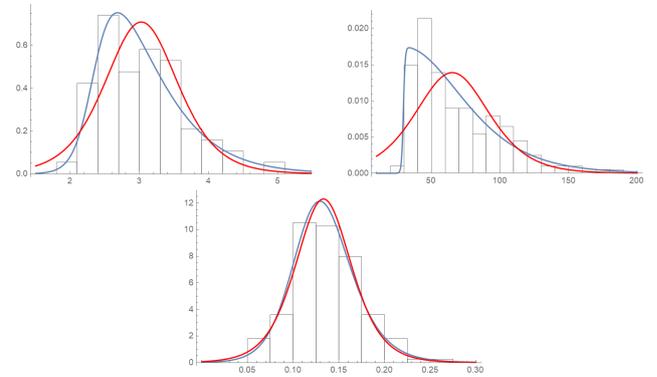


図1 各データのヒストグラムと当てはめた密度関数のグラフ。左上から順に Fiber data, 南山大学内の放射線量データ, アスリートの肩甲骨データ (SSF) である。

7 終わりに

Azzalini の手法を用いた歪ロジスティック分布の提案を行い、その分布の数学的性質を調べた。また、この分布を用いたパラメータ推定の例として 3 つの実データを用いた。実データのうち、放射線データについては学内でガイガーカウンターを使って計測した。データ解析の結果、非対称性がないことを尤度比検定で確認し、直感と合った結果が得られた。これによって、学内の放射線の分布の平均と分散がわかった。

参考文献

- [1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.
- [2] Cook, R. D. & Weisberg, S. (1994). *An Introduction to Regression Graphics*. New York: Wiley.
- [3] Gupta, R. & Kundu, D. (2010). Generalized Logistic Distributions. *Journal of Applied Statistical Science*, 1–23.