

ベイズ法による von-Mises 分布のパラメータ推定

Innovative Bayesian Estimation in the von-Mises Distribution

鎌倉研究室

16N7100009H 神谷勇太

1 はじめに

角度データは多岐にわたる科学分野で多く用いられている。Mardia (1972a) でも風の吹く方向を角度データと捉え、風向きのモデリングをしている。また、Aki-moto, *et al.* (2016) では活断層の向きに着目し、解析を行うことで災害予測に貢献しようと試みている。近年では、角度データを基にした時系列モデルが考案され、時系列解析の分野でも着目を浴びている。このように角度データの解析は様々な分野で行われ、その解析に対する関心も増えてきている。解析の際に最も多く用いられるのは von-Mises 分布である。これは角度の正規分布とも呼ばれる代表的な分布の一つである。von-Mises 分布を用いた角度データの解析に関心が増える一方で、それに対するパラメータ推定の研究は非常に少ない。代表的な von-Mises 分布の推定量として Maximum Likelihood Estimator や Maximum Marginal Likelihood Estimator (Schou, 1978) 等が挙げられる。そこで本研究では、ベイズ法を用いた新しい推定量の提案を行う。Yanagimoto and Onishi (2009) によると、指数型分布族において自然母数の事後平均はベイズ的最適性が保証されるとある。本論文では、指数型分布族の一つである von-Mises 分布の自然母数の事後平均を考える。この事後平均から各パラメータの推定量を導出し、それを本研究の提案推定量とする。提案推定量と3つの既存推定量とリスクの比較をシミュレーションにて行い、提案推定量の優位性を示す。また、提案推定量と既存推定量それぞれを実際のデータに適用し、推定量の選択によって解析結果に差異が生じることを確認する。

2 新しい推定量の提案

ベイズ法を基にした新しい推定量の提案を本研究の目的とする。個々のパラメータの事後平均 $\hat{\eta}, \hat{\tau}$ を推定するのではなく、自然母数の事後平均 θ_1, θ_2 を推定する。 θ_1, θ_2 は (Yanagimoto and Onishi, 2009) よりベイズ的最適性をもつことが確かめられている。この2つの事後平均より、個々のパラメータ $\hat{\eta}$ 及び $\hat{\tau}$ を導出し、それを提案推定量とする。

2.1 事前分布

本研究では無情報事前分布である reference prior (Garvan and Ghosh, 1997) を用いる。

$$\pi(\eta, \tau) \propto \left\{ 1 - \frac{A(\tau)}{\tau} - A^2(\tau) \right\}^{\frac{1}{2}} \quad (1)$$

ここで、 $A(x) = I_1(x)/I_0(x)$ である。

2.2 標本密度

von-Mises 分布の標本密度は以下のように表せる。

$$\begin{aligned} p(\mathbf{x}|\eta, \tau) &= \prod_{i=1}^n \frac{1}{2\pi I_0(\tau)} \exp\{\tau \cos(x_i - \eta)\} \\ &= \frac{1}{2\pi I_0(\tau R)} \exp\{\tau R \cos(\eta - \eta^*)\} \\ &\quad \cdot \frac{I_0(\tau R)}{2\pi^{n-1} I_0(\tau)^n} \end{aligned}$$

ここで、

$$\begin{aligned} \sum \cos x_i &= C, \quad \sum \sin x_i = S \\ R(\mathbf{x}) &= \sqrt{C^2 + S^2} \end{aligned} \quad (2)$$

$$\eta^* = \tan^{-1}(S/C) \quad (3)$$

である。

2.3 事後分布

上記の事前密度と標本密度を用い、事後平均を以下のように表す。ここで、 $\pi(\tau)$ は事前分布である。本来、事前分布は $\pi(\eta, \tau)$ で表記されるが、式 (1) から分かるように、これは η に依存しない。よって、以降は事前分布を $\pi(\tau)$ と表記する。

$$\begin{aligned} p(\eta, \tau|\mathbf{x}) &= \frac{p(\mathbf{x}|\eta, \tau) \cdot \pi(\tau)}{\int \int p(\mathbf{x}|\eta, \tau) \cdot \pi(\tau) d\eta d\tau} \\ &= \frac{1}{2\pi I_0(\tau R)} \exp\{\tau R \cos(\eta - \eta^*)\} \\ &\quad \cdot \frac{I_0(\tau R(\mathbf{x}))\pi(\tau)/I_0^n(\tau)}{\int I_0(\tau R(\mathbf{x}))\pi(\tau)/I_0^n(\tau) d\tau} \end{aligned}$$

2.4 事後平均

それぞれの自然母数 (θ_1, θ_2) の事後平均は

$$E(\tau \cos(\eta); p(\eta, \tau | \mathbf{x})) \\ = \cos(\eta^*) \int \tau A(\tau R(\mathbf{x})) \pi_m(\tau | \mathbf{x}) d\tau$$

$$E(\tau \sin(\eta); p(\eta, \tau | \mathbf{x})) \\ = \sin(\eta^*) \int \tau A(\tau R(\mathbf{x})) \pi_m(\tau | \mathbf{x}) d\tau$$

Yanagimoto and Onishi (2009) により、指数分布族において自然母数の事後平均 $(\hat{\eta}, \hat{\tau})$ はベイズ的最適性をもつことが立証されている。なお、最適性とは Yanagimoto and Onishi (2009) によると、Kullback-Leibler Divergence の e システムのリスクを最小にする予測子とある。個々のパラメータの提案推定量をこれらの式から導出し、

$$\hat{\eta} = \eta^* \quad (4)$$

$$\hat{\tau} = \int \tau A(\tau R(\mathbf{x})) \pi_m(\tau | \mathbf{x}) d\tau \\ = \int \tau A(\tau R(\mathbf{x})) p(\mathbf{x}; \eta, \tau) \pi(\eta, \tau) d\eta d\tau \quad (5)$$

が得られる。 $\hat{\eta}$ は既存の推定量である η^* と一致する。よって、本論文では τ の推定量に着目し既存推定量との比較を行う。比較の際は、頻度論的な観点からみたりスク比較を行う。

3 既存推定量とのリスク比較

本章では、提案推定量と3つの既存推定量のシミュレーションによるリスク比較を行う。はじめにリスク比較の際に用いる損失関数を示す。シミュレーションを行う際の条件を示した後、その結果をまとめる。

なおリスクは、推定量で得られた推定値を $\hat{\theta} = (\eta^*, \hat{\tau})$ 、損失関数を Φ としたとき

$$E(\Phi(\hat{\theta}, \theta)) \quad (6)$$

で表せる。

3.1 損失関数

リスク比較を行う際の損失関数は以下のものを用いる。

- 二乗誤差

$$(\theta - \hat{\theta})^2 \quad (7)$$

- Kullback-Leibler (e -divergence)

$$Le((\hat{\eta}, \hat{\tau}), (\eta, \tau)) = E \left\{ \log \frac{p(\mathbf{x} | \hat{\eta}, \hat{\tau})}{p(\mathbf{x} | \eta, \tau)}; p(\mathbf{x} | \eta, \tau) \right\} \quad (8)$$

- Dual Kullback-Leibler (m -divergence)

$$Lm((\hat{\eta}, \hat{\tau}), (\eta, \tau)) = E \left\{ \log \frac{p(\mathbf{x} | \eta, \tau)}{p(\mathbf{x} | \hat{\eta}, \hat{\tau})}; p(\mathbf{x} | \eta, \tau) \right\} \quad (9)$$

また、本論文では式 (8) 及び式 (9) をそれぞれ e KL と m KL と表記する。

3.2 シミュレーション条件

本研究では Mathematica を用いてシミュレーションを行う。サンプルサイズを $\{10, 25, 50, 70, 100\}$ の5通り、真値 τ を $\{0.25, 0.5, 1, 2, 5, 10, 30\}$ の7通り、計35通りのシミュレーションを行う。真値 η は0で固定し、試行回数は10000回とする。

3.3 既存推定量

von-Mises 分布のパラメータ推定に関して以下のような推定量が主に使用されている。これらの推定量をリスク比較の対象として用いる。

- MLE

$$\tau_{MLE} = A^{-1} \left(\frac{R(\mathbf{x})}{n} \right) \quad (10)$$

- τ の事後期待値

$$\tilde{\tau} = \int \tau \pi_m(\tau | \mathbf{x}) d\tau \quad (11)$$

- The Maximum Marginal Likelihood Estimator by Schou(1978)

$$\begin{cases} \tau_{MM} = 0 & R(\mathbf{x}) \leq \sqrt{n} \\ R(\mathbf{x})A(\tau R(\mathbf{x})) - A(\tau) = 0 & \text{otherwise} \end{cases} \quad (12)$$

本論文では以降、式 (12) を MMLE と表記する。

3.4 推定値の振る舞い

リスク比較を行う前に、それぞれの推定量で τ を推定した際に、推定値がどのような振る舞いをするのかをグラフで示す。以下のグラフは横軸に合成ベクトル長である R 、縦軸に推定値 τ をとったものである。なお、ここでの n の値は5とする。

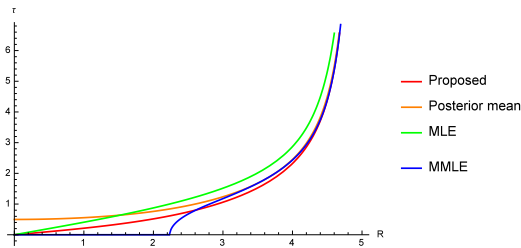


図 1: 統計量 R に伴う推定値の振る舞い

赤い線、黄色い線、緑の線、青い線はそれぞれ提案推定量、事後平均、MLE、MMLE を示す。

青い線 (MMLE を示す線) を見ると、 $\tau \approx 2.2$ 付近で急なカーブがあることが分かる。

3.5 シミュレーション

提案推定量とそれぞれの既存推定量との比較結果を以下に示す。損失関数を二乗誤差としたときの提案推定量と既存推定量のリスクの差をまとめる。本来ならば、損失関数を eKL 及び mKL としたときの比較結果も示すが、ここでは割愛する。

ここでリスクの差は、提案推定量によって得られた推定値を $\hat{\theta} = (\eta^*, \hat{\tau})$ 、既存推定量で得られた推定値を $\tilde{\theta} = (\eta^*, \tilde{\tau})$ 、損失関数を Φ としたとき

$$E(\Phi(\hat{\theta}, \theta)) - E(\Phi(\tilde{\theta}, \theta)) \quad (13)$$

と定義する。上から順に MLE、事後平均、MMLE との比較結果となっている。

表 1: MLE とのリスクの差

τ	n				
	10	25	50	70	100
0	-0.25	-0.09	-0.04	-0.03	-0.02
0.25	-0.17	-0.04	-0.01	-0.01	0.00
0.5	-0.13	-0.01	0.00	0.01	0.00
1	-0.13	0.00	0.00	0.00	0.00
2	-0.58	-0.04	-0.01	0.00	0.00
5	-5.59	-0.44	-0.10	-0.05	-0.02
10	-24.59	-1.99	-0.40	-0.20	-0.10
30	-232.13	-19.10	-3.76	-1.82	-0.86

表 2: PM とのリスクの差

τ	n				
	10	25	50	70	100
0	-0.16	-0.06	-0.03	-0.02	-0.01
0.25	-0.07	-0.01	0.00	0.00	0.00
0.5	-0.01	0.01	0.01	0.00	0.00
1	0.02	0.01	0.00	0.00	0.00
2	-0.02	0.00	0.00	0.00	0.00
5	-0.13	-0.01	0.00	0.00	0.00
10	-0.26	-0.03	-0.01	0.00	0.00
30	-0.86	-0.11	-0.03	-0.01	-0.01

表 3: MMLE とのリスクの差

τ	n				
	10	25	50	70	100
0	0.26	0.09	0.04	0.03	0.02
0.25	0.06	-0.04	-0.02	-0.02	-0.01
0.5	-0.12	-0.05	-0.01	0.00	0.00
1	-0.11	0.00	0.00	0.00	0.00
2	-0.12	0.00	0.00	0.00	0.00
5	-0.43	-0.03	0.00	0.00	0.00
10	-2.07	-0.06	-0.01	0.00	0.00
30	-17.17	-1.32	-0.02	-0.01	-0.01

これらの表は提案推定量のリスクから既存推定量のリスクの差をとったものなので、表中の値が負ならば、提案推定量の方がリスクが低いと言える。

表 1, 2 を見ると、ほとんどの値が負となっている。正の値もいくつか見受けられるが、その値の絶対値は非常に小さい。つまり、全体的に見て、提案推定量の方が MLE 及び事後平均 と比べリスクが低いということが分かる。一方で、表 3 を見ると $\tau = 0$ のとき提案推定量のリスクが高くなってしまっているが、それ以外ではリスクが低いと分かる。

4 実データ解析

提案推定量および既存推定量を 2 つの実際のデータに適用し、集中度の推定を行う。各推定量によって推定値にどれほどの差異があるのかを確認する。また、各推定値より密度関数を描写し、その形状の違いについても考察する。

4.1 風向きデータ

はじめに、風向きのデータに対する推定問題を扱う。このデータは Mardia (1972a) より得たもので、1968 年のゴーレストンの季節による風向きの違いを示したデータである。このデータに対して、提案推定量及び 3

つの既存推定量で集中度の推定を行う。それぞれの推定量により算出されたパラメータ値から導出される密度関数は以下ようになる。

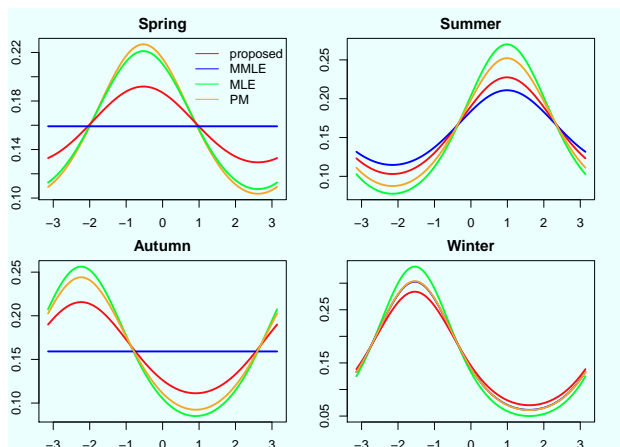


図 2: 推定された密度関数 1

次に、Batschelet (1981) より引用した交通事故が発生した時刻のデータを扱う。このデータは事故の発生した時刻が角度データへと変換されたものである。それぞれの推定量により算出されたパラメータ値から導出される密度関数は以下ようになる。

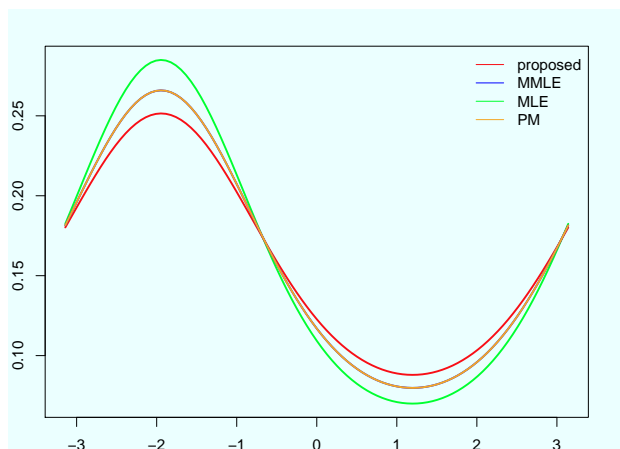


図 3: 推定された密度関数 2

図 2 および図 3 を見ると、各推定量によって描かれた密度関数には大きな差があることが分かる。つまりこれらの図は、どの推定量を解析に用いるかで結果が大きく違ってしまふことを意味する。これは、推定量の選出の重要性を示しているといえる。

5 議論

図 1 より、各推定量で推定値 τ についての挙動が大きく違うことが分かる。その中で最も特徴的なのが青

い線で示された MMLE である。この線は $R \cong 2.2$ 付近で急なカーブを有している。推定量において、このように推定値が微小な区間で急激に変化することは好ましくない。これは、MMLE の改善すべき点であるといえる。

表 1-表 2 を見ると、提案推定量の方が MLE や事後分布よりもリスクの少ない推定量であると分かる。一方で、表 3 を見ると、ある条件下の下では、提案推定量のリスクが大きくなっている。しかし、その条件以外の範囲では提案推定量のリスクの方が少ない。加えて、図 1 から分かるように MMLE は前述のような欠点を持っている。これらのことを考慮すると提案推定量のほうが有用な推定量であると考えられる。

参考文献

- [1] Akimoto, Y., Sakumura, T., & Kamakura, T. (2015). Directional data analysis for line segments. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 2).
- [2] Batschelet, E. (1981). Circular statistics in biology. *ACADEMIC PRESS*, 111 FIFTH AVE., NEW YORK, NY 10003, 1981, 388.
- [3] Fisher, N.I.(1995). *Statistical analysis of circular data*, Cambridge University Press.
- [4] Garvan, C. W., & Ghosh, M. (1997). Noninformative priors for dispersion models. *Biometrika*, 84(4), 976-982.
- [5] Mardia, K. V. (1972). Probability and mathematical statistics: statistics of directional data. *London: Academic*.
- [6] Mardia, Kanti V., and S. A. M. El-Atoum. (1976). Bayesian inference for the von Mises-Fisher distribution. *Biometrika* 63.1: 203-206.
- [7] Schou, G. (1978). Estimation of the concentration parameter in von Mises—Fisher distributions. *Biometrika*, 65(2), 369-377.
- [8] Yanagimoto, T. and Ohnishi (2009). Bayesian prediction of a density function in terms of e -mixture.” *J. Stat. Plann. Inf.* 139, 3064-3075. Amounts of Information, Journal of the Japan statistical society, 45(1), 171-191. in Japanese.