

## von Mises 分布における経験 Bayes 推定

叶 雄<sup>1</sup>・大西 俊郎<sup>2</sup>

(受付 2005 年 7 月 20 日; 改訂 2005 年 10 月 20 日)

### 要 旨

von Mises 分布の共役解析において、事後モードがエントロピー型損失関数の下で最適、すなわち、Bayes 推定量になっていることが知られている。本論文では、von Mises 事前分布の超パラメータのうち位置パラメータを既知とし拡散パラメータを周辺分布に対するモーメント法で推定することにより経験 Bayes 推定量を構成した。様々な設定の下でシミュレーションを行い、経験 Bayes 推定量のパフォーマンスを最尤推定量と比較評価したところ、パラメータが高次元であるとき実際の状況下で経験 Bayes 推定量が最尤推定量を圧倒的に優越することが分かった。

キーワード：共役事前分布，事後モード，ピタゴラス関係。

### 1. はじめに

von Mises 分布は、角度データを解析するとき標準的な分布として使用されている。ニースの高さを考えると非常に重要な分布の一つである。円周上の正規分布と呼ばれることもあり、その密度関数は

$$p(x; \mu, \tau) \propto \exp\{\tau \cos(x - \mu)\}$$

で与えられ、位置パラメータ  $\mu$  および拡散パラメータ  $\tau$  をもつ。拡散パラメータ  $\tau$  が既知定数のとき、von Mises 分布は曲指数型分布族に属する。

共役解析は、Bayes 統計学の一分野として重要な役割を果たしてきた。指数型分布族の共役解析が代表的であり、多くの研究者によって様々な研究がなされている。一方、いくつかの曲指数型分布族でも共役事前分布が存在する。実際、Mardia and El-Atoum (1976) が示したように、拡散パラメータ  $\tau$  が既知の場合の von Mises 分布においても共役解析が可能である。von Mises 分布の共役解析を論じた論文としては、Guttorp and Lockhart (1988)、Bagchi (1994) および Rodrigues et al. (2000) などがある。

本論文では、von Mises 分布において共役解析をベースにした経験 Bayes 推定量を構成する。具体的には、超パラメータに対する推定方程式を明示的に与え、これを解くことにより経験 Bayes 推定量を導出する。また、そのパフォーマンスをシミュレーションにより評価する。

<sup>1</sup> 総合研究大学院大学 複合科学研究科統計科学専攻：〒106-8569 東京都港区南麻布 4-6-7

<sup>2</sup> 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

## 2. 共役解析

確率変数  $x$  は次の密度関数をもつ von Mises 分布に従うとする .

$$(2.1) \quad p(x; \mu, \tau_0) = \frac{1}{2\pi I_0(\tau_0)} \exp\{\tau_0 \cos(x - \mu)\}, \quad 0 \leq x < 2\pi.$$

ここで,  $I_0(\tau_0)$  は第一種修正ベッセル関数であり, 拡散パラメータ  $\tau_0 > 0$  は既知とする . 位置パラメータ  $\mu \in [0, 2\pi)$  に対して, 次の von Mises 事前分布を仮定する .

$$(2.2) \quad \pi(\mu; m, \delta) = \frac{1}{2\pi I_0(\delta)} \exp\{\delta \cos(\mu - m)\}.$$

ここで,  $m$  および  $\delta$  は超パラメータである . Mardia and El-Atoum (1976) は次の命題を証明した .

命題 2.1. (Mardia and El-Atoum (1976)) (i) 事前分布 (2.2) は標本分布 (2.1) に対して共役であり, 事後密度は  $\pi(\mu; \hat{\mu}_{\text{MAP}}, \delta_*)$  と書ける . ただし,  $\hat{\mu}_{\text{MAP}}$  は事後モードであり,  $\delta_* = \{\tau_0^2 + \delta^2 + 2\tau_0\delta \cos(x - m)\}^{1/2}$  である .

(ii) 事後モード  $\hat{\mu}_{\text{MAP}}$  は損失関数  $L(\hat{\mu} - \mu) = 1 - \cos(\hat{\mu} - \mu)$  の下で最適である .

事後モードは具体的には次式で与えられる :

$$(2.3) \quad \hat{\mu}_{\text{MAP}} (= \hat{\mu}_{\text{MAP}}(x, m, \delta)) = \begin{cases} \cos^{-1} \left( \frac{\tau_0 \cos x + \delta \cos m}{\delta_*} \right) & \tau_0 \sin x + \delta \sin m \geq 0 \text{ のとき,} \\ 2\pi - \cos^{-1} \left( \frac{\tau_0 \cos x + \delta \cos m}{\delta_*} \right) & \text{その他 .} \end{cases}$$

後続の節で記述を簡単化するために, 事後モードが  $x, m$  および  $\delta$  に依存していることを明示的に表すときには,  $\hat{\mu}_{\text{MAP}}(x, m, \delta)$  と記す .

円周上のデータを扱っているので, 損失関数  $L(\hat{\mu} - \mu) = 1 - \cos(\hat{\mu} - \mu)$  は自然な選択と思われる . この損失関数はいわゆるエントロピー型損失関数であることを示す . まず, 2 つの標本密度  $p(x; \mu_1, \tau_0)$  および  $p(x; \mu_2, \tau_0)$  の間の Kullback-Leibler 分離度を計算する . 三角関数の加法定理により

$$\begin{aligned} \cos(x - \mu_2) &= \cos(x - \mu_1 + \mu_1 - \mu_2) \\ &= \cos(x - \mu_1) \cos(\mu_1 - \mu_2) - \sin(x - \mu_1) \sin(\mu_1 - \mu_2) \end{aligned}$$

であることに注意する . これを用いると,

$$\begin{aligned} \frac{1}{\tau_0} \text{KL}(p(x; \mu_1, \tau_0), p(x; \mu_2, \tau_0)) &= \int_0^{2\pi} \{\cos(x - \mu_1) - \cos(x - \mu_2)\} p(x; \mu_1, \tau_0) dx \\ &= \{1 - \cos(\mu_1 - \mu_2)\} \int_0^{2\pi} \cos(x - \mu_1) p(x; \mu_1, \tau_0) dx \\ &\quad + \sin(\mu_1 - \mu_2) \int_0^{2\pi} \sin(x - \mu_1) p(x; \mu_1, \tau_0) dx \end{aligned}$$

である . 等式  $\int_0^{2\pi} p(x; \mu, \tau_0) dx = 1$  の両辺を  $\mu$  および  $\tau_0$  で偏微分すると, それぞれ

$$(2.4) \quad \int_0^{2\pi} \sin(x - \mu) p(x; \mu, \tau_0) dx = 0,$$

$$(2.5) \quad \int_0^{2\pi} \cos(x - \mu) p(x; \mu, \tau_0) dx = \frac{I_1(\tau_0)}{I_0(\tau_0)}$$

を得る．ここで， $I_0'(\tau_0) = I_1(\tau_0)$  を用いた．したがって，

$$(2.6) \quad \text{KL}(p(x; \mu_1, \tau_0), p(x; \mu_2, \tau_0)) = \tau_0 \frac{I_1(\tau_0)}{I_0(\tau_0)} L(\mu_1 - \mu_2)$$

となる．

事後モードが他の推定量を優越する様子を明快に示す変形ピタゴラス関係を導く．

命題 2.2. 任意の推定量  $\hat{\mu}$  に対して，次の変形ピタゴラス関係が成り立つ．

$$(2.7) \quad \int_0^{2\pi} \{L(\hat{\mu} - \mu) - L(\hat{\mu}_{\text{MAP}} - \mu)\} \pi(\mu; \hat{\mu}_{\text{MAP}}, \delta_*) d\mu = \frac{I_1(\delta_*)}{I_0(\delta_*)} L(\hat{\mu} - \hat{\mu}_{\text{MAP}}).$$

証明．三角関数の加法定理を用いると次の式変形ができる：

$$\begin{aligned} L(\hat{\mu} - \mu) - L(\hat{\mu}_{\text{MAP}} - \mu) &= \cos(\hat{\mu}_{\text{MAP}} - \mu) - \cos(\hat{\mu} - \mu) \\ &= L(\hat{\mu} - \hat{\mu}_{\text{MAP}}) \cos(\hat{\mu}_{\text{MAP}} - \mu) + \sin(\hat{\mu} - \hat{\mu}_{\text{MAP}}) \sin(\hat{\mu}_{\text{MAP}} - \mu). \end{aligned}$$

ここで，式(2.4)および式(2.5)において  $x, \mu$  および  $\tau_0$  をそれぞれ  $\mu, \hat{\mu}_{\text{MAP}}$  および  $\delta_*$  に置換したものをを用いると，変形ピタゴラス関係(2.7)が得られる．□

変形ピタゴラス関係(2.7)の右辺の意味を議論する．2つの事前密度  $\pi(\mu; m_1, \delta)$  および  $\pi(\mu; m_2, \delta)$  の Kulback-Leibler 分離度は(2.6)と同様に

$$\text{KL}(\pi(\mu; m_1, \delta), \pi(\mu; m_2, \delta)) = \delta \frac{I_1(\delta)}{I_0(\delta)} L(m_1 - m_2)$$

となる．したがって，変形ピタゴラス関係(2.7)は次のように書き換えられる：

$$\int_0^{2\pi} \{L(\hat{\mu}, \mu) - L(\hat{\mu}_{\text{MAP}}, \mu)\} \pi(\mu; \hat{\mu}_{\text{MAP}}, \delta_*) d\mu = \frac{1}{\delta_*} \text{KL}(\pi(\mu; \hat{\mu}_{\text{MAP}}, \delta_*), \pi(\mu; \hat{\mu}, \delta_*)).$$

事後リスクの差が2つの密度の Kulback-Leibler 分離度を用いて表現できるというこの関係は，von Mises 分布に限られたものではない．ある一定の曲指数型分布族の共役解析において成立することがわかっている(大西・柳本, 2005)．

### 3. 経験 Bayes 推定

確率変数  $x_i (1 \leq i \leq p)$  が互いに独立に von Mises 分布  $p(x_i; \mu_i, \tau_0)$  に従っているとし，パラメータ  $\mu_i (1 \leq i \leq p)$  が互いに独立に von Mises 事前分布  $\pi(\mu_i; m_i, \delta)$  に従っていると仮定する．超パラメータ  $m = (m_1, \dots, m_p)^T$  は既知とし， $\delta$  を未知とする．周辺分布によるモーメント法を用いて超パラメータ  $\delta$  を推定し，経験 Bayes 推定量を構成する．ここで，推定すべきパラメータ  $\mu = (\mu_1, \dots, \mu_p)^T$  の次元  $p$  が大きい場合を想定している，すなわち，高次元パラメータの推定問題を考える．次節のシミュレーションでは，高次元の場合として  $p = 100$  のケースを扱う．また，比較のために低次元の例として  $p = 10$  の場合を計算する．

von Mises 分布における経験 Bayes 推定において，Bagchi (1994) はモーメント法を用いて  $m$  および  $\delta$  を推定している．本研究で  $m$  を既知としたのは，次の2つの理由による：(1) 実際の推定では，事前データなどがありパラメータ  $\mu$  のおおよその値が知れていることが多い(2) 正規分布における James-Stein 推定量は，任意に与えた位置パラメータについて拡散パラメータを推定して構成される経験 Bayes 推定量として理解できるが，von Mises 分布においても Stein 現象と似たような現象が予想される．

超パラメータ  $\delta$  に対する推定方程式を導くとき, 次の変形ピタゴラス関係が重要な役割を果たす. 証明は命題 2.2 と本質的に同じなので省略する.

補題 3.1. 次の変形ピタゴラス関係が成り立つ:

$$(3.1) \quad \int_0^{2\pi} \left\{ L(x_i - m_i) - L(x_i - \mu_i) - \frac{I_1(\tau_0)}{I_0(\tau_0)} L(\mu_i - m_i) \right\} p(x_i; \mu_i, \tau_0) dx_i = 0 \quad (1 \leq i \leq p).$$

超パラメータ  $\delta$  をモーメント法で推定するときの基礎となる等式を命題の形で与える.

命題 3.1. 次の等式が成立する.

$$(3.2) \quad E_{\text{marg}} \left[ \frac{1}{p} \sum_{i=1}^p \cos(x_i - m_i) \right] = \frac{I_1(\tau) I_1(\delta)}{I_0(\tau) I_0(\delta)}.$$

ただし,  $E_{\text{marg}}[\cdot]$  は周辺分布による期待値を表す.

証明. 周辺分布を  $p_{\text{marg}}(x; \tau_0, \mathbf{m}, \delta) = \prod p_{\text{marg}}(x_i; \tau_0, m_i, \delta)$  と書く. ただし,  $x = (x_1, \dots, x_p)^T$  であり,

$$p_{\text{marg}}(x_i; \tau_0, m_i, \delta) = \int_0^{2\pi} p(x_i; \mu_i, \tau_0) \pi(\mu_i; m_i, \delta) d\mu_i$$

である. 積分の順序を交換することにより,

$$\begin{aligned} & \int_0^{2\pi} L(x_i - m_i) p_{\text{marg}}(x_i; \tau_0, m_i, \delta) dx_i \\ &= \int_0^{2\pi} L(x_i - m_i) \left\{ \int_0^{2\pi} p(x_i; \mu_i, \tau_0) \pi(\mu_i; m_i, \delta) d\mu_i \right\} dx_i \\ &= \int_0^{2\pi} \left\{ \int_0^{2\pi} L(x_i - m_i) p(x_i; \mu_i, \tau_0) dx_i \right\} \pi(\mu_i; m_i, \delta) d\mu_i \end{aligned}$$

を得る. 補題 3.1 および等式

$$\int_0^{2\pi} L(x_i - \mu_i) p(x_i; \mu_i, \tau_0) dx_i = 1 - \frac{I_1(\tau_0)}{I_0(\tau_0)}$$

を用いると,

$$\begin{aligned} & \int_0^{2\pi} L(x_i - m_i) p_{\text{marg}}(x_i; \tau_0, m_i, \delta) dx_i \\ &= \int_0^{2\pi} \left[ \int_0^{2\pi} \left\{ L(x_i - \mu_i) + \frac{I_1(\tau_0)}{I_0(\tau_0)} L(\mu_i - m_i) \right\} p(x_i; \mu_i, \tau_0) dx_i \right] \pi(\mu_i; m_i, \delta) d\mu_i \\ &= \int_0^{2\pi} \left[ 1 - \frac{I_1(\tau_0)}{I_0(\tau_0)} + \frac{I_1(\tau_0)}{I_0(\tau_0)} L(\mu_i - m_i) \right] \pi(\mu_i; m_i, \delta) d\mu_i \end{aligned}$$

となる. ここで,  $\int_0^{2\pi} L(\mu_i - m_i) \pi(\mu_i; m_i, \delta) d\mu_i = 1 - I_1(\delta)/I_0(\delta)$  であるから,

$$\int_0^{2\pi} L(x_i - m_i) p_{\text{marg}}(x_i; \tau_0, m_i, \delta) dx_i = 1 - \frac{I_1(\tau_0) I_1(\delta)}{I_0(\tau_0) I_0(\delta)}$$

となる.

$$E_{\text{marg}} \left[ \frac{1}{p} \sum_{i=1}^p \cos(x_i - m_i) \right] = \frac{1}{p} \sum_{i=1}^p E_{\text{marg}} [\cos(x_i - m_i)]$$

$$= \frac{1}{p} \sum_{i=1}^p \int_0^{2\pi} \{1 - L(x_i - m_i)\} p_{\text{marg}}(x_i; \tau_0, m_i, \delta) dx_i$$

に注意すると，命題の等式が得られる．□

超パラメータ  $\delta$  を推定するために次の推定方程式を考える．

$$(3.3) \quad \frac{I_1(\delta)}{I_0(\delta)} = \frac{I_0(\tau_0)}{I_1(\tau_0)} \frac{1}{p} \sum_{i=1}^p \cos(x_i - m_i).$$

関数  $I_1(\delta)/I_0(\delta)$  は  $\delta \in \mathbb{R}$  について狭義単調増加であり，値域は区間  $(-1, 1)$  である．狭義単調増加性は次のように証明できる．

$$\frac{d}{d\delta} \frac{I_1(\delta)}{I_0(\delta)} = \frac{I_1'(\delta)I_0(\delta) - I_1(\delta)I_0'(\delta)}{I_0^2(\delta)}$$

であるから，右辺の分子が正であることを示せばよい．積分表現

$$I_n(\delta) = \frac{1}{\pi} \int_0^\pi \cos(n\theta) \exp(\delta \cos \theta) d\theta, \quad n \text{ は整数}$$

(Abramowitz and Stegun, 1970, p. 376)を用いると，右辺の分子は

$$\frac{1}{\pi^2} \left\{ \int_0^\pi \cos^2 \theta e^{\delta \cos \theta} d\theta \int_0^\pi e^{\delta \cos \theta} d\theta - \left( \int_0^\pi \cos \theta e^{\delta \cos \theta} d\theta \right)^2 \right\}$$

となる．ここで微分と積分の順序交換を行っている．Schwarz の不等式からこれは正である．

積分表現において  $\phi = \pi - \theta$  と置換し積分を実行すると，関数  $I_1(\delta)/I_0(\delta)$  は奇関数であることがわかる．奇関数なので  $I_1(0)/I_0(0) = 0$  である．また，修正ベッセル関数の漸近展開 (Abramowitz and Stegun, 1970, p. 377)を用いると， $\delta$  が十分大きいときの近似式

$$\frac{I_1(\delta)}{I_0(\delta)} \approx 1 - \frac{1}{2\delta}$$

が得られる．

狭義単調性および値域が  $(-1, 1)$  であることから，

$$\left| \frac{I_0(\tau_0)}{I_1(\tau_0)} \frac{1}{p} \sum_{i=1}^p \cos(x_i - m_i) \right| < 1$$

のときは， $\hat{\delta}$  は一意に定まる．パラメータ  $\mu_i$  に対する経験 Bayes 推定量は，式(2.3)を用いて， $\hat{\mu}_{\text{EB}, i} = \hat{\mu}_{\text{MAP}}(x_i, m_i, \hat{\delta})$  ( $1 \leq i \leq p$ ) となる．

推定方程式(3.3)が解を持たないときは次の手続きを行う．

$$\frac{I_0(\tau_0)}{I_1(\tau_0)} \frac{1}{p} \sum_{i=1}^p \cos(x_i - m_i) \geq 1$$

のとき， $\hat{\mu}_{\text{EB}, i} = m_i$  ( $1 \leq i \leq p$ ) とする．この不等式が成立するのは，データ  $x$  と超パラメータ  $m$  は十分近いときである．パラメータ  $\mu$  の推定量として  $m$  を採用するのは自然である．一方，

$$\frac{I_0(\tau_0)}{I_1(\tau_0)} \frac{1}{p} \sum_{i=1}^p \cos(x_i - m_i) \leq -1$$

のとき， $\hat{\mu}_{\text{EB}, i} = m_i + \pi$  とする．このケースでは，データ  $x$  と超パラメータ  $m$  は各成分が 180 度近く離れていると考えられるので， $m + \pi \mathbf{1}$  を推定量とすることは妥当である．ただし，

$\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^p$  である。ここで,

$$\begin{aligned}\lim_{\hat{\delta} \rightarrow \infty} \hat{\mu}_{\text{MAP}}(x_i, m_i, \hat{\delta}) &= m_i, \\ \lim_{\hat{\delta} \rightarrow -\infty} \hat{\mu}_{\text{MAP}}(x_i, m_i, \hat{\delta}) &= m_i + \pi\end{aligned}$$

が成り立つ。したがって、推定方程式(3.3)が解を持たないときも含めて経験 Bayes 推定量は形式的に  $\hat{\mu}_{\text{MAP}}(x_i, m_i, \hat{\delta})$  ( $1 \leq i \leq p$ ) と書ける。

ここで、一つ注意すべきことは、 $\hat{\delta}$  として負の値を許していることである。 $\delta \cos(\mu_i - m_i) = -\delta \cos(\mu_i - m_i - \pi)$  であり、 $\pi(\mu_i; m_i, \delta)$  および  $\pi(\mu_i; m_i + \pi, -\delta)$  は分布としては同一であるので、これは自然な取扱いといえる。

#### 4. リスクの比較

経験 Bayes 推定量  $\hat{\mu}_{\text{EB}}$  のリスクを数値シミュレーションし、最尤推定量  $\hat{\mu}_{\text{ML}} = x$  のリスクと比較する。最尤推定量のリスクは次のように計算できる。

$$\begin{aligned}E \left[ \sum_{i=1}^p L(x_i - \mu_i) \middle| \prod_{i=1}^p p(x_i; \mu_i, \tau_0) \right] &= \sum_{i=1}^p E[L(x_i - \mu_i) | p(x_i; \mu_i, \tau_0)] \\ &= p \left\{ 1 - \frac{I_1(\tau_0)}{I_0(\tau_0)} \right\}.\end{aligned}$$

ここで、 $E[f(x)|p(x)]$  は密度関数  $p(x)$  の下での  $f(x)$  の期待値を表す。パラメータ  $\mu$  の値によらず一定であることに注意する。当然のことながら、 $m$  にも依存しない。

シミュレーションは以下の 4 つの場合に分けて行う。反復回数はすべて 50,000 である。固定したパラメータ  $\mu$  に対して超パラメータ  $m$  を様々に変化させ、経験 Bayes 推定量のリスクを計算しグラフを描く。以下のグラフにおいて、経験 Bayes 推定量のリスクは '+' でプロットし、最尤推定量のリスクは実線で表示する。ケース 3 以外では  $\tau_0 = 1$  とし、ケース 3 では拡散パラメータ  $\tau_0$  を 2, 5, 10 と変化させて 2 つの推定量のリスクを比較している。

##### 4.1 ケース 1

経験 Bayes 推定量のパフォーマンスの全体像をつかむために、まず最初に次のような最も簡単な設定の場合を扱う：

$$\mu = \mu \mathbf{1} \quad \text{かつ} \quad m = m \mathbf{1}.$$

図 1(a) は、 $p = 10, \mu = \mu \mathbf{1}$  に対して  $m = m \mathbf{1}$  の  $m$  を 0 から  $2\pi$  まで変化させたときの経験 Bayes 推定量のリスクをグラフにしたものである。実際の計算では  $\mu = 0$  としているが、location family がもつ対称性により結果は  $\mu - m$  にのみ依存することが分かる。さらに、損失関数が偶関数であるからリスクは  $\mu - m$  の符号によらず絶対値のみによることが分かる。この 2 点について略証を与える。確率変数  $e_i = x_i - \mu$  ( $1 \leq i \leq p$ ) は互いに独立に von Mises 分布  $vM(0, \tau_0)$  に従う。ここで、記号  $vM(\mu, \tau_0)$  は密度関数(2.1)をもつ von Mises 分布を表す。加法定理を用いると、経験 Bayes 推定量に対する損失の値は次のように表現できる。

$$1 - \cos(\hat{\mu}_{\text{EB}, i} - \mu) = 1 - \hat{\delta}_{*i}^{-1} \{ \tau_0 \cos e_i + \hat{\delta} \cos(\mu - m) \}.$$

ただし、

$$\frac{I_1(\hat{\delta})}{I_0(\hat{\delta})} = \frac{1}{p} \frac{I_0(\tau_0)}{I_1(\tau_0)} \left\{ \cos(\mu - m) \sum_{i=1}^p \cos e_i - \sin(\mu - m) \sum_{i=1}^p \sin e_i \right\},$$

$$\hat{\delta}_{*i} = [\tau_0^2 + \hat{\delta}^2 + 2\tau_0\hat{\delta}\{\cos(\mu - m)\cos e_i - \sin(\mu - m)\sin e_i\}]^{1/2}.$$

これらの等式から，適当な関数  $f_i$  を用いて

$$1 - \cos(\hat{\mu}_{EB,i} - \mu) = f_i(e_1, \dots, e_p; \mu - m)$$

と表現できることが分かる．これはリスク  $\sum_{i=1}^p E[1 - \cos(\hat{\mu}_{EB,i} - \mu)]$  が  $\mu - m$  にのみ依存することを意味している．

リスクが  $\mu - m$  の絶対値のみに依存することは， $\mu - m < 0$  のとき

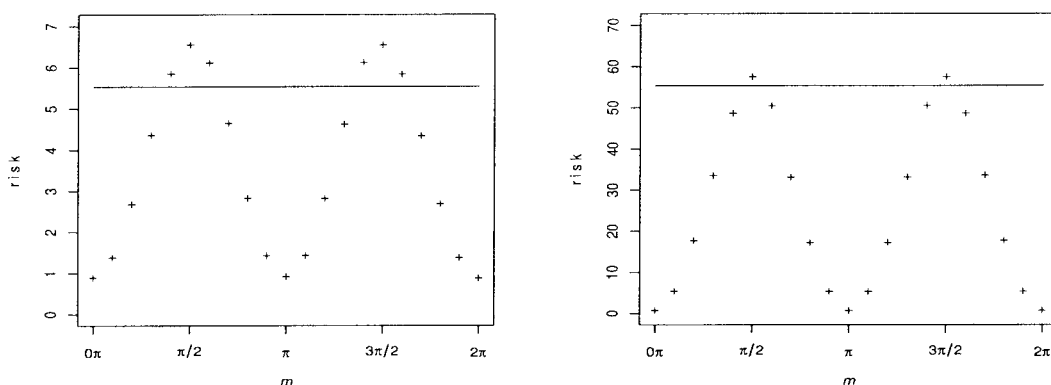
$$\cos(\mu - m)\cos e_i = \cos|\mu - m|\cos(-e_i),$$

$$\sin(\mu - m)\sin e_i = \sin|\mu - m|\sin(-e_i).$$

と書き換えられることおよび確率変数  $-e_i$  が von Mises 分布  $vM(0, \tau_0)$  に従うことから証明できる．

グラフから読み取れることは，次の 2 点である。(1)  $|\mu - m| \approx 0$  あるいは  $|\mu - m| \approx \pi$  のとき，圧倒的に経験 Bayes 推定量がよい(2)最尤推定量に対してリスクが 20%程度大きくなる  $|\mu - m| \approx \pi/2$  のときを除けば，経験 Bayes 推定量のパフォーマンスは最尤推定量に勝っている．第 1 点については次のように説明できる：超パラメータ  $m$  はパラメータ  $\mu$  の予想値という解釈ができる．予想と真の値が近い，すなわち  $|\mu - m| \approx 0$  のとき，経験 Bayes 推定量はとてもよい振る舞いをする訳である．もっと詳しく言えば， $\delta$  の推定値が高い確率で大きな値になり，経験 Bayes 推定値も高い確率で  $m$  に近くなるのである．一方，予想と真の値が 180 度近く外れているとき， $\delta$  の推定値は高い確率で絶対値が大きな負の値になり，推定値が  $(m + \pi)$  に近くなるのである．経験 Bayes 推定量のパフォーマンスが必ずしも良くないのは  $|\mu - m| \approx \pi/2$  のときである．

図 1(b)は，図 1(a)において  $p = 100$  の場合である．2 つのグラフを比較すると，高次元になれば経験 Bayes 推定量のパフォーマンスがよくなることが分かる． $p = 10$  のときでも最尤推定量を優越していた  $|\mu - m| \approx 0$  および  $|\mu - m| \approx \pi$  の場合には，その優越度がさらに高まっている．また， $p = 10$  のときに最尤推定量に劣っていた  $|\mu - m| \approx \pi/2$  の場合でも差がほとんど



(a)  $p = 10$

(b)  $p = 100$

図 1. ケース 1 における経験 Bayes 推定量および最尤推定量のリスク．

なくなっている．圧倒的に優越するか，劣るとしてもごく僅かということになって，提案した経験 Bayes 推定量のよい振る舞いが示されている．

高次元になったときのよい振る舞いを定性的に説明する．超パラメータ  $\delta$  の推定方程式は，加法定理を用いると，

$$\frac{I_1(\delta)}{I_0(\delta)} = \frac{I_0(\tau_0)}{I_1(\tau_0)} \left\{ \cos(\mu - m) \frac{1}{p} \sum_{i=1}^p \cos(x_i - \mu) - \sin(\mu - m) \frac{1}{p} \sum_{i=1}^p \sin(x_i - \mu) \right\}$$

と書ける．確率変数  $e_i = x_i - \mu$  ( $1 \leq i \leq p$ ) は独立に同一に von Mises 分布  $vM(0, \tau_0)$  に従う．大数の強法則を適用すると，確率 1 で

$$\lim_{p \rightarrow \infty} \frac{1}{p} \sum_{i=1}^p \cos(x_i - \mu) = \frac{I_1(\tau_0)}{I_0(\tau_0)} \quad \text{および} \quad \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{i=1}^p \sin(x_i - \mu) = 0$$

が成立することが分かる．したがって，確率 1 で

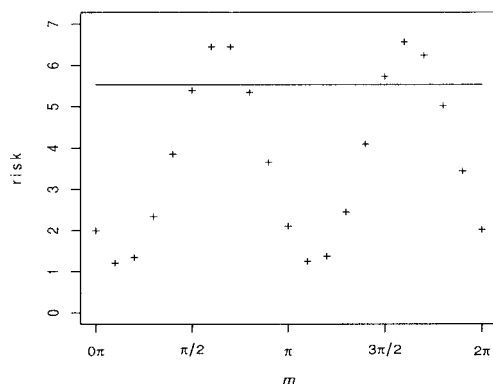
$$\lim_{p \rightarrow \infty} \frac{I_1(\hat{\delta})}{I_0(\hat{\delta})} = \cos(\mu - m)$$

が成り立つ．この等式は，高次元のとき高い確率で  $I_1(\hat{\delta})/I_0(\hat{\delta}) \approx \cos(\mu - m)$  が成り立つことを意味している．したがって， $\lim_{\delta \rightarrow \infty} \{I_1(\delta)/I_0(\delta)\} = 1$  に注意すると  $|\mu - m| \approx 0$  のとき高い確率で  $\hat{\delta}$  は非常に大きな正の値になる．経験 Bayes 推定量の式  $\hat{\mu}_{\text{EB}, i} = \hat{\mu}_{\text{MAP}}(x_i, m_i, \hat{\delta})$  および (2.3) から  $\hat{\mu}_{\text{EB}} \approx m$  となる．次に  $|\mu - m| \approx \pi$  の場合を考える．関数  $I_1(\delta)/I_0(\delta)$  が奇関数であることに注意すると，上と同様の議論により  $\hat{\delta}$  は絶対値が非常に大きな負の値になり， $\hat{\mu}_{\text{EB}} \approx m + \pi$  となることが分かる．最後に  $|\mu - m| \approx \pi/2$  の場合を考える．等式  $I_1(0)/I_0(0) = 0$  に注意すると，高い確率で  $\hat{\delta} \approx 0$  となることが分かる．したがって， $\hat{\mu}_{\text{EB}} \approx \hat{\mu}_{\text{ML}}$  となる．

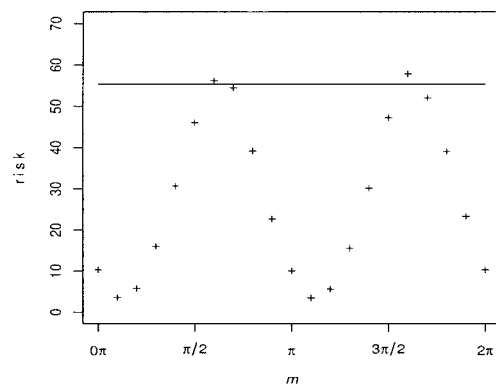
#### 4.2 ケース 2

実際的な設定で計算を行う．すなわち，

$$\mu_i = \frac{\pi}{4} \frac{i}{p} \quad (1 \leq i \leq p) \quad \text{かつ} \quad m = m_1$$



(a)  $p = 10$



(b)  $p = 100$

図 2. ケース 2 における経験 Bayes 推定量および最尤推定量のリスク．



の場合を考える．このケースでは，第 3 節で述べたように事前データなどから  $m \approx \pi/8$  と知れていることが多い．図 2 は， $\mu_i = (\pi/4) \cdot (i/p)$  ( $1 \leq i \leq p$ ) に対して  $m = m_1$  の  $m$  を 0 から  $2\pi$  まで変化させたときの経験 Bayes 推定量のリスクをグラフにしたものである．大雑把に言えば， $\mu = (\pi/8)1$  の場合と似たような結果になることが予想される．図 1 と比較するとそのことが確認できる． $p = 10$  のとき， $m = 5\pi/8$  あるいは  $m = 13\pi/8$  あたり以外では経験 Bayes 推定量のパフォーマンスは最尤推定量に勝っている．高次元になると，これまでと同様にほとんどの  $m$  の値に対して経験 Bayes 推定量は最尤推定量を優越する．最尤推定量に劣る  $m = 5\pi/8$  あるいは  $m = 13\pi/8$  においてもその差は極めて小さい．

### 4.3 ケース 3

拡散パラメータ  $\tau_0$  以外はケース 1 と同じ設定で， $\tau_0$  を 2, 5, 10 と変化させる．図 3 はそれぞれの  $\tau_0$  について経験 Bayes 推定量および最尤推定量のリスクをグラフにしたものである．これまでと同様に，高次元のとき経験 Bayes 推定量のパフォーマンスはよい．一方，10 次元の場合でも  $\tau_0$  が大きくなるにつれて経験 Bayes 推定量の最尤推定量に対する相対的なパフォーマンスはよくなっていくことが分かる．この事実の理由として考えられるのは， $\tau_0$  が大きくなると von Mises 分布が正規分布に近づくことである．計算は省略するが，第 3 節で導出した経験 Bayes 推定量は  $\tau_0$  が十分大きく，かつ， $\|\mu - m\| \approx 0$  のとき，James-Stein 推定量と近い形になる．第 3 節で少し述べたように，Stein 現象に似た状況が現れたと言える．

### 4.4 ケース 4

経験 Bayes 推定量にとってかなり不利な設定で計算してみる．具体的には次のような設定にする：

$$\mu_i = \begin{cases} \mu & (1 \leq i \leq p/2) \\ \mu' & (p/2 + 1 \leq i \leq p) \end{cases} \quad \text{かつ } m = m_1.$$

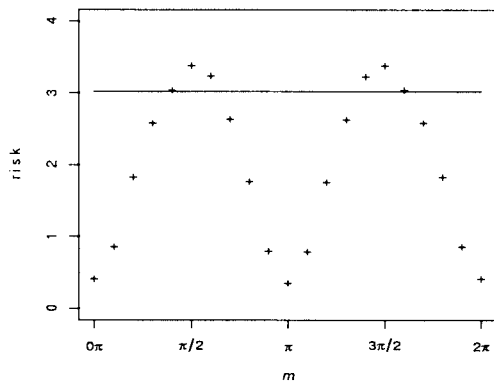
図 4(a) および図 4(b) は， $\mu = 0$ ， $\mu' = \pi$  に対して  $m$  を 0 から  $2\pi$  まで変化させたときの経験 Bayes 推定量のリスクをグラフにしたものである．この設定ではすべての  $m$  に対して経験 Bayes 推定量は最尤推定量に劣っている．パラメータの配置がこのような場合，経験 Bayes 推定量のパフォーマンスはよくない．しかし，そのような不利な状況でも次元が高くなるとリスクの差はほとんどないことが分かる．

第 4.1 節と同様に，高次元のときの振る舞いを調べてみる．推定方程式は

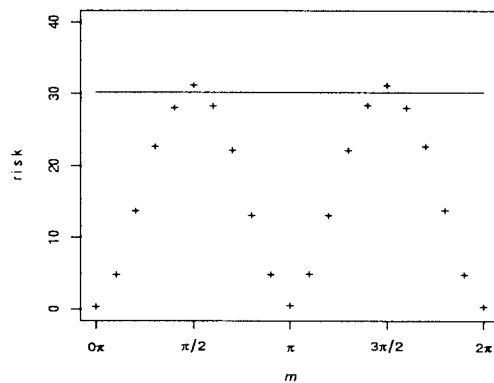
$$\frac{I_1(\delta)}{I_0(\delta)} = \frac{I_0(\tau_0)}{I_1(\tau_0)} \left\{ \cos(\mu - m) \frac{1}{p} \sum_{j=1}^{p/2} \cos(x_j - \mu) - \sin(\mu - m) \frac{1}{p} \sum_{j=1}^{p/2} \sin(x_j - \mu) \right. \\ \left. + \cos(\mu' - m) \frac{1}{p} \sum_{j=p/2+1}^p \cos(x_j - \mu') - \sin(\mu' - m) \frac{1}{p} \sum_{j=p/2+1}^p \sin(x_j - \mu') \right\}$$

となる．大数の強法則から，確率 1 で

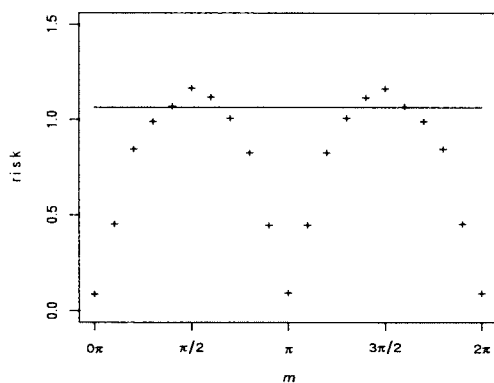
$$\lim_{p \rightarrow \infty} \frac{2}{p} \sum_{j=1}^{p/2} \cos(x_j - \mu) = \frac{I_1(\tau_0)}{I_0(\tau_0)} \quad \text{および} \quad \lim_{p \rightarrow \infty} \frac{2}{p} \sum_{j=1}^{p/2} \sin(x_j - \mu) = 0, \\ \lim_{p \rightarrow \infty} \frac{2}{p} \sum_{j=p/2+1}^p \cos(x_j - \mu') = \frac{I_1(\tau_0)}{I_0(\tau_0)} \quad \text{および} \quad \lim_{p \rightarrow \infty} \frac{2}{p} \sum_{j=p/2+1}^p \sin(x_j - \mu') = 0$$



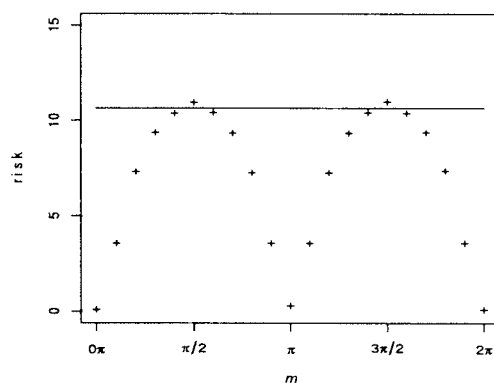
(a)  $p = 10, \tau_0 = 2$



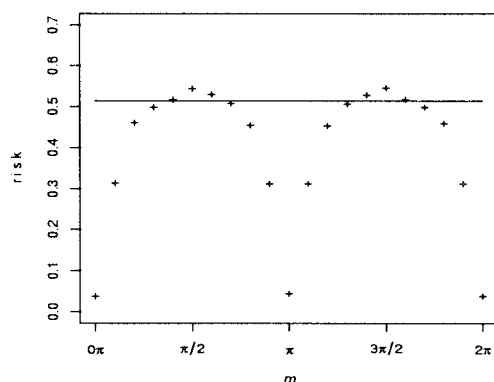
(b)  $p = 100, \tau_0 = 2$



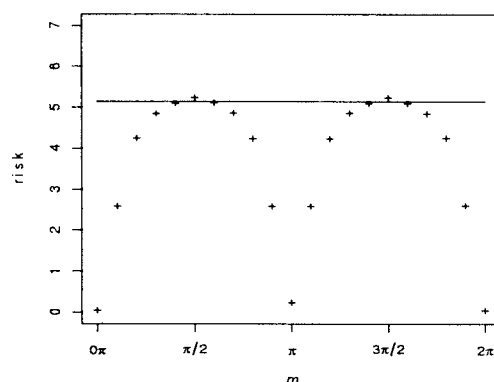
(c)  $p = 10, \tau_0 = 5$



(d)  $p = 100, \tau_0 = 5$

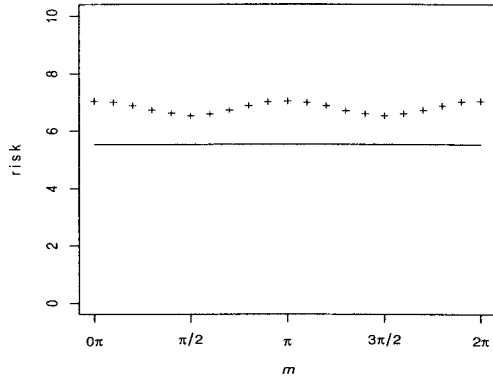


(e)  $p = 10, \tau_0 = 10$

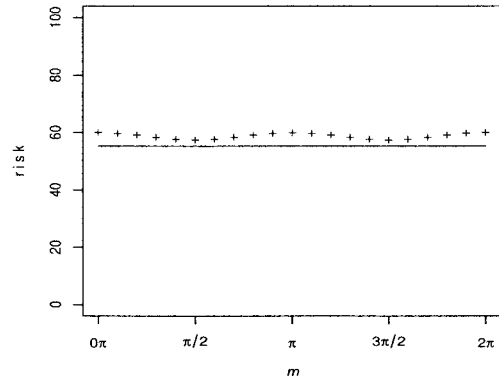


(f)  $p = 100, \tau_0 = 10$

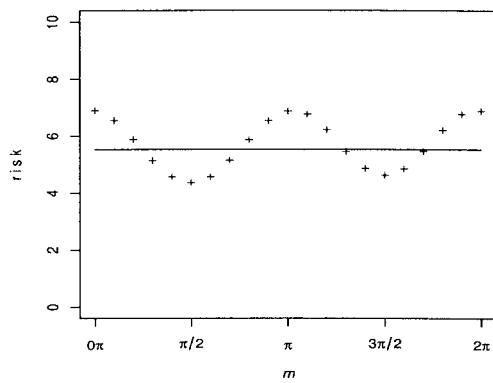
図 3. ケース 3 における経験 Bayes 推定量および最尤推定量のリスク.



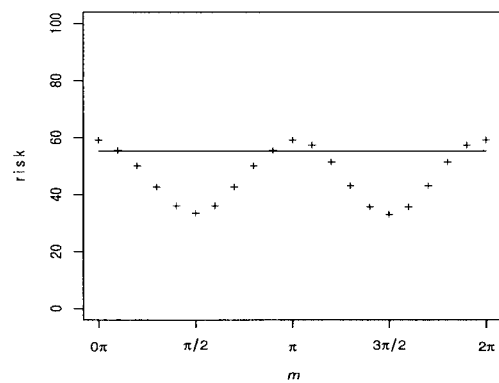
(a)  $p = 10, (\mu, \mu') = (0, \pi)$



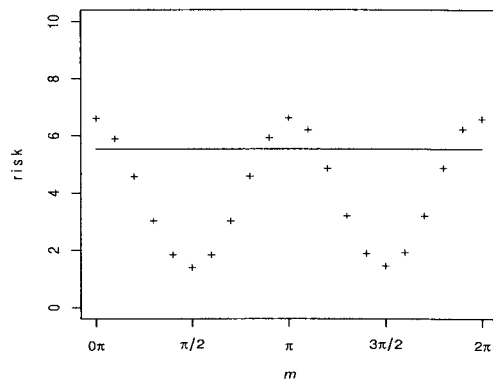
(b)  $p = 100, (\mu, \mu') = (0, \pi)$



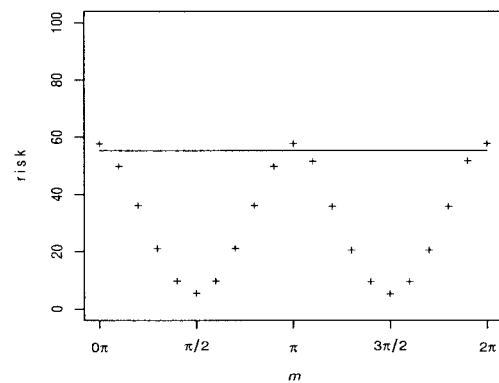
(c)  $p = 10, (\mu, \mu') = (0.2\pi, 0.8\pi)$



(d)  $p = 100, (\mu, \mu') = (0.2\pi, 0.8\pi)$



(e)  $p = 10, (\mu, \mu') = (0.4\pi, 0.6\pi)$



(f)  $p = 100, (\mu, \mu') = (0.4\pi, 0.6\pi)$

図 4. ケース 4 における経験 Bayes 推定量および最尤推定量のリスク。

が成立する．したがって，確率1で

$$(4.1) \quad \lim_{p \rightarrow \infty} \frac{I_1(\hat{\delta})}{I_0(\hat{\delta})} = \frac{1}{2} \{ \cos(\mu - m) + \cos(\mu' - m) \}$$

が成り立つ． $|\mu - \mu'| = \pi$  のとき，右辺は  $m$  の値によらず 0 になる．つまり， $p$  が十分大きいとき，経験 Bayes 推定量は確率1で最尤推定量に一致する．これは図4(b)のグラフはほぼ一定であることに合致する．

図4(c)および図4(d)は， $\mu = 0.2\pi$ ， $\mu' = 0.8\pi$  に対して  $m = m_1$  の  $m$  を 0 から  $2\pi$  まで変化させたときの経験 Bayes 推定量のリスクをグラフにしたものである．図4(a)および図4(b)に比べると少し経験 Bayes 推定量にとって有利な設定である．10次元の場合，経験 Bayes 推定量のパフォーマンスは最尤推定量に優越するときもあれば逆のときもある．しかし，100次元になるとたいていの場合で経験 Bayes 推定量のパフォーマンスが最尤推定量を優越する．劣る場合でも差はごく僅かである．

式(4.1)を用いて高次元のときの振る舞いを調べてみる．加法定理から，(4.1)の右辺において  $\cos(\mu - m) + \cos(\mu' - m) = (\cos \mu + \cos \mu') \cos m + (\sin \mu + \sin \mu') \sin m$  と変形できる．今の場合， $\mu + \mu' = \pi$  であるから，確率1で

$$(4.2) \quad \lim_{p \rightarrow \infty} \frac{I_1(\hat{\delta})}{I_0(\hat{\delta})} = \sin \mu \sin m$$

となる．したがって，高次元であり，かつ， $m = 0$  あるいは  $m = \pi$  のとき，経験 Bayes 推定量と最尤推定量の振る舞いは似ている．

図4(e)および図4(f)は， $\mu = 0.4\pi$ ， $\mu' = 0.6\pi$  に対して  $m = m_1$  の  $m$  を 0 から  $2\pi$  まで変化させたときの経験 Bayes 推定量のリスクをグラフにしたものである．図4(c)および図4(d)に比べてさらに経験 Bayes 推定に有利な設定である．また，近似的に  $\mu = 0.5\pi$  という設定と理解することもできる．10次元のときでもすでに経験 Bayes 推定のパフォーマンスはよい．高次元になると経験 Bayes 推定は最尤推定量よりよい推定量だと考えてほぼ差し支えない．

先と同様に，高次元のとき(4.2)を用いて定性的な振る舞いを理解できる．経験 Bayes 推定量の振る舞いが最もよくなるのは， $m = \pi/2$  あるいは  $m = 3\pi/2$  のときである．このとき， $\hat{\delta}$  の絶対値が大きくなっている．図4(d)と図4(f)の場合を比較すると，前者では(4.2)の右辺の値が  $\sin 0.2\pi$  であるのに対し，後者では  $\sin 0.4\pi$  となる．これは図4(d)および図4(f)のグラフにおける  $m = \pi/2$  あるいは  $m = 3\pi/2$  近辺での谷の深さに合致している．

謝 辞

本論文をまとめるにあたり，統計数理研究所の柳本武美教授に有益なコメントをいただいた．

#### 参 考 文 献

- Abramowitz, M. and Stegun, I. A. (1970) *Handbook of Mathematical Functions*, Dover, New York.  
 Bagchi, P. (1994) Empirical Bayes estimation in directional data, *Journal of Applied Statistics*, **21**, 317–326.  
 Guttorp, P. and Lockhart, R. A. (1988) Finding the location of a signal: A Bayesian analysis, *Journal of the American Statistical Association*, **83**, 322–330.  
 Mardia, K. V. and EL-Atoum, S. A. M. (1976) Bayesian inference for the von Mises-Fisher distribution, *Biometrika*, **63**, 203–206.

- 大西俊郎, 柳本武美(2005). Dual structure in the conjugate analysis of curved exponential families, *京都大学数理解析研究所講究録*, **1439**, 202–220.
- Rodrigues, J., Leite, J. G. and Milan, L. A. (2000). An empirical Bayes inference for the von Mises distribution, *Australian & New Zealand Journal of Statistics*, **42**, 433–440.

## Empirical Bayes Estimation in von Mises Distribution

Xiong Ye<sup>1</sup> and Toshio Ohnishi<sup>2</sup>

<sup>1</sup>Department of Statistical Science, The Graduate University for Advanced Studies

<sup>2</sup>The Institute of Statistical Mathematics

In the conjugate analysis of the von Mises distribution, it is known that the posterior mode is the optimal estimator, that is, the Bayes estimator under the Kullback-Leibler loss function. This paper derives an empirical Bayes estimator by estimating the dispersion hyper-parameter of the von Mises prior distribution under the assumption that the location hyper-parameter is known. The estimation of the dispersion hyper-parameter is based on the method of moment for the marginal density. We compare the risk of the empirical Bayes estimator with that of the maximum likelihood estimator (MLE) by a numerical-simulation method. In the practical situation, we find that the empirical Bayes estimator is far superior to the MLE in high-dimension cases.