Estimating the concentration parameter of a von Mises distribution: a systematic simulation benchmark

Guillaume Marrelec & Alain Giron

⊞ View supplementary material ↗

📅 Published online: 07 Apr 2022.

✎ Submit your article to this journal ↗

📊 Article views: 39

🔍 View related articles ↗

CrossMark View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Check for updates

Estimating the concentration parameter of a von Mises distribution: a systematic simulation benchmark

Guillaume Marrelec[a,b] and Alain Giron[a,b]

[a]Laboratoire d'imagerie biomédicale, LIB, Sorbonne Université, CNRS, INSERM, Paris, France; [b]Centre de recherches et d'études en sciences des interactions, CRÉSI, Center for Interaction Science, CIS, Paris, France

ABSTRACT

In directional statistics, the von Mises distribution is a key element in the analysis of circular data. While there is a general agreement regarding the estimation of its location parameter μ, several methods have been proposed to estimate the concentration parameter κ. We here provide a thorough evaluation of the behavior of 12 such estimators for datasets of size N ranging from 2 to 8192 generated with a κ ranging from 0 to 100. We provide detailed results as well as a global analysis of the results, showing that (1) for a given κ, most estimators have behaviors that are very similar for large datasets ($N \geq 16$) and more variable for small datasets, and (2) for a given estimator, results are very similar if we consider the mean absolute error for $\kappa \leq 1$ and the mean relative absolute error for $\kappa \geq 1$.

1. Introduction

In directional statistics, the von Mises distribution is a key element in the analysis of circular data. Its distribution is given by

$$f(x) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)}, \qquad x \in [0, 2\pi), \tag{1}$$

where $\mu \in [0, 2\pi)$ is the location parameter, $\kappa \geq 0$ the concentration parameter, and $I_0(\kappa)$ the incomplete Bessel integral of the first kind,

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(x)} \, dx. \tag{2}$$

The von Mises distribution vM(μ, κ) is a unimodal distribution with its mode at $x = \mu$. $\kappa = 0$ corresponds to the uniform distribution, and the larger κ, the more concentrated the distribution is around its mode. The von Mises distribution is a convenient way to model unimodal circular or directional data in many fields of science (Fisher 1995) and naturally arises in a wide variety of cases (Mardia and Jupp 2000, §3.5.4). It can be obtained by conditioning a bivariate normal distribution on the circle, is close to the wrapped normal distribution, and tends to a normal distribution when $\kappa \to \infty$. Also, it is a maximum-entropy distribution and can be characterized by the fact that the maximum likelihood of its location parameter is the sample mean direction. As such, it is an analogue on the circle of the Gaussian distribution. It also appears as the equilibrium distribution of a von Mises process.

CONTACT Guillaume Marrelec ✉ guillaume.marrelec@inserm.fr 🏛 Laboratoire d'imagerie biomédicale, 15 rue de l'École de Médecine, 75006 Paris, France.

ⓑ Supplemental data for this article is available online at https://doi.org/10.1080/03610918.2021.2011923

While there is a general agreement regarding the estimation of μ (Mardia and Jupp 2000, §12.4.1), estimation of κ has generated more research, and several methods have been proposed so far. For a partial review, see, e.g., Mardia and Jupp (2000, §12.4.2). Here, we provide a systematic benchmark of existing estimators. More specifically, we considered 12 estimators: seven frequentist estimators, including the maximum-likelihood estimator (Mardia and Jupp 2000, §5.3.1), the marginal maximum-likelihood estimator (Schou 1978), two estimators with bias correction (Best and Fisher 1981), two estimators based on the circular median (Lenth 1981; Ko 1992), and an estimator based on the normal approximation (Abeyasekera and Collett 1982); and five Bayesian estimators (Dowe et al. 1996), including three maximum a posteriori (MAP) estimators and two minimum message length (MML) estimators. We assessed the behavior of these 12 estimators for datasets of size N ranging from 2 to 8192 generated with a κ ranging from 0 to 100. We provide detailed results as well as a more global view of the estimators' performance. Our two main findings are that (1) for a given κ, most estimators have behaviors that are very similar for large datasets ($N \geq 2^4$) and much more variable for small datasets, and (2) for a given estimator, results are very similar if one considers the mean absolute error for $\kappa \leq 1$ and the mean relative absolute error for $\kappa \geq 1$.

The outline of the manuscript is the following. In Sec. 2, we provide a quick description of the 12 estimators used here. The simulation study itself is detailed in Sec. 3. Further issues are discussed in Sec. 4.

2. Overview of existing estimators

We here quickly review existing approaches, including frequentist (Sec. 2.1) and Bayesian (Sec. 2.2) estimators. All approaches start from a sample $\{x_1, ..., x_N\}$ of N independent and identically distributed (i.i.d.) realizations of a vM(μ, κ) distribution with μ and κ unknown.

2.1. Frequentist estimators

Most frequentist estimators of κ rely on the likelihood of the parameters. Letting $\bar{R}e^{im}$ be the sample circular mean of the data,

$$\bar{R}e^{im} = \frac{1}{N}\sum_{n=1}^{N} e^{ix_n}, \tag{3}$$

the likelihood can be expressed from Eq. (1) and the properties of $\bar{R}$ as

$$\ln L(\mu, \kappa) = p(D|\mu, \kappa) = [2\pi I_0(\kappa)]^{-N} \exp\left[\kappa N\bar{R}\cos\left(\mu - m\right)\right]. \tag{4}$$

The maximum-likelihood estimator $\hat{\kappa}$ is the value of κ that cancels the derivative of $\ln L(\mu, \kappa)$. Setting

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{5}$$

$\hat{\kappa}$ is the solution of (Mardia and Jupp 2000, §5.3.1)

$$A(\hat{\kappa}) = \bar{R}. \tag{6}$$

A marginal maximum-likelihood estimate $\tilde{\kappa}$ was also proposed. It is based on the expression of the density of $R = N\bar{R}$ with respect to Lebesgue measure (Watson and Williams 1956; Mardia 1975),

$$h_N(R; \kappa) = \frac{I_0(\kappa R)}{[I_0(\kappa)]^N} R h_N(R; 0). \qquad (7)$$

The maximum $\tilde{\kappa}$ of this expression depends on the relationship between $\bar{R}$ and N (Schou 1978)

- For $0 \le \bar{R} \le 1/\sqrt{N}$, $\tilde{\kappa} = 0$;
- For $1/\sqrt{N} < \bar{R} < 1$, $\tilde{\kappa}$ is the solution of

$$A(\tilde{\kappa}) = \bar{R}A(N\bar{R}\tilde{\kappa}); \qquad (8)$$

- For $\bar{R} = 1$, there is no maximum (as it would correspond to $\tilde{\kappa} = \infty$).

Using both a calculation based on a Taylor expansion around $E(\bar{R})$ of $\hat{\kappa} = A^{-1}(\bar{R})$ from Eq. (6) and a simulation study, Best and Fisher (1981) showed that the (regular) maximum-likelihood estimator $\hat{\kappa}$ of Eq. (6) can be strongly biased for small κ and N. They proposed to correct for this bias using an approximate expansion, leading to

$$\hat{\kappa}_1^* = \begin{cases} \max\left\{\hat{\kappa} - \dfrac{2}{N\hat{\kappa}}, 0\right\} & \text{for } \hat{\kappa} < 2 \\[2mm] \dfrac{(N-1)^3}{N^3 + N}\hat{\kappa} & \text{for } \hat{\kappa} \ge 2. \end{cases} \qquad (9)$$

They also proposed a jackknife correction of the bias,

$$\hat{\kappa}_2^* = \max\left\{N\hat{\kappa} - \frac{N-1}{N}\sum_{n=1}^{N}\hat{\kappa}_{-n}, 0\right\}. \qquad (10)$$

By analogy with the median absolute deviation (Rousseeuw and Croux 1993) proposed an estimator based on the median

$$\hat{\kappa}_L = \frac{0.6724}{\text{median}_n\{2[1 - \cos(x_n - \hat{\mu}_{\text{med}})]\}}, \qquad (11)$$

where $\hat{\mu}_{\text{med}}$ is the circular median (Mardia and Jupp 2000, §2.2.2). Ko (1992) improved this approach by replacing the expectation and average of Eq. (6) with median and sampling median (see also Ducharme and Milasevic 1990), yielding an estimator $\hat{\kappa}_{\text{med}}$ such that

$$\int_{-1}^{\text{median}_n[\cos(x_n - \hat{\mu}_{\text{SB}})]} \frac{1}{\pi I_0(\hat{\kappa}_{\text{med}})}\frac{e^{\hat{\kappa}_{\text{med}}t}}{\sqrt{1 - t^2}} = \frac{1}{2}, \qquad (12)$$

where $\hat{\mu}_{\text{SB}}$ is any estimator of μ that is standardized bias (SB) robust, such as the circular median or the least median square (LMS) estimator.

Finally, Abeyasekera and Collett (1982) made use of the fact that a $\text{vM}(\mu, \kappa)$ distribution is well approximated by the normal distribution $\mathcal{N}(\mu, \kappa^{-1})$ for large κ and therefore proposed as estimator of κ the unbiased estimator for σ^{-2}, leading to the so-called linear estimator for $N > 3$

$$\hat{\kappa}_{\text{lin}} = \left[\frac{1}{N-3}\sum_{n=1}^{N}(x_n - \bar{x})^2\right]^{-1}, \qquad (13)$$

where $\bar{x}$ is the linear mean, that is, the usual arithmetic mean

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N}x_n.$$

Note that there is an issue induced by the discontinuity at 0, which is dealt with by using a technical trick. Abeyasekera and Collett (1982) also proposed an improved estimator $\hat{\kappa}^*_{\text{lin}}$ where the bias is corrected by jackknifing the linear estimator.

2.2. Bayesian estimators

In a Bayesian setting, all the information that can be inferred about the parameters μ and κ from a dataset D is summarized in the posterior distribution $\mathrm{p}(\mu, \kappa|D)$, which, according to Bayes' theorem, can be expressed as

$$\mathrm{p}(\mu, \kappa|D) \propto \mathrm{p}(\mu, \kappa) \ \mathrm{p}(D|\mu, \kappa), \tag{14}$$

where "$\propto$" relates two expressions that are proportional. $\mathrm{p}(D|\mu, \kappa)$ is the data likelihood, already expressed in Eq. (4). $\mathrm{p}(\mu, \kappa)$ is the prior distribution, which translates the information about the parameters that is available independently of the data. A common approach is to assume no prior dependence between parameters, so that $\mathrm{p}(\mu, \kappa) = \mathrm{p}(\mu) \ \mathrm{p}(\kappa)$. $\mathrm{p}(\mu)$ is classically set as a noninformative uniform distribution on the circle, $\mathrm{p}(\mu) = 1/2\pi$. As for $\mathrm{p}(\kappa)$, Dowe et al. (1996) considered the following two prior distributions

$$h_2(\kappa) = \frac{2}{\pi(1 + \kappa^2)} \qquad \text{and} \qquad h_3(\kappa) = \frac{\kappa}{(1 + \kappa^2)^{\frac{3}{2}}}, \qquad \kappa \geq 0. \tag{15}$$

They calculated the posterior distribution $\mathrm{p}(\mu, \kappa|D)$

$$\mathrm{p}(\mu, \kappa|D) \propto h_i(\kappa) \ [2\pi I_0(\kappa)]^{-N} \exp\left[\kappa N \bar{R} \cos(\mu - m)\right], \tag{16}$$

and proposed to use the maximum a posteriori (MAP) estimators

$$\kappa^{\kappa, \mu}_{\mathrm{MAP}, i} = \arg_\kappa \max_{\mu, \kappa} \mathrm{p}(\mu, \kappa|D) = \arg_{\kappa \geq 0} \max\left\{\ln h_i(\kappa) - N \ln[2\pi I_0(\kappa)] + N\bar{R}\kappa\right\}, \qquad i = 2, 3. \tag{17}$$

For $h_3(\kappa)$, they also proposed to consider estimating κ as the MAP of the distribution of $[x = \kappa \cos(\mu), y = \kappa \sin(\mu)]$, leading to

$$\kappa^{x, y}_{\mathrm{MAP}, 3} = \max_{\kappa \geq 0}\left\{\ln\left[\frac{h_3(\kappa)}{\kappa}\right] - \ln[2\pi I_0(\kappa)] + N\bar{R}\kappa\right\}. \tag{18}$$

Finally, they proposed two minimum message length (MML) estimators. Minimizing the message length is equivalent to maximizing

$$\frac{\mathrm{p}(\mu, \kappa|D)}{\sqrt{\det F(\mu, \kappa)}}, \tag{19}$$

where $\det F(\mu, \kappa)$ is the determinant of the Fisher matrix for the von Mises distribution. The resulting estimators for $h_2(\kappa)$ and $h_2(\kappa)$ are

$$\kappa_{\mathrm{MML}, 2} = \arg\max_{\kappa \geq 0} \frac{h_2(\kappa) \ \mathrm{p}(\boldsymbol{x}|\mu, \kappa)}{\sqrt{\left[\kappa A(\kappa) + \frac{3}{N\pi^2}\right]A'(\kappa)}} \tag{20}$$

and

$$\kappa_{\mathrm{MML}, 3} = \arg\max_{\kappa \geq 0} \frac{h_3(\kappa) \ \mathrm{p}(\boldsymbol{x}|\mu, \kappa)}{N\sqrt{\kappa A(\kappa)A'(\kappa)}}, \tag{21}$$

respectively.

Table 1. Existing simulation studies.

Article	κ	N	M	Estimators	Measures of fit
Best and Fisher (1981)	$\{0.1, 0.5, 1, 2.5, 5\}$	$\{10, 20, 100\}$	≥ 1000	$\hat{\kappa}$	Qualitative on sampling features
	$\{0.1, 0.5, 1, 2.5, 5, 10, 100, 500\}$	$\{10, 20\}$	≥ 1000	$\hat{\kappa}_1^*, \hat{\kappa}_2^*$	Qualitative on sampling features
	$\{0.1, 0.5, 1, 2.5, 5\}$	100	≥ 1000	$\hat{\kappa}_1^*, \hat{\kappa}_2^*$	Qualitative on sampling features
	$\{0.01, 0.2, 0.3, 0.4, 0.75\}$	$\{10, 20\}$	≥ 1000	$\hat{\kappa}_1^*, \hat{\kappa}_2^*$	Qualitative on sampling features
	$\{0.01, 0.2, 0.3\}$	100	≥ 1000	$\hat{\kappa}, \hat{\kappa}_1^*, \hat{\kappa}_2^*$	Qualitative on sampling features
Abeyasekera and Collett (1982)	$\{0.1, 0.5, 1, 2, 3, 4, 5, 7.5\}$	$\{10, 20, 30, 100\}$	1000	$\hat{\kappa}, \tilde{\kappa}, \hat{\kappa}_2^*, \tilde{\kappa}^*$	Sample bias and MSE
Dowe et al. (1996)	$\{0, 0.5, 1, 10\}$	$\{2, 5\}$	1000	$\hat{\kappa}, \tilde{\kappa}, \kappa_{MAP,2}, \kappa_{MAP,3}$	MAE, MSE, MKL
	$\{0, 1, 10\}$	$\{25, 100\}$	1000	$\hat{\kappa}, \tilde{\kappa}, \kappa_{MAP,2}, \kappa_{MAP,3}$	MAE, MSE, MKL
Our simulation	$\{0, 0.01, 0.1, 1, 10, 100\}$	$\{2, 4, 8, 16, ..., 8192\}$	1000	see Table 2	MAE, MRAE

Summary of scope and methodology. N is the sample size and M the number of samples used to compute mean errors. MAE: mean absolute error; MSE: mean squared error; MKL: mean Kullback–Leibler divergence; MRAE: mean relative absolute error.

2.3. Existing simulation studies

We found three existing simulations studies assessing the relative performance of estimators of the concentration parameter: Best and Fisher (1981), Abeyasekera and Collett (1982), and Dowe et al. (1996). Their scope and methodology are summarized in Table 1. Best and Fisher (1981), found that $\hat{\kappa}_1^*$ and $\hat{\kappa}_2^*$ were similar and better than $\hat{\kappa}$ for small values of κ and N. $\hat{\kappa}_1^*$ was less biased than $\hat{\kappa}$ and $\hat{\kappa}_2^*$ was less median biased. However, the bias remained for small κ and N. Abeyasekera and Collett (1982) found that $\hat{\kappa}_2^*$ was better than $\hat{\kappa}$, $\tilde{\kappa}$ and $\tilde{\kappa}^*$; that $\tilde{\kappa}$ was better than $\hat{\kappa}$ for small values of N and worse when N was large; and, finally, that $\tilde{\kappa}$ and $\tilde{\kappa}^*$ had similar behaviors. Finally, Dowe et al. (1996) found that Bayesian methods outperformed frequentist methods, with a limited effect of the prior distribution on the behavior.

3. Simulation study

3.1. Data

We generated samples from von Mises distributions. More specifically, for each of the $Q = 6$ values of $\kappa \in \{0, 10^{-2}, 10^{-1}, 1, 10, 10^2\}$, we generated $M = 1000$ "maximal" datasets. Each dataset was composed of $N_{\max}$ i.i.d. realizations of a $vM(\mu, \kappa)$ distribution with μ uniformly distributed on the circle and $N_{\max} = 2^L$ with $L = 13$ (i.e., $N_{\max} = 8192$). Each sample was generated using Berens (2009)'s CircStat. This procedure therefore generated a total of $Q \times M = 6000$ maximal datasets of size 8192. From each maximal dataset, we then extracted L datasets composed of the 2^l first data points with $l \in \{1, ..., L\}$. This gave us a total of $Q \times M \times L = 78,000$ datasets on which inference was performed. On each dataset, estimation of κ was performed by application of each of the $J = 12$ estimators mentioned above (see also Table 2 for a summary). We therefore ended with a set of $Q \times M \times L \times J = 936,000$ estimates.

3.2. Evaluation

To assess the behavior of the various estimations, we proceeded as follows. For estimation κ_{jlmq}, obtained by application of estimator j to the 2^l first data points of simulation m associated with the qth value of κ, we first considered the absolute error of the estimation compared to the true value

$$\epsilon_{jlmq} = |\hat{\kappa}_{jlmq} - \kappa_q|. \tag{22}$$

Table 2. Simulation study: Summary of estimators used.

Estimator	Equation	Identification	Computation time (mean ± std dev)
$\hat{\kappa}$	(6)	jML	23.9 ± 5.6 ms
$\tilde{\kappa}$	(8)	mML	23.9 ± 5.6 ms
$\hat{\kappa}_1^*$	(9)	BF1	46.9 ± 8.5 ms
$\hat{\kappa}_2^*$	(10)	BF2	see Figure 1
$\hat{\kappa}_L$	(11)	median1	see Figure 1
$\hat{\kappa}_{\mathrm{med}}$	(12)	median2	see Figure 1
$\hat{\kappa}_{\mathrm{lin}}$	(13)	Linear	24.0 ± 5.6 ms
$\hat{\kappa}_{\mathrm{MML},2}$	(20)	MML-2	24.4 ± 5.7 ms
$\hat{\kappa}_{\mathrm{MML},3}$	(21)	MML-3	24.5 ± 5.7 ms
$\hat{\kappa}_{\mathrm{MAP},2}^{\kappa,\mu}$	(17)	BayesEst-2-jMAP-km	24.0 ± 5.5 ms
$\hat{\kappa}_{\mathrm{MAP},3}^{\kappa,\mu}$	(17)	BayesEst-3-jMAP-km	24.0 ± 5.4 ms
$\hat{\kappa}_{\mathrm{MAP},3}^{x,y}$	(18)	BayesEst-3-jMAP-xy	24.0 ± 5.3 ms

We then summarized the results over the M simulations by computing the mean absolute error (MAE), defined as a sampling average

$$\bar{\epsilon}_{jlq} = \frac{1}{M} \sum_{m=1}^{M} \epsilon_{jlmq}. \tag{23}$$

In the same fashion, we also considered the relative error, where the above error of Eq. (22) is normalized by the true value

$$\epsilon'_{jlmq} = \frac{\left| \hat{\kappa}_{jlmq} - \kappa_q \right|}{\kappa_q}, \tag{24}$$

and the mean relative absolute error (MRAE) as its sampling average over the M simulations

$$\overline{\epsilon'}_{jlq} = \frac{1}{M} \sum_{m=1}^{M} \epsilon'_{jlmq}. \tag{25}$$

3.3. Results

Detailed results can be found in the supplemental material, in the form of summary statistics (Tables 1–12, supplemental material) and graphs (Figures 1–6, supplemental material). We here focus on some key features of the behavior: computation time, algorithmic failures, the global trend as a function of N, the behavior for large samples ($N \geq 2^4$), and the small-sample behavior.

3.3.1. Computation time

The detailed computation times can be found in the supplemental material, Tables 1–12, columns 3 and 4. For all estimators, the actual value of κ seemed to have very limited influence on the computation time. Similarly, for most estimators (jML, mML, BF1, linear, MML-3, BayesEst-2-jMAP-km, BayesEst-3-jMAP-km, and BayesEst-3-jMAP-xy), the computation time was not a function of the sampling size; it was of the order of 40 ms for BF1 and of 20 ms for the others (see Table 2). Three estimators had computation times that were a function of data size: BF2, median1, and median2 (see Figure 1). For BF2, the computation time was roughly linear, with a doubling of the data size corresponding to a doubling of the computation time. For median1 and median2, the computation time was roughly constant up to $N = 2^7$ and then increased exponentially.
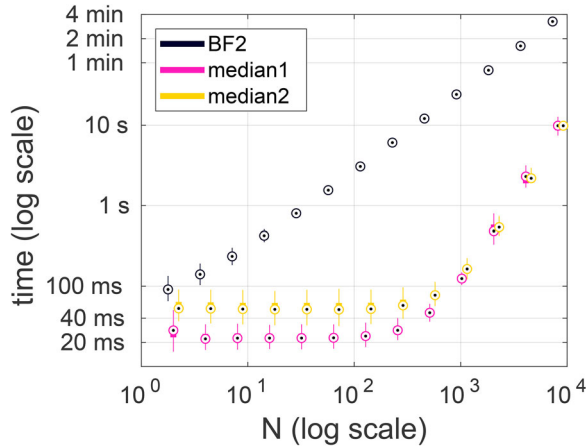
Figure 1. Simulation study. Computation time for estimators that were influenced by sample size. Boxplots (median and $[25\%, 75\%]$ percentile interval) across values of κ and simulations.

3.3.2. Failures

We first report cases where methods failed to compute an estimate (see Table 3). This happened for median2 (corresponding to $\hat{\kappa}_{med}$) and linear (corresponding to $\hat{\kappa}_{lin}$). By definition, $\hat{\kappa}_{lin}$ of Eq. (13) is not defined for $N = 2$. Also, the second median estimator, $\hat{\kappa}_{med}$, was not always defined, as there were cases for which the integral of Eq. (12) was smaller than 1/2 for all $\kappa \geq 0$ (and, in particular for $\kappa = 0$). Cases corresponding to failures were removed to compute the summary statistics $\bar{\epsilon}$ and $\overline{\epsilon'}$.

3.3.3. Evolution as a function of N

Graphs corresponding to $\bar{\epsilon}$ and $\overline{\epsilon'}$ for all estimators gathered by value of κ can be found in Figure 2. Globally, most estimators improved as the sample size increased. One clear exception was BF2 for small N and low τ. In that particular case, the method tended to estimate τ as equal to 0, leading to an error of the order of τ, with a correction toward a more normal trend as the data size increased. This behavior was also observed to a much lesser extend for the maximum message length estimators (MML-2 and MML-3) and the Bayesian estimators (BayesEst-2-jMAP-km, BayesEst-3-jMAP-km, and BayesEst-3-jMAP-xy).

3.3.4. Large-sample behavior

We observed quite homogeneous results for large datasets ($N \geq 2^4$), both within- and between estimators. First, the mean absolute error roughly decreased linearly in log-log coordinates for most estimators. The two exceptions were median1 for low τ, where the estimator seemed to reach a plateau independently of N, and linear, for which data samples of increasing size did not seem to improve estimation. Also, the performance of a given method for different values of κ were quite similar when one considered the mean absolute error $\bar{\epsilon}$ for $\kappa \leq 1$ and the mean relative absolute error $\overline{\epsilon'}$ for $\kappa \geq 1$. The main exception was median1 due to the difference in behavior for low values versus high values of τ mentioned earlier in the paragraph. For the other estimators, the slope of $\log_{10}(\bar{\epsilon})$ as a function of $\log_{10}(N)$ was found to be around $-1/2$, corresponding to a decrease of mean (relative) absolute error in $1/\sqrt{N}$. Finally, most estimators behaved quite similarly in this large-sample regime, at the exception of median1, linear and, to a lesser extend, median2.

To quantify these large-sample behaviors, we fitted to each curve a model of linear regression, either

Table 3. Simulation study: Summary of number of failures encountered for all τ and N.

τ	N	median2	linear
0	2	0	1000
	4	73	0
	8	109	0
	16	165	0
	32	175	0
	64	194	0
	128	201	0
	256	200	0
	512	211	0
	1024	214	0
	2048	193	0
	4096	241	0
	8192	229	0
0.01	2	0	1000
	4	71	0
	8	105	0
	16	150	0
	32	157	0
	64	189	0
	128	185	0
	256	193	0
	512	194	0
	1024	211	0
	2048	216	0
	4096	232	0
	8192	249	0
0.1	2	0	1000
	4	51	0
	8	114	0
	16	142	0
	32	138	0
	64	179	0
	128	187	0
	256	187	0
	512	203	0
	1024	193	0
	2048	205	0
	4096	210	0
	8192	231	0
1	2	0	1000
	4	29	0
	8	34	0
	16	54	0
	32	46	0
	64	42	0
	128	52	0
	256	57	0
	512	61	0
	1024	67	0
	2048	58	0
	4096	53	0
	8192	57	0
10	2	0	1000
100	2	0	1000

No failure was encountered for jML, mML, BF1, BF2, median1, MML-2, MML-3, BayesEst-2-km, BayesEst-3-km, and BayesEst-3-xy.
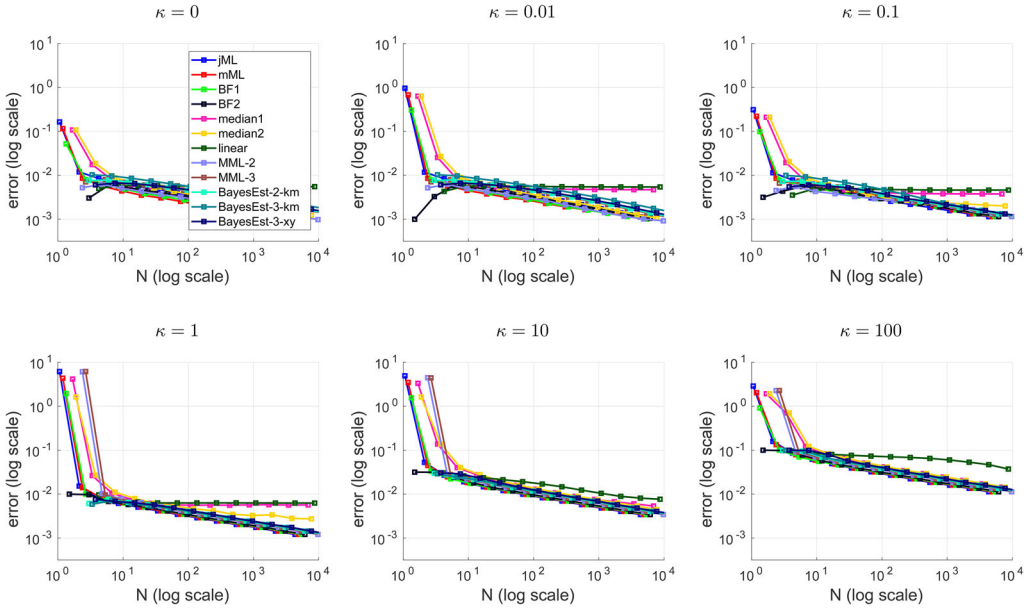
Figure 2. Simulation study. Average error $\bar{\epsilon}$ for all values of κ and all estimators.

$$\log_{10}(\bar{\epsilon}) = \alpha \log_{10}(N) + \beta + \eta, \qquad N \geq 2^4, \tag{26}$$

for $\kappa \leq 1$, or

$$\log_{10}(\bar{\epsilon'}) = \alpha \log_{10}(N) + \beta + \eta, \qquad N \geq 2^4, \tag{27}$$

for $\kappa \geq 1$. The results are summarized in Figure 3, confirming generally homogeneous behaviors (at the exception of median1, median2 and linear) that tended to depend more on the value of κ than on the specifics of the estimation method.

3.3.5. Small sample behavior

Behaviors were much more variable for small sample sizes, with both a variability between methods and, for a given method, between values of κ. A majority of estimators (jML, mML, BF1, median1, median2, MML-2, MML-3) tended to have very large average errors for very small sample sizes ($N \in \{2, 4\}$). It was found that these extremely large values were due to a few very large errors. For instance, in the case of MML-2, $\tau = 1$ and $N = 2$, all but one of the 1000 repetitions had errors smaller than 1, the remaining error being equal to 3.72×10^8.

To quantify this behavior, we modeled deviation from the linear trend as an exponential decay, either

$$\log_{10}(\bar{\epsilon}) = \hat{\alpha} \log_{10}(N) + \hat{\beta} + \gamma \left(\frac{1}{10}\right)^{\frac{\log_{10}(N) - \log_{10}(2)}{\tau}} + \eta \tag{28}$$
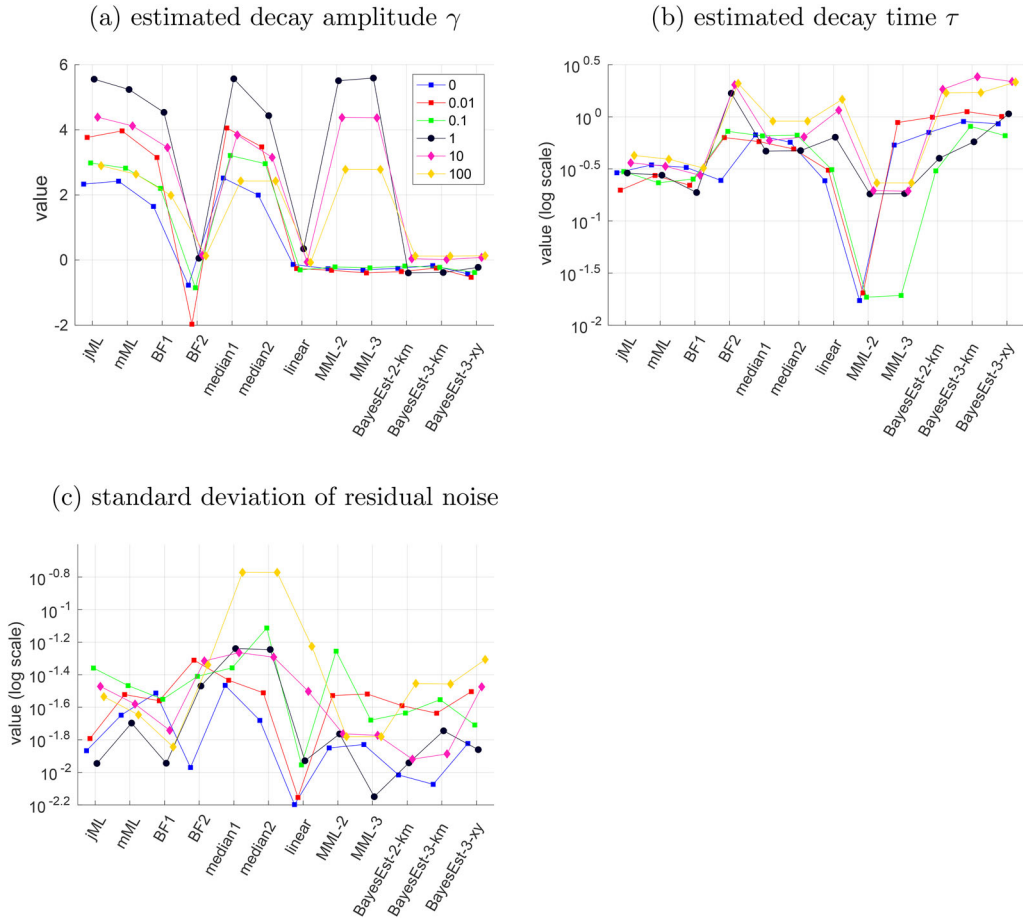
for $\kappa \leq 1$ or

$$\log_{10}(\bar{\epsilon'}) = \hat{\alpha} \log_{10}(N) + \hat{\beta} + \gamma \left(\frac{1}{10}\right)^{\frac{\log_{10}(N) - \log_{10}(2)}{\tau}} + \eta \tag{29}$$

for $\kappa \geq 1$. γ is the maximum departure of the mean (relative) absolute error from the linear trend and corresponds to the increase in mean (relative) absolute error at $N = 2$. A positive value indicates a mean (relative) absolute error that is larger than the linear expectation (and therefore a
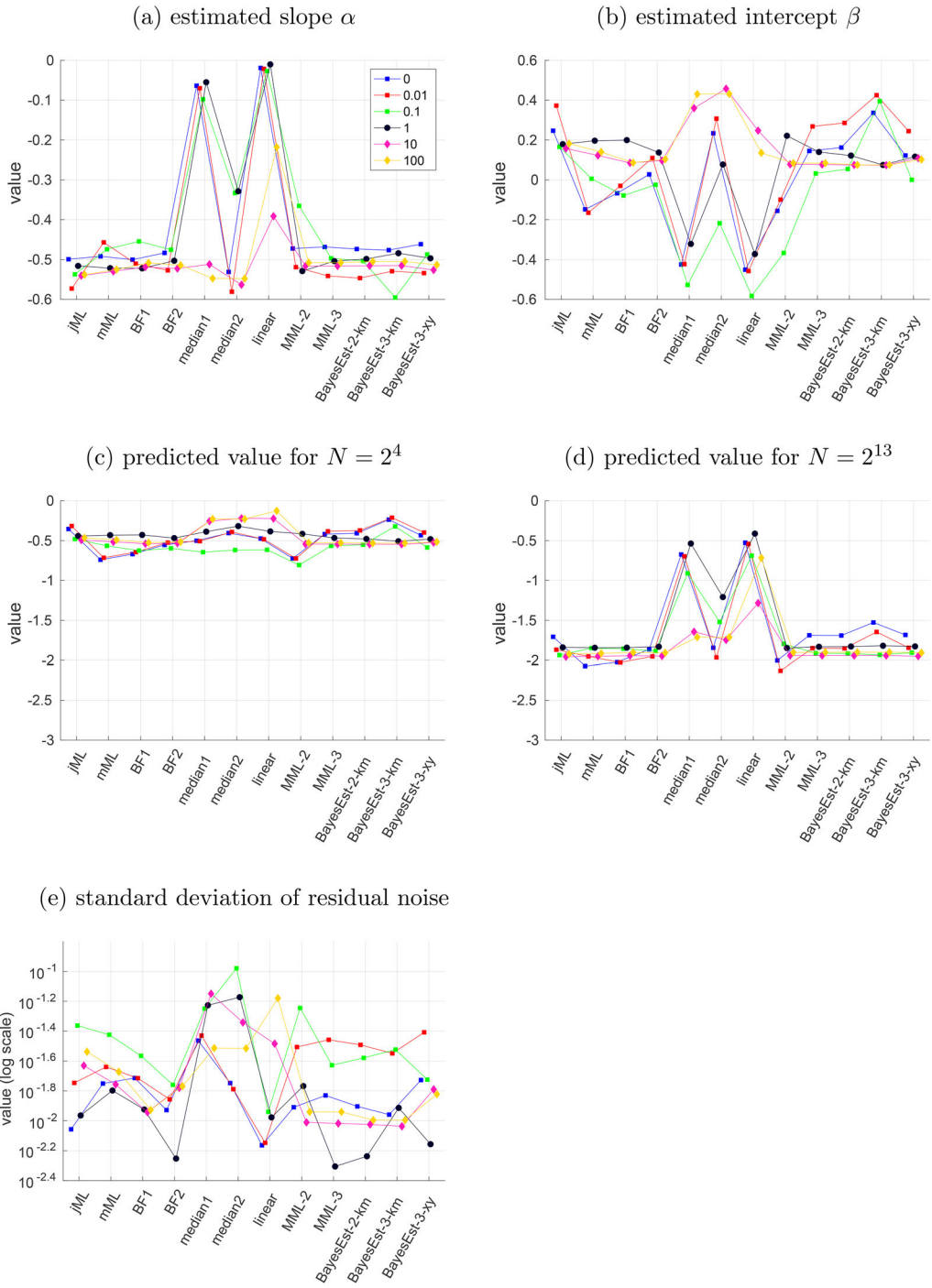
Figure 3. Simulation study. Result of linear regression analysis ($N \geq 2^4$) for either Eq. (26) (for $\tau \leq 1$) or Eq. (27) (for $\tau \geq 1$). (a) Estimated slope α. (b) Estimated intercept β. (c) Predicted values of $\log_{10}(\bar{\epsilon})$ (for $\kappa \leq 1$) or $\log_{10}(\bar{\epsilon}')$ (for $\kappa \geq 1$) at $N = 2^4$. (d) Predicted values of $\log_{10}(\bar{\epsilon})$ (for $\kappa \leq 1$) or $\log_{10}(\bar{\epsilon}')$ (for $\kappa \geq 1$) at $N = 2^{13}$. (e) Standard deviation of residual noise.

(a) estimated decay amplitude γ

(b) estimated decay time τ

(c) standard deviation of residual noise

Figure 4. Simulation study. Result of nonlinear regression analysis for either Eq. (28) (for $\tau \leq 1$) or Eq. (29) (for $\tau \geq 1$). (a) Estimated decay amplitude γ. (b) Estimated decay time τ. (c) Standard deviation of residual noise (bottom left).

decreased performance), while a negative value indicates a smaller mean (relative) absolute error (and therefore an increased performance). τ is such that the departure from the linear trend is divided by 10 for $\log_{10}(N) = \log_{10}(2) + \tau$. The larger the value of τ, the slower the decay. The results are summarized in Figure 4.

Regarding γ, we found several groups of estimators. A first group (jML, mML, BF1, median1, median2) had large positive values of γ regardless of κ. A second group (MML-1 and MML-2) had large positive values of γ only for $\kappa \geq 1$, and small negative values for $\kappa < 1$. A third group (linear, BayesEst-2-km, BayesEst-3-km, BayesEst-3-xy) had low (positive and negative) values of γ for all values of κ. A last estimator (mML) had small positive values of γ for $\kappa \geq 1$ and negative, small yet larger values of γ for $\kappa < 1$.

Regarding τ, all values were found to be in the range $[0.01, 3.2]$. All estimators with large values of γ (corresponding to the first two groups mentioned earlier) had values in the range $[0.01, 1]$, indicating a very fast exponential decay—for practical purposes, it mostly affected datasets of size $N = 2$.

4. Discussion

In the present manuscript, we assessed the behavior of 12 estimators of κ, the concentration parameter of a von Mises distribution. We used synthetic datasets of size N ranging from 2 to 8192

generated with a κ ranging from 0 to 100. We provided detailed results as well as a more global view of the estimators' performance. We found that, for a given estimator, results were very similar across values of κ if we considered the mean absolute error for $\kappa \leq 1$ and the mean relative absolute error for $\kappa \geq 1$. We also found that, for a given κ, most estimators had similar behaviors for large datasets ($N \geq 2^4$), while the behaviors differed more strongly for small datasets.

The fact that the mean absolute error for $\kappa \leq 1$ was similar to the mean relative absolute error for $\kappa \geq 1$ has two consequences. It first shows that, for $\kappa \geq 1$, error increases linearly as a function of τ. Larger values of τ are therefore expected to lead to larger estimation errors. While it also means that smaller values of τ tend to yield smaller estimation errors, this behavior cannot be used to our advantage for $\kappa \rightarrow 0$, where it would mean that error vanishes as $\kappa \rightarrow 0$. By contrast, it is the mean absolute error that remains stable for $\kappa \leq 1$, i.e., error reaches a plateau when κ decreases below $\kappa = 1$.

For large datasets, the common behavior of most estimators was found to be that the mean (relative) absolute error roughly decreased linearly in log-log coordinates, with a slope around $-1/2$, corresponding to a decrease of mean (relative) absolute error in $1/\sqrt{N}$. This is in line with the general estimation theory, where errors are often found to decrease as $1/\sqrt{N}$.

In the light of our results, we are able to give general practical guidelines regarding the use of the estimators tested here. First, we would not recommend the use of median1, median2 and linear unless there is evidence that they might perform well in the specific context of interest. For large datasets ($N \geq 2^4$), all other estimators perform in a similar fashion and there is no obvious reason to recommend one or the other. In this context, BF2's computational burden is a disadvantage with no compensation in terms of performance. By contrast, for small datasets, BF2 was found to be the estimator with the best behavior, still at the price of a high computational cost as well as a consistent underestimation of the concentration parameter when it was small but different from 0. For instance, for $N = 2$, it estimated κ as 0 regardless of its actual value (Table 5, supplemental material). The Bayesian estimators, which only performed slightly worse than BF2 but had computational costs similar to the other estimators, could provide valuable tradeoffs.

Importantly, all estimators considered here require a sample $\{x_1, ..., x_N\}$ of N independent and identically distributed (i.i.d.) realizations of a vM(μ, κ) distribution with μ and κ unknown. Our simulation study faithfully respected this requirement. As a consequence, it did not explore the behavior of estimators with respect to the presence of outliers or model misspecification. Another consequence of the fact that we did not depart from the von Mises model is that we did not consider estimators assuming that one or several observations are identified as outliers (e.g., Winsorized estimate of Fisher 1982). Also, some of the estimators proposed in the literature depend on parameters (see, e.g., Lenth 1981; Kato and Eguchi 2016). Since the behavior of these estimators strongly depends on the choice of the parameters, they were not incorporated in the analysis.

As mentioned earlier, there were cases where some of the estimators were not defined. This happened for the linear estimator and $N = 2$, since this estimator requires at least 4 data points. The median-based estimator $\hat{\kappa}_{\mathrm{med}}$ also failed sometimes, in particular for small values of κ, as Equation (12) had no solution (the integral was smaller than 0.5 for all values of $\kappa \geq 0$). In such cases, we could have decided to set the estimator to 0. However, we believe that such problem is symptomatic of a deeper problem whose investigation goes beyond the scope of the present manuscript. So we just took the estimator as is.

As a final note, the routines calculating the various estimators were written by us and were not optimized. As such, the computation times presented here should only serve as a rough indicator of the time required to apply each of them. In particular, it is possible that optimal coding of BF2 with a compiled language could greatly improve its computation time.

References

Abeyasekera, S., and D. Collett. 1982. On the estimation of the parameters of the von Mises distribution. *Communications in Statistics - Theory and Methods* 11 (18):2083–90. doi:10.1080/03610928208828371.

Berens, P. 2009. CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software* 31 (10):21. doi:10.18637/jss.v031.i10.

Best, D. J., and N. I. Fisher. 1981. The bias of the von Mises–Fisher concentration parameters. *Communications in Statistics - Simulation and Computation* 10 (5):493–502. doi:10.1080/03610918108812225.

Dowe, D. L., J. J. Oliver, R. A. Baxter, and C. S. Wallace. 1996. Bayesian estimation of the von Mises concentration parameter. In *Maximum entropy and Bayesian methods. Fundamental theories of physics, Vol. 79: An international book series on the fundamental theories of physics: Their clarification, development and application*, ed. K. M. Hanson and R. N. Silver, 51–60. Dordrecht: Springer.

Ducharme, G. R., and P. Milasevic. 1990. Estimating the concentration of the Langevin distribution. *Canadian Journal of Statistics* 18 (2):163–9. doi:10.2307/3315565.

Fisher, N. I. 1982. Robust estimation of the concentration parameter of Fisher's distribution on the sphere. *Applied Statistics* 31 (2):152–4. doi:10.2307/2347979.

Fisher, N. I. 1995. *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

Kato, S., and S. Eguchi. 2016. Robust estimation of location and concentration parameters for the von Mises–Fisher distribution. *Statistical Papers* 57 (1):205–34. doi:10.1007/s00362-014-0648-9.

Ko, D. 1992. Robust estimation of the concentration parameter of the von Mises–fisher distribution. *The Annals of Statistics* 20 (2):917–28. doi:10.1214/aos/1176348663.

Lenth, R. V. 1981. Robust measure of location for directional data. *Technometrics* 23 (1):77–81. doi:10.1080/00401706.1981.10486240.

Mardia, K. V. 1975. Distribution theory for the von Mises-Fisher distribution and its application. In *A modern course on statistical distributions in scientific work. Vol. 1 – Models and structures, Vol. 17 of NATO Advanced study institutes series. Series C: Mathematical and physical sciences*, ed. G. P. Patil, S. Kotz, and J. K. Ord, 113–30. Dordrecht, Holland: D. Reidel Publishing Company.

Mardia, K. V., and P. E. Jupp. 2000. *Directional statistics. Wiley series in probability and statistics*. Chichester: Wiley.

Rousseeuw, P. J., and C. Croux. 1993. Alternatives to the median absolute deviation. *Journal of the American Statistical Association* 88 (424):1273–83. doi:10.1080/01621459.1993.10476408.

Schou, G. 1978. Estimation of the concentration parameter in von Mises-Fisher distributions. *Biometrika* 65 (2):369–77. doi:10.2307/2335217.

Watson, G. S., and E. J. Williams. 1956. On the construction of significance tests on the circle and the sphere. *Biometrika* 43 (3–4):344–52. doi:10.2307/2332913.