

A Preliminary Test Estimator in Circular Statistics

The von Mises distribution plays a pivotal role in circular data analysis. This paper addresses the question as to how much better one can do in estimating its concentration parameter if partial information is available on the mean direction, either (i) through a prior or (ii) through a “pre-test”. These two alternate scenarios and the resulting estimators are compared with the standard maximum likelihood estimator, and we explore when one estimator is superior to the other.

Keywords: Preliminary test estimator; Circular statistics; Circular normal distribution; Concentration parameter; Maximum likelihood estimates; Bayesian; Mean square error; Nuisance parameters

Introduction

Saleh AME [1], provides an introduction and thorough review on PTEs and Stein-type estimators for various linear models. In statistical inference, the use of prior information on other parameters in a statistical model, usually leads to improved inference on the parameter of interest. Prior information may be (i) known and deterministic which is then incorporated into the model in the form of constraints on the parameter space, leading to a restricted model, or (ii) uncertain and specified in the form of a prior distribution or a verified null hypothesis. In (ii), choosing certain restricted estimators may be justified when the prior information can be quantified i.e. comes with known confidence levels.

In some statistical models, certain parameters are of primary interest while other parameters may be considered as nuisance parameters. One procedure to mitigate the presence of nuisance parameters is to assess what value(s) such nuisance parameter(s) take, by a preliminary test with a null hypothesis restricting the nuisance parameter values. The null hypothesized value(s) of the nuisance parameter are either used or not, depending on whether the observed preliminary test statistic falls in the acceptance or rejection region of the hypothesis. That is, our final estimator for the parameter of interest is thus a linear combination, conditional on whether the preliminary test statistic is in the acceptance or rejection region of the test, and is called a Preliminary Test Estimator (PTE).

[2,3], and [4] were among the first to implement the idea of preliminary test estimation (PTE) in an analysis of variance (ANOVA) framework to analyze the effect of the preliminary test on the estimation of variance. The idea goes back to a suggestion in [5], which considers testing differences between two means after testing for the equality of variances; then using the usual t-test with the pooled estimate for variance, if the variance test shows equality; otherwise, it falls into the category of Behrens-Fisher problem. In these problems it became clear that the performance of the PTE depended heavily on the significance level of the preliminary test. Han c [6] were the first to attempt to find an optimum size of significance level for the preliminary test for this two-sample problem.

Research Article

Volume 6 Issue 3 - 2017

Michael M. Nava and Sreenivasa R. Jammalamadaka*

Dept. of Statistics & Applied Probability, University of California, USA

***Corresponding author:** S Rao Jammalamadaka, Department of Statistics and Applied Probability, University of California, USA, Email: rao@pstat.ucsb.edu

Received: July 26, 2017 | **Published:** August 30, 2017

All Stein-type estimators [7,8] involve appropriate test statistics for testing the adequacy of uncertain prior information on the parameter space, which is incorporated into the actual formulation of the estimator. Stein-type estimators adjust the unrestricted estimator by an amount of the difference between unrestricted and restricted estimators scaled by the adjusted test statistics for the uncertain prior information. Usually, the test statistics are the normalized distance between the unrestricted and restricted estimators and follow a noncentral chi-square or an F -distribution. The risk or the MSE of Stein-type estimators depends on the non-centrality parameter, which represents the distance between the full model and restricted model. The PTE may be considered a precursor of the Stein-type estimator. A simple replacement of the indicator function that we will see in the PTE with a multiple of the test statistic, leads to a Stein-type estimator.

Other Estimators

The CND is the most widely used circular distribution in circular statistics. It plays as central role as the Normal distribution does in usual ‘linear’ statistics. The probability density for a CND denoted by $CND(\mu, \kappa)$, is:

$$\frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\alpha - \mu)), 0 \leq \alpha < 2\pi \quad (1)$$

The mean direction is also referred to as the preferred direction and the concentration parameter can be thought of as the inverse of variance as it is a measure of concentration around the mean direction. A larger value for implies that observations are more concentrated around the mean direction, while a value of close to 0 implies there may not be a strongly preferred direction. We consider now the maximum likelihood estimates (MLEs) for the parameter in a classical and Bayesian setting.

Maximum Likelihood Estimate for Concentration Parameter

Given a random sample $\alpha_1, \dots, \alpha_n$ from a $CND(\mu, \kappa)$, the MLE

for κ when μ in unknown is given by Jammalamadaka SR [9]:

$$\hat{\kappa}_{MLE} \text{ is the solution to: } \frac{1}{n} \sum_{i=1}^n \cos(\alpha_i - \bar{\alpha}_0) = \frac{I_1(\kappa)}{I_0(\kappa)} \quad (2)$$

When the mean direction μ is known, then the MLE for κ is obtained by substituting μ in place of $\bar{\alpha}_0$ in 2. Since the estimation of concentration parameter is of main interest here, we will denote $\hat{\kappa}_{MLE}$ and $\hat{\kappa}_\mu$ the MLEs for κ when sample mean direction is used (if μ unknown), and when the mean direction μ is known, respectively. In both cases the MLEs carry the usual asymptotic properties. Analogous to the case of a linear Normal distribution, $\hat{\kappa}_\mu$ is superior (has smaller MSE) than $\hat{\kappa}_{MLE}$, [9].

If the sample comes from a population with population mean μ then the $MSE(\hat{\kappa}_\mu) < MSE(\hat{\kappa}_{MLE})$ [9]:

$$\frac{1}{n(1-\bar{V}_0/\hat{\kappa}_\mu - \bar{V}^2)} \leq \frac{1}{n(1-\bar{R}/\hat{\kappa}_{MLE} - \bar{R}^2)} \quad (3)$$

where we have inequality if and only if $\mu = \bar{\alpha}_0$. We denote $\bar{R} = R/n$ and $\bar{V}_0 = V_0/n$, where $V_0 = \sum_{i=1}^n \cos(\alpha_i - \mu)$, $R = \sum_{i=1}^n \cos(\alpha_i - \bar{\alpha}_0)$, and $V_0 \leq R$.

This raises the question whether we can go somewhere in between if we have partial information on μ .

MLE for κ when there is a prior on μ

In this semi-Bayesian setting we will place a prior on the nuisance mean direction μ , a convenient choice being a CN:

$$\pi(\mu) = \frac{1}{2\pi I_0(\tau)} \exp(\tau \cos(\mu - \mu_0)), \quad 0 \leq \mu < 2\pi \quad (4)$$

where μ_0 and τ are the mean direction and concentration parameters for the prior. The value for τ measures confidence in the prior mean direction μ_0 . A larger value of τ makes the prior distribution have higher concentration around μ_0 . A value of $\tau = 0$ implies a uniform prior on $[0, 2\pi)$ for μ .

In this context, the parameter μ has a prior distribution, while the parameter κ is an unknown parameter as in the classical setting. The parameter κ is of interest, while μ is the nuisance parameter. We thus blend together classical and Bayesian methods to get an estimate for κ .

We begin with the usual likelihood given the data $(\alpha_1, \dots, \alpha_n)$ independent and identically distributed:

$$L(\mu, \kappa | \alpha) = \left(\frac{1}{2\pi I_0(\kappa)} \right)^n \exp\left(\kappa \sum_{i=1}^n \cos(\alpha_i - \mu) \right), \quad 0 \leq \alpha_i < 2\pi \quad (5)$$

Given the prior distribution on μ , we wish to estimate the concentration parameter κ . We derive the likelihood function for κ by first averaging out our prior knowledge on μ . The result is the likelihood for κ given by:

$$\int_0^{2\pi} L(\mu, \kappa | \alpha_1, \dots, \alpha_n) \pi(\mu) d\mu = L(\kappa | \alpha_1, \dots, \alpha_n) \quad (6)$$

In 6, we begin with joint likelihood for the μ and κ which is just the joint density of the data. We then derive marginal distribution for the observations by integrating with respect to μ . After incorporating our prior knowledge on μ and integrating with respect to μ , we obtain a valid likelihood $L(\kappa | \alpha_1, \dots, \alpha_n)$ for κ which we want to maximize with respect to κ .

$$L(\kappa | \alpha_1, \dots, \alpha_n) = \int_0^{2\pi} \frac{\exp((\kappa \sum \cos(\alpha_i) \cos(\mu) + \sum \sin(\alpha_i) \sin(\mu)) + \tau \cos(\mu) \cos(\mu_0) + \sin(\mu) \sin(\mu_0))}{(2\pi)^n (I_0(\kappa))^n 2\pi I_0(\tau)} d\mu \quad (7)$$

where setting $\sum \cos(\alpha_i) = R \cos(\bar{\alpha})$ and $\sum \sin(\alpha_i) = R \sin(\bar{\alpha})$ gives

$$= \int_0^{2\pi} \frac{\exp((\kappa R \cos(\bar{\alpha}) + \tau \cos(\mu_0)) \cos(\mu) + \kappa R \sin(\bar{\alpha}) \sin(\mu) + \tau \sin(\mu_0) \sin(\mu))}{(2\pi)^n (I_0(\kappa))^n 2\pi I_0(\tau)} d\mu \quad (8)$$

Putting $\kappa R \cos(\bar{\alpha}) + \tau \cos(\mu_0) = \gamma \cos(\alpha^*)$ and

$\kappa R \sin(\bar{\alpha}) + \tau \sin(\mu_0) = \gamma \sin(\alpha^*)$, and by the definition of the Bessel function $I_0(x)$, the resulting integral in 8 is our likelihood for κ which is given by:

$$L(\kappa | \alpha_1, \dots, \alpha_n) = \frac{2\pi I_0(\sqrt{\kappa^2 R^2 + 2\kappa R \tau \cos(\bar{\alpha} - \mu_0)})}{(2\pi)^n (I_0(\kappa))^n 2\pi I_0(\tau)}$$

The likelihood is a ratio of Bessel functions. Given the likelihood, prior distribution on μ , and data we can find the MLE for κ . There is not a simple analytical solution for the MLE, so numerical methods are required for the maximization of 9 with respect to κ leading to the semi-Bayesian MLE $\hat{\kappa}_{Bay}$.

One interesting comparison would be of the frequentist MLE for κ as in 2 with the semi-Bayesian MLE obtained from 9, using a circular uniform prior distribution on μ in the latter, i.e. setting $\tau = 0$. In some cases, placing uniform priors result in Bayes estimates that are similar to classical MLEs. Using a circular uniform prior distribution on μ in 9, we derive the Fisher Information to find the variance of our semi-Bayesian MLE. From 9, with a circular uniform prior, the log-likelihood is,

$$\ell = \ln I_0(\kappa R) - n \ln 2\pi - n \ln I_0(\kappa), \quad (10)$$

and the semi-Bayesian MLE for κ is the solution to setting $\dot{\ell} = 0$ where $\dot{\ell} = \frac{\delta \ell}{\delta \kappa}$,

$$\dot{\ell} = R \frac{I_1(\kappa R)}{I_0(\kappa R)} - n \frac{I_1(\kappa)}{I_0(\kappa)} \quad (11)$$

The solution for MLE κ in this case is found by,

$$\hat{\kappa}_{Bay} \text{ is the solution to: } \frac{R}{n} = \frac{A(\kappa)}{A(\kappa R)} \quad (12)$$

Immediately we notice a difference when comparing $\hat{\kappa}_{MLE}$ in

2. Taking another derivative of 11 we have obtained the Hessian where $\ddot{\ell} = \frac{\partial^2 \ell}{\partial \kappa^2}$,

$$\begin{aligned} \ddot{\ell} &= R^2 \frac{I_0(\kappa R) I_2(\kappa R) - I_1^2(\kappa R)}{I_0^2(\kappa R)} + n \frac{I_1^2(\kappa) - I_0(\kappa) I_2(\kappa)}{I_0^2(\kappa)} \\ &= R^2 \left(\frac{I_2(\kappa)}{I_0(\kappa)} - A^2(\kappa R) \right) + n \left(A^2(\kappa) - \frac{I_2(\kappa)}{I_0(\kappa)} \right) \quad (13-15) \\ &= \frac{I_2(\kappa)}{I_0(\kappa)} (R^2 - n) + n A^2(\kappa) - R^2 A^2(\kappa R) \end{aligned}$$

Then the Fisher Information (I) is given by,

$$I = \frac{I_2(\kappa)}{I_0(\kappa)} (n - R^2) - n A^2(\kappa) + R^2 A^2(\kappa R), \quad (16)$$

where substituting the semi-Bayes MLE, the asymptotic variance (V) of $\hat{\kappa}_{Bay}$:

$$V = \left(\frac{I_2(\hat{\kappa}_{Bay})}{I_0(\hat{\kappa}_{Bay})} (R^2 - n) + n A^2(\hat{\kappa}_{Bay}) - R^2 A^2(\hat{\kappa}_{Bay} R) \right)^{-1} \quad (17)$$

Therefore the asymptotic variance of the MLE can be found using 17. Next, we can compare the two MLE's via their respective large-sample confidence intervals. The $(1 - \gamma) \times 100\%$ confidence interval for κ is given by,

$$\begin{aligned} &\left(\hat{\kappa}_{MLE} \pm Z_{(\gamma/2)} \sqrt{\frac{1}{n(1 - \bar{R}/\hat{\kappa}_{MLE} - \bar{R}^2)}} \right) \\ &\left(\hat{\kappa}_{Bay} \pm Z_{(\gamma/2)} \sqrt{\frac{I_2(\hat{\kappa}_{Bay})}{I_0(\hat{\kappa}_{Bay})} (R^2 - n) + n A^2(\hat{\kappa}_{Bay}) - R^2 A^2(\hat{\kappa}_{Bay} R)} \right)^{-1} \end{aligned}$$

Figure 1 displays histograms for $\hat{\kappa}_{MLE}$ and $\hat{\kappa}_{Bay}$ based on 1000 simulations from the CND with $\kappa = 1, 3$, using sample size of $n = 30$. In each κ setting the histograms of estimated values are nearly identical for $\hat{\kappa}_{MLE}$ and $\hat{\kappa}_{Bay}$.

Preliminary Test Estimators

A preliminary test estimator (PTE) is a method of estimation that introduces sample-based prior information via a hypothesis test on the nuisance parameter to aid in estimating the parameter of interest [1]. If we fail to reject the null, then we use an estimator evaluated using the null hypothesis value. If we reject the null hypothesis, we use an estimator based directly on the sample, the usual MLE. The parameter value in the null hypothesis represents our prior knowledge. The idea is when the true parameter value is in or near the null hypothesis value, the PTE will provide a better estimator in terms of mean squared error (MSE), or any other risk function.

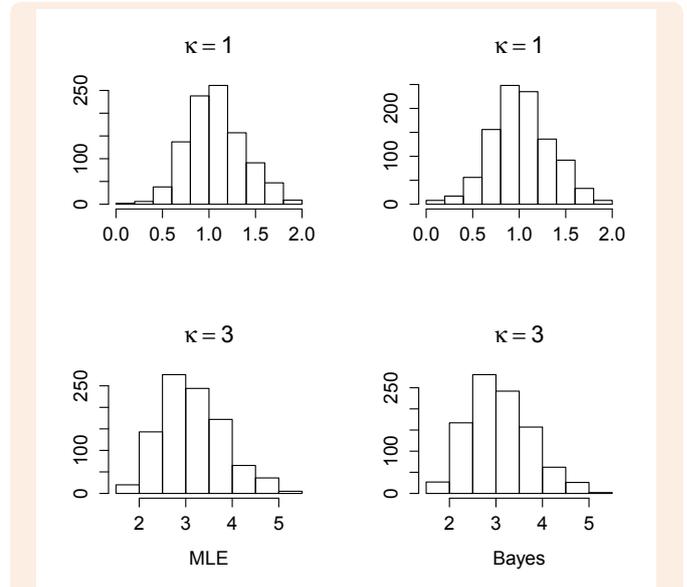


Figure 1: Histograms of $\hat{\kappa}_{MLE}$ (MLE) and $\hat{\kappa}_{Bay}$ (Bayes) with circular uniform prior for 1000 simulations from CND(μ, κ).

We observe data from a CND with unknown mean direction and concentration parameter. We are interested in estimating the concentration parameter, with the mean direction being a nuisance parameter. Our preliminary test has null hypothesis of mean direction equal to a pre-specified direction, versus a two-sided alternative.

Our PTE for the concentration performs better than the usual MLE and Bayesian estimates for the parameter. The result is similar to the linear case where we have a normal distribution with unknown mean and variance, [10]. This methodology can be used to improve the estimation accuracy in many existing applications since the CND is one of the most commonly used distributions in circular statistics.

Test for assumed mean direction

Suppose we have observations $\alpha_1, \dots, \alpha_n$ from a CND with both mean direction and concentration parameter unknown. We want to test:

$$H_0: \mu = \mu_0 \text{ vs } H_1: \mu \neq \mu_0 \quad (18)$$

In the linear case with data from a Normal distribution, this is parallel to the standard Student's t-test. In [9], the Likelihood Ratio Test (LRT) is based on the test statistic:

$$V_0 = \sum_{i=1}^n \cos(\alpha_i - \mu_0) \text{ or } = R \cos(\bar{\alpha}_0 - \mu_0) \frac{V_0}{R} = \text{Cos}(\bar{\alpha}_0 - \mu_0) \quad (19)$$

where we reject the null hypothesis for small values of the test statistic. Note the distribution for V_0 and $\frac{V_0}{R}$ depend on the nuisance parameter κ . However, the exact conditional test for the

mean direction of the CN can be obtained by using the conditional distribution of $R | V_0$, which is independent of κ . V_0 is the length of the projection of sample resultant vector, R , towards the null hypothesized mean direction, $\mu_0 = (\cos(\mu_0), \sin(\mu_0))$. In the conditional test we reject null if V_0 is too small for a given R , or equivalently, we reject the null if R is too large for a given V_0 .

To illustrate the geometry of the test, suppose we have polar vector given by the null hypothesis, $(\cos(\mu_0), \sin(\mu_0))$. Next, we have n observations and we calculate the length of projection, c , of the sample resultant vector on the polar vector. Conditioning on the value of c , we find the probability of observing our sample resultant vector, R , and larger values when the null direction is true, conditional on the observed value of $V_0 = V$.

The space consists of sample resultant vectors that have projection length, c , on the polar vector. Suppose R_1 and R_2 are two resultant vector with equal projection length and $R_1 > R_2$. Then the direction of R_1 is further away from μ_0 , than R_2 's direction.

For significance level γ , we find the rejection region via the exact conditional distribution of $R | V_0$. That is, r_0 is the solution to the equation that satisfies:

$$P(R > r_0 | V_0 = v) = \gamma \quad (20)$$

As shown in [9], this critical point r_0 is the solution to:

$$\int_{r_0}^n f(r|v) dr = \int_{r_0}^n \frac{r \Psi_n(r)}{r_0 \sqrt{r^2 - v^2} f_0(v) \pi} dr = \alpha \quad (21)$$

where we solve for r_0 , for a given v and n . Equations for $\Psi_n(r)$ and $f_0(v)$ can be found in [9]. There is no analytical solution for r_0 in this case, and [11] provides a table of rejection regions for various values of v_0 . To simplify our hypothesis test we use results in [12], where approximate confidence intervals for the mean direction are provided. Our test statistic derived from the approximate LRT is broken into two cases:

For $\bar{R} \leq 0.9$, we reject H_0 if:

$$R_2 > V_0^2 + \frac{1}{4n}(2n^2 - V_0^2)Z_\gamma \quad (22)$$

Where $\bar{R} = \frac{R}{n}$ and Z_γ is the upper quantile of the standard Normal distribution.

For $\bar{R} \geq 0.9$, we reject H_0 if:

$$n \log \left(\frac{n^2 - V_0^2}{n^2 - R^2} \right) > X_{1,\gamma}^2 \quad (23)$$

These approximations hold well for even small sample sizes when the concentration is high.

The PTE for the concentration parameter

Now we introduce our PTE for estimating the concentration parameter, where the mean direction is a nuisance parameter. Given observations, $\alpha_1, \dots, \alpha_n$, with unknown mean direction and concentration parameter we test our null hypothesized mean direction via the aforementioned hypothesis test. Our PTE is given by:

(i) For $\bar{R} \leq 0.9$,

$$\hat{\kappa}_{PTE} = \hat{\kappa}_{MLE} \mathbb{I}(Z > Z_\gamma) + \hat{\kappa}_{\mu_0} \mathbb{I}(Z < Z_\gamma) \quad (24)$$

where Z is found by using 22 and solving for Z_γ .

(ii) For $\bar{R} \geq 0.9$,

$$\hat{\kappa}_{PTE} = \hat{\kappa}_{MLE} \mathbb{I}(X^2 > X_{1,\gamma}^2) + \hat{\kappa}_{\mu_0} \mathbb{I}(X^2 > X_{\gamma,r}^2) \quad (25)$$

where χ^2 is found by using 23.

we break the estimator into the two cases according to our hypothesis test. The PTE in either case selects only one of the two estimators according to the result of the hypothesis test. The performance of the PTE depends on the level of the test and the proximity of the true mean direction to the null hypothesized value. We measure performance in terms of mean squared error (MSE) of our estimator over different significance levels γ , and different true differences between the mean directions $\delta = \mu - \mu_0$.

In Figure 2, we observe the simulation-based MSE of the PTE and MLE for the concentration parameter. We perform 1000 simulations of $(\alpha_1, \dots, \alpha_{20}) \sim CN(\delta_j, \kappa)$, for $j = 1, \dots, 50$. Here $\delta_1, \dots, \delta_{50}$ represent the 50 equally spaced points between 0 and π . For each δ_j , we record the MSE. Each line represents MSE of an estimator over values of $\delta = \mu - \mu_0$, where δ represents the true difference between the population mean direction and the null hypothesized mean direction.

MSE Comparisons: MLE vs. Various PTE

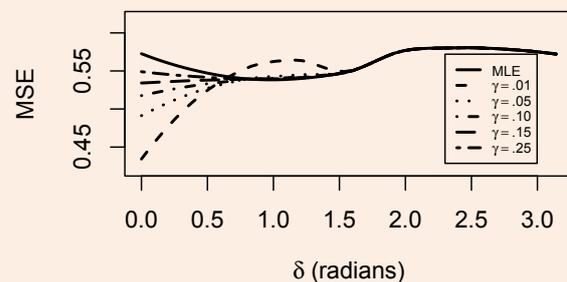


Figure 2: Simulation-based MSE of MLE and PTE for different significance levels γ .

For significance levels $\gamma = .10, .15, .25$, the PTE performs at least as good as the MLE, and performs better when the true mean direction is closer to the null hypothesized value. For larger significance levels the test requires less evidence to reject the null

hypothesis, and when we reject the null the PTE is equivalent to the MLE, $\hat{\kappa}_{MLE}$. In Figure 2, we observe that as the significance level increases the PTE is more likely to use $\hat{\kappa}_{MLE}$ for smaller values of δ . To show the vast improvement in our PTE, we examine the mean-square relative efficiency (MRE) of the 2 estimators

$$e(\hat{\kappa}_{PTE} = \hat{\kappa}_{MLE}) = \frac{MSE(\hat{\kappa}_{MLE})}{MSE(\hat{\kappa}_{PTE})} \quad (26)$$

Values larger than unity imply that the PTE performs better than the MLE. In Figure 3, we have the MRE of the MLE and PTE with $\gamma = 0.01$ across all values of δ . The relative efficiency is greater than 1 for all δ less than approximately 0.65 radians. In this example, the PTE can reduce the MSE by 20% when the true difference in mean directions is small. For $0.65 < \delta < 1.5$, the MRE is less than 1 implying the MLE has the smaller MSE. This is due to our preliminary test failing to reject the null hypothesis. For $\delta > 1.5$, the preliminary test will almost always reject the null hypothesis value and the PTE will be the same as the MLE resulting in the MRE being equal to one.

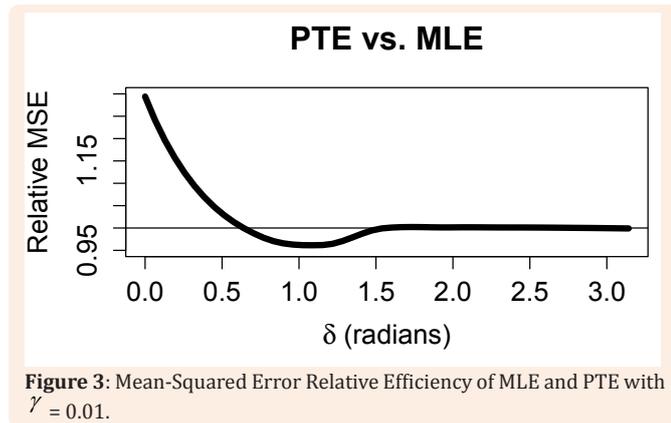


Figure 3: Mean-Squared Error Relative Efficiency of MLE and PTE with $\gamma = 0.01$.

In Equation 3, the MRE is maximum for $\delta = 0$, and when the PTE will almost always reject the null for large enough δ the MRE is equal to one. For $0 < \delta < \pi$ the PTE may reject or fail to reject the null hypothesis depending on the sample observed. In the case it fails to reject, $V_0 = \sum_{i=1}^n \cos(\alpha_i - \mu_0)$ is no longer minimized at μ_0 since μ is the population mean (Recall $\delta = \mu - \mu_0$). Therefore $V_0 < R$ if μ_0 is closer to μ than $-\alpha_0$, and $V_0 > R$ if $-\alpha_0$ is closer to μ than μ_0 . If the latter case appears more often than the former case for some intermediate values of δ , then the MRE will be less than 1.

In Figure 2, we compare the MLE and PTE with $\gamma = 0.01$ from Figure 2. The PTE with $\gamma = 0.01$ has the best results for smaller values of δ , but could perform worse than the MLE for intermediate values of δ . PTEs with $\gamma = 0.1, 0.15, 0.25$ perform at least as good as the MLE. Now, we illustrate other possibilities that can occur and the performance of the PTE.

In Figure 4, we simulate from four different realities and examine the performance of our PTE for the same significance

levels as used in Figure 2. Note that the lines have same labels as in Figure 2. For each plot we have simulated-based MSEs for each line. We perform 1000 simulations of $(\alpha_1, \dots, \alpha_{20}) \sim CN(\delta_j, \kappa)$, for $j = 1, \dots, 50$. Here $\delta_1, \dots, \delta_{50}$ represent the 50 equally spaced points between 0 and π . For each δ_j , we record the MSE which creates our MSE curve over δ for each scenario.

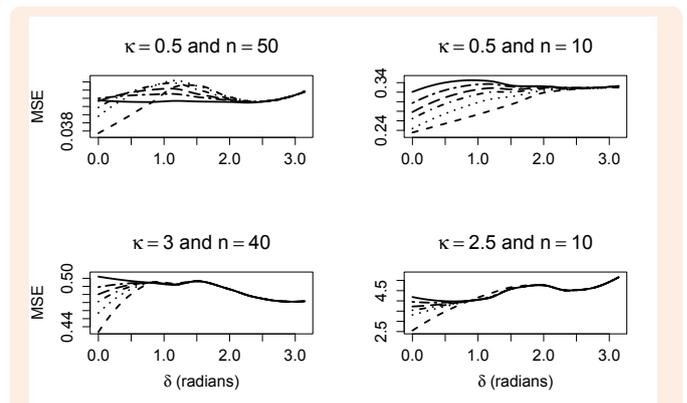


Figure 4: Simulation-based Comparison of PTE Performances for Sample Sizes $n = 10, 40, 50$ and Concentration Parameters $\kappa \in \{0.5, 2.5, 3\}$. Lines are labeled as in Figure 2.

First examine that in all scenarios, the PTE with significance level $\gamma = 0.01$ performs the best when the true difference in mean direction is null or small. In the top-left plot we have $n = 50$ simulated observations from $CN(\delta, \kappa = 0.5)$; top-right plot we have $n = 10$ simulated observations from $CN(\delta, \kappa = 0.5)$; bottom-left plot we have $n = 40$ simulated observations from $CN(\delta, \kappa = 3)$; bottom-right plot we have $n = 10$ simulated observations from $CN(\delta, \kappa = 2.5)$.

In the top-right plot all of the PTE's in this simulation performed uniformly better (over δ) than $\hat{\kappa}_{MLE}$. In the remaining three plots there are values of δ where the $\hat{\kappa}_{MLE}$ has better performance. This occurs when our preliminary test fails to reject the null hypothesis for intermediate values of δ . The difference becomes more obvious when we have a large sample size and the value of κ is small as in the top left plot. Here the PTE's MSE increases for intermediate values of δ , for relatively smaller significance levels. There is a similar pattern in the bottom two plots. This pattern is to be expected, since smaller significance level will require more evidence to reject the null hypothesis of the preliminary test.

In applications, the values of δ and κ are unknown. So how do we select the optimal significance level given n observations from $CN(\mu, \kappa)$? Following the work of [1], we create tables to find a PTE with minimum and maximum MREs.

Tables were constructed through simulations. Given a sample size n and value for κ , we generate values from a $CN(\delta, \kappa)$ distribution to estimate the MRE over a grid of γ and δ values, where $0 \leq \delta \leq \pi$. For each γ , we compute the maximum MRE,

$$E_{\max} \text{ minimum MRE, } E_{\min}, \text{ over all } \delta, \text{ and record the } \delta$$

where E_{min} is located, Δ_{min} . For almost all cases the location of the maximum MRE is located at $\delta = 0$ and the function $MRE(\delta)$ is monotone decreasing from $\delta = 0$ to $\delta = E_{min}$. For values $\delta > E_{min}$, the function $MRE(\delta)$ increases back to unity since the PTE will reject the null hypothesized values for larger δ . We then repeated this procedure for different parameter values for κ .

The mean resultant vector is the normalized length of R since $0 \leq \bar{R} \leq 1$ and is a measure of concentration for a sample of observations. A value close to 1 implies high concentration and a value close to 0 implies little to no concentration around any single direction. This estimate does not depend on the knowledge of κ or of the mean μ of the distribution. For the CND, there is a one-to-one correspondence between statistic \bar{R} and the concentration parameter κ . Given a sample size n and κ , we observe the average

\bar{R} over our simulations and use the average as an indication of strength of concentration. In practice, we advise the user to find the sample observed \bar{R} of the n observations, and then use the column of the table with the nearest \bar{R} value.

In Table 1, we provide a list of potential PTEs for $n = 20$. The rows list various significance levels γ for the PTE ranging from 1% to 50%. The columns list the different observed values for \bar{R} . Suppose we have a sample size of 20 observations and observe \bar{R} close to 0.779. Following the procedure in [1], we then decide the minimum MRE preferred is $E_{min} = 0.471$. Then using the Table 1, the optimal PTE corresponds to using $\alpha = 0.20$. Upon request, we provide tables for various sample sizes, where the tables require only knowledge of sample size, \bar{R} , and the predetermined E_{min} .

Table 1: n = 20: Mailmen and Minimum Guaranteed Efficiencies for the PTE.

		\bar{R}									
γ		0.196	0.198	0.202	0.471	0.617	0.710	0.779	0.818	0.871	0.900
0.01	E_{max}	1.016	1.027	1.676	1.38	1.244	1.075	1.081	1.116	1.033	1.018
	E_{min}	1.008	1.015	0.995	0.688	0.695	0.869	0.928	0.951	0.936	0.915
	Δ_{min}	0	2.949	2.757	1.154	0.833	0.769	0.769	0.833	0.705	0.449
0.02	E_{max}	1.055	1.103	1.555	1.231	1.105	1.049	1.07	1.118	1.048	1.012
	E_{min}	1.043	1.046	0.993	0.689	0.807	0.921	0.961	0.98	0.968	0.95
	Δ_{min}	0	3.142	2.629	1.09	0.769	0.769	0.769	0.833	0.769	0.449
0.05	E_{max}	1.127	1.228	1.313	1.08	1.02	1.031	1.06	1.111	1.061	1.015
	E_{min}	1.116	1.102	0.991	0.743	0.924	0.968	0.988	0.995	0.988	0.983
	Δ_{min}	0	3.142	2.5	0.962	0.769	0.769	0.833	0.898	0.833	0.513
0.1	E_{max}	1.138	1.246	1.202	1.012	1.004	1.017	1.042	1.08	1.051	1.006
	E_{min}	1.134	1.103	0.984	0.825	0.964	0.988	0.994	0.997	0.995	0.994
	Δ_{min}	0	3.142	1.795	0.833	0.705	0.769	0.833	0.898	0.833	0.641
0.15	E_{max}	1.103	1.195	1.139	1.007	1.002	1.02	1.031	1.066	1.044	1.007
	E_{min}	1.099	1.076	0.975	0.889	0.979	0.994	0.997	0.997	0.997	0.997
	Δ_{min}	3.142	3.142	1.667	0.769	0.705	0.833	0.833	0.898	0.898	0.769
0.2	E_{max}	1.084	1.144	1.085	1.005	1.001	1.015	1.026	1.054	1.031	1.007
	E_{min}	1.076	1.06	0.968	0.935	0.987	0.996	0.998	0.998	0.998	0.998
	Δ_{min}	1.988	3.077	1.603	0.641	0.641	0.833	0.898	0.898	0.898	0.709
0.25	E_{max}	1.064	1.111	1.045	1.005	1	1.017	1.018	1.046	1.023	1.006
	E_{min}	1.059	1.040	0.965	0.96	0.991	0.997	0.998	0.998	0.999	0.999
	Δ_{min}	0	3.077	1.346	0.385	0.449	0.833	0.833	0.898	0.898	0.833

0.3	E_{max}	1.049	1.08	1.015	1.006	1	1.013	1.016	1.029	1.02	1.007
	E_{min}	1.046	1.032	0.963	0.971	0.995	0.998	0.999	0.999	0.999	0.999
	Δ_{min}	0	3.142	1.346	0.128	0.577	0.833	0.833	0.898	0.898	0.833
0.35	E_{max}	1.036	1.059	1.002	1.004	1	1.01	1.012	1.02	1.015	1.006
	E_{min}	1.033	1.024	0.965	0.979	0.996	0.999	0.999	0.999	0.999	1
	Δ_{min}	1.859	3.142	1.282	0	0.449	0.833	0.898	0.962	0.898	0.833
0.4	E_{max}	1.026	1.044	0.992	1.002	1	1.008	1.008	1.013	1.012	1.007
	E_{min}	1.024	1.016	0.965	0.982	0.996	0.999	1	1	0.999	1
	Δ_{min}	0	3.142	1.154	0	0	0.898	0.898	0.962	0.898	0.898
0.45	E_{max}	1.019	1.031	0.99	1.001	1	1.007	1.008	1.009	1.01	1.005
	E_{min}	1.017	1.011	0.969	0.987	0.998	1	1	1	0.999	1
	Δ_{min}	0	3.142	1.154	0	0	0.898	0.962	0.898	0.898	0.898
0.5	E_{max}	1.015	1.023	0.993	1.001	1	1.005	1.003	1.005	1.006	1.003
	E_{min}	1.013	1.009	0.971	0.99	0.999	1	1	1	1	1
	Δ_{min}	0	3.142	0.962	0	0.449	0.898	0.898	0.898	0.898	1.218

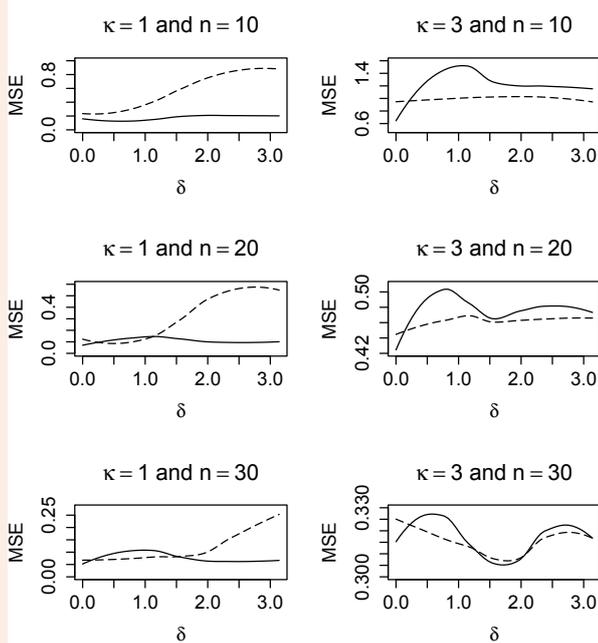


Figure 5: MSE of PTE and Bayes Estimators over δ : $\hat{\kappa}_{PTE}$ (—) and $\hat{\kappa}_{Bay}$ (---).

Comparison of the PTE and Bayes Estimators

Both the PTE and Bayes estimators in 12, use prior information on the mean direction μ to aid in estimation of the concentration parameter. A smaller significance level for the PTE requires stronger evidence to reject the null hypothesized

value μ_0 . A smaller significance level may be chosen to coincide with a stronger belief in the mean direction μ_0 . In the previously mentioned Bayesian setting of this chapter, a larger value for the concentration parameter τ focuses our prior distribution around the mean direction μ_0 . A larger value in parameter τ represents a stronger belief in prior mean direction μ_0 .

In Figure 5 we make a comparison of the MSE of our PTE with significance level of 1% with the Bayes estimator with CN prior centered around the null hypothesis value μ_0 and with $\tau = 4$. We plot the MSE curve of each estimator over values of δ .

In each plot, the solid line is MSE curve for the PTE and the dashed line is the MSE curve for the Bayes estimator. For $\kappa = 1$, $\hat{\kappa}_{PTE}$ performs better overall for all sample sizes. For $n = 10$, $\hat{\kappa}_{PTE}$ performs uniformly better than $\hat{\kappa}_{Bay}$. For $n = 20$ and $n = 30$ the estimators have similar performances for small values of δ , but the MSE for $\hat{\kappa}_{Bay}$ is much larger for large values of δ .

If $\kappa = 3$, we have different results when comparing the MSEs. In all sample sizes of $n = 10, 20, 30$, the MSE of $\hat{\kappa}_{PTE}$ is best for small values of δ . Also, for all sample sizes, $\hat{\kappa}_{Bay}$ has the smaller MSE for the larger values of δ . In this case for large value of κ , $\hat{\kappa}_{Bay}$ would be the preferred estimator since the performance is better overall.

In reality we do not know the value of κ , so need a data driven way to select $\hat{\kappa}_{Bay}$ versus $\hat{\kappa}_{PTE}$. If we suspect a high concentration

then we suggest to use $\hat{\kappa}_{\text{Bay}}$, and for a weak concentration then use $\hat{\kappa}_{\text{PTE}}$. If given a sample size n , go to the corresponding PTE table for the same sample size. In the table, go to the 7th column which gives the expected \bar{R} under $\kappa = 3$ simulations. From your observed sample of size n , calculate R in column 7, and compare to the value from the PTE table. If less than the PTE table value, then use $\hat{\kappa}_{\text{PTE}}$, otherwise use $\hat{\kappa}_{\text{Bay}}$.

Conclusion

In our work we have shown a superior estimator for the concentration parameter when the mean direction is unknown of a CND. In all cases the PTE has better performance in terms of MSE if δ is null or small. Another interesting result is in comparison for the PTE we developed an MLE for the concentration parameter based from the prior distribution of mean direction.

References

1. Saleh AME (2006) Theory of preliminary test and Stein-type estimation with applications. volume 517. John Wiley & Sons pp. 656.
2. Bancroft TA (1944) On biases in estimation due to the use of preliminary tests of significance. The Annals of Mathematical Statistics 15(2): 190-204.
3. Bancroft TA (1964) Analysis and inference for incompletely specified models involving the use of preliminary test (s) of significance. Biometrics 20(3): 427-442.
4. Bancroft TA (1965) Inference for incompletely specified models in physical sciences. Bulletin of the International Statistical Institute 41(1): 497-515.
5. Snedecor GW (1938) Statistical Methods. Collegiate Press, Iowa, USA.
6. Han C, Bancroft TA (1968) On pooling means when variance is unknown. Journal of the American Statistical Association 63(324): 1333-1342.
7. Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proceedings of the Third Berkeley symposium on mathematical statistics and probability, University of California Press, Berkeley, USA, volume 1 pp. 197-206.
8. James W, Stein C (1961) Estimation with quadratic loss. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability 1: 361-379.
9. Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics. volume 5. World Scientific, Singapore.
10. Ohtani K (1991) Estimation of the variance in a normal population after the one-sided pretest for the mean. Communications in statistics-theory and methods 20(1): 219-234.
11. Stephens MA (1962) Exact and approximate tests for directions. Biometrika 49(3-4): 463-477.
12. Upton GJG (1986) Approximate confidence intervals for the mean direction of a von mises distribution. Biometrika, 73(2): 525-527.