Statistical Methodology 6 (2009) 280–289

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

A general censoring scheme for circular data

S. Rao Jammalamadaka [a,*], Vasudevan Mangalam [b]

[a] *University of California, Santa Barbara, USA*
[b] *Universiti Brunei Darussalam, Brunei*

ARTICLE INFO

Article history:
Received 30 May 2008
Accepted 31 October 2008

Keywords:
Directional data
Generalized censoring scheme
von Mises distribution
Maximum likelihood estimators
Large-sample properties
Self consistent estimator
Nonparametric MLE
Monte Carlo performance

ABSTRACT

A generalized censoring scheme in the survival analysis context was introduced by the authors in Jammalamadaka and Mangalam [S. Rao Jammalamadaka, V. Mangalam, Nonparametric estimation for middle censored data, J. Nonparametr. Stat. 15 (2003) 253–265]. In this article we discuss how such a censoring scheme applies to circular data and in particular when the original data is assumed to come from a parametric model such as the von Mises. Maximum likelihood estimation of the parameters as well as their large-sample properties are considered under this censoring scheme. We also consider nonparametric estimation of the circular probability distribution under such a general censoring scheme and use Monte Carlo methods to investigate its effects on the estimation of the mean direction and concentration.

1. Introduction

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be a set of independent and identically distributed (i.i.d.) measurements on two-dimensional directions. Such measurements, called angular or circular data, can be represented as points on the circumference of a circle with unit radius. They may represent wind directions, the vanishing angles at the horizon for a group of birds, or the times of arrival at a hospital emergency room where the 24 h cycle is represented as a circle. In assigning numerical values to such directions, one has to keep in mind the arbitrary choice of the zero direction, as well as the sense of rotation. For definiteness, all throughout this paper, we measure angles in the range $[0, 2\pi)$ and use anti-clockwise direction as positive. However, the statistical measures as well as methods should be independent of these choices.

* Corresponding address: University of California, Department of Statistics, 93106 Santa Barbara, CA, USA.
E-mail address: rao@pstat.ucsb.edu (S. Rao Jammalamadaka).

1572-3127/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2008.10.002

We consider the censoring problem where one may not be able to observe all the data points. For instance, a bird's vanishing angle at the horizon might be obscured by a passing cloud or a (fixed) hill range so that one sees some actual observations α_i while the others may be simply noted as falling inside a random interval (L_i, R_i). A similar thing happens when the registration counter of a hospital emergency room is closed for a temporary period and all patients who arrive during that period are registered as having come during that interval. We consider both parametric and nonparametric estimation problems in this context and evaluate the loss in efficiency because of such censoring. See for instance, Iyer et al. [1] who consider censoring in an exponential model.

In Section 1, we consider a von Mises (or Circular Normal) model for the original data. Maximum likelihood estimation of the mean direction and concentration parameter are fairly standard and the reader is referred to books such as Mardia and Jupp [2], or Jammalamadaka and SenGupta [3]. The last reference also includes S-Plus based software for computing these parameters. We consider various censoring distributions that generate an interval of censoring, (L_i, R_i) and if an observation falls in this interval, then we do not observe its actual value but just note this interval. We consider maximum likelihood estimation under such censoring and show that in large samples, such estimators follow a normal distribution, allowing one to find confidence intervals etc. In the following section, we drop the parametric model assumption and provide an iterative scheme to estimate the circular probability distribution which results in a "Self-Consistent Estimator (SCE)" which is also most often the "Nonparametric Maximum Likelihood Estimator (NPMLE)".

Remark. The idea of middle-censoring contained here can be easily adapted to higher dimensions say for example for the von Mises–Fisher distribution in 3-dimensions, with spherical caps replacing the arcs discussed here(like clouds covering the earth). However it entails considerably more work, theoretical as well as computational, as can be surmised from the results in the next few sections, and we plan to pursue this in a future project. Even the nonparametric work needs extensions to higher dimensions.

2. Censoring with parametric models

Formally, let $\alpha_1, \ldots, \alpha_n$ be a set of angular measurements and suppose they follow a von Mises distribution. Recall that a random angle A is said to follow a von Mises distribution with mean direction μ and concentration parameter κ, to be denoted by a $vM(\mu, \kappa)$, if it has the probability density function

$$f(\alpha; \boldsymbol{\theta}) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\alpha - \mu)}, \quad 0 \le \alpha < 2\pi$$

where $\boldsymbol{\theta} = (\mu, \kappa) \in \Theta = [0, 2\pi) \times [0, \infty)$. Here I_ν is the *modified Bessel function of the first kind and order* ν (also called Bessel function of purely imaginary argument), and is given by

$$\begin{aligned} I_\nu(z) &= \frac{1}{2\pi} \int_0^{2\pi} \cos \nu t e^{z \cos t} dt \\ &= \sum_{r=0}^{\infty} \frac{(z/2)^{\nu+2r}}{r! \Gamma(\nu + r + 1)}. \end{aligned}$$

Our goal is to estimate $\boldsymbol{\theta}$ when some of the observations are censored by intervals of the type (l, r) where l and r forms an arc on the circumference. Denoting the observed arcs as (l_i, r_i), the likelihood function takes the form

$$L_n(\boldsymbol{\alpha}; \boldsymbol{\theta}) = \frac{1}{(2\pi I_0(\kappa))^n} \exp\left[\kappa \sum_{i=1}^n \delta_i \cos(\alpha_i - \mu)\right] \prod_{i=1}^n \left[\int_{l_i}^{r_i} \exp[\kappa \cos(t - \mu)] dt\right]^{1-\delta_i}$$

where δ_i takes the value "0" if the observation is censored and the value "1" if it is uncensored. If the arc straddles the origin $0 = 2\pi$ so that $r < 0 < l$, the integral over the arc starting from l and ending at r is interpreted, throughout this paper, as the sum of integrals from l to 2π and 0 to r. When convenient, we indicate this arc by $A_{l,r}$ and its complement by $\bar{A}_{l,r}$

Thus the log-likelihood function is given by

$$l_n(\boldsymbol{\alpha};\boldsymbol{\theta}) = -n\log 2\pi - n\log I_0(\kappa) + \kappa\sum_{i=1}^{n}\delta_i\cos(\alpha_i-\mu) + \sum_{i=1}^{n}(1-\delta_i)\log\left[\int_{A_{l_i,r_i}}\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t\right]. \tag{2.1}$$

The $\hat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}_n} = \left(\hat{\mu}_n, \hat{\kappa}_n\right)$ which maximizes this likelihood function, is clearly the MLE of $\boldsymbol{\theta}$. Computational aspects of the MLE are discussed below.

Before deriving the first and second derivatives for the solution of the ML equation and computation of the information matrix, we introduce the following notations. Let

$$A_0(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$$
$$B_{0i}(\mu,\kappa) = \int_{l_i}^{r_i}\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t$$
$$B_{1i}(\mu,\kappa) = \int_{l_i}^{r_i}\sin(t-\mu)\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t$$
$$B_{2i}(\mu,\kappa) = \int_{l_i}^{r_i}\cos(t-\mu)\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t$$
$$B_{3i}(\mu,\kappa) = \int_{l_i}^{r_i}\sin^2(t-\mu)\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t$$
$$B_{4i}(\mu,\kappa) = \int_{l_i}^{r_i}\cos^2(t-\mu)\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t$$
$$B_{5i}(\mu,\kappa) = \int_{l_i}^{r_i}\sin(t-\mu)\cos(t-\mu)\exp[\kappa\cos(t-\mu)]\,\mathrm{d}t.$$

Then it is easy to check

$$\frac{\partial B_{0i}}{\partial\mu} = \kappa B_{1i}$$
$$\frac{\partial B_{0i}}{\partial\kappa} = B_{2i}$$
$$\frac{\partial B_{1i}}{\partial\mu} = \kappa B_{3i} - B_{2i}$$
$$\frac{\partial B_{1i}}{\partial\kappa} = B_{5i}$$
$$\frac{\partial B_{2i}}{\partial\kappa} = B_{4i}.$$

The derivatives of the log-likelihood function are given by

$$\frac{\partial l_n}{\partial\mu} = \kappa\sum_{i=1}^{n}\delta_i\sin(\alpha_i-\mu) + \kappa\sum_{i=1}^{n}(1-\delta_i)\frac{B_{1i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)} \tag{2.2}$$

and

$$\frac{\partial l_n}{\partial\kappa} = -nA_0(\kappa) + \sum_{i=1}^{n}\delta_i\cos(\alpha_i-\mu) + \sum_{i=1}^{n}(1-\delta_i)\frac{B_{2i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)}. \tag{2.3}$$

The second derivatives are given by

$$\frac{\partial^2 l_n}{\partial \mu^2} = -\kappa \sum_{i=1}^{n} \delta_i \cos(\alpha_i - \mu) + \kappa \sum_{i=1}^{n} (1-\delta_i) \left[\frac{\kappa B_{3i}(\mu,\kappa) - B_{2i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)} - \kappa \left(\frac{B_{1i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)} \right)^2 \right]$$

$$\frac{\partial^2 l_n}{\partial \kappa^2} = -nA_0'(\kappa) + \sum_{i=1}^{n} (1-\delta_i) \left[\frac{B_{4i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)} - \left(\frac{B_{2i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)} \right)^2 \right]$$

$$\frac{\partial^2 l_n}{\partial \kappa \partial \mu} = \sum_{i=1}^{n} \delta_i \sin(\alpha_i - \mu) + \sum_{i=1}^{n} (1-\delta_i) \left[\frac{B_{1i}(\mu,\kappa) + B_{5i}(\mu,\kappa)}{B_{0i}(\mu,\kappa)} - \frac{B_{1i}(\mu,\kappa) B_{2i}(\mu,\kappa)}{B_{0i}^2(\mu,\kappa)} \right].$$

Expressions (2.2) and (2.3) are equated to zero and solved numerically. By substituting these solutions into the information matrix, one obtains the "observed information" matrix viz.

$$\hat{I} = \left[\begin{array}{cc} \dfrac{\partial^2 l_n}{\partial \mu^2} & \dfrac{\partial^2 l_n}{\partial \kappa \partial \mu} \\ \dfrac{\partial^2 l_n}{\partial \kappa \partial \mu} & \dfrac{\partial^2 l_n}{\partial \kappa^2} \end{array} \right] \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} .$$

Then, provided censoring is not too strong, $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)$ will be asymptotically normal with mean zero and covariance $\hat{I}^{-1}$ as shown in the next section.

3. Large-sample properties of the MLE

We assume that the random censoring mechanism is independent of the variable of interest and does not involve $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_0$ denote the true value of the parameter and let

$$p(\boldsymbol{\theta}, l, r) = P_{\boldsymbol{\theta}}[A \in (l, r)] = \int_l^r f(t; \boldsymbol{\theta})\, \mathrm{d}t.$$

Let

$$g_1(\boldsymbol{\theta}, l, r) = -\log(2\pi I_0(\kappa)) + \kappa \int_{\bar{A}_{l,r}} \cos(t-\mu) f(t; \boldsymbol{\theta}_0)\, \mathrm{d}t$$
$$+ p(\boldsymbol{\theta}_0, l, r) \log \left[\int_{A_{l,r}} \exp[\kappa \cos(t-\mu)]\, \mathrm{d}t \right] \tag{3.4}$$

and define a function g on the parameter space as $g(\boldsymbol{\theta}) = \int g_1(\boldsymbol{\theta}, l, r)\, \mathrm{d}F_{LR}$.

Lemma 3.1. *Under $P_{\boldsymbol{\theta}_0}$ measure, $\frac{1}{n} l_n(\boldsymbol{\alpha}; \boldsymbol{\theta}) \rightarrow^{a.s.} g(\boldsymbol{\theta})$.*

Proof. Define

$$X_i(\boldsymbol{\theta}) = -\log[2\pi I_0(\kappa)] + \kappa \delta_i \cos(A_i - \mu) + (1-\delta_i) \log \left[\int_{L_i}^{R_i} \exp[\kappa \cos(t-\mu)]\, \mathrm{d}t \right]. \tag{3.5}$$

Then X_i's are i.i.d. random variables with mean $g(\boldsymbol{\theta})$ under $P_{\boldsymbol{\theta}_0}$. So by the strong law of large numbers $\frac{1}{n} l_n(\boldsymbol{\alpha}; \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow^{a.s.} E_{\boldsymbol{\theta}_0}[X_1] = g(\boldsymbol{\theta})$.

Lemma 3.2. *Let f_1 and f_2 be density functions w.r.t. a measure ν. Then*

$$\int f_1(x) \log f_1(x)\, \mathrm{d}\nu(x) \geq \int f_1(x) \log f_2(x)\, \mathrm{d}\nu(x)$$

with equality holding if and only if $f = g$ a.s. (ν)

For a proof see (1e.6.6) of Rao [4].

Lemma 3.3. *If l and r are two distinct arbitrary points in $[0, 2\pi)$, then $g_1(\boldsymbol{\theta}, l, r) \leq g_1(\boldsymbol{\theta}_0, l, r)$ for all $\boldsymbol{\theta} \in \Theta$ with equality holding only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

Proof. Let β be the midpoint of the arc from l to r. Let ν be a measure on $[0, 2\pi)$ defined as the sum of Lebesgue measure and the point mass at β. Define

$$h(t, \boldsymbol{\theta}) = I[t \notin (l, r)] f(t, \boldsymbol{\theta}) + I[t = \beta] p(\boldsymbol{\theta}, l, r).$$

Then $\int h(t, \boldsymbol{\theta})\, d\nu(t) = 1$ for all $\boldsymbol{\theta} \in \Theta$. Thus by Lemma 3.2, it follows that $\int h(t, \boldsymbol{\theta}_0) \log h(t, \boldsymbol{\theta})\, d\nu(t) \leq \int h(t, \boldsymbol{\theta}_0) \log h(t, \boldsymbol{\theta}_0)\, d\nu(t)$ for all $\boldsymbol{\theta} \in \Theta$. Now it is fairly straightforward to verify that $\int h(t, \boldsymbol{\theta}_0) \log h(t, \boldsymbol{\theta})\, d\nu(t) = g_1(\boldsymbol{\theta}, l, r)$. Thus it follows that $g_1(\boldsymbol{\theta}, l, r) \leq g_1(\boldsymbol{\theta}_0, l, r)$

If $g_1(\boldsymbol{\theta}_1, l, r) = g_1(\boldsymbol{\theta}_0, l, r)$, for some $\boldsymbol{\theta}_1 \in \Theta$, then by Lemma 3.2, $h(t, \boldsymbol{\theta}_1) = h(t, \boldsymbol{\theta}_0)$ *a.s.* (ν) which immediately implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$.

Theorem 3.1. *If the identifiability condition*

$$p(\boldsymbol{\theta_0}) = P_{\boldsymbol{\theta_0}}\{A \in (L, R)\} < 1$$

is satisfied, then $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. $\left(P_{\boldsymbol{\theta_0}}\right)$.

Proof. First, note that from Lemma 3.3, it follows that $g(\boldsymbol{\theta}) \leq g(\boldsymbol{\theta}_0)$ for all $\boldsymbol{\theta} \in \Theta$ with equality holding only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Also note that under the identifiability condition $g(\mu, \kappa)$ goes to $-\infty$ as $\kappa \to \infty$. Let Ω_0 be a set of $P_{\boldsymbol{\theta_0}}$ measure 1 where $\frac{1}{n} l_n(\boldsymbol{\theta}) \to g(\boldsymbol{\theta})$. The discussion below is for an arbitrary point $\omega \in \Omega$. If $\widehat{\boldsymbol{\theta}}_n \nrightarrow \boldsymbol{\theta}_0$, then there is a subsequence n_k through which $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_1 = (\mu_1, \kappa_1)$ where $\kappa_1 \in [0, \infty]$.

If $\kappa_1 < \infty$, then $\frac{1}{n_k} l_{n_k}\left(\widehat{\boldsymbol{\theta}}_{n_k}\right) \to g(\boldsymbol{\theta}_1)$. But $\frac{1}{n_k} l_{n_k}\left(\widehat{\boldsymbol{\theta}}_{n_k}\right) \geq \frac{1}{n_k} l_{n_k}(\boldsymbol{\theta}_0) \to g(\boldsymbol{\theta}_0)$ and it follows that $g(\boldsymbol{\theta}_1) \geq g(\boldsymbol{\theta}_0)$ contradicting Lemma 3.3.

If $\kappa_1 = \infty$, then $\lim_{k\to\infty} \frac{1}{n_k} l_{n_k}\left(\widehat{\boldsymbol{\theta}}_{n_k}\right) = \lim_{\kappa\to\infty} g(\mu_1, \kappa) = -\infty$. As earlier, $\frac{1}{n_k} l_{n_k}\left(\widehat{\boldsymbol{\theta}}_{n_k}\right) \geq \frac{1}{n_k} l_{n_k}(\boldsymbol{\theta}_0) \to g(\boldsymbol{\theta}_0)$ leading to a contradiction.

Theorem 3.2. *Assume that the conditions of Theorem 3.1 hold. Let $\Sigma_1(\boldsymbol{\theta})$ be the dispersion of $\mathbf{X}_1'(\boldsymbol{\theta}) = \left(\frac{\partial X_1(\boldsymbol{\theta})}{\partial \mu}, \frac{\partial X_1(\boldsymbol{\theta})}{\partial \kappa}\right)$ where $X_1(\boldsymbol{\theta})$ is as in (3.5) and let $\Sigma(\boldsymbol{\theta}) = \left[g''(\boldsymbol{\theta})\right]^{-1} \Sigma_1(\boldsymbol{\theta}) \left[g''(\boldsymbol{\theta})\right]^{-1}$. Then $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \Rightarrow N_2(\mathbf{0}, \Sigma(\boldsymbol{\theta}_0))$.*

Proof. Let

$$\mathbf{l}_n'(\boldsymbol{\theta}) = \left(\frac{\partial l_n}{\partial \mu}, \frac{\partial l_n}{\partial \kappa}\right)$$

and

$$l_n''(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial^2 l_n}{\partial \mu^2} & \dfrac{\partial^2 l_n}{\partial \kappa \partial \mu} \\ \dfrac{\partial^2 l_n}{\partial \kappa \partial \mu} & \dfrac{\partial^2 l_n}{\partial \kappa^2} \end{bmatrix}.$$

By applying the standard multivariate Taylor expansion (see (8.19), Vol 2 of Apostol [5]), we get

$$\mathbf{l}_n'\left(\widehat{\boldsymbol{\theta}}_n\right) = \mathbf{l}_n'(\boldsymbol{\theta}_0) + \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) l_n''(\boldsymbol{\theta}_0) + \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| E\left(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0\right)$$

where the function E is such that $\lim_{x\to y} E(x, y) = \mathbf{0}$. As $\widehat{\boldsymbol{\theta}}_n$ is the maximizer of the likelihood function, $\mathbf{l}_n'\left(\widehat{\boldsymbol{\theta}}_n\right) = \mathbf{0}$. From this it follows that

$$\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) l_n''(\boldsymbol{\theta}_0) = -\mathbf{l}_n'(\boldsymbol{\theta}_0) - \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| E\left(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0\right)$$

and hence

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)=\frac{-\mathbf{l}'_n(\boldsymbol{\theta}_0)}{\sqrt{n}}\left[\frac{l''_n(\boldsymbol{\theta}_0)}{n}\right]^{-1}-\frac{1}{\sqrt{n}}\|\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|E\left(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0\right)\left[\frac{l''_n(\boldsymbol{\theta}_0)}{n}\right]^{-1}. \tag{3.6}$$

By the strong law of large numbers, $\frac{l''_n(\boldsymbol{\theta}_0)}{n}$ converges to a constant matrix given by $E_{\boldsymbol{\theta}_0}\left[X''_1(\boldsymbol{\theta}_0)\right]=g''(\boldsymbol{\theta}_0)$. By Theorem 3.1, $\frac{1}{\sqrt{n}}\|\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|E\left(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0\right)$ goes to zero a.s., and hence so does the second term on the RHS of (3.6). As a consequence of the standard multivariate central limit theorem for i.i.d. random variables, $\frac{\mathbf{l}'_n(\boldsymbol{\theta}_0)}{\sqrt{n}}$ goes to a multivariate normal distribution with mean zero and dispersion Σ_1 and the desired result follows.

4. Nonparametric estimation for censored circular data

4.1. Self-consistent estimator

Let $A_i, i = 1,\dots,n$, be a sequence of independent identically distributed (i.i.d.) circular random variables with an unknown distribution F_0. Let (L_i, R_i) be a sequence of i.i.d. random vectors, independent of $A_i's$ and both components taking values in $[0, 2\pi)$, with unknown bivariate distribution. While A denotes the variable of interest, (L_i, R_i) represents the censoring mechanism. We observe A_i when $A_i \notin (L_i, R_i)$ and the censoring arc (L_i, R_i) when $A_i \in (L_i, R_i)$ i.e. we either observe the original value A_i if there is no censoring or the censoring arc (L_i, R_i) when there is censoring.

Note 1: We assume that A_i, L_i as well as R_i take values in the interval $[0, 2\pi)$. Since there is no natural ordering on the circle, an R_i may have a smaller value than the corresponding L_i, as when the censoring interval straddles the zero direction. To cover this situation when $r < l$, we say an angle α belongs to the arc (l, r) if either $r < l < \alpha$ or $\alpha < r < l$. In the more standard situation when $l < r$, it is clear that an angle α belongs to an arc (l, r) if $l < \alpha < r$. With this definition, for instance the observed angular values of 10 or 356 still fall inside the censoring arc which goes from 355 to 18.

Note 2: Any distribution function on the circle is first defined on $[0, 2\pi)$ and then extended to the rest of the real line by the relationship $F(x + 2\pi) = 1 + F(x)$ for all $x \in \mathbb{R}$.

In many censoring situations, if we were to try to estimate the distribution function via the EM algorithm, the result is that of equating F with the conditional expectation under F of the empirical distribution function given the data. The resulting equation in our case takes the form

$$F(t) = E_F\left[(\mathbf{E_n})(\mathbf{t})\,|\mathbf{A_i}, (\mathbf{L_i}, \mathbf{R_i})\right]$$

where $\mathbf{E_n}$ is the empirical distribution function. This equation was referred to as self-consistency equation by Efron [6]. In the case of censored circular data, the SCE $\hat{F}$ satisfies the equation

$$F(t)=\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i I(A_i\le t)+\bar{\delta}_i I(R_i\le t)+\bar{\delta}_i I[t\in(L_i,R_i)]\frac{F(t)-F(L_i)}{F(R_i-)-F(L_i)}\right\} \tag{4.7}$$

where $\delta_i = I[A_i \notin (L_i, R_i)]$ and $\bar{\delta}_i = 1 - \delta_i$. As is the case with many types of censored data, there is no explicit closed form solution to the equation and has to be computed by the iterative formula

$$\hat{F}^{(m+1)}(t)=\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i I(A_i\le t)+\bar{\delta}_i I(R_i\le t)+\bar{\delta}_i I[t\in(L_i,R_i)]\frac{\hat{F}^{(m)}(t)-\hat{F}^{(m)}(L_i)}{\hat{F}^{(m)}(R_i-)-\hat{F}^{(m)}(L_i)}\right\}.$$

The convergence of the algorithm is assured by Theorem 2.1 of Tsai and Crowley [7] provided that the initial estimator gives positive mass to all observed points. For a general discussion on self-consistency and its relation to EM algorithm, see Tarpey and Flury [8].

In order to avoid the confusion that stems from having to deal with an arbitrary starting point when we talk about distribution functions on the circle, we convert our self-consistency equation to the probability mass function form. (We refer to a censoring interval as "empty" if it contains no other uncensored observations.) Before doing this, we need the following

Proposition 4.1. *Any distribution function that satisfies (4.10) attaches all its mass on the uncensored observations and empty censoring intervals. If empty intervals are replaced by arbitrary points in them, any SCE for the new system will also be an SCE for the old system. Consequently, if each observed censored interval (L_i, R_i) contains at least one uncensored observation X_j, $j \neq i$, then any distribution function that satisfies (4.10) attaches all its mass on these uncensored observations.*

Proof is similar to the proof of Proposition 1 in Jammalamadaka and Mangalam [9].

In the light of the above proposition, we may replace all the empty intervals by their midpoints and treat them as uncensored observations. Now all mass of any SCE will be concentrated on these points. Let these points be denoted by $x_1, x_2, \ldots, x_m$ and let $p_1, p_2, \ldots, p_m$ be the mass attached to them by the SCE. From the self-consistency equation it follows that

$$p_j = \frac{1}{n} + \frac{p_j}{n} \sum_{i=1}^{n} \left\{ \frac{(1-\delta_i) I[x_j \in (L_i, R_i)]}{\sum\limits_{\{k: x_k \in (L_i, R_i)\}} p_k} \right\}. \tag{4.8}$$

This can also be rewritten as

$$1/p_j = n - \sum_{i=1}^{n} \left\{ \frac{(1-\delta_i) I[x_j \in (L_i, R_i)]}{\sum\limits_{\{k: x_k \in (L_i, R_i)\}} p_k} \right\}. \tag{4.9}$$

Now we find the self-consistent estimator by iterating this equation.

If it so happens that a censored arc contains no uncensored observation, we are in a situation similar to that of right-censored data where the largest observation is censored. While in the right-censored case the extra mass is usually left unassigned, for censored circular data there is a natural way of handling this. When a censored arc contains no uncensored points, we let the mass that corresponds to that arc be assigned to its midpoint. Thus our initial estimator may give equal mass to all uncensored observations and to the midpoints of the censoring arcs that contain no uncensored observations. But convergence of the iteration would be faster if no mass is assigned to the midpoints of non-empty intervals.

Remark. In the light of above discussion, it would be sub-optimal (does not lead to maximization of the likelihood) to simply replace each censored observation by the midpoint of the censoring interval irrespective of whether it contains any other points, although this could be lead to a crude and simple estimate.

4.2. Nonparametric maximum likelihood estimation

The SCE, being a result of convergence of the EM algorithm, provides a local maximum of the likelihood equation [see, for example, Ref. [10]] and may not coincide with the NPMLE. Trivial examples of cases when an SCE is not the NPMLE can be constructed by considering situations where two empty censoring intervals overlap. For instance, if we have 1, 2, (3, 4), (3.8, 5) as the data, we could assign 0.25 mass to 1, 2, 3.5 and 4.4 to get an SCE. The NPMLE will assign 0.25 each on 1 and 2, but assign 0.5 on some point, say 3.9, on the overlap area (3.8, 4). Both estimators are self-consistent, but the latter has higher likelihood. This happens whenever there are empty, overlapping intervals. Let $\mathcal{F}$ denote the set of all distribution functions on the line. For $F \in \mathcal{F}$ the likelihood of the sample is given by

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)]^{\delta_i} [F(R_i-) - F(L_i)]^{1-\delta_i}.$$

The distribution that maximizes the likelihood function is the Nonparametric Maximum Likelihood Estimator (NPMLE). Interestingly, the maximizer of the Nonparametric likelihood function will automatically be a self-consistent estimator.

Table 1
Simulation results for $\hat{\boldsymbol{\theta}}$ when $\kappa = 1$.

Length of censoring arc	Estimation of μ				Estimation of κ	
	Average		Mean direction			
	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\kappa}$	$\tilde{\kappa}$
1	1.98504	1.98512	1.98522	1.9853	1.04659	1.0464
3	2.00533	2.0033	2.00561	2.00331	1.03721	1.03091

Table 2
Simulation results for $\hat{\boldsymbol{\theta}}$ when $\kappa = 3$.

Length of censoring arc	Estimation of μ				Estimation of κ	
	Average		Mean direction			
	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\kappa}$	$\tilde{\kappa}$
1	1.99318	1.99461	1.99318	1.99461	3.10974	3.10108
3	2.0081	2.00651	2.0081	2.0065	3.08278	3.07627

Theorem 4.1. *The NPMLE satisfies the equation*

$$F(t) = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i I(X_i \le t) + \bar{\delta}_i I(R_i \le t) + \bar{\delta}_i I[t \in (L_i, R_i)]\frac{F(t) - F(L_i)}{F(R_i -) - F(L_i)}\right\}. \tag{4.10}$$

For proof see Theorem 1 in Jammalamadaka and Mangalam [9].

Consider the following example with $n = 5$ where the data set is $\{0.2, 0.4, 0.6, (0.1, 0.5), (0.3, 0.7)\}$. Let p_1, p_2, p_3 be the masses to be assigned to 0.2, 0.4, 0.6 respectively. The likelihood function is given by

$$p_1 \cdot p_2 \cdot p_3 \cdot (p_1 + p_2) \cdot (p_2 + p_3)$$

and, as p_i's add up to 1 and the roles of p_1 and p_3 are interchangeable, we can simplify the problem to that of maximizing $(x^2)(1-2x)(1-x)^2$ with $p_1 = p_3 = x$ and $p_2 = 1-2x$. The solution, then, is given by $x = (5-\sqrt{5})/10$ so that $p_1 = p_3 = (5-\sqrt{5})/10$ and $p_2 = 1/\sqrt{5}$ is the solution to the NPMLE. In this example the iterations of the self-consistency equation rapidly converged to the NPMLE.

From Theorem 3.1 and from Proposition 4.1, it follows that the NPMLE will put all its mass on uncensored observations and censored intervals that are empty. But if we replace empty intervals by points in them, say by their midpoints, the resulting system may have a different NPMLE.

4.3. *Data simulation*

Extensive simulation studies were conducted to check the performance of the estimators of the mean direction μ and the concentration parameter κ. In all cases, the data were generated from $vM(2, \kappa)$ and the censoring was done by arcs of fixed length whose starting points were generated from a uniform distribution over the circle. Simulations were done for various combinations of κ and a, the length of the censoring arc. We did not vary the values of μ because of its location invariance property. For each combination of the choice of a and κ, we generated 200 samples of size 100 each and computed the value of the estimators. Average and the mean direction for the 200 values of the MLE of μ were computed along with the average for the MLE of κ. (With the choice of $\mu = 2$, all the estimators obtained were between 0 and π and there were no straddling zero, so computing the average value made sense.) For comparison, the corresponding values for $\tilde{\boldsymbol{\theta}}$, the uncensored estimator are given. The results, given in Tables 1 and 2, are quite favorable and the estimators performed well even under strong censorship.

For censoring of this sort, there is a natural competitor for the MLE of μ. One may replace all the censoring arc by its midpoint, treat them as part of the uncensored data and compute the estimates of

Table 3
Comparison between MLE and MME for μ when $\kappa = 0.1$.

Length of censoring arc	Average deviation			Maximum deviation		
	MLE	MME	$\tilde{\mu}$	MLE	MME	$\tilde{\mu}$
1	1.08785	1.09174	1.07432	3.10065	3.10629	3.13887
3	1.13732	1.16591	1.02863	3.0934	3.14097	3.00418

Table 4
Comparison between MLE and MME for μ when $\kappa = 1$.

Length of censoring arc	Average deviation			Maximum deviation		
	MLE	MME	$\tilde{\mu}$	MLE	MME	$\tilde{\mu}$
1	0.118486	0.118316	0.11699	0.47819	0.481383	0.468343
3	0.147316	0.147739	0.110262	0.567545	0.534483	0.40235

Table 5
Comparison between MLE and MME for μ when $\kappa = 4$.

Length of censoring arc	Average deviation			Maximum deviation		
	MLE	MME	$\tilde{\mu}$	MLE	MME	$\tilde{\mu}$
1	0.0411071	0.0414139	0.0388204	0.156435	0.155863	0.1573
3	0.0506923	0.0658454	0.040262	0.182545	0.252105	0.162144

the parameters from this data set in the usual way. We investigated the performance of this estimator that we call the MME, *the method of midpoints estimator* and found that the MLE does better, though only slightly when the censoring is mild. As censoring gets stronger, the MLE does significantly better than the MME. Tables 3–5 give the average and maximum deviations from the true value of μ for these estimators for various values of κ. It also gives these for the uncensored estimator $\tilde{\mu}$, calculated before censoring is applied. All generated data come from $vM(0, \kappa)$, where κ takes values 0.1, 1 and 4. $m = 200$ samples of size $n = 100$ were taken and each sample is censored by arcs of uniformly distributed starting points and fixed length a. For each of these samples, MLE, MME and $\tilde{\mu}$ were computed and the average and maximum deviations of these estimators from 0 were calculated. For all theses cases, estimation of μ is done with the assumption that κ is known. The lengths of the censoring arcs were taken to be 1 and 3.

We also looked into the situation where data from von Mises distribution were censored by a fixed arc of length a for various starting points on the circle. The essential difference here from the previous study is that here the censoring arc is the same for all observations. We took a sample of size 10 000 observations from $vM(2, 0.5)$ and censored it with fixed intervals of length 2, spanning the circle. The average deviation and maximum deviation from μ were computed. The average deviations for MLE and MME were 0.03887 and 0.14750 respectively while the maximum deviations for MLE and MME were 0.08949 and 0.34939 respectively. This shows that the MLE is startlingly efficient compared to the MME.

Even though our simulations assumed that κ was known when the estimate of μ was computed, this turns out to be not a critical issue. When κ is unknown, one may use any reasonable value k in its place for the purpose of estimating μ. In our algorithm, varying the value of k had very little effect on the estimated value of μ. This means that in the cases where μ is all that we are interested in and κ is an unknown nuisance parameter, we can take a moderate value k, say $k = 1$, in place of true κ and the results for μ are not seriously affected. This robustness with respect to κ was demonstrated in various simulations which are not reproduced here.

Nonparametric estimation of $\hat{\boldsymbol{\theta}}$ also gave us good results, though, as expected, not as good as the parametric estimation. We conclude from this that if we know that the data come from von Mises distribution, then it is better to use the parametric method.

Acknowledgments

Authors thank the anonymous referees for valuable comments which improved the presentation.

References

[1] S.K. Iyer, S. Rao Jammalamadaka, D. Kundu, Analysis of middle-censored data with exponential lifetime distributions, J. Statist. Plann. Inference 138 (2008) 3550–3560.
[2] K.V. Mardia, P.E. Jupp, Directional Statistics, Wiley, New York, 2000.
[3] S. Rao Jammalamadaka, A. SenGupta, Topics in Circular Statistics, World Scientific Press, Singapore, 2001.
[4] C.R. Rao, Linear Statistical Inference and its Applications, second ed., John Wiley and Sons, New York, 1973.
[5] T.M. Apostol, Calculus, vol. II, John Wiley and Sons, New York, 1967.
[6] B. Efron, The two-sample problem with censored data, in: Proc. Fifth Berkeley Symp., in: Math. Stat. Probab., vol. 4, University of California Press, Berkeley, 1967, pp. 831–853.
[7] W.Y. Tsai, J. Crowley, A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency, Ann. Statist. 13 (1985) 1317–1334.
[8] T. Tarpey, B. Flury, Self-consistency: A fundamental concept in statistics, Statist. Science 11 (1996) 229–243.
[9] S. Rao Jammalamadaka, V. Mangalam, Nonparametric estimation for middle censored data, J. Nonparametr. Stat. 15 (2003) 253–265.
[10] P.A. Mykland, J. Ren, Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data, Ann. Statist. 24 (1996) 1740–1764.