Circular-circular regression model with a spike at zero

Jayant Jha

and

Atanu Biswas

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Circular-circular regression model with a spike at zero

Jayant Jha and Atanu Biswas

Applied Statistics Unit, Indian Statistical Institute

203 B. T. Road, Kolkata - 700 108, India

E-mails: jayantjha@gmail.com and atanu@isical.ac.in

Abstract

With reference to a real data on cataract surgery, we discuss the problem of zero-inflated circular-circular regression when both covariate and response are circular random variables and a large proportion of the responses are zeros. The regression model is proposed and the estimation procedure for the parameters is discussed. Some relevant test procedures are also suggested. Simulation studies and real data analysis are carried out to illustrate the applicability of the model.

Some keywords and phrases: Möbius transformation; EM algorithm; von Mises distribution; Wrapped Cauchy distribution.

1 Introduction

Circular random variables are topologically quite different from the linear random variables, and hence the statistical techniques developed for the linear random variables can not be directly applied to the case of circular random variables. The circular random variables can be considered to be the directions which can be represented as angles as well as unit complex numbers corresponding to the angles. The need of developing statistical methods for the study of circular random variables is important for many biomedical problems due to its wide application in the field of ophthalmology and orthopaedics. The paper deals with the case of one such example where the dataset is the angle of astigmatism induced by two

popular methods of Small Incision Cataract Surgery (SICS), namely the irrigating Vectis technique (see Masket [1]) and the Snare technique (see Basti *et al.* [2]). Quite naturally, here the data is of circular nature, and we observe a quite high proportion (13 out of the 37 responses) of zeros (indicating no astigmatism). The methodology for analysing such a data set is considered in this present paper. We develop a general theory for zero-inflated circular-circular regression in this context.

Circular-circular regression deals with the prediction of a circular random variable conditional on another circular random variable. Link function of a circular-circular regression is of different nature than the classical linear regression setup. In Rivest [3], the circular dependent variable is regressed on an independent variable by decentering the dependent variable. Other models for this purpose are explained in Downs and Mardia [4], Minh and Farnum [5] and Hussin *et al.* [6]. Kato *et al.* [7] discussed a circular-circular regression by using the Möbius transformation link function where the antipodal point of the covariate is projected via a point to another point on the circle. The rotational model is a subset of this model and the Möbius Transformation has some nice distributional and geometric properties as discussed in Kato *et al.* [7]. In this present paper, we use the link function of the circular-circular regression model proposed in Kato *et al.* [7] and extended it to the case when many of the responses are exactly the same (here zero).

Zero-inflated regression models for discrete linear data (when there are many zeros) are available in the literature. Such type of model was introduced in Lambert [8]. See Angers and Biswas [9], Dietz and Böhning [10], and the references therein for the most popular zero-inflated discrete model, namely the zero-inflated Poisson. In contrast, not much works have been done in the continuous setup, see Tian [11] and Labrecque-Synnott and Angers [12] among the very few exceptions in this context. The zero-inflated setup for circular data is much more complicated due to the circular nature of the response variable, the zeros can not be assumed to be on one end of the support. Thus, a different procedure is applied to approximate the true model in the case of circular-circular regression. So far our knowledge goes, the recent work by Biswas *et al.* [13] is the only work in this direction, that too in the absence of covariates. In the present paper, we consider regression modelling of circular data with a spike at a particular point (say, at zero) where the covariate is also considered in the model, and moreover the covariate is of circular nature.

The rest of the paper is organized in the following way. Section 2 introduces the Möbius transformation based circular-circular regression model with a spike at zero. In Section 3, estimation of the underlying model parameters is discussed. Some test procedures are also discussed in this context. Simulation studies are carried out in Section 4. The cataract surgery data set is analysed in Section 5. Section 6 concludes.

2 The model

In the discrete count data setup, zero-inflated Poisson regression was introduced in [8] with more zeros among the responses than can be explained by a Poisson model. In the zero-inflated regression model, response is assumed to follow a mixture distribution; a mixture of a degenerate distribution at zero with some probability p, and a distribution having mean dependent on the covariates with probability $(1 - p)$.

In the setup of linear continuous distributions, the zero-spike model can be defined as a mixture of a Dirac mass at 0 with probability p, and a density $f_1(y)$ defined on $\mathbb{R}^+$ with probability $1 - p$. There are very few works in this setup; most of the works are on the lognormal and gamma models. A short literature review is available in [11]. Subsequently Labrecque-Synnott and Angers [12] used the model for the Laplace distribution truncated on $[0, \infty)$. In the regression setup with zero-inflated continuous distribution, the model proposed in [14] has been used extensively. The assumption of the Tobit model is that there is a latent variable Y^* taking values on the real line and when $Y^* \leq 0$, the response Y is denoted as 0. Since then, there have been various generalisations of the Tobit model such as [15], [16], [17], [18] and [19]. As the zeros are only on one side of the responses in the Tobit model, this model can not be applied in the circular case directly. Also, the reason for the zero-inflation of the responses in the Tobit model of [14] is that the zeros are coming from the same regression function but due to the left-censoring of the latent variable, the observed zero is a non-positive value of the unobserved latent variable which is very different from our assumption regarding the model. In this paper, we have assumed that some of the zeroes occur independently of the covariates while some are coming due to the measurement error where the response is noted as 0 whenever the true response is within a small range around 0.

Zero-inflated circular distribution is a newly studied research topic. Motivated by the model of [12], recently Biswas *et al.* [13] introduced von Mises distribution with a spike at zero. But, regression problem is not considered in [13]; no covariate is considered in the modelling.

Circular-circular regression can be defined as follows. Let $\Omega = \{z : z \in \mathbb{C}, |z| = 1\}$ and $\theta_x, \theta_y \in [0, 2\pi)$ are the covariate and the response, respectively, and let $x, y \in \Omega$ be the unit complex numbers corresponding to θ_x, θ_y. Then, the circular-circular regression model of [7] is defined as

$$y = \beta_0 \frac{x + \beta_1}{1 + \overline{\beta_1} x} \epsilon, \tag{2.1}$$

where $\beta_0, \epsilon \in \Omega$ and $\beta_1 \in \mathbb{C}$. Kato *et al.* [7] used the Wrapped Cauchy distribution due to some of its elegant properties. The regression function is a form of Möbius transformation; it is a mapping of unit circle $|z| = 1$ onto itself. In the model (2.1), β_0 is a rotation parameter while β_1 is a fixed point in the complex plane which converts the point x to a point x_β on the unit circle which is the intersection of the unit circle with the line joining $-x$ and β_1. See Figure 1. The case of $|\beta_1| > 1$ is also discussed in [7]. In this case, first the fixed point is taken to be $\frac{1}{\overline{\beta_1}}$ and then it is joined to $\frac{\beta_1}{|\beta_1|} \frac{\beta_1}{|\beta_1|} \overline{x}$. The intersection of this line with the unit circle is taken as x_β. Thus, if $|\beta_1|$ is closer to 1, the function results in the resultant values of x_β concentrating around $\frac{\beta_1}{|\beta_1|}$. The role of β_0 is then to induce a rotation of x_β to an angle α on the unit circle. The resultant point is γ, say. Then, y has the same error distribution as ϵ, but with the circular mean γ. For example, if ϵ follows a Wrapped Cauchy distribution with circular mean 0, then y follows a Wrapped Cauchy distribution with circular mean γ. See Figure 1 for the illustration of the model 2.1. Clearly, when $|\beta_1| = 0$, we have $\gamma = \beta_0 x$; which is just a rotation.

In the present paper, we extend the theory of modelling of Biswas *et al.* [13] to the regression case. Let $(\theta_{y_i}, \theta_{x_i})$, $i = 1, \dots, n$, be the observations such that a large number of the responses θ_y's are 0. Then, using the circular-circular regression model of [7], as given in (2.1), the density of θ_{y_i} given θ_{x_i} can be written as

$$f(\theta_{y_i} | \theta_{x_i}) = p.f_0(\theta_{y_i}) + (1 - p)f_1(\theta_{y_i} | \theta_{x_i}), \tag{2.2}$$

where $p \in [0, 1]$, and $f_0(\theta_{y_i})$ is a known density with mean direction 0 and a very high concentration parameter (i.e. approximately a degenerate distribution at 0), and $f_1(\theta_{y_i} | \theta_{x_i})$
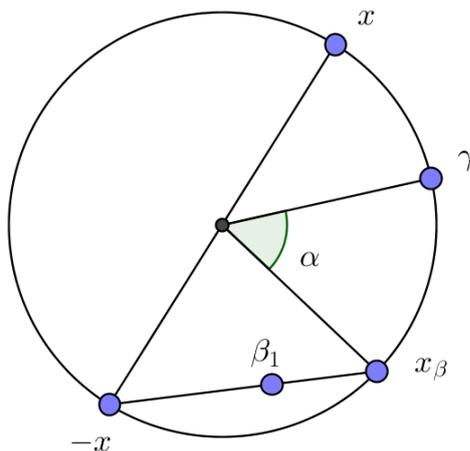
5

Figure 1: Circular-circular regression model.

has a distribution with unknown parameters. Here we consider $f_1(\theta_{y_i}|\theta_{x_i}) = vM(\mu_{y_i,x_i}, \kappa)$, the von Mises distribution with mean direction μ_{y_i,x_i} and concentration parameter κ. Also, $\mu_{y_i,x_i} = \arg\left(\beta_0 \frac{x_i + \beta_1}{1 + \beta_1 x_i}\right)$. Any other suitable distribution like the Wrapped Cauchy distribution can be assumed for $f_1(\cdot|\cdot)$.

An equivalent representation of the aforementioned regression link function is given in [4]. The regression model in [4] can be written as:

$$\theta_y = \beta + 2\tan^{-1}(\omega \tan \frac{(\theta_x - \alpha)}{2}) + \theta_\epsilon$$

where, $\alpha, \beta \in [0, 2\pi), \omega \in [-1, 1]$ and ϵ is the angular error. The theory given in the paper is equally valid if we take this regression model.

3 Related inference

3.1 Estimation of parameters

EM algorithm (cf. Dempster *et al.* [20]) can be used to find the MLE of the parameters of a mixture regression model . Suppose we have to estimate the parameters of the mixed regression model where

$$f(y|x) = pf_1(y|x, \theta_1) + (1 - p)f_2(y|x, \theta_2)$$

We consider a latent variable Z which takes the value 1 with probability p and 0 with probability $1-p$, and represent Z to indicate whether an observation Y has density $f_1(\cdot|x, \theta_1)$

or $f_2(\cdot|x, \theta_2)$. Then,

$$f(y, z|x, \theta) = [p.f_1(y|x, \theta_1)]^z [(1 - p)f_2(y|x, \theta_2)]^{1-z} .$$

where $\theta_j = (\beta_0(j), \beta_1(j), \kappa_j)$ for j=1,2. Writing

$$A_i^k = P(Z_i = 1|p^{(k-1)}, \beta_1^{(k-1)}, \kappa^{(k-1)}, \beta_0^{(k-1)}),$$

where $p^{(k-1)}$, for example, is the estimate of p obtained at the $(k-1)$th iteration, by EM algorithm, the parameters can be estimated iteratively as follows. Given the parameter estimates $p^{(k-1)}$ and $\theta_1^{(k-1)}, \theta_2^{(k-1)}$ obtained from the $(k-1)$th iteration, let

$$P^{(k)}(p, \theta_1, \theta_2) = \int_z [\log L(p, \theta_1, \theta_2|x, y, z)] f\left(z|x, y, p^{(k-1)}, \theta_1^{(k-1)}, \theta_2^{k-1}\right) dz$$

be the expectation of the log-likelihood with respect to the latent variables. Then

$$A_i^{(k)} = \frac{p^{(k-1)} \ f_1(y_i|x_i, \theta_1^{(k-1)})}{p^{(k-1)} \ f_1(y_i|x_i, \theta_1^{(k-1)}) + (1 - p^{(k-1)})f_2(y_i|x_i, \theta_2^{(k-1)})}.$$

Taking partial derivative of $P^{(k)}$ with respect to the parameters we get

$$\frac{\partial P^{(k)}}{\partial p} = \frac{\sum_{i=1}^n A_i}{p} - \frac{\sum_{i-1}^n (1 - A_i)}{1 - p},$$

$$\frac{\partial P^{(k)}}{\partial \theta_j} = \sum_{i=1}^n (1 - A_i)\frac{\partial \log f_j(y_i|x_i, \theta_j)}{\partial \theta_j}.$$

for $j = \{1, 2\}$. Solving these we get $\widehat{p}^{(k)}$, the kth iteration of p, as

$$\widehat{p}^{(k)} = \frac{\sum_{i=1}^n A_i^{(k)}}{n}.$$

Also,

$$\frac{\partial P^{(k)}}{\partial \kappa_1} = \sum_{i=1}^n (1 - A_i^{(k)})\frac{\partial \log f_1(y_i|x_i, \theta_1)}{\partial \kappa_1}$$

and

$$\frac{\partial P^{(k)}}{\partial \kappa_2} = \sum_{i=1}^n (1 - A_i^{(k)})\frac{\partial \log f_2(y_i|x_i, \theta_2)}{\partial \kappa_2}$$

implies

$$A(\kappa^{(k_j)}) = \frac{\sum_{i=1}^n (1 - A_i^{(k)}) \cos\left(\theta_{y_i} - \mu_{j,y_i,x_i}^{(k)}\right)}{\sum_{i=1}^n (1 - A_i^{(k)})},$$

for j=1,2. where $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$, with $I_j(\cdot)$ is the modified Bessel function of order j. So, given $\beta_1^{(k)}, \beta_0^{(k)}$, one can find $\kappa^{(k)}$ at each stage of the iteration. Finally, the estimates are obtained upon convergence.

In our case, we have taken $\theta_1 = (1, 1, \kappa_0)$ where, κ_0 is very large and is assumed to be known, $\theta = \theta_2$ and p have to be estimated. Thus, the steps of EM algorithm can be rewritten as:

$$f(y, z | x, \theta) = [p.f_0(y)]^z [(1 - p) f_1(y|x, \theta)]^{1-z}.$$

Given the parameter estimates $p^{(k-1)}$ and $\theta^{(k-1)}$ obtained from the $(k-1)$th iteration, let

$$P^{(k)}(p, \theta) = \int_z [\log L(p, \theta | x, y, z)] f\left(z | x, y, p^{(k-1)}, \theta^{(k-1)}\right) dz$$

be the expectation of the log-likelihood with respect to the latent variables. Then

$$A_i^{(k)} = \frac{p^{(k-1)} f_0(y_i)}{p^{(k-1)} f_0(y_i) + (1 - p^{(k-1)}) f_1(y_i | x_i, \theta^{(k-1)})}.$$

Writing

$$A_i^k = P(Z_i = 1 | p^{(k-1)}, \theta_1^{(k-1)}, \theta_2^{(k-1)}),$$

where $p^{(k-1)}$, for example, is the estimate of p obtained at the $(k-1)$th iteration, by EM algorithm, the parameters can be estimated iteratively as follows. As p is non-negative, we write

$$f(y|x, p, \theta) = p.f_0(y) + (1 - p) f_1(y|x, \theta),$$

where $\theta = (\beta_0, \beta_1, \kappa)$. Given the parameter estimates $p^{(k-1)}$ and $\theta_1^{(k-1)}$ obtained from the $(k-1)$th iteration, let

$$P^{(k)}(p, \theta) = \int_z [\log L(p, \theta | x, y, z)] f\left(z | x, y, p^{(k-1)}, \theta^{(k-1)}\right) dz$$

be the expectation of the log-likelihood with respect to the latent variables. Then

$$A_i^{(k)} = \frac{p^{(k-1)} f_0(y_i)}{p^{(k-1)} f_0(y_i) + (1 - p^{(k-1)}) f_1(y_i | x_i, \theta^{(k-1)})}.$$

Taking partial derivative of $P^{(k)}$ with respect to the parameters we get

$$\frac{\partial P^{(k)}}{\partial p} = \frac{\sum_{i=1}^n A_i}{p} - \frac{\sum_{i-1}^n (1 - A_i)}{1 - p},$$

$$\frac{\partial P^{(k)}}{\partial \theta} = \sum_{i=1}^n (1 - A_i) \frac{\partial \log f_1(y_i | x_i, \theta)}{\partial \theta}.$$

Solving these we get $\widehat{p}^{(k)}$, the kth iteration of p, as

$$\widehat{p}^{(k)} = \frac{\sum_{i=1}^n A_i^{(k)}}{n}.$$

Also,

$$\frac{\partial P^{(k)}}{\partial \kappa} = \sum_{i=1}^{n}(1 - A_i^{(k)})\frac{\partial \log f_1(y_i|x_i,\theta)}{\partial \kappa}$$

implies

$$A(\kappa^{(k)}) = \frac{\sum_{i=1}^{n}(1 - A_i^{(k)})\cos\left(\theta_{y_i} - \mu_{y_i,x_i}^{(k)}\right)}{\sum_{i=1}^{n}(1 - A_i^{(k)})},$$

where $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$, with $I_j(\cdot)$ is the modified Bessel function of order j. So, given $\beta_1^{(k)}, \beta_0^{(k)}$, one can find $\kappa^{(k)}$ at each stage of the iteration. Finally, the estimates are obtained upon convergence.

3.2 Test for p

The testing for $H_{0p} : p = 0$ versus $H_{1p} : p > 0$ can be done to check whether the proportion of zeros is significant or not in the data, and whether or not the zero-inflated model is needed at all.

Let L_{0p} is the log-likelihood under H_{0p}, and L_{1p} is the log-likelihood under $H_{0p} \cup H_{1p}$. Taking $\Lambda_p = -2(L_{0p} - L_{1p})$, as in Self and Liang [21], asymptotically $\Lambda_p \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Here, χ_0^2 denotes a degenerate distribution at 0.

3.3 Test for $\arg(\beta_0\beta_1) = 0$

The approximation of taking a unimodal distribution with mean direction 0 and a high concentration parameter is more accurate if, in the mixture distribution (2.2), the probability of 0-values coming from the unknown distribution is very low compared to the fixed distribution, and also the probability of an angle other than zero coming from the unknown distribution is very high compared to the fixed distribution. As explained in [7], the regression link function in (2.1) is such that the concentration of points around $\arg(\beta_0\beta_1)$ is more if the covariate is uniformly distributed. Thus, it can be argued that the approximation will be better if $\arg(\beta_0\beta_1)$ is far from 0. This can also be seen from the simulation results in Section 4; see Table 1 for example. Thus, a test of $H_{0\beta} : \arg(\beta_0\beta_1) = 0$ against $H_{1\beta} : \arg(\beta_0\beta_1) \neq 0$ can be performed to check if the chosen concentration parameter is good enough to approximate the true distribution.

Let $L_{0\beta}$ is the log-likelihood under $H_{0\beta}$, and $L_{1\beta}$ is the log-likelihood under $H_{0\beta} \cup H_{1\beta}$. Then $\Lambda_\beta = -2(L_{0\beta} - L_{1\beta})$ can be used as the test statistic, where asymptotically $\Lambda_\beta \sim \chi_1^2$.

4 Simulation

In this Section, we carry out a detailed simulation study for the zero-inflated circular-circular regression model. We consider different values of the parameters at three sample sizes (n) to check the performance of the proposed model. We considered the worst cases in the simulations when zero's are randomly generated from a degenerate distribution with probability p and with (1-p) probability responses come from the regression model. Thus, we study the behaviour of our approximation for estimation of parameters in the worst possible scenarios as some values near zero may come from the regression function also. The small sample behavior is studied with $n = 35$, which is close to the sample size of our data example of Section 5. The effect of moderate sample size is studied for $n = 70$, while $n = 200$ gives the large sample behavior. In the estimation, the fixed density f_0 is considered to be the von Mises density with mean 0 and the concentration parameter 370 (the largest value that can be taken in R which enabled us to estimate the parameters by EM algorithm). The covariates are generated uniformly between $[0, 2\pi)$ in all the cases. The simulation results are presented in Tables 1-4. In the Tables, we consider $\beta_1 = b_1 + ib_2$ and $\beta_0 = \exp(i\theta_0)$. Then, the mean value of the estimates are taken over 10,000 simulations and are reported (with standard errors in parentheses) for the parameters b_1, b_2, p and κ. For θ_0, the circular mean is reported and the circular variance (which is *one minus mean resultant length*) is reported in parentheses.

In Table 1, we vary $\arg(\beta_1)$, keeping other parameter values fixed, for $n = 35$ and 70. It can be seen that, for the fixed $\beta_0 = 1$, as $\arg(\beta_1)$ goes far from 0, the approximation gets better. Table 2 shows the results for the simulation when the regression model is the one given in [4]. It can be seen from the table, that for this model also, the results are generally good. It can be seen from the table that at $\beta = 0$, when ω moves away from zero, the fit gets better. We also found that as β moves away from 0, the estimates improve as the predicted values from $f_1(.)$ move away from 0. It was also observed that as p gets smaller, the approximation gets better. By varying θ_0 at a fixed value of $\arg(\beta_1) = 0$, we observed that results become more accurate as θ_0 deviates from the spike, that is 0, the observations from the unknown density is most clearly distinguishable in that case. The best results are obtained when $\theta_0 = \pi$ when $\arg(\beta_1) = 0$. We also observed that as κ gets larger, the approximation is improved. The simulation results also showed that the approximation gets better as n

increases, which is an expected scenario. We performed some simulations for tests mentioned in Section 3.2 and 3.3. The results showed that even for a very small positive value of mixing parameter p, the p-value for H_{0p} against H_{1p} is large enough so that we cannot reject the null hypothesis. The results for testing $H_{0\beta} : \arg(\beta_0\beta_1) = 0$ against $H_{1\beta} : \arg(\beta_0\beta_1) \neq 0$ showed very small p-values when $\arg(\beta_0\beta_1)$ is as far as $\pi/2$ or π.

The angular error in the circular-circular regression model developed in [7] follows the wrapped Cauchy distribution. We have used the same link function with the angular error following von Mises distribution. When the error follows Wrapped Cauchy distribution, then a model mixed with a Wrapped Cauchy distribution with the known value of parameter ρ very close to 1 may also be used for the fixed parameter part in the zero-inflated regression setup. As the Wrapped Cauchy distribution has a heavy tail, thus, we have compared the two distributions based on CPRS as mentioned in [22] to check which one of the distribution goes better with the approximation for the true model. We have taken training data at two sample sizes $n = 35$ and $n = 70$ and test data of size 40% of the training data. 10000 simulations are performed taking two values for the κ, when the true error is von Mises and two values of ρ when the true model is Wrapped Cauchy. The result of the comparison is shown in Table 3. It may be interpreted from the Table that the von Mises distributional assumption for the fixed distribution outperforms the Wrapped Cauchy assumption in each case which is probably due to heavy-tail of the Wrapped Cauchy distribution. The fixed value of the Wrapped Cauchy known parameter is considered to be 0.999.

4.1 Fixing the high-valued concentration parameter

In the paper, it has been mentioned through simulations that we can approximate the zero-spike model by taking a distribution with high concentration having a fixed value and a regression function with unknown parameters. EM algorithm can then easily be used to estimate the parameters. The validity of the approximation can be shown theoretically in the following way:

Our approximate likelihood is

$$L = \Pi_{i=1}^{n}(pf_0(y) + (1-p)f_1(y_i/x_i; \beta_0, \beta_1, k)$$

We have taken the meaning of the spike as when $Y_i \in (-\alpha, \alpha)$, we observe $Y_i = 0$. Also, in the EM algorithm, the event $Z_i = 1$ represents that the observation comes from f_0 and

11

$Z_i = 0$ represents that the observation comes from $f_1()$. When $Y_i = 0$,

$$P_a(Z_i = 1) = \frac{pf_0(y)}{pf_0(y) + (1-p)f_1(y_i/x_i; \beta_0, \beta_1, k)} \qquad (4.1)$$

approximates the true likelihood. For the true likelihood,

$$P_t(Z_i = 1) = \frac{p}{p + (1-p)\int_{-\alpha}^{\alpha} f_1(y_i/x_i; \beta_0, \beta_1, k)} \qquad (4.2)$$

When $Y_i = 0$, equations (4.1) and (4.2) are equal if

$$\int_{-\alpha}^{\alpha} f_1(y_i/x_i; \beta_0, \beta_1, k) = \frac{f_1(0/x_i; \beta_0, \beta_1, k)}{f_0(0)}$$

. Now,

$$\frac{f_1(0/x_i; \beta_0, \beta_1, k)}{f_0(0)} = \frac{I_0(k_0)}{I_0(k)} exp(\kappa \cos \mu - k_0)$$

When the concentration parameter k_0 is very large compared to k, then the RHS goes to 0. Thus, the approximation is valid when LHS goes to 0. As α goes to 0, the LHS also goes to 0. Also, when μ is far from 0, then also LHS goes to 0 for small values of α. If we take k_0 such that the probability of the fixed density vanishes when $Y \neq 0$, then the approximation and exact likelihood are same in both the cases for $Y \neq 0$ and hence, a high value of k_0 should be considered such that $P\{Y \notin (-\alpha, \alpha)\} = 1 - \epsilon$, where ϵ is a predefined very small value near to 0. Thus, the approximation can be justified and hence, the estimation by taking a fixed high value of concentration parameter can be considered.

If we try to estimate the fixed concentration parameter , then it will take the value infinity as many observations have the exact value as 0. Hence, the fixing of the high-valued known concentration parameter(κ_0) is considered and then, the estimates approximate the true model in a satisfactory way as illustrated by the simulations. Table 4 shows that there is very small changes in the estimates when κ_0 gets higher.

It is recommended that κ_0 should be chosen in such a way that it should incorporate the measurement error. Thus, we should take κ_0 such that the probability that $Y \in (-\alpha, \alpha) = 1 - \epsilon$, where ϵ is close to 0. For the ease of computation and in consistency with the simulation results, we have taken $\alpha = 6$ degrees and $\epsilon = 0.05$ in data analysis and thus, if we take $\kappa_0 = 370$, $Pr(Y \in (-\alpha, \alpha)) > 0.95$. Thus, we have fixed $\kappa_0 = 370$.

5 Data Analysis

A study on comparison of four available methods of cataract surgeries was conducted at the Disha Eye Hospital and Research Center, Barrackpore, West Bengal, India, over a period of two years (2008-10), where a randomized unbiased allocation resulted 19 patients to undergo SICS with Snare technique (see [2]) and 18 patients to undergo SICS with Vectis technique (see [1]). Bakshi [23] has given a detailed description of the setup, study, associated data structure and the data description. It is well-known that one common side-effect of the cataract surgery is that the incision causes unwanted changes to the natural corneal shape causing an astigmatic eye. Also, it is preferred that the axis of astigmatism is 0, $\pi/2$ or π. The data is of circular nature. A multiplication by 4 (modulo 2π) results in a single preferred direction (which is 0) for the responses. This dataset is considered by Biswas *et al.* [24, 25] for testing which method of cataract surgery gives the best protection against the problem of astigmatism and also for some data-dependent allocation. Biswas *et al.* [13] again considered the data and illustrated that a zero-inflated circular distribution fits the data; however all the earlier studies ignored the associated covariates.

In the cataract surgery dataset, the data is available for the axis of astigmatism, 7 days after the surgery and also 1 month after the surgery. It was observed that when the surgery was carried out from Conventional Phacoemulsification or Torsional Phacoemulsification, the problem of astigmatism did not occur; the problem specially occurred when the technique used for the surgery was SICS: whether it is Vectis or or Snare. Thus, in our regression analysis, we only use the data obtained from SICS, both by Vectis and Snare techniques.

Taking the covariate (x) as the axes of astigmatism 7 days after the surgery and the response (y) as the axes 1 month after the surgery, a circular-circular regression can be modelled. Considering the axis of astigmatism to be 0 if the axis is within 7 degrees of the desired axes, i.e. 0^0, 90^0 or 180^0, it is found that 13 out of the 37 responses are 0. The left panel of Figure 2 shows the scatter plot of 37 points on the unit circle, while the right panel gives the plot by converting the points within 7 degrees of the desired axes, i.e. 0^0, 90^0 or 180^0, to be 0; all the data points are plotted after multiplication by 4 (modulo 360^0). Hence, in place of simple circular-circular regression, the zero-spike circular-circular regression is performed on the dataset.

Testing of $H_{0p} : p = 0$ versus $H_{1p} : p > 0$ for the dataset can be carried out as mentioned

in Section 3.2. From the data, we observe that the value of the test statistic Λ_p comes out to be 39.482, and the corresponding p-value is 0.013. Hence, H_{0p} is rejected at 1% level of significance, and we can conclude that the zero-spike model is a better fit for the data than the model with no zero-spike. Similarly, to test $H_{0\beta} : \arg(\beta_0\beta_1) = 0$ against $H_{1\beta} : \arg(\beta_0\beta_1) \neq 0$, the value of the test statistic Λ_β comes out to be 122.024, and $H_{0\beta}$ is rejected as the corresponding p-value is less than 0.0001 for the dataset. This also indicates that the circular-circular regression model with zero-spike is a good approximation for the data set, and hence the analysis of the data can be performed by taking the approximate model mentioned in the paper. The estimates of the parameters with estimated standard deviations in parentheses (circular variance in case of $\widehat{\theta}_0$) for the zero-inflated model are given by $\widehat{b}_1 = -0.472$ (0.143), $\widehat{b}_2 = 0.607$ (0.114), $\widehat{\theta}_0 = -1.067$ (0.229), $\widehat{\kappa} = 2.208$ (0.544), and $\widehat{p} = 0.330$ (0.081). Without considering zero-inflation in the model (i.e. for standard circular-circular model, as in [7]), the estimates for the parameters come out to be $\widehat{b}_1 = 0.572$ (0.101), $\widehat{b}_2 = -0.261$ (0.139), $\widehat{\theta}_0 = 0.460$ (0.212), and $\widehat{\kappa} = 2.141$ (0.430).

In Figure 3, we present two QQ-plots based on 37 observations for the data generated by assuming $x \sim vM(\mu_1, \kappa_1)$, where μ_1, κ_1 are the MLEs obtained from the x-data. Then we generate y-values from the the model of Kato *et al.* [7], and also from our model. In the left hand side of the Figure 3, we present the QQ-plot between the observed y-values and the y-values generated from model [7]; while, in the right hand side of Figure 3, we present the QQ-plot between the observed y-values and the y-values generated from our zero-inflated model. In the QQ-plots, the quantiles of the observed data are along the horizontal axis, while the quantiles of the predicted values are along the vertical axis. Clearly, the proposed zero-inflated model gives a much better fit for the data; smaller values (which are close to zero) are much better captured by our proposed model with zero-spike. The AIC for our model and the regression model used in [7] are 60.871 and 98.352 respectively. The BIC for our model and the regression model used in [7] are 86.980 and 119.240 respectively. Hence, AIC and BIC also mentions the better fitting of our model than the model mentioned in [7].

In Figure 4, the *spokeplots*, as introduced by Zubairi *et al.* [26], are drawn. The left panel shows the spokeplot for the predicted values with von Mises model and responses (without considering zero-inflation in the model), while the right panel gives the spokeplot with predicted values based on zero-inflated model and responses. The predicted values are
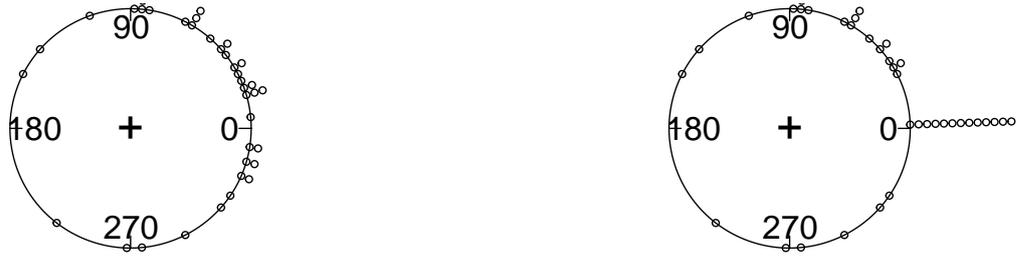
Figure 2: Scatter plot of 37 points on the unit circle. (Left:) Original points; (Right:) Making points within 7 degrees of the desired axes, i.e. 0^0, 90^0 or 180^0, to be 0.

on the inner circle, while the observed responses are on outer circle, each predicted point is joined by a straight line to the corresponding observed point.

Another objective way of comparing two models is the The continuous ranked probability score (CRPS) for circular variables as mentioned in [25]. Lower the CRPS, better is the fit. The mean CRPS (CPRS/no. of observations) value for our predicted model and the model fitted by using the regression model of Kato et.al. corresponding to the observations are 0.128 and 0.452 respectively. Hence, our model fits better in terms of CRPS also. In terms of CRPS, we also fitted a model with wrapped cauchy fixed parameter with mean 0 and $\rho = 0.992$ (equivalent to the value of $\kappa = 370$ in terms of symmetric 95% coverage about the mean) and wrapped cauchy error distribution. The mean CPRS in this case is 0.326. Thus, in consistency with the simulation result, von Mises distribution is better choice for fixed density than Wrapped Cauchy distribution.

6   Concluding remarks

We observe that the circular-circular regression with zero-inflated responses can be modelled satisfactorily by taking a mixture density with the known density having a huge concentration. The same procedure can also be applied when the unknown density is considered to be Wrapped Cauchy by taking the fixed parameter of Wrapped Cauchy in a way that the probability of responses around zero is very high. In this case also, the estimates are again
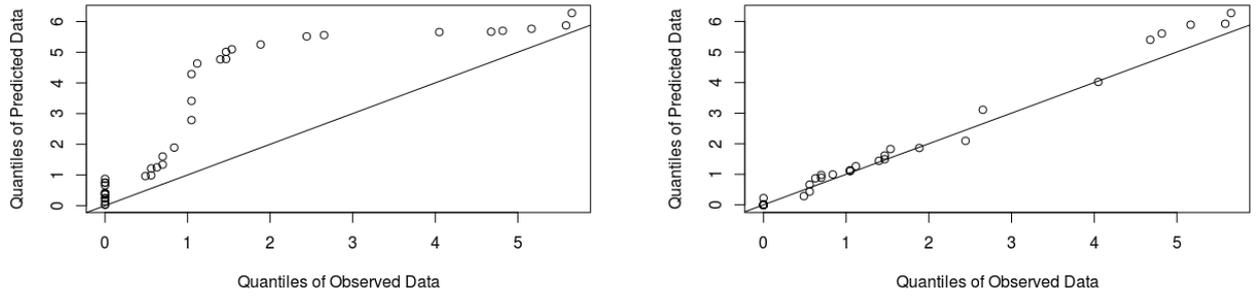
Figure 3: QQ-plots of 37 points in radian scale. (Left:) Observed points versus predicted points by the model of Kato *et al.*; (Right:) Observed points versus predicted points by the proposed model.
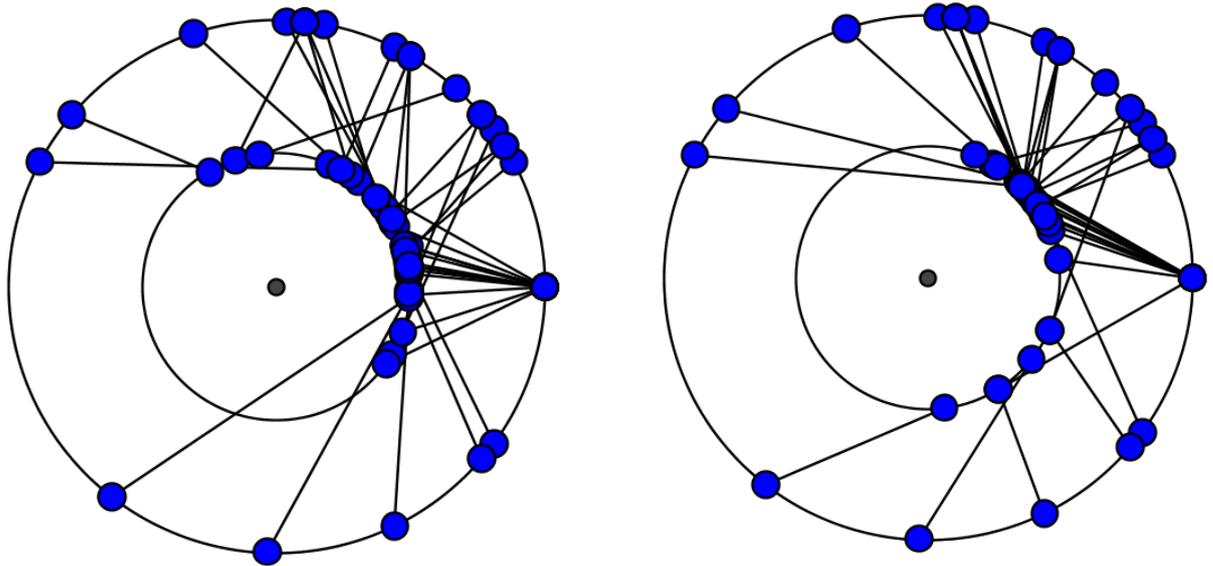

Figure 4: Spokeplots. The predicted values are on the inner circle, while the observed responses are on outer circle. (Left:) Spokeplot for the predicted values with von Mises model and responses (no zero-inflation); (Right:) Spokeplot with predicted values based on zero-inflated model and responses.

satisfactory but CPRS shows that fixing von Mises as the known fixed distribution gives better fit for the test data than the Wrapped Cauchy distribution.

The regression model can also be used when there are s (≥ 2) spikes in the dataset by considering fixed densities at different spikes and then the final density as the mixture of $(s + 1)$ densities.

In this paper ,we have considered the case when the reponse values are zero at random and hence, p does not depend on the covariate. If p depends on the covariate, then we can estimate p by taking $p(x; \beta) = \frac{g(x;\beta)}{1+g(x;\beta)}$ where $g(x; \beta) \geq 0 \forall x$ and then, along with β_0, β_1, κ, we can also estimate β. For example, let $\beta = (a, b, c)$ and $g(x; \beta) = exp\{a + b\cos(\theta_x) + c\sin(\theta_x)\}$, then our model is a special case of this model with $b = c = 0$. The functional form in this case bears a little similarity to the model mentioned in [19] as in this case also, the assumption behind the occurrence of zeros is different from the regression function used for the prediction of the response given that the response is non-zero. But, still the difference between the two models is the left censoring assumption used in [19]. The concept of censoring may be extended to the circular case where the latent variable θ_y^* is such that whenever $\theta_y^* \in (-\alpha, \alpha)$, then $\theta_y = 0$ else, $\theta_y = \theta_y^*$ for a pre-specified value of α which specifies the angle at which censoring has occurred and then the circular-circular regression of θ_y^* on θ_x may be performed.

References

[1] Masket S. The beginning of modern cataract surgery. The evaluation of small incision cataract surgery – A short history of ophthalmologists in progress. *Cataract and Refractory Surgeries Today* 2004; 77-80.

[2] Basti S, Vasavada AR, Thomas R, Padhmanabhan P. Extracapsular cataract extraction: Surgical techniques. *Indian J. Ophthalmol.* 1993; **41**: 195-210.

[3] Rivest L-P. A decentred predictor for circular-circular regression. *Biometrika* 1997; **84**: 318-324.

[4] Downs TD, Mardia KV. Circular regression. *Biometrika* 2002; **89**, 683-697.

[5] Minh DLP, Farnum NR. Using bilinear transformations to induce probability distributions. *Comm. Statist - Theory Methods* 2003; **32**: 1-9.

[6] Hussin AG, Fieller NRJ, Stillman EC. Linear Regression for circular variables with application to directional data. *Journal of Applied Science and Technology* 2004; **9**: 1-6.

[7] Kato S, Shimizu K, Shieh GS. A circular-circular regression model. *Statistica Sinica* 2008; **18**: 633-645.

[8] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; **34**: 1-14.

[9] Angers J-F, Biswas A. A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis* 2003; **42**: 37-46.

[10] Dietz E, Böhning D. On estimation of the Poisson parameter in zero-modified Poisson models. *Computational Statistics & Data Analysis* 2000; **34**: 441-459.

[11] Tian L. Inferences on the mean of zero-inflated lognormal data: the generalized variable approach. *Statistics in Medicine* 2005; **24**: 3223-3232.

[12] Labrecque-Synnott F, Angers J-F. An extension of zero-modified models to the continuous case. CRM-3289, October 2009.

[13] Biswas A, Jha J, Dutta S. Modelling circular random variables with a spike at zero. *Statistics and Probability Letters* 2016; **109**: 194-201.

[14] Tobin, J. Estimation of relationships for limited dependent variables. *Econometrica* 1958; **26**: 24-36.

[15] Cragg, JG. Some statistical models for limited depedent variables with application to the demand for durable goods. *Econometrica* 1971; **39**: 829-844.

[16]Anemiya, T. Regression analysis when the dependent variable is truncated normal. *Econometrica* 1973; **41**: 997-1016.

[17] Gronau, R. Wage comparison- a selectivity bias.*Journal of Political Economy* 1958; **82**: 1119-1144.

[18] Heckman, J. Shadow prices, market wages and labor supply. *Econometrica* 1974; **42**: 679-694.

[19] Heckman, J. Sample selection bias as a specification error. *Journal of the Royal statistical Society, Series B, Methodological* 1979; **49**: 127-145.

[20] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1977; **39**: 1-38.

[21] Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 1987; **82**: 605-610.

[22] Grimit EP, Gneiting T, Berrocal VJ, Johnson NA. The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification *Quarterly Journal of Royal Meteorological Society* 2006; **132**: 2925-2942.

[23] Bakshi P. Evaluation of various surgical techniques in brunescent cataracts. Unpublished thesis, Disha Eye Hospital, India, 2010.

[24] Biswas A, Dutta S, Laha AK, Bakshi PK. Response-adaptive allocation for circular data. *Journal of Biopharmaceutical Statistics* 2015; **25**: 830-842.

[25] Biswas A, Dutta S, Laha AK, Bakshi PK. Comparison of treatments in a cataract surgery with circular response. *Statistical Methods in Medical Research* 2016; **5**: 2238-2249.

[26] Zubairi YZ, Hussain F, Hussain AG. An alternative analysis of two Circular variables via graphical representation: an application to the Malaysian wind data. *Computer and Information Science* 2008; **1**: 3-8.

Table 1: Simulation results: Estimates of parameters (standard errors in parentheses) for zero-inflated circular-circular regression for $n = 35, 70$, $\beta_0 = 1, \kappa = 3.5, p = 0.5$, and different values of $\arg(\beta_1)$ are taken.

$n = 35$			
β_1	0.3	$0.3i$	-0.3
b_1	0.289 (0.001)	-0.007 (0.001)	-0.302 (0.001)
b_2	0.002 (0.001)	0.302 (0.001)	-0.000 (0.001)
p	0.515 (0.001)	0.508 (0.001)	0.507 (0.001)
κ	4.824 (0.019)	4.815 (0.018)	4.795 (0.018)
θ_0	-0.002 (0.014)	0.003 (0.013)	0.002 (0.014)
$n = 70$			
β_1	0.3	$0.3i$	-0.3
b_1	0.290 (0.001)	-0.008 (0.001)	-0.304 (0.001)
b_2	0.001 (0.001)	0.302 (0.001)	0.000 (0.001)
p	0.514 (0.001)	0.508 (0.001)	0.505 (0.001)
κ	4.053 (0.011)	4.064 (0.010)	4.041 (0.010)
θ_0	0.001 (0.006)	0.004 (0.006)	0.000 (0.001)

Table 2: Simulation results: Estimates of parameters (standard errors in parentheses) for zero-inflated circular-circular regression for $n = 35, 70$, $\beta = 0$, $\alpha = 0$,$\kappa = 3.5$, $p = 0.5$, and different values of ω are taken.

$n = 35$			
ω	0.25	0.5	0.75
α	0.001 (0.022)	-0.002 (0.068)	-0.009 (0.217)
β	-0.001 (0.029)	-0.002 (0.078)	-0.010 (0.225)
ω	0.256 (0.001)	0.498 (0.001)	0.711 (0.002)
p	0.523 (0.001)	0.516 (0.001)	0.512 (0.001)
κ	4.863 (0.025)	4.956 (0.026)	4.876 (0.024)
$n = 70$			
ω	0.25	0.5	0.75
α	0.001 (0.006)	-0.003 (0.026)	0.005 (0.142)
β	-0.001 (0.012)	-0.003 (0.032)	0.006 (0.148)
ω	0.258 (0.001)	0.507 (0.001)	0.738 (0.001)
p	0.524 (0.001)	0.516 (0.001)	0.512 (0.001)
κ	3.988 (0.011)	4.042 (0.011)	4.048 (0.011)

Table 3: Simulation results: CPRS for comparing von Mises and Wrapped Cauchy distributions for $\beta_1 = 0.3, \beta_0 = 1$, $p = 0.5$ and different choices of κ and ρ are taken.

	$\kappa = 1.5$	$\kappa = 3.5$	$\rho = 0.5$	$\rho = 0.65$
n=35				
CPRS(WC)	5.902 (0.027)	6.401 (0.021)	5.877 (0.029)	6.493 (0.025)
CPRS(VM)	3.561 (0.024)	4.001 (0.021)	3.469 (0.025)	3.667 (0.024)
n=70				
CPRS(WC)	10.105 (0.038)	11.786 (0.028)	9.775 (0.040)	11.974 (0.035)
CPRS(VM)	6.927 (0.034)	7.759 (0.030)	6.621 (0.034)	7.077 (0.034)

Table 4: Simulation results: Estimates of parameters (standard errors in parentheses) for zero-inflated circular-circular regression for $n = 35, 70$, $\beta_0 = 1, \kappa = 3.5, p = 0.5, \beta_1 = 0.3$, and different κ_0's are taken.

$n = 35$				
κ_0	100	200	300	370
b_1	0.279 (0.001)	0.284 (0.002)	0.288 (0.001)	0.289 (0.001)
b_2	-0.001 (0.001)	0.002 (0.002)	0.000 (0.001)	0.002 (0.001)
p	0.525 (0.001)	0.519 (0.001)	0.516 (0.001)	0.515 (0.001)
κ	4.925 (0.020)	4.856 (0.019)	4.818 (0.019)	4.824 (0.019)
θ_0	0.003 (0.014)	0.001 (0.014)	-0.002 (0.013)	-0.002 (0.014)
$n = 70$				
κ_0	100	200	300	370
b_1	0.283 (0.001)	0.286 (0.001)	0.289 (0.001)	0.290 (0.001)
b_2	-0.000 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.001 (0.001)
p	0.524 (0.001)	0.519 (0.001)	0.517 (0.001)	0.514 (0.001)
κ	4.121 (0.011)	4.081 (0.011)	4.053 (0.011)	4.053 (0.011)
θ_0	0.001 (0.006)	0.002 (0.006)	-0.001 (0.006)	0.001 (0.006)