Article can be accessed online at http://www.publishingindia.com

Distribution of Traffic Accident Times in India - Some Insights using Circular Data Analysis

Arnab Kumar Laha*, Pravida Raja A. C.*, Dilip Kumar Ghosh**

Abstract

Traffic accidents are a major hazard for travellers on Indian roads. These are caused by a variety of reasons including the bad condition of roads, traffic density, lack of proper training of drivers, slack in enforcement of traffic rules, poor road lighting etc. It is further known that certain times of the day are more prone to traffic accidents than others. In this paper we investigate the distribution of traffic accident times using the data published annually by the National Crime Records Bureau (NCRB) over the period 2001-2014 using the tools of circular data analysis. It is seen that the observed distribution of the traffic accident times in most years is bimodal. Thus, several modelling strategies for bimodal distributions are tried which include fitting of mixture of von-Mises distributions and mixture of Kato-Jones distribution. It is seen from this analysis that the distribution of the traffic accident times are changing over the years. Notably, the proportion of accidents happening in late night has reduced over the years while the same has increased for late evening hours. Some more insights obtained from this analysis are also discussed.

Keyword: Circular Statistics, Kato-Jones Distribution, Mixture Distribution, Traffic Accidents, Von-Mises Distribution

Introduction

Increasing incidents of road traffic accidents pose a major societal problem in India and other developing countries. According to Aderamo (2012), road traffic accidents are decreasing in developed countries and increasing in developing nations. Many researchers have paid attention in determining the factors that significantly affect injury. It is mentioned in research by David and Hyder (2006) that the application of policies and interventions to control traffic accidents can decrease the societal cost. Petrol rationing, an improvement in traffic enforcement, setting up of speed bumps, legislation and the enforcement of the use of helmets for cyclists and motorcyclists are examples of such interventions. There are many factors which can increase the risk of traffic accidents such as construction and maintenance of roads and vehicles, driver's behaviour, speed of vehicle, highway characteristics, traffic characteristics, and weather condition. Cools, Moons, and Wets (2010) focussed on the effect of weather conditions on daily traffic intensities (the number of cars passing a specific segment of a road) in Belgium and the results of their analysis indicates that snowfall, rainfall and wind speed reduces the traffic density but high temperature increases the traffic density. Statistical modelling for predicting road accidents is gaining popularity in the literature on road safety. Kong, Lekawa, Navarro, McGrath, Cohen, Margulies, & Hiatt (1996) studied bicyclist accidents in China and Germany during 2001 to 2006 and the analysis shows there were similarities and differences between the two countries especially for the frequency, age distribution of the fatalities and the road environment where accidents occurred. The paper also suggests the importance of the usage of helmet and improvement of road environment for reduction of accidents and fatalities in China.

The time of day has an important role in traffic accidents. It is believed that even though the traffic density is less in the night compared to the day time, the number of accidents is more in the night time. According to studies, reduced visibility is an important contributor to the night time traffic accidents. Owens and Sivak (1993) studied the role of reduced visibility in night time road fatalities recorded by the U.S Fatal Accident Reporting Systems from 1980 through 1990.

* Indian Institute of Management, Ahmedabad, Gujarat, India.Email: arnab@iima.ac.in
** Saurashtra University, Rajkot, Gujarat, India.

Plainis, Murray, and Pallikaris (2006) compared the road injury data under dim and bright conditions for two EU countries and showed that low luminance is likely to contribute to the disproportionate number of road traffic injuries occurring at night. According to them, the presence of road lighting leads to substantial decrease in the severity of injuries in both countries, despite the fact that they have dramatically different injury rates.

Data

The yearly data for the period 2001-2014 are obtained from the reports entitled Accidental Deaths and Suicides in India (ADSI) published by National Crime Records Bureau, India. The data consist of number of traffic accidents at different times of the day at national level in 53 Indian cities. The given data are grouped in 3 hourly time intervals 0 – 3am, 3 – 6 am, 6 – 9 am, 9-12 noon, 12-3 pm, 3-6 pm, 6-9 pm and 9-12 midnight aggregated over the years. The snapshot of the final data retrieved from http://ncrb.gov.inis given in Table 1.

Table 1: Snapshot of Data. Each Figure Indicates the Number of Accidents in the Specified Time Interval for a given Year

Time of occurrence	*2001*	*2002*	*2008-2012*			*2013*	*2014*
0-3	23869	23894	-	-	-	28332	26068
3-6	30949	28485	-	-	-	35385	32554
6-9	41612	41110	-	-	-	52771	52279
9-12	50293	51904	-	-	-	67224	69042
12-15	47291	48794	-	-	-	65974	68918
15-18	49925	52026	-	-	-	73141	77830
18-21	46663	48287	-	-	-	74411	76334
21-24	33118	34934	-	-	-	45763	47873
Total	323720	329434	-	-	-	443001	450898

We map each accident time in 24-hour period onto a point on the unit circle i.e., an angle between 0 to 2π radians. Every 45 degrees ($\pi/4$ radians) on the circle denotes 3 hours in real time. 03:00 a.m. is mapped to 0 degree on the circle. The histogram of the time of accident occurrence for the years 2001, 2008 and 2014 are given in the Fig. 1.

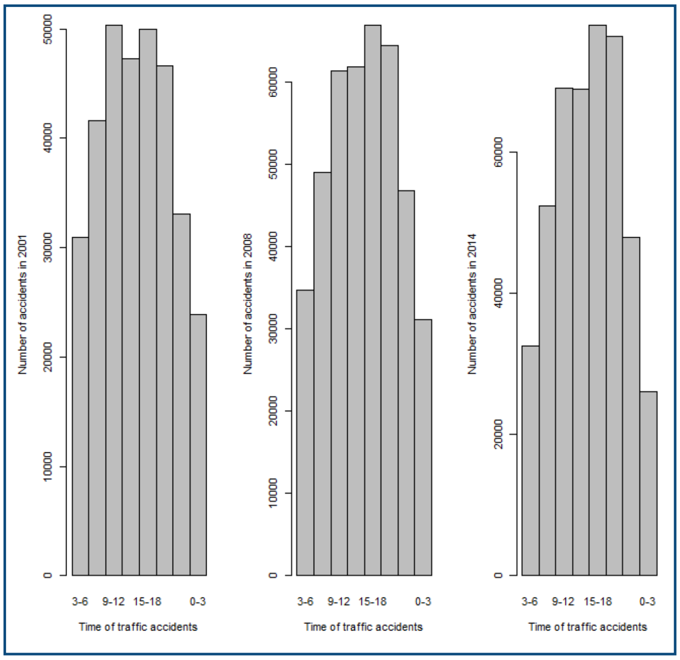

Fig. 1: Histogram of the Time of Traffic Accidents for the Years 2001, 2008 and 2014

Circular Data

Circular data are data measured in angles and occur in a variety of fields. They are commonly summarised as locations on a unit circle or as angles over a 360°or 2π radians range. Angular data arise in two ways, natural angles and observations which can be converted to angles. In this paper, we are mainly interested in the time of the day when the accident occurred. The time of accident is converted to an angle in the following manner. Let x be the time of the day recorded in hours. Then the corresponding angle would be $\frac{x}{24} * 2\pi$ radian. For example, if an accident occurred at 4 a.m., then the corresponding angle would be $\frac{4}{24} * 2\pi = \frac{\pi}{3}$ radians.

The purpose of this work is to analyse Indian road traffic accident times data using circular data analysis. Recently several authors have used Circular Statistics to analyse and model distributions of random variables that are cyclic in nature. Brunsdon and Corcoran (2006) used circular statistics to analyse time patterns in crime incidence. They analysed a data set related to the reports of criminal damage in the city of Cardiff, Wales during the period July 1999 to June 2001. The circular plot of the data (see Fig. 4 of Brunsdon & Corcoran, 2006) shows a bimodality in which the frequency of reporting crimes peaks around 11 PM and 10 AM rounded to the nearest hour. Faggian, Corcoran, and McCann (2013) introduced the use of circular statistics to study the interregional graduate migration flows in Britain. Corcoran, Chhetri, and Stimson (2009) applied circular statistics to analyse journey to work data. They calculated the direction and frequency of each journey using bespoke tools developed in a Geographic Information System (GIS) environment. They used the geographical angle of journey from an origin to a destination as a central variable for analysis which is circadian in nature (see Fig. 1 of Corcoran *et al.*, 2009). The application of circular statistics in particular circular mean direction of travel and circular spread gives an indication of the modality direction of the commuter from any movements given origin zone. Gill and Hangartner (2010) studied an interesting application of circular data in political science. They developed a circular regression model for terrorism events.

Circular Distributions

The most popular circular distributions used in applied work is the von-Mises (vM) or Circular Normal distribution (CN), which is described in sub-section below. However, there are many alternative circular distributions and a comprehensive account of the properties of these distributions can be found in Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001). Recently an extension of the Circular Normal distribution known as Kato-Jones distributions is finding increasing use in applied work. We discuss these distributions in sub-sections below.

Circular Normal (von-Mises) Distribution

Circular Normal (CN) distribution plays a central role in the analysis of circular data. This distribution was introduced as a statistical model by von-Mises (1918). A circular random variable Θ is said to have a CN distribution with mean direction parameter μ and concentration parameter κ if it has the probability density function (p.d.f.)

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta-\mu)),$$

$$0 \le \theta < 2\pi,\; 0 \le \mu < 2\pi, \kappa > 0$$

Where $I_0(.)$ is the modified Bessel function of order 0. This distribution is symmetric about μ and unimodal. We will denote this distribution as $CN(\mu,\kappa)$. Another interesting property of $CN(\mu,\kappa)$ distribution is that, for sufficiently large κ, the CN distribution can be approximated by a linear normal distribution with mean μ and variance $\frac{1}{\sqrt{\kappa}}$.

Kato-Jones (KJ) Distributions

Kato and Jones (2010) proposed a family of four parameter distributions on the circle that contains von-Mises and wrapped Cauchy distributions as special cases. This family of distributions is derived by transforming von-Mises distribution through Mobius transformation. The density function of this distribution is

$$f(\theta;\, \mu,\, \nu,\, r,\, \kappa) = \frac{1-r^2}{2\pi\, I_0(\kappa)} \exp\left\{ \frac{\kappa\left(\xi \cos(\theta-\eta) - 2r\cos\nu\right)}{1+r^2-2r\cos(\theta-\gamma)} \right\}$$

$$\times \frac{1}{1+r^2-2r\cos(\theta-\gamma)};\; 0 \le \theta < 2\pi$$

where $\gamma = \mu + \nu$, $\xi = \sqrt{r^4 + 2r^2\cos 2\nu + 1}$ and $\eta = \mu + \arg\left\{r^2\cos 2\nu + 1 + ir^2 \sin 2\nu\right\}$

such that $0 \le \mu, \nu < 2\pi$, $\kappa > 0$, and $0 \le r < 1$.

A three parameter family of distributions can be derived as a special case when $\nu = 0$ or $\nu = \pi$. In this case the above four parameter distribution reduces to

$$f(\theta;\mu,\kappa,r) = \frac{1-r^2}{2\pi I_0(\kappa)} \exp\left\{\frac{\kappa\left(1+r^2\right)\cos(\theta-\mu) - 2r}{1+r^2-2r\cos(\theta-\mu)}\right\} \times \frac{1}{1+r^2-2r\cos(\theta-\mu)};\ 0 \le \theta < 2\pi .$$

The distribution is symmetric about $\theta = \mu$ and $\mu + \pi$ and is unimodal when $0 \le r < 1$. The parameter μ is the directional mean. Symbolically, we will write $\theta \sim KJ(\mu, \kappa, \mathrm{r})$. The above model involves the von-Mises $(r = 0)$, wrapped Cauchy $(\kappa = 0)$ and uniform distributions $(\kappa = r = 0)$ as special cases. As $\kappa \to \infty$, the Kato-Jones distribution tends to $\mathrm{N}(\mu, \omega_{\mathrm{r}})$ where the standard deviation $\omega_{\mathrm{r}} = \frac{1-\mathrm{r}}{1+\mathrm{r}}$.

Finite Mixture of Distributions

Finite mixtures of distributions (FM) has seen many applications in the linear data context. Some applications in the circular data context have also been reported in the literature. A random variable X is said to follow a k-component mixture distribution of densities $f_1, f_2, \ldots, f_k$ if its p.d.f. is of the form

$$p(x) = \sum_{j=1}^{k} \pi_j f_j(x)$$

where π_j s is a set of probabilities also known as mixing weights such that $\sum_{j=1}^{k} \pi_j = 1$ and $f_j(x)$, $\mathrm{j} = 1, 2, \ldots, \mathrm{k}$ are the component densities. An up-to-date brief overview of the developments in FM models can be seen in Zhang and Huang (2015). Roy et al. (2012) designed mixture model based colour image segmentation in the LCH colour space using a Circular- Linear distribution. Mooney, Helms, and Jollife (2003) analysed Sudden Infant Death Syndrome (SIDS) data for the UK from 1983 to 1998 and in their study, they pointed out that for some years, there seems to be more than one mode. Later Jiang (2009) analysed this data set by fitting a von-Mises distribution and a mixture of two von-Mises distributions and reported that for most years the data could be fitted using a mixture of two von-Mises distributions. Jiang (2009) also analysed the fatal traffic crash time data in the United States in 2007 and showed that for Washington and the District of Columbia a mixture of two von-Mises distributions fitted the dataset.

Modelling Traffic Accident Times

In this paper we model the time of traffic accidents data using a mixture of two circular distributions. We consider two models

(1) a two component mixture of Circular Normal distributions $\alpha\, CN(\mu_1, \kappa_1) + (1-\alpha)\, CN(\mu_2, \kappa_2)$ and

(2) a two component mixture of Kato and Jones distributions $\alpha\, KJ(\mu_1, \kappa_1, r) + (1-\alpha)\, KJ(\mu_2, \kappa_2, r)$.

In both these cases we restrict μ_1 in $[0, \pi)$ and μ_2 in $[\pi, 2\pi)$. Let $f_{CN}(\theta; \mu_1, \mu_2, \kappa_1, \kappa_2, \alpha)$ be the pdf of mixture of Circular Normal distributions $\alpha\, CN(\mu_1, \kappa_1) + (1-\alpha)\, CN(\mu_2, \kappa_2)$. We define

$$p_1 = \int_0^{\frac{\pi}{4}} f_{CN}(\theta)d\theta,\quad p_2 = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} f_{CN}(\theta)d\theta,\ \ldots$$

$$p_8 = \int_{\frac{7\pi}{4}}^{2\pi} f_{CN}(\theta)d\theta. \tag{1}$$

Note that p_j's(j=1,…,8) depend on the unknown parameters $\mu_1, \mu_2, \kappa_1, \kappa_2, \alpha$. Now to estimate the unknown parameters we apply the minimum chi-square method (Berkson, 1980) which is briefly discussed below.

Let $x_1, x_2, \ldots, x_n$ be the given data. Define $e_j = np_j$ and $\mathrm{n}_j = \sum_{i=1}^{n} \varepsilon_{ij}$ where

$$\varepsilon_{ij} = \begin{cases} 1 & \text{if} \quad \mathrm{x_i} \in \left[(j-1)\frac{\pi}{4}, j\frac{\pi}{4}\right) \\ 0 & \text{otherwise} \end{cases}$$

for i=1,…,n and j=1,…,8. Consider the function $g(\mu_1, \mu_2, \kappa_1, \kappa_2, \alpha) = \sum_{j=1}^{8} \frac{(n_j - e_j)^2}{e_j}$. To obtain the estimates of the parameters we minimize the function g subject to the conditions $0 \le \mu_1, \mu_2 < 2\pi$, $\kappa_1 > 0,\ \kappa_2 > 0,\ 0 < \alpha < 1$. Since this function g is difficult to minimize analytically, we adopt a direct numerical minimisation approach using the function DEOptim in R (Mullen, Ardia, Gil, Windover, and Cline, 2011).Fig. 2(a)–(c) show estimated parameters of mixture of von-Mises distribution.

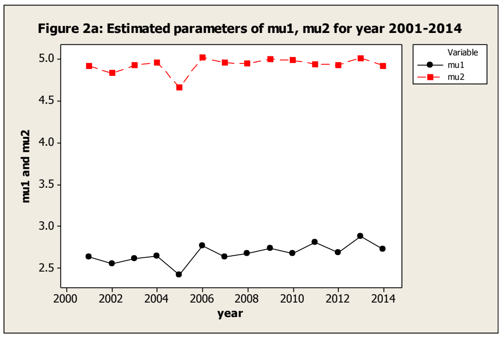

Fig. 2 (a): Estimated parameters of mu1, mu2 for year 2001-2014

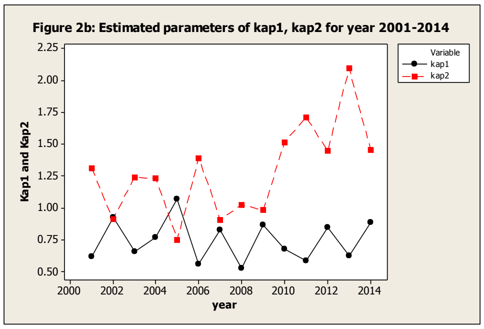

Fig. 2 (b): Estimated parameters of kap1, kap2 for year 2001-2014

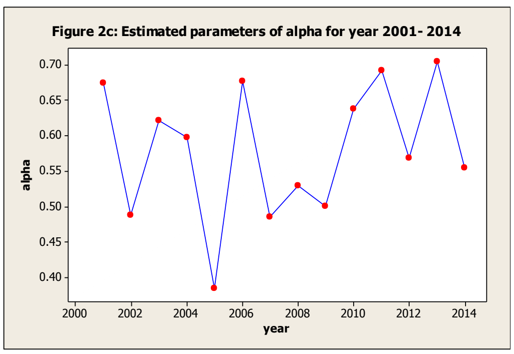

Fig. 2 (c): Estimated parameters of alpha for year 2001-2014

Let $f_{KJ}(\theta;\mu_1,\mu_2,\kappa_1,\kappa_2,r,\alpha)$ be the pdf of mixture of Kato-Jones distributions $\alpha\ \mathrm{KJ}(\mu_1,\kappa_1,r)+(1-\alpha)\ KJ(\mu_2,\kappa_2,r)$.We define p_j's (j=1,…, 8) in a manner analogous to (1) and obtain the minimum chi-square estimates of the parameters $(\mu_1,\mu_2,\kappa_1,\kappa_2,r,\alpha)$ using direct numerical minimisation of g subject to the conditions $0 \le \mu_1, \mu_2 < 2\pi$, $\kappa_1 > 0, \kappa_2 > 0,\ 0 \le r < 1, 0 < \alpha < 1$.Fig. 3(a)–(d) show estimated parameters of mixture of Kato-Jones distributions for year 2001 to 2014.

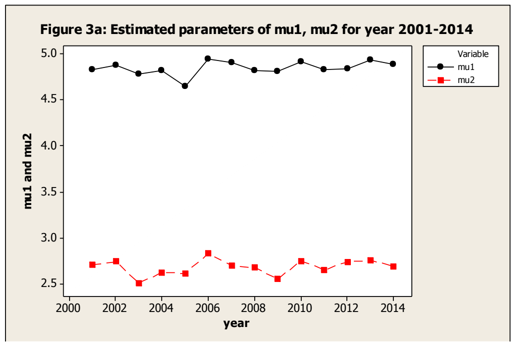

Fig. 3 (a): Estimated parameters of mu1, mu2 for year 2001-2014

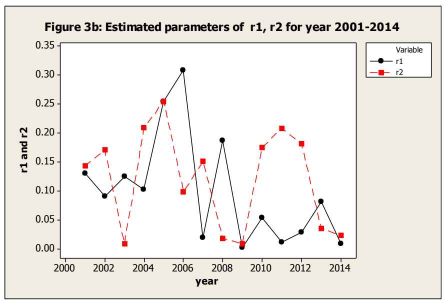

Fig. 3 (b): Estimated parameters of r1, r2 for year 2001-2014

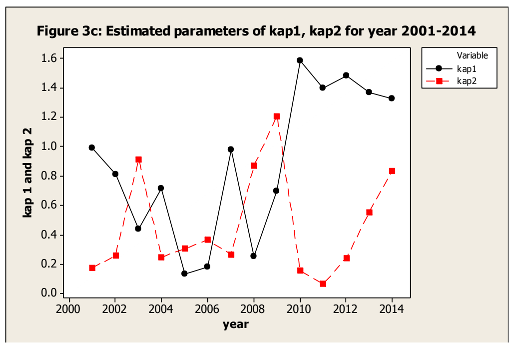

Fig. 3 (c): Estimated Parameters of kap1, kap2 for year 2001-2014

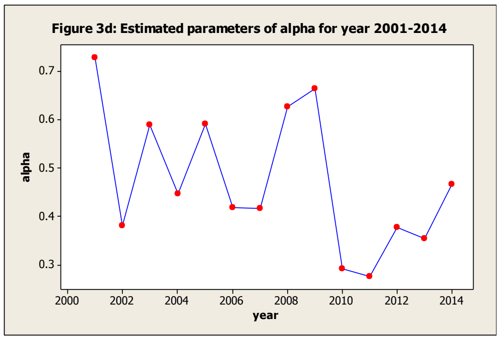

Fig. 3 (d): Estimated Parameters of alpha for year 2001-2014

Using these estimated parameters, we calculated the proportion of accidents happening in certain times of the day from year 2001-2014 which has been shown in Table 2.

Table 2: Proportion of Accidents Happening in Certain Times of the Day from Year 2001 to 2014 using Mixture of Circular Normal Distributions(CN) and Mixture of Kato-Jones (KJ) Distributions

		Time of occurrence							
Year	*Model*	*0-3*	*3-6*	*6-9*	*9-12*	*12-15*	*15-18*	*18-21*	*21-24*
2001	Actual	0.073	0.096	0.129	0.155	0.146	0.154	0.144	0.102
	vM	0.076	0.093	0.131	0.154	0.149	0.153	0.145	0.100
	KJ	0.079	0.093	0.128	0.153	0.147	0.155	0.147	0.097
2002	Actual	0.073	0.086	0.125	0.158	0.148	0.158	0.147	0.106
	vM	0.073	0.085	0.128	0.155	0.151	0.157	0.148	0.103
	KJ	0.076	0.084	0.122	0.158	0.151	0.155	0.149	0.103
2003	Actual	0.073	0.092	0.124	0.156	0.144	0.157	0.148	0.106
	vM	0.075	0.089	0.127	0.149	0.147	0.156	0.152	0.105
	KJ	0.077	0.090	0.128	0.149	0.146	0.158	0.149	0.102
2004	Actual	0.070	0.088	0.122	0.160	0.144	0.159	0.151	0.106
	vM	0.072	0.086	0.131	0.154	0.146	0.156	0.151	0.104
	KJ	0.075	0.084	0.125	0.158	0.146	0.158	0.152	0.103
2005	Actual	0.073	0.089	0.122	0.161	0.146	0.163	0.146	0.101
	vM	0.073	0.087	0.129	0.152	0.153	0.161	0.145	0.101
	KJ	0.076	0.082	0.123	0.159	0.149	0.165	0.147	0.098
2006	Actual	0.077	0.090	0.199	0.150	0.144	0.157	0.154	0.110
	vM	0.078	0.088	0.121	0.146	0.148	0.156	0.155	0.109
	KJ	0.080	0.087	0.118	0.149	0.147	0.154	0.157	0.106
2007	Actual	0.077	0.086	0.116	0.150	0.147	0.157	0.153	0.114
	vM	0.079	0.083	0.119	0.147	0.147	0.156	0.154	0.113
	KJ	0.078	0.085	0.118	0.149	0.150	0.151	0.156	0.112
2008	Actual	0.075	0.084	0.118	0.148	0.149	0.161	0.155	0.112
	vM	0.076	0.079	0.123	0.146	0.145	0.162	0.157	0.112
	KJ	0.079	0.084	0118	0.146	0.149	0.162	0.155	0.107
2009	Actual	0.069	0.083	0.115	0.152	0.146	0.163	0.152	0.120
	vM	0.075	0.078	0.116	0.149	0.152	0.159	0.158	0.114
	KJ	0.076	0.079	0.119	0.151	0.150	0.159	0.153	0.111

		Time of occurrence							
Year	Model	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24
2010	Actual	0.068	0.088	0.121	0.151	0.145	0.160	0.161	0.107
	vM	0.071	0.084	0.124	0.149	0.145	0.158	0.162	0.107
	KJ	0.073	0.084	0.119	0.153	0.146	0.157	0.165	0.103
2011	Actual	0.067	0.085	0.118	0.152	0.150	0.166	0.160	0.102
	vM	0.069	0.083	0.121	0.149	0.151	0.165	0.161	0.101
	KJ	0.076	0.093	0.131	0.154	0.149	0.152	0.145	0.100
2012	Actual	0.063	0.078	0.118	0.153	0.148	0.167	0.166	0.107
	vM	0.063	0.075	0.121	0.152	0.149	0.167	0.168	0.105
	KJ	0.067	0.075	0.114	0.151	0.149	0.172	0.169	0.102
2013	Actual	0.064	0.080	0.119	0.152	0.149	0.165	0.168	0.103
	vM	0.063	0.078	0.119	0.151	0.150	0.164	0.171	0.103
	KJ	0.067	0.080	0.119	0.148	0.146	0.166	0.172	0.102
2014	Actual	0.058	0.072	0.116	0.153	0.153	0.173	0.169	0.106
	vM	0.059	0.070	0.117	0.153	0.153	0.171	0.171	0.104
	KJ	0.061	0.068	0.114	0.154	0.155	0.169	0.172	0.106

It can be seen from Table 2 that the proportion of accidents happening in late night (9pm – 3am) has reduced over the years (0.175 in 2001 to 0.164 in 2014) while the same has increased for late evening hours (6-9pm) which has been captured by both the models under consideration. We use the Schwarz Information Criterion (SIC) (Schwarz,1978) to choose the best model among these two. The SIC is defined as $SIC = -2 \log L\left(\hat{\theta}\right) + k \ln(n)$, where $L\left(\hat{\theta}\right)$ is the likelihood function for the model evaluated at the estimated parameter value $\hat{\theta}$, k is the number of parameters and n is the sample size. The likelihood is calculated for the two models and the corresponding SIC year wise values are given in Table 3 for the years 2005 - 2009.

Table 3: SIC for Mixture of Von-Mises and Mixture of Kato-Jones Distributions from 2005 to 2009

Year	SIC (Mix vM)	SIC (Mix KJ)
2005	569.3544	365.9092
2006	277.066	258.6512
2007	252.944	232.7688
2008	186.6506	197.3286
2009	820.9194	823.3113

We see that in some of the cases the SIC is minimum for mixture of Kato-Jones distributions whereas in some other cases it is minimum for mixture of von-Mises distributions. Since the family of two component mixture of Kato-Jones distributions contains the family of two component mixture of von-Mises distributions, we consider the former for modelling the time of the accidents.

Change Point Problem

The change point problem is introduced in statistics by Page (1955) in the context of statistical quality control. It has been discussed quite extensively in the literature for linear data. Let $x_1, x_2, \ldots, x_n$ be independent observations. It is often of interest to know if there exists a(unknown) point s; $1 \le s \le n-1$ such that $x_1, x_2, \ldots x_s$ are independently and identically distributed (i.i.d) F_0 and $x_{s+1}, x_{s+2}, \ldots x_n$ are i.i.d. F_1 $(F_0 \ne F_1)$. Here s is called the change point of the data. This formulation is usually referred to as at most one (or single) change point problem (AMOC). If $s = n$ then all observations are from F_0 or we say that there is no change point. In change point problem, one is interested to test

$H_0 : x_i{'}s$ are i.i.d F_0 against the alternative

H1 : there exist s, $1 \le s \le n-1$, such that $x_1, x_2, \ldots, X_s$ are i.i.d. F_0 and $x_{s+1}, x_{s+2}, \ldots, x_n$ are i.i.d. F_1

Here F_i , $\mathrm{i}=0,1$ may be known or unknown and may contain one or more unknown parameters.

Lombard (1986) proposed rank based non-parametric procedures to test the presence of change point in a sequence of angular observations. Ghosh, Jammalamadaka, and Vasudaven (1999) considered a generalised likelihood ratio test procedure and a Bayes procedure for change-point problems of the mean direction of the CN distribution. Grabovsky and Horvath (2001) suggested a modified procedure to detect changes in circular data. Sengupta and Laha (2008a) introduced a likelihood integrated method for exploratory graphical analysis of change point problem with directional data. Sengupta and Laha (2008b) also discussed the problem of detecting change in the mean direction of the circular normal distribution when the concentration parameter is unknown using Bayesian analysis.

Chen and Gupta (1997) approached the change point problem as a model selection problem. Specifically they consider the models

M_0 : The accident time distribution is F_0 for all the years against

M_s: The accident time distribution is F_0 for the first s years and is F_1 for the years $s+1$ to n.

They then propose to use SIC to choose the best model among these models. We assume F_0 belongs to the family of two-component mixture of Kato-Jones distribution $\alpha\ \mathrm{KJ}\ (\mu_{1b},\kappa_1,r)+(1-\alpha)\ KJ\ (\mu_{2b},\kappa_2,r)$, where μ_{1b} and μ_{2b} are the mean directions of the two component Kato-Jones distributions before change point occurred. We also assume that F_1 is a member of the above family but with different parameters. i.e., F_1 is $\alpha\ \mathrm{KJ}\ (\mu_{1a},\kappa_1,r)+(1-\alpha)\ KJ\ (\mu_{2a},\kappa_2,r)$ where μ_{1a} and μ_{2a} are the mean directions of the two component Kato-Jones distributions after the change point. Since there are unknown parameters in both F_0 and F_1 we only investigate the presence of change point in the period 2005 to 2009. Following Chen and Gupta (1997), we use the minimum SIC criterion for choosing the best model amongst $M_0,\ldots,M_{14}$. To compute the SIC for M_0 we need to compute the likelihood L_0. It is not difficult to observe $L_0=\prod_{s=1}^{14}\frac{n_s\,!}{\mathrm{n}_{1s}\,!n_{2s}\,!\ldots n_{8s}\,!}p_1^{\mathrm{n}_{1s}}p_2^{\mathrm{n}_{2s}}\ldots p_8^{\mathrm{n}_{8s}}$

where $p_1=\int_0^{\frac{\pi}{4}} f_{KJ_M}(\theta)d\theta$, $p_2=\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} f_{KJ_M}(\theta)d\theta$, … , $p_8=\int_{\frac{7\pi}{4}}^{2\pi} f_{KJ_M}(\theta)d\theta$; n_{1s} is the number of accidents occurring between 3 - 6 am in years, $n_s=\sum_{j=1}^{8}n_{js}$ and $f_{KJ_M}(\theta;\mu_1,\mu_2,\kappa_1,\kappa_2,r,\alpha)$ be the pdf of mixture of Kato-Jones distributions $\alpha\ \mathrm{KJ}(\mu_1,\kappa_1,r)+(1-\alpha)\ KJ(\mu_2,\kappa_2,r)$. For computing SIC for M_s, we need to compute the likelihood L_s which is given by $L_s=\prod_{k=1}^{s}\frac{n_k\,!}{\mathrm{n}_{1k}\,!n_{2k}\,!\ldots n_{8k}\,!}p_{1b}^{n_{1k}}p_{2b}^{n_{2k}}\ldots p_{8b}^{n_{8k}}$ k

$\prod_{k=s+1}^{14}\frac{n_k\,!}{\mathrm{n}_{1k}\,!n_{2k}\,!\ldots n_{8k}\,!}p_{1a}^{n_{1k}}p_{2a}^{n_{2k}}\ldots p_{8a}^{n_{8k}}$ where

$p_{1b}=\int_0^{\frac{\pi}{4}} f_{KJ_{bM}}(\theta)d\theta$ $p_{2b}=\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} f_{KJ_{bM}}(\theta)d\theta$,….,

$p_{8b}=\int_{\frac{7\pi}{4}}^{2\pi} f_{KJ_{bM}}(\theta)d\theta$ and $p_{1a}=\int_0^{\frac{\pi}{4}} f_{KJ_{aM}}(\theta)d\theta$

, $p_{2a}=\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} f_{KJ_{am}}(\theta)d\theta$,….., $p_{8a}=\int_{\frac{7\pi}{4}}^{2\pi} f_{KJ_{aM}}(\theta)d\theta$

. Here $f_{KJ_{bM}}(\theta;\mu_{1b},\mu_{2b},\kappa_1,\kappa_2,r,\alpha)$ and $f_{KJ_{aM}}(\theta;\mu_{1a},\mu_{2a},\kappa_1,\kappa_2,r,\alpha)$ are the p.d.f. of mixture of Kato-Jones distributions before and after the change point respectively.

Table 4 gives SIC of the models we considered for year 2005-2009.

Table 4: Model for 2005-2009 and the corresponding SIC values

Model	SIC KJ
M_0	13282.21
M_5	11716.88
M_6	10755.03
M_7	13286.83
M_8	10743.92
M_9	10787.72

It can be seen from Table 4 that M_8 has the lowest SIC which indicates that there is a change point and the year of the change is 2008. The estimated parameters of the mixture of Kato-Jones distributions for the years 2001-2008 are μ_{1b} = 1.82 radian (104.19°), μ_{2b} = 4.13 radian (236.62°), κ_1 = 0.6, κ_2 = 1.42, r = 0.01 and α = 0.64, and the same for the years 2009 -2014 are μ_{1l} = 1.98 radian (113.68°), μ_{2a} = 4.16 radian (238.36°), κ_1 = 0.6, κ_2 = 1.42, r = 0.01 and α = 0.64. Thus it can be observed that the accident time during the years 2001-2008 has modes at 10:34am and 7:18pm, and the same during the years 2009-2014 are at 10:58am and 7:29pm.

Conclusion

The observed distribution of the traffic accident times in India for most years in the period 2001-14 is seen to be bimodal. This bimodal distribution has been modelled using a two component mixture of Kato-Jones distributions. The distribution is seen to be a decent fit to the observed data. It is further seen that the distribution of the traffic accident times are changing over the years. Notably, the proportion of accidents happening in late night has reduced over the years while the same has increased for late evening hours. A formal change point analysis indicates the presence of a change point in the year 2008.

References

Aderamo, A. J. (2012). Assessing the trends in road traffic accident casualties on Nigerian roads. *Journal of Social Sciences*, 31, 19-25.

Berkson, J. (1980). Minimum Chi-Square, not maximum likelihood. *Annals of Statistics*, 8(3) , 457-487.

Brunsdon, C., & Corcoran, J. (2006). Using circular statistics to analyse time patterns in crime incidence. *Computers, Environment and Urban Systems*, 30, 300-319.

Chen, J., & Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, 92, (438), 739-747.

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *American Meteorological Society, Weather, Climate and Society*, 2, 60-68.

Corcoran, J. , Chhetri, P., & Stimson, R. (2009). Using circular statistics to explore the geography of the journey to work, *Progress in Regional Science*, 88(1), 119-132.

David, M. B, & Hyder, A. A. (2006). Modelling the cost effectiveness of injury interventions in lower and middle income countries: Opportunities and challenges. *Cost Effectiveness and Resource Allocation,* 4, 2, 1-11.

Faggian, A., Corcoran, J., & McCann, P. (2013). Modelling geographical graduate job search using circular statistics, *Papers in regional Science*, 92(2), 329-343.

Ghosh, K., Jammalamadaka, S. R., & Vasudaven, M. (1999). Change-point problems for the von-Mises distribution. *Journal of applied statistics*, 26(4), 423-434.

Gill, J., & Hangartner, D. (2010). Circular data in Political Science and how to handle It, *Political Analysis*, 18 (3), 316-336.

Grabovsky, I., & Horvath, L. (2001). Change-point detection in angular data. *Annals of the Institute of Statistical Mathematics*, 53(3), 552-556.

Jammalamadaka, S. R., & SenGupta, A. (2001). *Topics in circular statistics*. Singapore: World scientific.

Jiang, Q. (2009). *On fitting a mixture of two von-Mises distributions, with applications*. M.Sc Project, Department of Statistics and Actuarial Science, Simon Fraser University, Canada.

Kato, S., & Jones, M. C. (2010). A family of distributions on the circle with links to, and applications arising from mobius transformation. *Journal of American Statistical Association*, 105 (489), 249-262.

Kong, L. B., Lekawa,M., Navarro, R. A,. McGrath, J., Cohen, M., Margulies, D. R., &Hiatt, J. R. (1996). Pedestrian-motor vehicle trauma: an analysis of injury profiles by age. *Journal of the American College of Surgeons*, 182, 17-23.

Lombard, F. (1986). The change-point problem for angular data: A nonparametric approach. *Technometrics*, 28, 391-397.

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Chichester,Wiley.

Mooney, J. A., Helms, P. J., & Jollife, I. T. (2003). Fitting mixtures of von-Mises distributions: A case study involving sudden infant death syndrome. *Computational Statistics & Data Analysis*, 41, 505-513.

Mullen, K. M., Ardia, D., Gil,D. L., Windover, D., & Cline, J. (2011) . DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6), 1-26.

Owens, D. A., & Sivak , M. (1993). *The role of reduced visibilityin night time road fatalities*. Report no. UMTRI-93-33, University of Michigan. Transportation Research Institute, U.S.A.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42, 523-527.

Plainis, S, Murray, I. J., & Pallikaris, I. G. (2006). Injury prevention, 12, 125-128. doi: 10.1136/ip.2005.011056.

Roy, A., Parui, S. K., & Roy, U. (2012). A mixture model of circular-linear distributions for color image segmentation. *International Journal of Computer Applications* (0975 - 8887), 58(9), 6-11.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.

Sengupta, A., & Laha, A. K. (2008a). A likelihood integrated method for exploratory graphical analysis of change point problem with directional data. *Communications in statistics. Theory and methods*, 37(11-12), 1783-1791.

Sengupta, A., & Laha, A. K. (2008b). A Bayesian analysis of the change-point problem for directional data. *Journal of Applied Statistics*, 35(6), 693-700.

von Mises, R. (1918). UJber die ‘Ganzzahligheit’ der Atomgewichte und vermandteFragen. Phys. Z. 19, 490-500.

Zhang, H., & Huang, Y. (2015). Finite mixture models and their applications: A review. *Austin Biometrics and Biostatistics*, 2(1), 1013, 1-6.