Journal of the Royal Statistical Society Series C: Applied Statistics, 2024, **00**, 1–21
https://doi.org/10.1093/jrsssc/qlae057

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society Series C Applied Statistics

Original Article

Traffic count data analysis using mixtures of Kato–Jones distributions

Kota Nagasaki[1], Shogo Kato[2], Wataru Nakanishi[3] and M. C. Jones[4]

[1]Department of Civil and Environmental Engineering, Institute of Science Tokyo, Tokyo, Japan
[2]Institute of Statistical Mathematics, Tokyo, Japan
[3]School of Geosciences and Civil Engineering, Kanazawa University, Ishikawa, Japan
[4]School of Mathematics and Statistics, The Open University, Milton Keynes, UK

Address for correspondence: Kota Nagasaki, Department of Civil and Environmental Engineering, Institute of Science Tokyo, W6-9, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan. Email: nagasaki.k.d116@m.isct.ac.jp

Abstract

We discuss the modelling of traffic count data that show the variation of traffic volume within a day. For the modelling, we apply mixtures of Kato–Jones distributions in which each component is unimodal and affords a wide range of skewness and kurtosis. We consider two methods for parameter estimation, namely, a modified method of moments and the maximum-likelihood method. These methods were seen to be useful for fitting the proposed mixtures to our data. As a result, the variation in traffic volume was classified into the morning and evening traffic whose distributions have different shapes, particularly different degrees of skewness and kurtosis.

Keywords: circular data, directional statistics, EM algorithm, maximum-likelihood estimation, method of moments estimation, traffic counter data

1 Introduction

Traffic flow volume is one of the most essential variables in the transport engineering field. Also, it is easy to observe by installing a traffic counter and counting the number of vehicles passing by at the observation site. Many traffic counters exist along arterial roads in Japan and other countries, and the traffic flow data are accumulated day by day (Anacleto et al., 2013; Leduc, 2008).

These data are mainly utilized in the following two manners. The first is evaluating road performance, or throughput, i.e. the maximum number of vehicles the road can accommodate within a unit time (Edie, 1963; Transportation Research Board, 2010). For example, this value is represented in units of vehicles per hour. Traffic congestion occurs if more vehicles than this value come to the road. The second is the so-called 'traffic state estimation', which comprises three essential variables: the traffic flow rate, the average vehicle density, and the average vehicle speed. Understanding the relationship among these three variables is quite important to predict traffic congestion that is spread along the road network (Muñoz et al., 2003; Wang & Papageorgiou, 2005). Both of road performance and traffic state estimation deal with the observed traffic flow volume as aggregated data such as 1—or 5 min flow volumes, i.e. the number of vehicles that pass by in such time intervals. In addition, some studies use such data without aggregation, i.e. the raw timestamp of each vehicle's passing. They also aim at analysing congestion and modelling the time intervals between successive vehicles (Chen et al., 2010; Li & Chen, 2017).

Utilizing the data in such ways is reasonable regarding the purpose of installing the traffic counter. Nonetheless, the raw data may have much potential to bring additional insights to transport engineering. For example, if we model the probability distribution of time of vehicles passing, we

Received: June 2, 2022. **Accepted:** October 3, 2024

can interpolate missing values, which may be caused by the failure of counters. Predicting the traffic volume in the whole road section including unobserved locations may also be possible by combining our model with interpolation methods. Additionally, it is useful for the verification and validation of microscopic traffic simulations that deal with each vehicle as an agent. To the best of the authors' knowledge, no study except one written in Korean (Na & Jang, 2011) estimates the probability distribution of traffic volume.

In this study, considering such potential applications, we estimate the complex distribution of the variation of traffic volume within an average weekday. Traffic volume is generally determined by the dynamic system of the interaction between the supply side (i.e. how many vehicles the road can accommodate) and the demand side (i.e. how many drivers wish to use the road). Due to this complexity, the distribution of traffic volume in a day is not expected to be unimodal or symmetric. Generally, the timestamp of each vehicle's passing does not follow a Poisson process because of some bottlenecks such as traffic congestion and traffic lights (Daganzo, 1997). Actually, it is empirically bimodal and asymmetric on weekdays (explained in detail in Section 2). In addition, if we consider an average weekday, time is periodic and represented as points on a circle. Thus, the time axis is circular in that 0 a.m. and 12 p.m. represent the same time. This suggests that we had better employ directional statistics.

Our study is the first attempt to use the raw data of a traffic counter and employ a mixture of the circular distributions of Kato and Jones (2015) to represent the distribution of traffic volume in an average weekday. Also, we contribute a method of parameter estimation for mixtures of Kato–Jones distributions as there is no established method thus far. The use of mixtures of the distributions of Kato and Jones (2015) has been briefly discussed for an analysis of another traffic dataset in the conference proceedings of Nagasaki et al. (2019). However, despite its simplicity, it is not the case that their inferential algorithm for the mixtures guarantees the consistency or any other optimal property of the proposed estimator.

Some mixtures of circular distributions have been proposed in the literature. The most attention has been paid to mixtures of von Mises distributions (e.g. Banerjee et al., 2005; Mooney et al., 2003; Mulder et al., 2020; Wallace & Dowe, 2003). The components of these mixtures, the von Mises distributions, are symmetric distributions with two parameters controlling the location and mean resultant length. Recently, Miyata et al. (2020) discussed mixtures of sine-skewed distributions whose components can adopt mildly asymmetric shapes. However, none of these existing models seem to be appropriate for our traffic data because one of the clusters in our data is strongly skewed. Therefore, for the modelling of our data, it seems reasonable to consider the distributions of Kato and Jones (2015) as possible components of the mixtures because these distributions can control a wide range of skewness as well as kurtosis.

The paper is organized as follows. The traffic count data of interest are introduced in Section 2. We define the mixtures of the distributions of Kato and Jones (2015), which are applied to the traffic data, and investigate basic properties of the mixtures and their components in Section 3. Two methods for parameter estimation for the proposed mixtures are presented in Section 4. A simulation study is conducted to compare the two proposed methods for parameter estimation in Section 5. The proposed mixtures are applied to the traffic data using the proposed inferential methods in Section 6. The interpretation of the estimated model is also discussed. The fit of the proposed mixtures to the data is compared with the fits of some other mixture models in Section 7. Finally, Section 8 concludes the paper. The online supplementary material provides the additional data analysis, a simulation study, proofs of the claims made in the main article, and Python codes for parameter estimation.

2 The data

First, we briefly explain about traffic counters in general. Traffic counters record the timestamps of all vehicles passing. Many counters are installed along the arterial roads in Japan, often at intervals of one kilometre. Usually, these data are aggregated to 1—or 5 min vehicle counts. Then, the aggregated data are utilized for traffic control such as detecting congestion and providing information to drivers. On the other hand, we use the data without aggregation in this paper, the data being raw timestamps, to model the probability of a vehicle passing at each particular time. In particular, this paper focuses on the averaged distribution for weekdays. Then, the time axis is

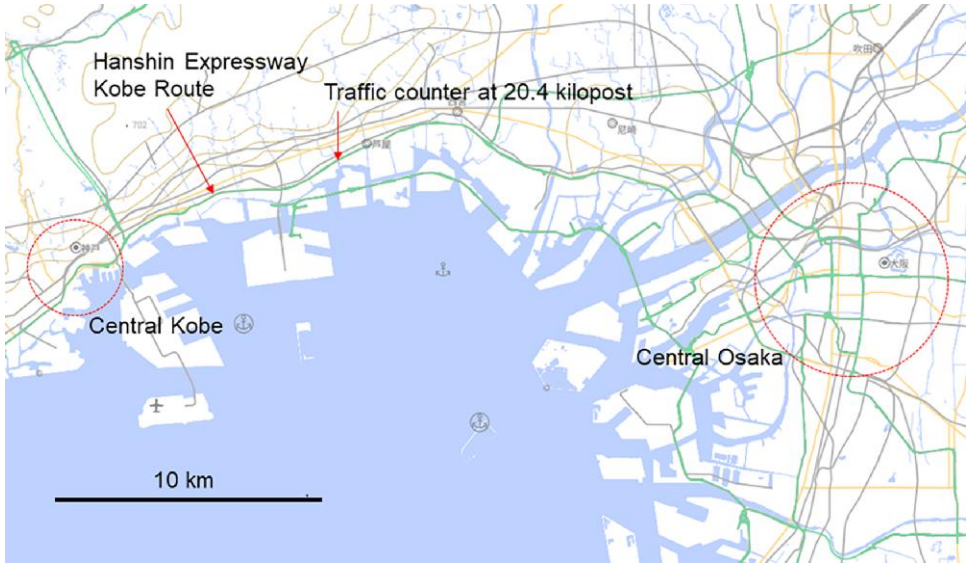

Figure 1. Map of target area including the location of the traffic counter and the Kobe route of the Hanshin Expressway. The background map is based on a Digital Map by the Geospatial Information Authority of Japan.

circular; 0:00 (0 a.m.) and 24:00 (12 p.m.) represent the same time, and 1:00 (1 a.m.) and 23:00 (11 p.m.) are not 22 hr apart, but only 2 hr apart. To consider these characteristics accurately, we employ the concept of directional statistics.

Concretely, we use the data of a traffic counter provided by the Hanshin Expressway Co. Ltd., Japan via personal communication. Specifically, the data are obtained at the 20.4 kilopost of the Kobe route, Hanshin Expressway. Figure 1 shows a map of the target area including the location of the traffic counter and the Kobe route of the Hanshin Expressway. The background map is based on a Digital Map by the Geospatial Information Authority of Japan. The Kobe route exists in the Osaka metropolitan area and connects two large cities, Osaka and Kobe. The Osaka metropolitan area is the second largest megacity in Japan following the Tokyo metropolis and its population is around 20 million. Osaka city is the main city of the Osaka metropolitan area with a population of around 2.8 million. Kobe city is the second largest city in the area with a population of 1.5 million and is located about 30 km west of Osaka city.

The counter records the timestamps of vehicles passing in the direction from Kobe to Osaka. The data period is 46 weekdays; from 6th June to 7th July and from 22nd August to 10th October 2016. It is not unnatural to aggregate data for all weekdays because the distribution of weekday traffic is generally similar (Transportation Research Board, 2010). A histogram of the data of the summation of all 46 days is shown in Figure 2. The bin width is set as 1 hr for this figure.

The total number of observed vehicles for 46 days is 1,121,262, which equals 24,375 per day on average. Here, the histogram has two peaks at around 7 a.m. and 3 p.m. We assume that the former peak represents the morning rush hour and the latter does the evening rush hour, from Kobe to Osaka, as the data are for weekdays. Note that all recorded traffic is in the same direction (i.e. from Kobe to Osaka). It means that vehicles commuting to their office and returning home are observed only once in the morning or evening.

In addition, the peak in the morning is higher and sharper than that in the evening. Also, each peak has an asymmetric shape. For example, as many commuters should arrive at their offices in Osaka at around the same time, typically, at 8:30 a.m. or 9:00 a.m., the histogram of the morning peak suggests a negatively skewed shape. A model that cannot represent flexible shapes cannot adequately describe traffic data's shape and characteristics. To fully represent the characteristics of this distribution, a mixture of probability density functions that allows a flexible shape is necessary. Hereafter, we assume that the data follow the IID property. Although this might be a strong assumption, it seems acceptable from the transportation engineering viewpoint. See online supplementary material, Appendix A for the details.

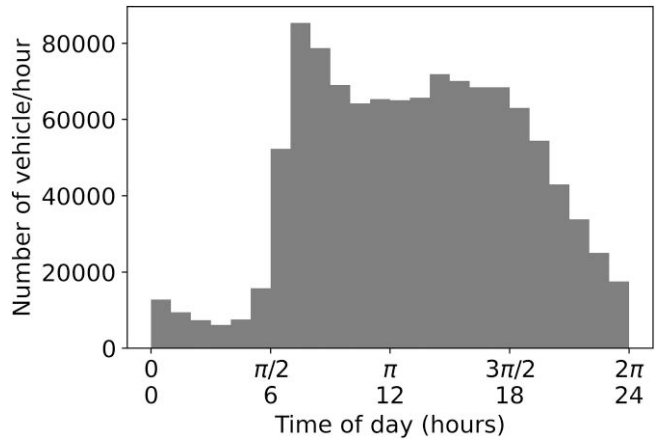

Figure 2. A histogram of the data for the summation of all 46 weekdays.

3 Model

3.1 Definition of the model

This section considers a model for the traffic data. As displayed in Figure 2, our data show two distinct features, namely, (i) bimodality and (ii) different degrees of skewness and peakedness in the clusters of the morning and evening rush hours. Therefore, it seems reasonable to adopt a mixture model whose components have flexibility in terms of skewness and peakedness. In this paper, we adopt the distribution of Kato and Jones (2015) as each component of the mixture. Its density is given by

$$g_{\mathrm{KJ}}(\theta;\mu,\gamma,\lambda,\rho)=\frac{1}{2\pi}\left\{1+2\gamma\frac{\cos(\theta-\mu)-\rho\cos\lambda}{1+\rho^2-2\rho\cos(\theta-\mu-\lambda)}\right\},\quad 0\le\theta<2\pi, \tag{1}$$

where the ranges of the parameters are $0\le\mu<2\pi$, $0\le\gamma<1$, $0\le\rho<1$, and $0\le\lambda<2\pi$ satisfying $(\rho\cos\lambda-\gamma)^2+(\rho\sin\lambda)^2\le(1-\gamma)^2$. Throughout the paper, the distribution (1) will be called the Kato–Jones distribution.

Then, we propose the following mixture with density

$$f(\theta)=\frac{1}{2\pi}\sum_{k=1}^{m}\pi_k g_{\mathrm{KJ}}(\theta;\mu_k,\gamma_k,\lambda_k,\rho_k),\quad 0\le\theta<2\pi, \tag{2}$$

where $m\in\mathbb{N}$ is the number of the components of the mixture and $0<\pi_1,\ldots,\pi_m<1$ are the weights of the components satisfying $\sum_{k=1}^{m}\pi_k=1$. The parameters of each component, namely, μ_k, γ_k, ρ_k, and λ_k, are defined as in (1).

3.2 Components of the mixture

Here, we introduce three important benefits of the Kato–Jones distribution (1), which motivates us to use this model as the components of the proposed mixture.

First, the Kato–Jones distribution is a flexible unimodal distribution that can provide a wide range of skewness and peakedness. As can be seen in Figure 1b and c of Kato and Jones (2015), the degrees of skewness and peakedness of the Kato–Jones distribution can be widely controlled by adjusting the parameters. Unimodality is also important to interpret each component of the mixture.

The second benefit is that each parameter of the Kato–Jones distribution is easily interpretable. The parameter μ controls the location or mean direction, while γ regulates the concentration or the mean resultant length. The parameters λ and ρ influence the skewness and kurtosis of the model. As

shown in Lemma 1 of Kato and Jones (2015), the circular skewness and kurtosis of Batschelet (1981) for the Kato–Jones distribution (1) are given by $\bar{\beta}_2 = \rho\gamma \sin\lambda$ and $\bar{\alpha}_2 = \rho\gamma\cos\lambda$, respectively. Note that kurtosis is related to the peakedness of the density (see Kato & Jones, 2015, Figure 1b). The Kato–Jones distribution can be reparametrized to have the following density with the parameters μ, γ, $\bar{\alpha}_2$, and $\bar{\beta}_2$:

$$g^*_{\mathrm{KJ}}(\theta;\mu,\gamma,\bar{\alpha}_2,\bar{\beta}_2) = \frac{1}{2\pi}\left[1 + 2\gamma^2\frac{\gamma\cos(\theta-\mu)-\bar{\alpha}_2}{\gamma^2+\bar{\alpha}_2^2+\bar{\beta}_2^2-2\gamma\{\bar{\alpha}_2\cos(\theta-\mu)+\bar{\beta}_2\sin(\theta-\mu)\}}\right].$$

Consequently, the density of the proposed mixture (2) can be written as

$$f(\theta) = \sum_{k=1}^{m}\pi_k g^*_{\mathrm{KJ}}(\theta;\mu_k,\gamma_k,\bar{\alpha}_{2k},\bar{\beta}_{2k}), \tag{3}$$

where $0 \le \mu_k < 2\pi$, $0 \le \gamma_k < 1$, and $(\bar{\alpha}_{2k}, \bar{\beta}_{2k}) \ne (\gamma_k, 0)$ satisfy $(\bar{\alpha}_{2k} - \gamma_k^2)^2 + \bar{\beta}_{2k}^2 \le \gamma_k^2(1-\gamma_k)^2$. With this representation of the density, all the parameters of each component of our mixture can be clearly interpreted; the parameters μ_k, γ_k, $\bar{\alpha}_{2k}$, and $\bar{\beta}_{2k}$ control the mean direction, mean resultant length, circular kurtosis, and circular skewness of the kth component of the mixture, respectively.

Finally, parameter estimation for the Kato–Jones distribution is straightforward by both the method of moments and maximum likelihood. This is mainly due to the fact that the trigonometric moments and probability density function of the Kato–Jones distribution (1) can be expressed in simple and closed form. In general, parameter estimation for mixture models is involved. However, the parameters of our mixture model (2) can be estimated with no great computational cost because of the simplicity of estimation for the Kato–Jones distribution.

The authors are not aware of other distributions which have all the three benefits above. Note also that the use of the Kato–Jones distribution is recommended in recent review papers of Ley et al. (2021) and Pewsey and García-Portugués (2021) as a flexible distribution on the circle.

3.3 Skewness of mixtures of symmetric distributions

In general, mixtures of symmetric distributions have skewed density shapes. Then, a natural question is whether it is necessary to use the mixtures of Kato–Jones distributions rather than mixtures of symmetric distributions for the modelling of skewed data. The following theorem provides an answer to this question.

Theorem 1 Let g_{KJ} be the density of the Kato–Jones distribution (1). Assume that $h_k(\theta)$ $(k = 1, \ldots, m)$ is the density of a distribution on the circle which is symmetric about $\theta = \mu_k (\in [0, 2\pi))$ and satisfies $\int_0^{2\pi}\cos\{p(\theta-\mu_k)\}h_k(\theta)d\theta > 0$ for any $p \in \mathbb{N}$. Suppose g_{KJ} can be expressed as $g_{\mathrm{KJ}}(\theta) = \sum_{k=1}^m \pi_k h_k(\theta)$, where $\sum_{k=1}^m \pi_k = 1$ and $\pi_k > 0$. Assume $\mu_k - \mu - \lambda$ is a rational number, i.e. $\mu_k - \mu - \lambda = a_k/b_k$ for some $a_k \in \mathbb{Z}$, $b_k = \mathbb{N}$. Then, g_{KJ} is a symmetric density with $\lambda = 0$.

See online supplementary material, Appendix D.1 for the proof.

This theorem suggests that an asymmetric Kato–Jones distribution essentially has a different skewed shape from mixture of symmetric distributions under certain conditions. Our experiences suggest that it is difficult to model strong skewness using mixtures of symmetric distributions. Indeed, as will be discussed later, mixtures of Kato–Jones distributions provide a much better fit to our data than those of the von Mises distributions because of the lack of fit in the skewed shape between 5 a.m. and 9 a.m. Also, it does not seem to be appropriate to use a mixture of symmetric distributions as a single component skewed distribution because a mixture of symmetric distributions is generally multimodal and therefore its interpretation is difficult.

Note that many well-known distributions satisfy the assumption $\int_0^{2\pi}\cos\{p(\theta-\mu_j)\}h_j(\theta)d\theta > 0$ for any $p \in \mathbb{N}$ in Theorem 1, including the von Mises, wrapped Cauchy, and wrapped normal distributions. In addition, a stronger statement can be made for mixtures of the wrapped Cauchy distributions as follows.

Theorem 2 Let g_{KJ} be the Kato–Jones density (1) with $\gamma < \rho$. Suppose that $h_{\mathrm{WC},k}(\theta)$ $(k = 1, \ldots, m)$ is the wrapped Cauchy distribution with density

$$h_{\mathrm{WC},k}(\theta) = \frac{1}{2\pi}\frac{1-\rho_k^2}{1+\rho_k^2-2\rho_k\cos(\theta-\mu_k)}, \quad 0 \le \theta < 2\pi,$$

where $0 \le \mu_k < 2\pi$, $0 \le \rho_k < 1$, and $(\mu_k, \rho_k) \neq (\mu_\ell, \rho_\ell)$ for $k \neq \ell$. Assume that $g_{\mathrm{KJ}}(\theta) = \sum_{k=1}^m \pi_k h_{\mathrm{WC},k}(\theta)$, where $\sum_{k=1}^m \pi_k = 1$ and $\pi_k > 0$. Then, $g_{\mathrm{KJ}}(\theta)$ is expressed as the following density of the symmetric two-component mixture

$$g_{\mathrm{KJ}}(\theta) = \frac{\gamma}{\rho}\frac{1}{2\pi}\frac{1-\rho}{1+\rho^2-2\rho\cos(\theta-\mu)} + \left(1-\frac{\gamma}{\rho}\right)\frac{1}{2\pi}, \quad 0 \le \theta < 2\pi. \tag{4}$$

See online supplementary material, Appendix D.2 for the proof. This theorem implies that any asymmetric case of the Kato–Jones distribution can not be expressed as mixtures of the wrapped Cauchy distributions.

4 Parameter estimation

We consider parameter estimation for the proposed mixture (2) by maximum likelihood and a modified method of moments. A key reparametrization is carried out to circumvent a problem in parameter estimation. In our data analysis, the maximum-likelihood estimator is mainly discussed because of its efficiency. However, the calculation of the modified method of moments estimate is faster and this estimate is used as an initial value of our algorithm for maximum-likelihood estimation.

4.1 Reparametrization

Before, we proceed to parameter estimation for the mixture (2), we reparametrize it to avoid potential problems associated with parameter estimation. First, we provide alternative expressions for the densities of Kato–Jones distributions and their mixture. See online supplementary material, Appendix D.3 for the proof.

Proposition 1 The following hold for the densities of Kato–Jones distributions (1) and their mixture (2).

(i) The probability density function of Kato–Jones distribution (1) can be expressed as

$$g_{\mathrm{KJ}}(\theta;\mu,\gamma,\rho,\lambda) = \pi' \cdot \frac{1}{2\pi}\left\{1 + 2\bar{\gamma}\frac{\cos(\theta-\mu)-\rho\cos\lambda}{1+\rho^2-2\rho\cos(\theta-\mu-\lambda)}\right\} + (1-\pi')\frac{1}{2\pi}, \tag{5}$$

where $\pi' = \gamma/\bar{\gamma} (\in (0, 1])$ and $\bar{\gamma} = (1-\rho^2)/\{2(1-\rho\cos\lambda)\}$ is the upper bound of the range of γ for given ρ and λ.

(ii) An alternative expression for the density of the mixture (2) is

$$f(\theta) = \frac{1}{2\pi}\sum_{k=1}^{m}\pi'_k \cdot g_{\mathrm{KJ}}(\theta;\mu_k,\bar{\gamma}_k,\rho_k,\lambda_k) + \frac{\pi'_{m+1}}{2\pi}, \tag{6}$$

where g_{KJ} is as in (5), $\pi'_k = \pi_k\gamma_k/\bar{\gamma}_k$ $(k = 1, \ldots, m)$, $\pi'_{m+1} = 1 - \sum_{k=1}^m \pi'_k$, and $\bar{\gamma}_k = (1-\rho_k^2)/\{2(1-\rho_k\cos\lambda_k)\}$.

Proposition 1 implies that the proposed mixture (2) can be viewed as a mixture of Kato–Jones submodels $g_{\mathrm{KJ}}(\mu_k, \bar{\gamma}_k, \lambda_k, \rho_k)$ and a uniform distribution. The expression (6) suggests that π_k and γ_k cannot be uniquely determined in parameter estimation because the information on both parameters is contained in the single parameter π'_k. This implies that the parameters of the proposed mixtures reduce to $(\mu_1, \ldots, \mu_m, \rho_1, \ldots, \rho_m, \lambda_1, \ldots, \lambda_m, \pi'_1, \ldots, \pi'_{m+1})$. With this parametrization, it is possible to prove the identifiability of the proposed family of mixtures. See online supplementary material, Appendix D.4 for the proof.

Theorem 3 Consider a family of mixture models with the density (6), where μ_k, ρ_k, λ_k, and π'_ℓ are the parameters of the family and $\bar{\gamma}_k = (1-\rho_k^2)/\{2(1-\rho_k \cos\lambda_k)\}$ $(k = 1, \ldots, m,\ \ell = 1, \ldots, m+1)$. Define the parameter space of the family (6) by

$$\Omega = \left\{ (\mu_1, \ldots, \mu_m, \rho_1, \ldots, \rho_m, \lambda_1, \ldots, \lambda_m, \pi'_1, \ldots, \pi'_{m+1}) \,\middle|\, \begin{array}{l} \mu_k, \lambda_k \in [0, 2\pi),\ \rho_k \in (0, 1),\ \rho_k > \rho_\ell,\ 1 \le k < \ell \le m, \\ \pi'_q > 0,\ \sum_{q=1}^{m+1} \pi'_q = 1,\ q = 1, \ldots, m+1,\ m \in \mathbb{N} \end{array} \right\}. \tag{7}$$

Then, the family of the mixtures (6) is identifiable under Ω.

Therefore, instead of estimating the parameters of (2) directly, we estimate the reparametrized mixture (6) to ensure the identifiability of the model.

In order to interpret the estimated model, we consider the following two approaches:

(a) The first approach is to interpret the reparametrized mixture (6) directly.
(b) Alternatively, we recover the original parameters, (2) or (3), from the reparametrized ones (6) and interpret the mixture with the recovered parameters.

For achieving the recovery of the parameters in approach (b), it is necessary to add an additional assumption on the parameters to allocate the mixing proportion, π'_{m+1}, of the uniform component to the other mixing proportions, $\{\pi'_k\}_{k=1}^m$. To this end, we propose the assumption $\pi_k = \pi'_k/(1-\pi'_{m+1})$ between the original and reparametrized parameters. Although it is mathematically impossible to find which cluster observations from the uniform distribution in (6) belong to, it seems natural to assume that the ratio of the uniform observations in the kth cluster is proportional to that of the observations from Kato–Jones distribution $g_{\mathrm{KJ}}(\mu_k, \bar{\gamma}_k, \lambda_k, \rho_k)$ in the kth cluster.

Summarizing the results above, we make the following steps to estimate the parameters of the mixture (1):

(i) Estimate the parameters $\{\mu_k, \rho_k, \lambda_k, \pi'_k\}$ of the mixture (6) to obtain their estimates $\{\hat{\mu}_k, \hat{\rho}_k, \hat{\lambda}_k, \hat{\pi}'_k\}$.
(ii) Recover the estimates of the original parameters $\{\hat{\mu_k}, \hat{\gamma}_k, \hat{\rho}_k, \hat{\lambda}_k, \hat{\pi}_k\}$ via $\hat{\pi}_k = \hat{\pi}'_k/(1-\hat{\pi}'_{m+1})$ and $\hat{\gamma}_k = \hat{\pi}'_k\hat{\bar{\gamma}}_k/\hat{\pi}_k$.

Then, we interpret the reparametrized mixture (6) estimated in Step (i) as in approach (a) or the original mixture (2) estimated in Step (ii) as in approach (b). Note that the mixture estimated via approach (a) or (b) has the same flexibility as the original mixture (2) although the former mixture has $m-1 (= (5m-1)-4m)$ fewer parameters than the latter mixture. In this paper, we mainly discuss the interpretation of the mixture (2) estimated in Step (ii).

In order to achieve the identifiability of the Kato–Jones mixtures with the original parametrization (2), an alternative assumption on the parameters is to adopt the same assumption as the asymmetric three-parameter Kato–Jones distributions (Kato & Jones, 2015, equation (7))

for each component of the mixture (2). This can be easily done by imposing $\gamma_k = \rho_k \cos\lambda_k$ in (2). However, the degree of kurtosis, which influences the peakedness, of each component cannot be regulated for this mixture. On the other hand, the reparametrized mixture (6) achieves much greater flexibility than the mixture of three-parameter asymmetric Kato–Jones submodels although the reparametrized mixture (6) has only one more parameter than latter mixture.

4.2 Modified method of moments estimation

Throughout this section, let random variables $\Theta_1, \ldots, \Theta_n$ be independent and identically distributed from the reparametrized mixture (6). Denote the parameters of the mixture (6) by

$$\boldsymbol{\Psi} = (\mu_1, \ldots, \mu_m, \rho_1, \ldots, \rho_m, \lambda_1, \ldots, \lambda_m, \pi'_1, \ldots, \pi'_{m+1}).$$

First, we discuss the modified method of moments estimation based on trigonometric moments. It is straightforward from Kato and Jones (2015, equation (2)) to see that the pth trigonometric moment of Θ_j is given by

$$E(e^{ip\Theta_j}) = \sum_{k=1}^{m} \pi'_k \bar{\gamma}_k (\rho_k e^{i\lambda_k})^{-1} \left\{\rho_k e^{i(\mu_k+\lambda_k)}\right\}^p, \quad p \in \mathbb{N}, \tag{8}$$

where i is the imaginary unit. Ordinary method of moments estimation based on the trigonometric moments is obtained by equating the theoretical trigonometric moments (8) to empirical ones, namely,

$$\frac{1}{n}\sum_{j=1}^{n} e^{ip\Theta_j} = \sum_{k=1}^{m} \pi'_k \bar{\gamma}_k (\rho_k e^{i\lambda_k})^{-1} \left\{\rho_k e^{i(\mu_k+\lambda_k)}\right\}^p, \tag{9}$$

for some selected values of p. However, there is a problem associated with the ordinary method of moments estimation for the proposed mixture (2). As is the case for $p = 1$, the solutions to equation (9) are not always within the range of λ_k and ρ_k (see Kato & Jones, 2015, Section 5.1). In order to circumvent this problem, we consider the following function to evaluate the weighted error between the empirical and theoretical trigonometric moments:

$$\begin{aligned}\mathrm{ETM}(\boldsymbol{\Psi}) &\equiv \sum_{p=1}^{q} w(p) \left| \frac{1}{n}\sum_{j=1}^{n} e^{ip\Theta_j} - E(e^{ip\Theta}) \right|^2 \\ &= \sum_{p=1}^{q} w(p) \left| \frac{1}{n}\sum_{j=1}^{n} e^{ip\Theta_j} - \sum_{k=1}^{m} \pi'_k \bar{\gamma}_k (\rho_k e^{i\lambda_k})^{-1} \left\{\rho_k e^{i(\mu_k+\lambda_k)}\right\}^p \right|^2,\end{aligned}$$

where $q \in \mathbb{N}$ and $w(p)$ is a weight function.

This function can also be expressed as

$$\begin{aligned}\mathrm{ETM}(\boldsymbol{\Psi}) = \sum_{p=1}^{q} w(p) \Bigg[&\left\{ \frac{1}{n}\sum_{j=1}^{n} \cos p\Theta_j - \sum_{k=1}^{m} \pi'_k \bar{\gamma}_k \cdot \rho_k^{p-1} \cos\big(p\mu_k + (p-1)\lambda_k\big) \right\}^2 \\ &+ \left\{ \frac{1}{n}\sum_{j=1}^{n} \sin p\Theta_j - \sum_{k=1}^{m} \pi'_k \bar{\gamma}_k \cdot \rho_k^{p-1} \sin\big(p\mu_k + (p-1)\lambda_k\big) \right\}^2 \Bigg].\end{aligned}$$

Then, we propose a modified method of moments estimator as the minimizer of ETM, namely,

$$\hat{\boldsymbol{\Psi}} = \underset{\boldsymbol{\Psi}\in\Omega}{\operatorname{argmin}}\, \mathrm{ETM}(\boldsymbol{\Psi}), \tag{10}$$

where Ω is the parameter space of $\boldsymbol{\Psi}$ defined in (7). If there exists a solution $\hat{\boldsymbol{\Psi}} \in \Omega$ to equation (9) for $p = 1, \ldots, q$, then $\mathrm{ETM}(\hat{\boldsymbol{\Psi}}) = 0$ and $\hat{\boldsymbol{\Psi}}$ becomes the same as the ordinary method of moments estimate. Therefore, the proposed estimator (10) is a generalization of the method of moments estimator. An advantage of the proposed estimator is that the estimate always belongs to the parameter space. For a single distribution $m = 1$ with $w(1) = c$ and $w(p) = 0$ $(p \geq 3)$, this estimator converges to the method of moments estimator of Kato and Jones (2015) as $w(2) \to 0$.

The proposed estimator (10) is somewhat related to the estimator for the stable distribution on $\mathbb{R}$ discussed in Press (1972). Both estimators are M-estimators based on the square of the absolute difference between theoretical and empirical functions related to the characteristic functions. However, our estimator is essentially different from the estimator of Press (1972) in the sense that our estimator is for a different distribution on a different manifold and avoids an integral over the infinite interval of the argument of the function.

The result of the estimation (10) depends on the choice of q and $w(p)$. Since the mixture (6) has essentially $4m$ free parameters, there are potentially multiple solutions to the equation $\mathrm{ETM}(\hat{\boldsymbol{\Psi}}) = 0$ for $q < 2m$. To avoid this potential problem, it is recommended to assume $q \geq 2m$ and, in our analysis, we assume that $q = 2m$. As for the choice of the weight function $w(p)$, we adopt $w(p) = c^p (0 < c < 1)$ in our data analysis to place much more importance on the low-order trigonometric moments than on the high-order ones. This choice is made in a somewhat similar spirit to the method of moments estimation of Kato and Jones (2015) in which a couple of estimates are determined only by the first trigonometric moment.

It is known that, under certain regularity conditions, consistency holds for M-estimators. The modified method of moments estimator (10) defined for the identifiable model given in Theorem 3 satisfies all the regularity conditions for consistency given in Case A of Section 6.2 of Huber (1981). One of the regularity conditions for consistency is the following convergence:

$$\mathrm{ETM}(\hat{\boldsymbol{\Psi}}) - \inf_{\boldsymbol{\Psi} \in \Omega} \mathrm{ETM}(\boldsymbol{\Psi}) \longrightarrow 0 \quad \text{almost surely as } n \to \infty.$$

This convergence does not hold if mixture models are not identifiable or objective functions for minimization have multiple solutions. However, as seen in Theorem 3, our reparametrized mixture (6) is identifiable. In addition, if q is appropriately selected, the equation $\mathrm{ETM}(\hat{\boldsymbol{\Psi}}) = 0$ has a unique solution (which is not necessarily within the parameter space), and the law of large numbers guarantees that the solution almost surely converges to the true value in the parameter space. The other regularity conditions for consistency given in Huber (1981, Section 6.2, (A.1)–(A.5)) are satisfied for our simple objective function and parameter space.

As for the asymptotic normality of the estimator (10), it does not seem straightforward to prove the regularity conditions of Huber (1981, Section 6.3). Specifically, we need to prove that $\partial E(\mathrm{ETM}(\boldsymbol{\Psi}))/\partial \boldsymbol{\Psi}$ has a nonsingular derivative matrix as assumed in Corollary 3.2 of Huber (1981), and this would be future work.

4.3 Maximum-likelihood estimation

Next, we consider the maximum-likelihood estimation. The log-likelihood function for the sample $(\theta_1, \ldots, \theta_n)$ is given by

$$\ell(\boldsymbol{\Psi}) = C + \sum_{j=1}^{n} \log \left[\sum_{k=1}^{m} \pi'_k \left\{ 1 + 2\bar{\gamma}_k \frac{\cos(\theta_j - \mu_k) - \rho_k \cos \lambda_k}{1 + \rho_k^2 - 2\rho_k \cos(\theta - \mu_k - \lambda_k)} \right\} + \pi'_{m+1} \right], \tag{11}$$

where $C = -n \log(2\pi)$. As is the case for $p = 1$, there is no closed-form expression for the maximum-likelihood estimator for general p. Therefore, we consider a numerical algorithm to estimate the maximum-likelihood estimate of the mixture (6). We apply the EM algorithm to estimate the mixing proportions of the mixture (e.g. McLachlan & Krishnan, 2008, Section 1.4.3). The following algorithm is established:

Algorithm 4.1

Step 1: Take an initial value $\boldsymbol{\Psi}^{(0)}$, where

$$\boldsymbol{\Psi}^{(0)} = (\mu_1^{(0)}, \ldots, \mu_m^{(0)}, \rho_1^{(0)}, \ldots, \rho_m^{(0)}, \lambda_1^{(0)}, \ldots, \lambda_m^{(0)}, \pi'^{(0)}_1, \ldots, \pi'^{(0)}_{m+1}).$$

Step 2: For $r = 1, \ldots, N$, compute the following until the value of $\boldsymbol{\Psi}^{(N)}$ is virtually unchanged from $\boldsymbol{\Psi}^{(N-1)}$:

$$\begin{aligned} \pi'^{(r)}_k &= \frac{1}{n}\sum_{j=1}^{n} w_{kj}^{(r-1)}, \quad \pi'^{(r)}_{m+1} = 1 - \sum_{k=1}^{m} \pi'^{(r)}_k, \\ \left(\mu_k^{(r)}, \lambda_k^{(r)}, \rho_k^{(r)}\right) &= \underset{(\mu_k,\lambda_k,\rho_k)}{\operatorname{argmax}} \left[\sum_{j=1}^{n} w_{kj}^{(r-1)} \log\{g_{\mathrm{KJ}}(\theta_j; \mu_k, \bar{\gamma}_k, \lambda_k, \rho_k)\}\right], \end{aligned} \tag{12}$$

where $k = 1, \ldots, m$ and

$$w_{kj}^{(r-1)} = \frac{\pi'^{(r-1)}_k g_{\mathrm{KJ}}(\theta_j; \mu_k^{(r-1)}, \bar{\gamma}_k^{(r-1)}, \lambda_k^{(r-1)}, \rho_k^{(r-1)})}{\sum_{h=1}^{m} \pi'^{(r-1)}_h g_{\mathrm{KJ}}(\theta_j; \mu_h^{(r-1)}, \bar{\gamma}_h^{(r-1)}, \lambda_h^{(r-1)}, \rho_h^{(r-1)}) + \pi'^{(r-1)}_{m+1}/(2\pi)}.$$

Step 3: Record $\hat{\boldsymbol{\Psi}}^{(N)}$ as the maximum-likelihood estimate of $\boldsymbol{\Psi}$.

An advantage of this algorithm is that the values of $\{\pi'^{(r)}_k\}$ can be expressed in closed form. In addition, although it is necessary to calculate $(\mu_k^{(r)}, \lambda_k^{(r)}, \rho_k^{(r)})$ numerically in (12), this maximization is essentially weighted maximum-likelihood estimation for a single Kato–Jones distribution and can be done in a similar manner as in Kato and Jones (2015, Section 5.2).

Our experiments suggest that the modified method of moments estimates (10) provide useful initial values of the algorithm. However, it is advisable to try multiple starting values to ensure the global maximum of the likelihood function. The modified method of moments estimation is faster than the maximum-likelihood estimation. There is, however, no great difficulty in implementing maximum-likelihood estimation using the algorithm above.

Consistency and asymptotic normality hold for the maximum-likelihood estimator under certain regularity conditions. One of the regularity conditions for consistency given in Miyata et al. (2020, (A1)) is the boundedness of the likelihood function. As the likelihood function (11) with certain conditions on the parameters tends to $-\infty$ as $\rho_k \to 1$, we need to replace the range of ρ_k of Ω in (7) by $0 < \rho \leq 1 - \varepsilon$ for a given small value of $\varepsilon > 0$. With this modification, the maximum-likelihood estimator is consistent in the sense of Theorem 1 of Miyata et al. (2020). In order to prove the asymptotic normality for the maximum-likelihood estimator of our model, we need to show that the Fisher information matrix of $\boldsymbol{\Psi}$ is positive definite (Redner & Walker, 1984, Condition 2) and that is our future work.

In our data analysis, we mainly discuss the maximum-likelihood estimate because the maximum-likelihood estimator has lower variance than the method of moments estimator (10) for a large sample size as demonstrated in a simulation study in Section 5.

5 Simulation study

We compare the performance of the modified method of moments estimation and the maximum-likelihood estimation by EM algorithm via a Monte Carlo simulation study. Random samples of sizes $n = 50$, 100, 500, 1,000, and 5,000 were generated from the estimated distribution whose parameters are shown in Table 4b. For each sample size, $r = 2{,}000$ samples were generated.

For comparison of the two estimators, we employ the generalized mean squared error which is defined by $\det(\Sigma)$, where $\Sigma = E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)^T\}$ and $\hat{\xi}$ is an estimator of $\xi = (\mu_1, \mu_2, \rho_1, \rho_2, \lambda_1, \lambda_2, \pi'_1, \pi'_2)^T$. An estimate of the generalized mean squared error is given by replacing Σ by its sample analogue $\hat{\Sigma} = r^{-1}\sum_{j=1}^{r}(\hat{\xi}_j - \xi)(\hat{\xi}_j - \xi)^T$, where the $\hat{\xi}_j$'s, $j = 1, \ldots, r$, are the

estimates from the r simulation samples. We also consider the estimated relative generalized mean squared error of the estimate via the modified method of moments with respect to the maximum-likelihood estimate, which is defined as

$$\widehat{\mathrm{RGMSE}} = \frac{\det(\hat{\Sigma}_{MM})}{\det(\hat{\Sigma}_{ML})}.$$

Here, $\hat{\Sigma}_{MM}$ and $\hat{\Sigma}_{ML}$ are sample estimates of the mean squared error matrices of modified method of moments estimation and maximum-likelihood estimation, respectively.

For numerical optimization of ETM, the scipy.optimize.minimize package in Python with SLSQP method is employed (Jones et al., 2001; Kraft, 1988). This method is a kind of quasi-Newton method and enables us to solve the constrained optimization problem. In this analysis, the optimization of ETM is terminated when the difference in the value of ETM between two successive steps is lower than 1×10^{-10}. The initial values of μ_k, ρ_k and λ_k are random samples from the uniform distribution on the intervals $[0, 2\pi)$, $[0, 1)$ and $[0, 2\pi)$, respectively. Also, those of π'_1, π'_2 and π'_3 are in $[0, 1)$ and $\pi'_1 + \pi'_2 + \pi'_3 = 1$. For this sampling, two random values are drawn from the uniform distribution on $[0, 1)$ and the smaller one is regarded as r_1 and the other as r_2. Then, $\pi'_1 = r_1$, $\pi'_2 = r_2 - r_1$, and $\pi'_3 = 1 - r_2$ are calculated. Furthermore, the moment weight parameter c in $w(p)$ is set to 0.9 because our experiments suggest that estimates with small c are not stable with regard to fitting the data.

100 estimates are obtained via this process, and the estimate whose value of ETM is smallest is set as the initial value for the maximum-likelihood estimation. For numerical optimization in the M-step, namely, Step 2 in Algorithm 4.1, the same Python package as in the optimization of ETM is employed. In this analysis, each M-step is terminated when the difference of the value of the objective function between two successive iterations is lower than 1×10^{-6}. Also, the whole EM algorithm is terminated when the difference of the log-likelihood function between two successive M-steps is lower than $n \times 10^{-6}$. In both estimates, it is assumed that $\hat{\mu}_1 < \hat{\mu}_2$.

Table 1 shows the value of $\det(\hat{\Sigma}_{MM})$, $\det(\hat{\Sigma}_{ML})$, and $\widehat{\mathrm{RGMSE}}$ for each sample size. In addition, Figure 3 shows the trend of the change of the value of $\det(\hat{\Sigma}_{MM})$ and $\det(\hat{\Sigma}_{ML})$ in different sample sizes. As the sample size n increases, the value of $\det(\hat{\Sigma}_{MM})$ and $\det(\hat{\Sigma}_{ML})$ decreases. The value of $\det(\hat{\Sigma}_{MM})$ at $n = 50$ is smaller than $\det(\hat{\Sigma}_{ML})$, however, the value of $\det(\hat{\Sigma}_{MM})$ at $n = 5{,}000$ is larger than $\det(\hat{\Sigma}_{ML})$ because the degree of decline in value with respect to sample size is greater for $\det(\hat{\Sigma}_{ML})$ than for $\det(\hat{\Sigma}_{MM})$. Therefore, the values of $\widehat{\mathrm{RGMSE}}$ are smaller than 1 for $n \leq 1{,}000$ and larger for $n = 5{,}000$. These results imply modified moments estimation is preferable for estimation from a small number of samples in terms of $\widehat{\mathrm{RGMSE}}$. On the other hand, maximum-likelihood estimation is preferable for estimation from a large number of samples such as the data in this paper. Note that modified moments estimation yields not only an estimate but also a good initial value for maximum-likelihood estimation. Although the value of $\widehat{\mathrm{RGMSE}}$ is relatively large for $n = 5{,}000$, the very small value of $\det(\hat{\Sigma}_{MM})$ for $n = 5{,}000$ indicates that the estimate via the modified moments estimation still has reasonable performance as the initial value for maximum-likelihood estimation. Therefore, modified moments estimation is useful to obtain the maximum-likelihood estimates efficiently for a large sample size as well as small and medium sample sizes.

It is noted that the degree of decline of $\det(\hat{\Sigma})$ with respect to sample size depends on the estimation method. Therefore, for $n = 500$, the value of $\widehat{\mathrm{RGMSE}}$ is slightly smaller than for $n = 100$. The

Table 1. The values of $\det(\hat{\Sigma}_{MM})$, $\det(\hat{\Sigma}_{ML})$, and $\widehat{\mathrm{RGMSE}}$

n	$n = 50$	$n = 100$	$n = 500$	$n = 1{,}000$	$n = 5{,}000$
$\det(\hat{\Sigma}_{MM})$	5.35×10^{-9}	7.00×10^{-10}	2.70×10^{-13}	4.25×10^{-15}	4.01×10^{-20}
$\det(\hat{\Sigma}_{ML})$	5.05×10^{-8}	6.41×10^{-9}	3.56×10^{-12}	1.33×10^{-14}	1.23×10^{-22}
$\widehat{\mathrm{RGMSE}}$	0.106	0.109	0.076	0.322	325.3

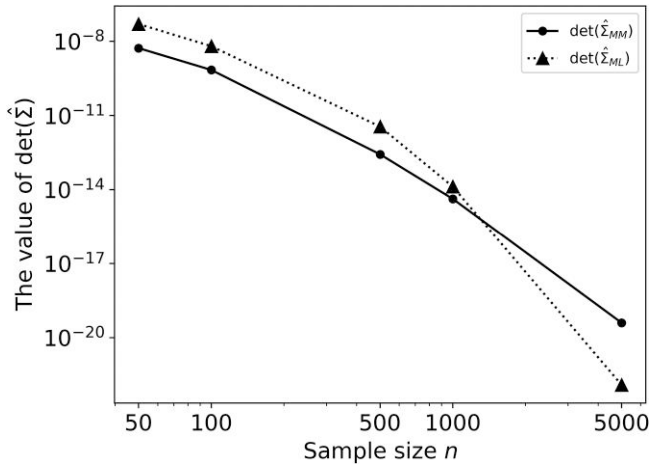

Figure 3. Plot of the value of $\det(\hat{\Sigma}_{MM})$ (bold) and $\det(\hat{\Sigma}_{ML})$ (dotted) in simulation study.

tendency for $\widehat{\mathrm{RGMSE}}$ to not increase monotonically with respect to n is also observed in the unimodal Kato–Jones distribution (Kato & Jones, 2015, online supplementary material).

6 Application

6.1 Application of modified method of moments estimation

The proposed inferential methods are applied to the traffic count data of interest. The number of samples is 1,121,262, as mentioned in Section 2. In this section, the number of components is fixed to two because two main peaks are recognized in the histogram (see Section 6.2 for results for other numbers of components). The conditions of the modified method of moments estimation are the same as in the previous section. Note that (Nagasaki et al., 2023) in Japanese traffic engineering journal also analysed traffic count data by applying the method proposed in this paper. However, Nagasaki et al. (2023) only discussed the ultimate results from the viewpoint of traffic engineering and did not discuss theoretical or statistical aspects of the method. Moreover, they just focused on comparing the results of data from many counters in different locations, whereas this paper interprets the results from a single location in more detail.

The optimization is carried out for 10,000 initial values. Figure 4 and Table 2 show the best result in the sense that the value of ETM is the smallest among 10,000 estimations. Table 3 provides the values of the empirical and theoretical trigonometric moments. The solid line in Figure 4 represents the estimated density of the mixture and the dashed lines refer to the estimated densities of the components. The same convention will be used in the following figures of fitted densities. The value of ETM is 7.06×10^{-13} and that of the log-likelihood function is $-1{,}840{,}939$. As shown in the right column of Table 3, the difference between the values of empirical and theoretical trigonometric moments of all the degrees are approximately 0, which means that this result is very close to the result of ordinary method of moments estimation. Nonetheless, we generally need the proposed modified method of moments to ensure that the parameters are within their ranges, as explained in Section 4.2. Actually, when the number of components m is set to be greater than two, much larger values of ETM are obtained for some cases and the ordinary method of moments estimates are out of range for some parameters (see Section 6.2). Also, the estimated densities in Figure 4 fit fairly well to the histogram. The sharp and negatively skewed peak around $\pi/2$ is represented by the component with $k = 1$, while the gentle peak around $3\pi/2$ is modelled by the component with $k = 2$.

The computation time for 10,000 trials is about 14 s. The calculation was run on a Windows 10 computer with an Intel Core i9-12900 processor running at 3.20 GHz using 64.0 GB of RAM. Furthermore, about 25% of 10,000 trials converge to almost the same value as the best result. In this case, 10,000 initial values were used to verify the performance of the modified method

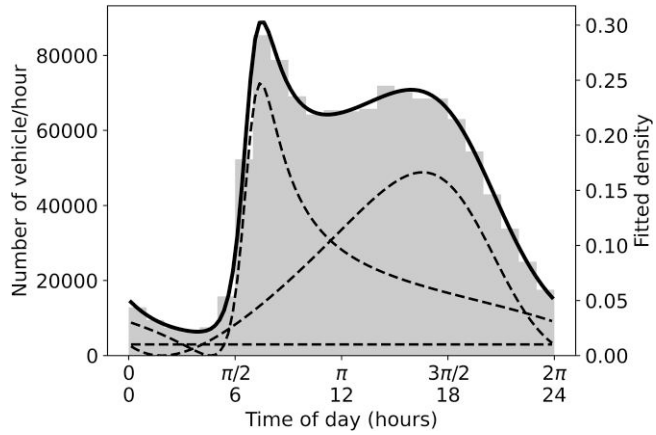

Figure 4. Plot of the densities of the reparametrized mixture (6) (bold) and its components (dashed) estimated via the modified method of moments estimation.

Table 2. Parameter estimates associated with the modified method of moments estimation

(a) Initial value					(b) Estimated parameter				
k	μ_k	ρ_k	λ_k	π'_k	k	$\hat{\mu}_k$	$\hat{\rho}_k$	$\hat{\lambda}_k$	$\hat{\pi}'_k$
1	2.2755	0.5322	4.9979	0.0316	1	2.7514	0.7322	5.3162	0.4543
2	3.9458	0.8131	2.0455	0.2331	2	4.0106	0.1947	1.1589	0.4820
3	–	–	–	0.7353	3	–	–	–	0.0637

Table 3. Values of the empirical and theoretical trigonometric moments (t.m.'s) and squares of their differences in the best result

		empirical t.m.	theoretical t.m.	square of difference
cosine	$p=1$	-3.29×10^{-1}	-3.29×10^{-1}	2.95×10^{-14}
	$p=2$	-7.07×10^{-2}	-7.07×10^{-2}	5.64×10^{-15}
	$p=3$	9.46×10^{-2}	9.46×10^{-2}	3.07×10^{-15}
	$p=4$	-1.61×10^{-2}	-1.61×10^{-2}	8.35×10^{-15}
sine	$p=1$	-1.23×10^{-1}	-1.23×10^{-1}	2.81×10^{-14}
	$p=2$	-1.18×10^{-1}	-1.18×10^{-1}	1.21×10^{-13}
	$p=3$	1.29×10^{-2}	1.29×10^{-2}	4.21×10^{-16}
	$p=4$	6.97×10^{-2}	6.97×10^{-2}	7.43×10^{-16}

of moments estimation. The estimates that provide the smallest value of ETM can be obtained from 25% of the initial values, which implies that estimation from a much smaller number of initial values (e.g. 100) would be sufficient.

Note that the modified method of moments estimation is carried out to obtain a good initial value for maximum-likelihood estimation, which is more appropriate for such a large sample size, as discussed in Section 5.

6.2 Maximum-likelihood estimation

The EM algorithm is carried out by setting the initial values as the best estimates from the modified method of moments to achieve the maximum-likelihood estimation efficiently. This choice of the

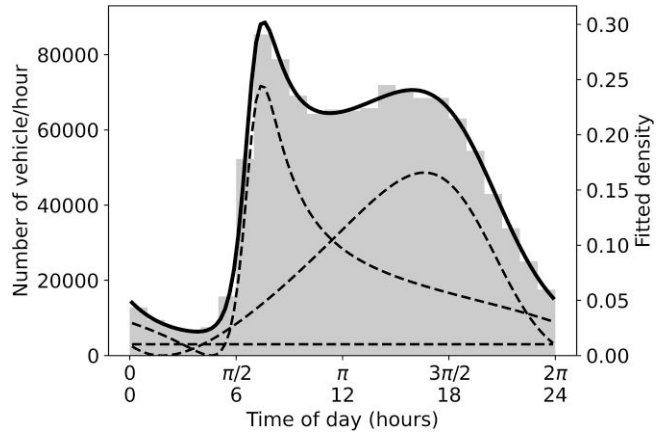

Figure 5. Plot of the densities of the reparametrized mixture (6) (bold) and its components (dashed) estimated via the EM algorithm.

initial values seems reasonable because of the large sample size of the data and the consistency of the modified method of moments estimator and maximum-likelihood estimator. The conditions of the maximum-likelihood estimation are the same as in the previous section except for the threshold for the termination of the EM algorithm, which is set to 1. This threshold is set by the following reasoning. Since the number of samples in this study is more than 1,000,000, the absolute value of the log-likelihood function is expected to be of that order of magnitude. Thus, we regard the threshold 1 as small enough to claim convergence.

The result of the EM algorithm is shown in Figure 5 and Table 4. It takes just two steps to terminate the EM algorithm. The value of ETM increases along with that of the log-likelihood during the EM algorithm. The difference between the parameter estimates by the modified method of moments and those by the maximum-likelihood estimation is not much. This implies that the modified method of moments estimate is close to the maximum-likelihood estimate, the estimate produced by another consistent estimator.

In addition, the standard errors for the parameter estimates by the modified method of moments and those by the maximum-likelihood estimation are calculated. The standard errors for the parameter estimates by the modified method of moments are calculated by the nonparametric bootstrap method (Davison & Hinkley, 1997), and those by the maximum-likelihood estimation are calculated by the Fisher information matrix for the log-likelihood function due to computation time.

The results are shown in Table 5. These results indicate that, for all parameters, the values of standard error of the maximum-likelihood estimates are smaller than those of the modified method of moments estimates. However, the modified method of moments estimates generally exhibits satisfactory performance, supporting our idea of using these estimates as the initial values of the EM algorithm. For both the modified method of moments and maximum-likelihood estimates, the values of the standard errors of $\hat{\lambda}_k$ are greater than those of the other parameters. The values of standard error for the parameter estimates by the maximum-likelihood estimation are in between the order of the second and third power of 10, and those by the modified method of moments are in between the order of the one and third power of 10.

As for the computation time, it takes 93 s. for the EM algorithm although only two M-steps are conducted in the calculation. The calculation environment is the same as that of the previous subsection. Compared to the modified method of moments that takes only about 14 s. for 10,000 trials, the EM algorithm requires a longer time. Therefore, setting the initial value close to the maximum-likelihood estimator is necessary to achieve the maximum-likelihood estimation by the EM algorithm in a feasible computation time. The proposed modified method of moments is one possible solution for this purpose.

Although two parameter estimation methods are introduced in this paper, the maximum-likelihood estimation is preferred in terms of the mean squared error of the estimators and usage for model comparison. The modified method of moments is useful to obtain a good initial value for the maximum-likelihood estimation with a short computation time for large sample sizes.

Table 4. Parameter estimates associated with the maximum-likelihood estimation for the reparametrized mixture (6) via the EM algorithm

k	$\hat{\mu}_k$	$\hat{\rho}_k$	$\hat{\lambda}_k$	$\hat{\pi}'_k$
1	2.7572	0.7266	5.3136	0.4536
2	4.0107	0.1970	1.1895	0.4825
3	–	–	–	0.0639

Table 5. The value of standard errors for the parameter estimates by the modified method of moments and those by the maximum-likelihood estimation

	k	$\hat{\mu}_k$	$\hat{\rho}_k$	$\hat{\lambda}_k$	$\hat{\pi}'_k$
the modified method of moments	1	5.77×10^{-2}	2.46×10^{-2}	1.87×10^{-1}	8.91×10^{-3}
	2	5.74×10^{-2}	2.47×10^{-2}	1.90×10^{-1}	5.92×10^{-3}
	3	–	–	–	6.59×10^{-2}
the maximum-likelihood estimation	1	6.36×10^{-3}	3.50×10^{-3}	2.88×10^{-2}	3.27×10^{-3}
	2	6.82×10^{-3}	1.98×10^{-3}	9.03×10^{-3}	3.76×10^{-3}
	3	–	–	–	3.30×10^{-3}

Here, the parameters are estimated in the form of the mixture of two Kato–Jones distributions and the uniform distribution (Table 4b and equation (4)). Actually, we can regard this mixture of three components as the estimated model. Nonetheless, we are also able to reparameterize and recover the estimates of the original parameters (2) as discussed in Section 4.1. As we initially aim to estimate the mixture of two components, we employ the latter in the following. The maximum-likelihood estimates obtained via the EM algorithm (Figure 5 and Table 4) are recovered to $\hat{\mu}_k$, $\hat{\gamma}_k$, $\hat{\rho}_k$, $\hat{\lambda}_k$ and $\hat{\pi}_k$ as shown in Figure 6 and Table 6. The shapes of the two estimated components with $k = 1, 2$ in Table 4 slightly change by adding the restored uniform component with $k = 3$.

In addition, $\hat{\mu}_k$, $\hat{\gamma}_k$, $\hat{\rho}_k$, $\hat{\lambda}_k$ are reparameterized into $\hat{\mu}_k$, $\hat{\gamma}_k$, $\hat{\bar{\alpha}}_{2k}$, $\hat{\bar{\beta}}_{2k}$ for interpretation because $\hat{\bar{\alpha}}_{2k}$ and $\hat{\bar{\beta}}_{2k}$ can be interpreted as the circular skewness and kurtosis of Batschelet (1981) as mentioned in Section 3.2. The recovered parameters are reparameterized as shown in Table 7. The discussion of the parameter values to follow is carried out only for the parameters in Table 7 because of the interpretability for $\hat{\bar{\alpha}}_{2k}$ and $\hat{\bar{\beta}}_{2k}$.

6.3 Discussion

Here, we discuss the interpretation of the estimated models with the recovered parametrizations (3) given in Table 7. As shown in Figure 6, the densities of the mixture fit fairly well to the histogram for the entire day, such as the sharp peak in the morning and the gentle peak in the evening. The numerically calculated morning mode and evening mode of the whole distribution of the mixture are 7:32 a.m. and 3:56 p.m., respectively.

The two components divide the mixture into 48.5% and 51.5% as denoted by each value of π_k. Each peak roughly represents the morning and evening traffic. The peak of the component with $k = 1$ is sharp and negatively skewed. Its mean, $\hat{\mu}_1$, is around 10:32, and its mode is around 7:28 (Kato & Jones, 2015, online supplementary material). The mode of the component with $k = 1$ is close to the morning mode of the mixture because the component with $k = 1$ is dominant around the morning. The shape is characterized by large $\hat{\bar{\alpha}}_{21}$ and negatively large $\hat{\bar{\beta}}_{21}$. On the other hand, the peak of the component with $k = 2$ is gentle and positively skewed. Its mean, $\hat{\mu}_2$, is around 3:19 p.m., and its mode is around 4:37 p.m. The mode of the component with $k = 2$ and the evening mode of the mixture are apart because the tail of the component with $k = 1$ is contained even in the evening part of the mixture. The shape is characterized by small $\hat{\bar{\alpha}}_{22}$ and positively small $\hat{\bar{\beta}}_{22}$.

The peak period of the component with $k = 1$ is from 7:00 a.m. to 8:00 a.m., and its shape is remarkably sharp and quick-rising, which is described by large $\hat{\bar{\alpha}}_{21}$ and negatively large $\hat{\bar{\beta}}_{21}$.

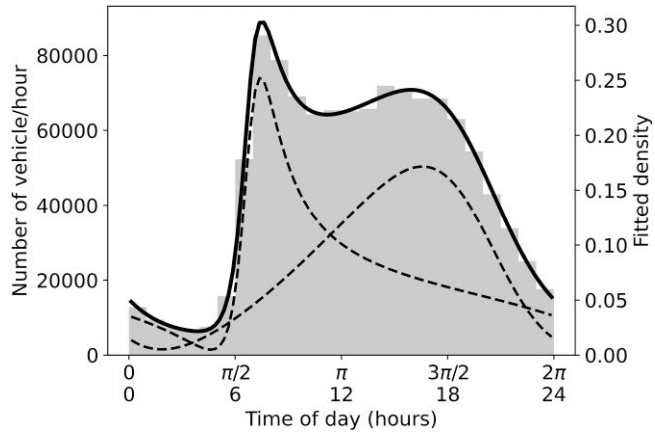

Figure 6. Plot of the estimated densities of the recovered mixture (2) (bold) and their components (dashed).

This is explained by the actual situation that many drivers would like to arrive in central Osaka around the same time, like 8:30 a.m. and 9:00 a.m., which is the typical start time for working in Japan. On the other hand, the peak period of the component with $k = 2$ is from 3:00 p.m. to 7:00 p.m., and its shape is gradual and positively skewed, which is described by small $\hat{\tilde{\alpha}}_{22}$ and positively small $\hat{\tilde{\beta}}_{22}$. This may be because drivers do not have to leave their office at the same time, which is different from morning hours. Some of them may decide to make their departure time earlier or later than peak hour to avoid traffic congestion. In addition, the positively skewed shape may imply that they also avoid returning home too late.

These interpretations are made by focusing on the shape around the mode of each peak. It is more difficult to understand the traffic flow volume around the tails of both components. For example, the first component still has a moderate probability at around 4:00 p.m., which cannot be included in the morning rush hour. One possible explanation is that the first component represents the outward trip, including morning commuting, commercial vehicles, etc. For example, many commercial vehicles travel from Kobe to Osaka during the daytime. In addition, some people make their first trip in the evening because they work at night or go shopping. People who work at night are more likely to use their vehicles for commuting than those who work in the daytime, as public transport does not operate after midnight. This could explain the first component having a larger probability than the second component after around 10:00 p.m. Similarly, the second component represents the day's return trips, including evening commuting. As round and migration trips during the daytime, for example, may be included in the second component, it already has a moderate value even in the morning hours. Note that outward and return trips are observed only once because all recorded traffics are in the same direction.

Finally, we compare our results with some phenomena typically mentioned in the traffic engineering field. First, the difference in the value of $\hat{\tilde{\alpha}}_2$ between the two components is consistent with the concept of departure time choice in traffic engineering (Noland & Small, 1995). In general, drivers choose their departure time by estimating backward from the time they want to arrive. Alexander et al. (2015) indicate that many drivers depart within short periods in the morning (i.e. traffic is concentrated), while drivers depart within longer periods in the evening and later. Second, 30% of Hanshin Expressway users are for commercial purposes (Hanshin Expressway, 2017, (in Japanese)). Assuming that people make outward commercial trips after they start working, the tail of the first components around noon is likely to contain commercial vehicles at a certain level. Therefore, the interpretation mentioned above that the first component is for the outward trips of the day and the second one is for the return trips is reasonable to some extent. Note that the share of large vehicles on this road is 13.6% in total and less than 30% even in the midnight period, although one might expect that the freight traffic is the main component after midnight (see online supplementary material, Appendix B). Therefore, no peak for freight traffic in the midnight period is found in the histograms, nor are such peaks found in the estimated model. Since these discussions are only derived from traffic volume data, it is necessary to further develop

Table 6. Maximum-likelihood estimates for the mixture with the recovered original parametrization (2)

k	$\hat{\mu}_k$	$\hat{\gamma}_k$	$\hat{\rho}_k$	$\hat{\lambda}_k$	$\hat{\pi}_k$
1	2.7572	0.3751	0.7267	5.3136	0.4845
2	4.0107	0.4855	0.1970	1.1895	0.5155

Table 7. Maximum-likelihood estimates for the reparametrized mixture (3)

k	$\hat{\mu}_k$	$\hat{\gamma}_k$	$\hat{\tilde{\alpha}}_{2k}$	$\hat{\tilde{\beta}}_{2k}$	$\hat{\pi}_k$
1	2.7572	0.3751	0.1542	−0.2248	0.4845
2	4.0107	0.4855	0.0356	0.0888	0.5155

the interpretations by combining trip purpose data and origin and destination data. Nonetheless, it would be a contribution of this paper that the proposed model that only uses traffic volume information is consistent with various concepts used in traffic engineering.

7 Comparison with other models

7.1 Mixtures of other distributions

We compare other multimodal distributions with our proposed model in terms of practicability and fit. The distributions considered here are the mixture of von Mises distributions (MovM), the mixture of wrapped Cauchy distributions (MowC), the mixture of sine-skewed von Mises distributions (MossvM), and the mixture of sine-skewed wrapped Cauchy distributions (MosswC) whose densities are given by

$$
\begin{aligned}
f_{\mathrm{MovM}}(\theta) &= \sum_{k=1}^{m} \frac{\pi_k}{2\pi I_0(\kappa_k)} \exp\left(\kappa_k \cos\left(\theta - \mu_k\right)\right), \\
f_{\mathrm{MowC}}(\theta) &= \sum_{k=1}^{m} \frac{\pi_k}{2\pi} \frac{1-\rho_k^2}{1+\rho_k^2 - 2\rho_k \cos\left(\theta - \mu_k\right)}, \\
f_{\mathrm{MossvM}}(\theta) &= \sum_{k=1}^{m} \frac{\pi_k (1 + \lambda_k \sin\left(\theta - \mu_k\right))}{2\pi I_0(\kappa_k)} \exp\left(\kappa_k \cos\left(\theta - \mu_k\right)\right), \\
f_{\mathrm{MosswC}}(\theta) &= \sum_{k=1}^{m} \frac{\pi_k}{2\pi} \frac{1-\rho_k^2}{1+\rho_k^2 - 2\rho_k \cos\left(\theta - \mu_k\right)} \left(1 + \lambda_k \sin\left(\theta - \mu_k\right)\right),
\end{aligned}
$$

respectively, where $\mu_k \in [0, 2\pi)$ is the mean direction, $\kappa_k \in [0, \infty)$ and $\rho_k \in [0, 1]$ are concentration parameters, $\lambda_k \in [-1, 1]$ is a skewing parameter, $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero, $m \in \mathbb{N}$ is the number of components of the mixture and $0 < \pi_1, \ldots, \pi_m < 1$ are the weights of the components satisfying $\sum_{k=1}^{m} \pi_k = 1$. The number of free parameters for MovM and MowC is $3m - 1$, and that for MossvM and MosswC is $4m - 1$. These distributions are popular mixtures discussed, e.g. in Wallace and Dowe (2003), Mooney et al. (2003), Banerjee et al. (2005), Mulder et al. (2020), and Miyata et al. (2020). Note that the mixture of sine-skewed wrapped Cauchy is a submodel of the proposed model (Kato & Jones, 2015, Supplementary Material).

The 50-fold cross-validation is conducted to compare the performance of the proposed model with that of these four existing models. The cross-validated log-likelihood is employed as the criterion for comparison (Smyth, 2000). The number of components is set to two, as in our proposed model.

Table 8 shows each model's average value of the cross-validated log-likelihood function. The predictive performance of the proposed model is superior to the other models because the value for the proposed model is the largest. As an example of a plot of the estimated densities,

Table 8. The average value of cross-validated log-likelihood function in 50-fold cross-validation for MovM, MowC, MossvM, MosswC, and proposed model

model	log-likelihood function
proposed	−36818.5
MovM	−37059.4
MowC	−37731.7
MossvM	−37026.3
MosswC	−36855.3

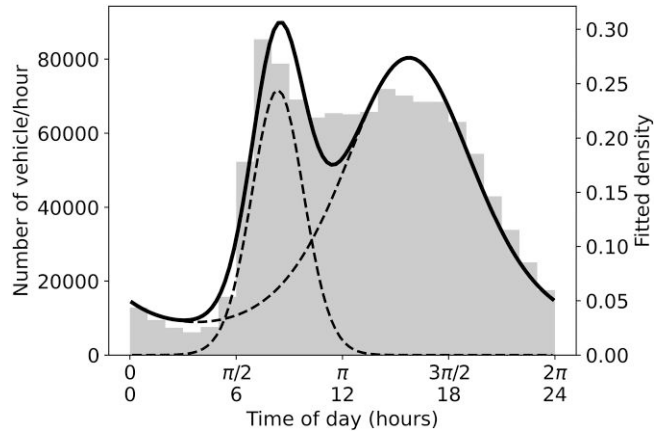

Figure 7. Plot of maximum-likelihood fit of the density of MovM with two components.

Figure 7 shows the maximum likelihood fits of the densities of MovM, the most well-known mixture of circular distributions, with two components. The goodness of fit looks significantly worse than in the proposed model shown in Figure 6. Particularly, the distorted peak around $\pi/2$ and the gradual peak around $3\pi/2$ are not represented. In contrast to the proposed model that allows varying kurtosis and skewness values, MovM and MowC do not allow both, and MossvM and MosswC do not allow varying kurtosis. In terms of the number of parameters, MovM and MowC have m fewer than the proposed model, and MossvM and MosswC have the same. The proposed model is superior to the other models in terms of predictive performance because the proposed model can vary its shape flexibly with a small number of parameters. In addition to the cross-validation, a comparison of the proposed model with other models using simulation studies is provided in online supplementary material, Appendix C.

7.2 Our mixture with more than two components

We estimate the mixed Kato–Jones distributions (3) whose number of components m is more than two. In this subsection, models with $m \leq 6$ are discussed because of stability of estimation; estimation for the model with $m > 6$ was carried out but the results were unstable. Recall that the histogram shows two peaks that can be interpreted as shown above and the assumption of more than two peaks is not necessarily insightful from the transport engineering viewpoint. As such, the main purpose of this subsection is not to find the most accurate number of components but to find out what shapes of distributions will be obtained.

The 50-fold cross-validations for the proposed model, whose number of components m is from 2 to 6, are conducted. The conditions for the estimation are the same as those in Section 5.

Table 9 shows the average value of the cross-validated log-likelihood function for each number of components. The model whose number of components $m = 4$ is the best because the value for the proposed model is the largest. The plot and value of the parameter for the estimated model

Table 9. The average value of cross-validated log-likelihood function in 50-fold cross-validation for proposed model whose number of components m from 2 to 6

m	log-likelihood function
2	−36818.5
3	−36818.7
4	−36808.6
5	−36817.6
6	−36820.2

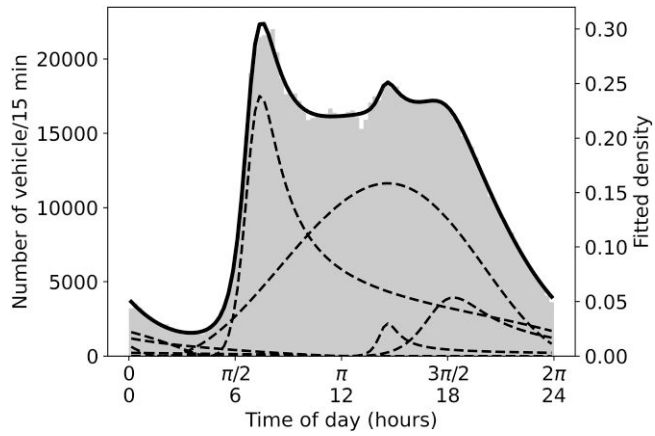

Figure 8. Plot of maximum-likelihood fits of the densities with the numbers of components $m = 4$.

Table 10. Maximum-likelihood estimates of the parameters of Kato–Jones mixtures (3) whose numbers of components $m = 4$

k	$\hat{\mu}_k$	$\hat{\gamma}_k$	$\hat{\tilde{\alpha}}_{2k}$	$\hat{\tilde{\beta}}_{2k}$	$\hat{\pi}_k$
1	2.6665	0.4376	0.2128	−0.2452	0.3855
2	3.7380	0.5065	0.0076	0.0225	0.4907
3	4.3642	0.4887	0.3402	−0.2284	0.0285
4	5.2889	0.6037	0.2628	−0.2166	0.0952

whose number of components $m = 4$ are shown in Figure 8 and Table 10. Note that the number of bins in Figure 8 is 96 (i.e. the width of the bin equals 15 min) to visualize the small peak around 3:00 p.m. described by the component with $k = 3$.

In the results of $m = 4$, two prominent components with $k = 1, 2$ have a large value of $\hat{\pi}_k$. As well as the results of $m = 2$, the components with $k = 1, 2$ represent the peak around $\pi/2$ and $3\pi/2$, respectively. However, the component with $k = 2$ for $m = 4$ is not skewed compared to the component with $k = 2$ for $m = 2$ shown in Figure 6. Not only the component with $k = 2$ but also that with $k = 4$ in the results of $m = 4$ represent the shape between $3\pi/2$ and 2π that is represented by only one component with $k = 2$ in the results of $m = 2$. In addition, the component with $k = 3$ has a small value of $\hat{\pi}_3$ and represents the small peak around 4:00 p.m. seen in the histogram.

The components with $k = 3, 4$ may not be meaningful from the transport engineering viewpoint; they only represent the trivial peak along with the improvement of the value of the log-likelihood function. The interpretation of the parameters for these smaller components is

difficult with only the traffic volume data. Some additional information, such as each vehicle's origin, destination, and trip purpose, may help the interpretation of the parameters.

8 Conclusion

In this paper, we estimated the probability distribution of the variation of traffic volume within an average weekday. We employed the mixture of Kato–Jones distributions, which can provide a wide range of skewness and kurtosis for each component. To estimate the parameters of these distributions, we developed the modified method of moments. This method ensures that the estimated parameters always belong to the parameter space. Then the maximum-likelihood method for the proposed mixture was established using the EM algorithm.

The proposed method was applied to some traffic counter data from Japan. As a result, the data were classified into two components, which were interpreted as the morning and evening traffic. Some reasonable explanations for these components are suggested, that the difference in the shapes of the components might be caused by the different behaviours underlying the commuting trips between morning and evening hours. In addition, the modified method of moments was seen to allow for fast calculation and provided a reasonable initial value for maximum-likelihood estimation. With this initial value, the EM algorithm provided reasonable maximum-likelihood estimates.

According to the cross-validated log-likelihood function, the model with the number of components $m = 4$ is the best. However, the interpretation of the trivial components in the model is difficult thus far. Further consideration of missing factors that generate third and later components may be necessary in future work. Additional information, such as each vehicle's origin, destination, and trip purpose, would be useful. In addition, we assume the IID property for all traffic data in the modelling, and we regard this assumption as acceptable, as mentioned in the online supplementary material. However, modelling the dependent characteristics of the arrival interval, if it exists, is one of the principal and important issues to be dealt with in the future.

Acknowledgments

We thank an associate editor and two reviewers for their instructive and insightful comments on an earlier version of this paper. The authors are grateful to Hironori Fujisawa for constructive comments on this work. The data used in this study were provided by the Hanshin Expressway Co., Ltd.

Conflicts of interest: None declared.

Funding

This research was partially supported by JSPS KAKENHI Grant Numbers JP18K13846, JP20H02267, JP20K03759, and JP24K21656.

Data availability

The python code used in the applications are available in a github repository https://github.com/Kota-Nagasaki/Estimation_MoKJ. The data cannot be made public due to confidentiality, but the code to generate random sample from the estimated parameters is included.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

References

Alexander L., Jiang S., Murga M., & González M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, *58*, 240–250. https://doi.org/10.1016/j.trc.2015.02.018

Anacleto O., Queen C., & Albers C. J. (2013). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(2), 251–270. https://doi.org/10.1111/j.1467-9876.2012.01059.x

Banerjee A., Dhillon I. S., Ghosh J., & Sra S. (2005). Clustering on the unit hypersphere using von Mises–Fisher distributions. *Journal of Machine Learning Research*, *6*, 1345–1382.

Batschelet E. (1981). *Circular statistics in biology*. Academic Press.
Chen X., Li L., & Zhang Y. (2010). A Markov model for headway/spacing distribution of road traffic. *IEEE Transactions on Intelligent Transportation Systems*, *11*(4), 773–785. https://doi.org/10.1109/TITS.2010.2050141
Daganzo C. F. (1997). *Fundamentals of transportation and traffic operations*. Pergamon Oxford.
Davison A. C., & Hinkley D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.
Edie L. C. (1963). *Discussion of traffic stream measurements and definitions*. Port of New York Authority.
Hanshin Expressway (2017). *CSR report on 2017*. viewed 25 May 2023 https://www.hanshin-exp.co.jp/company/csr/files/report/csrreport_201707_2.pdf.
Huber P. J. (1981). *Robust statistics*. Wiley.
Jones E., Oliphant T., & Peterson P. (2001). SciPy: Open source scientific tools for Python. (Available from http://www.scipy.org/).
Kato S., & Jones M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, *102*(1), 181–190. https://doi.org/10.1093/biomet/asu059
Kraft D. (1988). A software package for sequential quadratic programming. (Available from http://www.opengrey.eu/item/display/10068/147127).
Leduc G. (2008). *Road traffic data: Collection methods and applications*. Working Papers on Energy, Transport and Climate Change.
Ley C., Babić S., & Craens D. (2021). Flexible models for complex data with applications. *Annual Review of Statistics and Its Application*, *8*(1), 369–391. https://doi.org/10.1146/statistics.2021.8.issue-1
Li L., & Chen X. M. (2017). Vehicle headway modeling and its inferences in macroscopic/microscopic traffic flow theory: A survey. *Transportation Research Part C: Emerging Technologies*, *76*, 170–188. https://doi.org/10.1016/j.trc.2017.01.007
McLachlan G. J., & Krishnan T. (2008). *The EM algorithm and extensions* (2nd ed.). Wiley.
Miyata Y., Shiohama T., & Abe T. (2020). Estimation of finite mixture models of skew-symmetric circular distributions. *Metrika*, *83*(8), 895–922. https://doi.org/10.1007/s00184-019-00756-z
Mooney J. A., Helms P. J., & Jolliffe I. T. (2003). Fitting mixtures of von Mises distributions: A case study involving sudden infant death syndrome. *Computational Statistics & Data Analysis*, *41*(3–4), 505–513. https://doi.org/10.1016/S0167-9473(02)00181-0
Mulder K., Jongsma P., & Klugkist I. (2020). Bayesian inference for mixtures of von Mises distributions using reversible jump MCMC sampler. *Journal of Statistical Computation and Simulation*, *90*(9), 1539–1556. https://doi.org/10.1080/00949655.2020.1740997
Muñoz L., Sun X., Horowitz R., & Alvarez L. (2003). Traffic density estimation with the cell transmission model. In *Proceedings of the 2003 American control conference* (Vol. 5, pp. 3750–3755). IEEE.
Na J. H., & Jang Y. M. (2011). Modeling on daily traffic volume of local state road using circular mixture distributions. *The Korean Journal of Applied Statistics*, *24*(3), 547–557. https://doi.org/10.5351/KJAS.2011.24.3.547
Nagasaki K., Nakanishi W., & Asakura Y. (2019). Application of rose diagram to road network analysis. In *Proceedings of 24th international conference of Hong Kong Society for Transportation Studies* (pp. 169–176). Hong Kong Society for Transportation Studies.
Nagasaki K., Nakanishi W., & Asakura Y. (2023). Analysis of traffic volume by the estimated parameter of the mixture of circular distribution. *Journal of Japan Society of Civil Engineers, Series D3 (Infrastructure Planning and Management)*, *78*(5), I825–I831. https://doi.org/10.2208/jscejipm.78.5_I_825
Noland R. B., & Small K. A. (1995). Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation Research Record*, *1493*, 150–158.
Pewsey A., & García-Portugués E. (2021). Recent advances in directional statistics. *TEST*, *30*(1), 1–58. https://doi.org/10.1007/s11749-021-00759-x
Press S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, *67*(340), 842–846. https://doi.org/10.1080/01621459.1972.10481302
Redner R. A., & Walker H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, *26*(2), 195–239. https://doi.org/10.1137/1026034
Smyth P. (2000). Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*, *10*(1), 63–72. https://doi.org/10.1023/A:1008940618127
Transportation Research Board (2010). *Highway capacity manual*. National Research Council.
Wallace C. S., & Dowe D. L. (2003). MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Statistics and Computing*, *10*(1), 73–83. https://doi.org/10.1023/A:1008992619036
Wang Y., & Papageorgiou M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. *Transportation Research Part B: Methodological*, *39*(2), 141–167. https://doi.org/10.1016/j.trb.2004.03.003