Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 1163 – 1172

doi:10.3906/elk-1511-45

Traffic density estimation via KDE and nonlinear LS

Mikail ERDEM[1], **Mehmet Kemal ÖZDEMİR**[1,2,*]
[1]Department of Electronics and Computer Engineering, Graduate School of Natural and Applied Sciences, İstanbul Şehir University, İstanbul, Turkey
[2]Department of Electrical and Electronics Engineering, College of Engineering and Natural Sciences, İstanbul Şehir University, İstanbul, Turkey

Received: 04.11.2015 • **Accepted/Published Online:** 31.03.2016 • **Final Version:** 10.04.2017

Abstract: With increasing population, the determination of traffic density becomes critical in managing urban city roads for safer driving and low carbon emissions. In this study, kernel density estimation is utilized in order to estimate traffic density more accurately when the speeds of vehicles are available for a given region. For the proposed approach, as a first step, the probability density function of the speed data is modeled by kernel density estimation. Then the speed centers from the density function are modeled as clusters. The cumulative distribution function of the speed data is then determined by Kolmogorov–Smirnov test, whose complexity is less when compared to the other techniques and whose robustness is high when outliers exist. Then the mean values of clusters are estimated from the smoothed density function of the distribution function, followed by a peak detection algorithm. The estimates of variance values and kernel weights, on the other hand, are found by a nonlinear least square approach. As the estimation problem has linear and nonlinear components, the nonlinear least square with separation of parameters approach is adopted, instead of dealing with a high complexity nonlinear equation. Simulations are carried out in order to assess the performance of the proposed approach. It is observed that the error between cumulative distribution functions is less than 1%, an indication that the traffic densities are estimated accurately. For an assumed traffic condition that bears five speed clusters, the minimum mean square error of kernel weights is found to be less than 0.00004. The proposed approach was also applied to real data from sample road traffic, and the speed center and the variance were accurately estimated. By using the proposed approach, accurate traffic density estimation is realized, providing extra information to the municipalities for better planning of their cities.

Key words: Traffic density estimation, kernel density estimation, Kolmogorov–Smirnov tests, nonlinear least square

1. Introduction

Estimating traffic density and predicting it for a specified time interval, especially in megacities, has become an urgent issue that needs to be tackled. An efficient tool providing traffic density estimation and its prediction will help municipalities to better manage traffic and plan their infrastructure. Moreover, in their daily lives, drivers would make better decisions, thereby directly having impacts on gas and time savings, as well as low carbon emissions.

There are numerous ways to estimate traffic density. However, kernel density estimation (KDE), which is also known as the Parzen window method [1], is one of the most effective approaches, because KDE can better differentiate between cars that travel on the same road but at different speeds in different lanes [2]. Moreover,

*Correspondence: kemalozdemir@sehir.edu.tr

estimation approaches can be divided into two groups as parametric and nonparametric. While parametric approaches have a fixed number of parameters, nonparametric ones have an increasing number of parameters as the training data size becomes larger [3]. As we are dealing with continuously incoming data from roads, the latter, the nonparametric approach, is used with KDE, since the combination is a better fit to the current problem, not to mention the mathematical advantages offered by KDE [4].

Each sample of the received data from the field can be treated as normally distributed, and the many realizations of the normal distributions are generated. The probability density function (PDF) of the received data can be obtained either by bin allocation or histogram. Bin allocation involves splitting data into bins with certain step size. With this method, PDF value varies with the speed of vehicles on the road. As the number of vehicles with a velocity in the specified bin increases, the PDF value also increases. Alternative to this approach, KDE can be used to represent diverse traffic scenarios more explicitly and the estimation process via KDE is not too complicated since it makes use of the gaussian distribution. After determining the PDF, the next step would be to find the cumulative distribution function (CDF). There are numerous methods like Cramer-von Mises (CvM) and Kolmogorov–Smirnov (KS) tests to get the CDF but in this study the latter is preferred, since it is less affected by the existence of outliers [5].

The traffic data can be modeled by clusters, and each traffic cluster is represented by a mean, a bandwidth, and a coefficient that shows cluster's weight among all present clusters. Traffic density estimation via KDE has been studied in [4] by using CvM test for the determination of the kernel weights. However, the study in [4] assumes the variances and mean values of the traffic clusters as constants. PDF values of Dirac distribution and gaussian mixtures are compared by using CvM and localized cumulative distributions. Then optimized kernel weights are estimated by minimizing the distance between floating car position centers and points-of-interest through quadratic programming.

Similar to [4], in [6], which is the initial work of this study, the bandwidths and the speed centers are also counted as constants. The methodology of [6] is an alternative to [4] since it can determine the kernel weights more easily by using KDE, KS test, and linear least square (LS) approaches in fewer steps. We extend our initial study in [6] by assuming that the variances and the mean values of the traffic clusters are nonconstants. This, though, requires the estimation of the mean values and the variances of each cluster. The parameter that is fixed by the approach is the number of traffic clusters. However, if more clusters than the existing amount are made, then the kernel weights are simply found to be close to zero. For the estimation of the variances and the mean values, instead of linear LS, nonlinear LS (NLS) with separation of parameters approach is exploited. After determination of empirical CDF with KS test, the straightforward way of computing kernel weights is to take the derivative of the PDF. However, if the received data contain noise, it might not give the desired outcome. Hence, instead of using derivatives of PDF, the NLS approach is utilized. Before the estimation of parameters via NLS, the mean values are estimated first by using a peak detection algorithm over the smoothed version of the PDF curve. In order to estimate variance values and kernel weights, NLS is used successively. First, speed center's bandwidth and then its kernel weights are estimated. After these estimations, the next speed center's variance and its kernel weight are estimated, and so on. In applying the NLS, two methods are exploited: linear search and Newton–Raphson (N-R) methods. While the former generally gives the desired results in a longer time, the latter gives sufficiently close results in a considerably shorter amount of time [7].

In the following section, Section 2, the simulation model will be presented. In Section 3, the computation of KDE and kernel weights will be illustrated with numerical calculations. In Section 4, the proposed model and its practical results will be examined. Finally, in Section 5, a summary of conclusions, observations, and future work will be presented.

2. The model

2.1. Finding density distribution with KDE

For a given N independent samples, let $x \equiv \{X_1, \ldots, X_N\}$ come from a continuous PDF f, which is defined on X. Gauss KDE can then be defined as follows [8]:

When the mean of each data sample is X_i and the corresponding variance is σ, then the gauss kernel PDF is

$$\hat{f}(x;\sigma) = \frac{1}{N}\sum_{i=1}^{N}\varphi(x, X_i;\sigma), \qquad x \in R \tag{1}$$

where

$$\varphi(x, X_i;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-X_i)^2/(2\sigma)} \tag{2}$$

When the contributions of the gauss kernels are different, then (1) can be re-expressed by including kernel weights, α_i, as

$$\hat{f}(x;\sigma,\alpha) = \frac{1}{N}\sum_{i=1}^{N}\alpha_i\varphi(x, X_i;\sigma), \qquad x \in R \tag{3}$$

where

$$0 \le \alpha_i \le 1 \quad \text{and} \quad \sum \alpha_i = 1 \tag{4}$$

For a given set of data, PDF plots can be found in two ways: either through plotting or a closed-form expression. One of the easiest methods of plotting PDF is to use a histogram plot as seen in Figure 1.

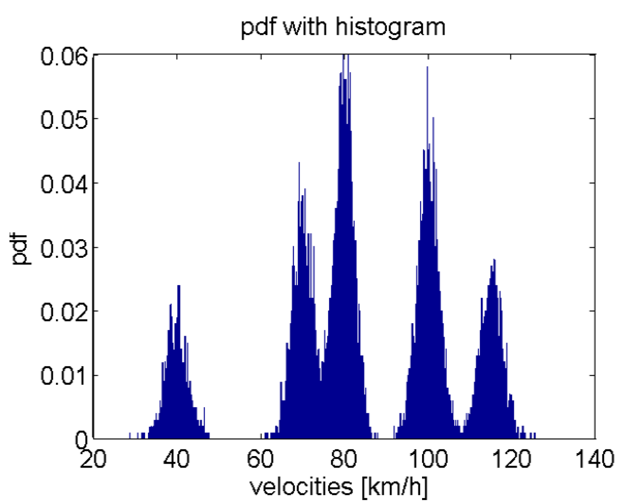

Figure 1. PDF plot via histogram.

Alternatively, bin allocation is another method for PDF plot without using a closed-form expression as seen in Figure 2. However, these methods give only figures, and hence do not provide closed form expressions that can be utilized for further theoretical developments. On the other hand, if KDE is used to determine the

PDF plots, such plots will be more convenient to visualize as in Figure 3 and will provide a better theoretical background for the estimation studies [8]. Since one of the main goals of KDE is to produce a smooth density surface over a 2-D geographical space, the smoothed version of the density is used to plot PDF with KDE [9].

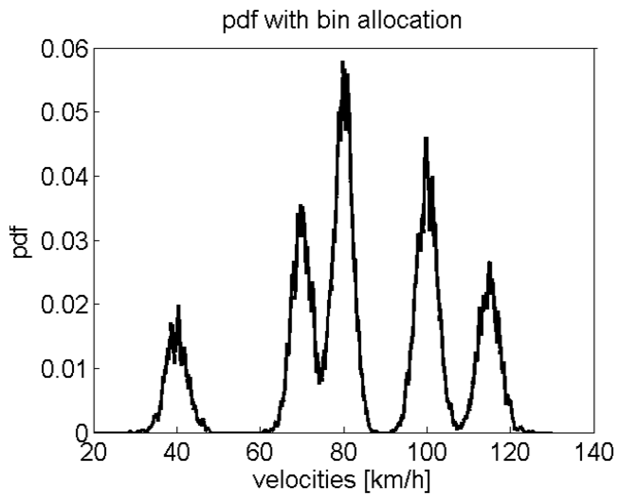

Figure 2. PDF plot via bin allocation.

2.2. Finding empirical CDF with KS test

For a one sample test in the KS test, definition of the CDF and the hypothesis can be developed as follows: $X_1, X_2, \ldots, X_n$ are random variables for $x_1, x_2, \ldots, x_n$ continuous, independent, and identically distributed (i.i.d) samples. When F is defined as CDF, $\hat{F}$ would be empirical CDF [10]. With F_0 is known, for all $x \in R$ values, the hypothesis

$$H_0 : F(x) = F_0(x) \tag{5}$$

$$\hat{F} = \frac{\#(i : x_i \leq x)}{n} \tag{6}$$

is empirical CDF. Then the KS test statistics D_ns are defined as follows:

$$D_n = \sup_x \left| \hat{F}(x) - F_0(x) \right|, \tag{7}$$

where *sup* is the supremum of the CDF values.

The KS test initially examines the difference between the empirical and the real values, and then checks how much they match up with each other. For this model, the distribution of the speeds can be obtained and checked if they have normal distribution or not. The expression in (6) arranges all the data in order according to their values, and then rescales them, and finally when the biggest value is encountered the CDF reaches unity. This approach can be criticized, as the number of operations would be huge when the number of samples is very large. However, this problem can be overcome by efficient and fast computer programs like MATLAB and high performing processors.

2.3. Determination of variance and kernel weights with nonlinear LS method

The mean values of the speed clusters resemble separate regions for the PDF. Hence, first of all they should be determined. By using a peak detection algorithm applied to the PDF, mean values can be determined straightforwardly. For this, first the derivative of CDF is determined, yielding the PDF. Then through a smoothing operation, the PDF is prepared for a peak detection algorithm, which then produces accurate speed center values. The peak detection algorithm checks all data in the data set and finds peak values. From the peak detection algorithm, we obtain the corresponding mean values of the peak values in the PDF.

We then need to find both variance and kernel weights by using the separability of parameters property of NLS. In our model, which is defined in (3), α is linear and σ is nonlinear with respect to the system. Then the approach is to find LS for parameters α in terms of σ, and then to determine σ. For this, the following equation (LS error) should be minimized for α [7]:

$$J(\sigma,\alpha) = (x - H(\sigma)\alpha)^T (x - H(\sigma)\alpha) \tag{8}$$

Here x values are F values of empirical CDF and so F will be used instead of x, and α is estimated as follows:

$$\hat{\alpha} = \left(H^T(\alpha) H(\alpha)\right)^{-1} H^T(\alpha) F \tag{9}$$

Then, by replacing $\hat{\alpha}$ in the above LS error (8), we get

$$J(\sigma,\hat{\alpha}) = F^T \left(I - H(\alpha)\left(H^T(\alpha) H(\alpha)\right)^{-1} H^T(\alpha)\right) F \tag{10}$$

Therefore, minimization of $J(\sigma,\hat{\alpha})$ is the same as maximization of the following equation over α:

$$\max_{\alpha} \left[F^T H(\alpha)\left(H^T(\alpha) H(\alpha)\right)^{-1} H^T(\alpha) F\right], \tag{11}$$

where

$$H_i = \frac{1}{2}\left(1 + erf\frac{x-\mu_i}{\sigma\sqrt{2}}\right) \tag{12}$$

$$F = \sum \alpha_i H_i \tag{13}$$

$$F_i = \left(\alpha_i - \sum_{j=1}^{i-1}\alpha_j\right) H_i \tag{14}$$

Then all α values are determined by

$$\alpha_i = \left(H_i^T H_i\right)^{-1} H_i^T F_i - \sum_{j=1}^{i-1}\alpha_j \tag{15}$$

Initially, the variance in the speed center is found with (11) by linear search method or N-R method. By inserting this into (15), the kernel weight of the corresponding speed center is found. Since the mean values are determined before the NLS method is invoked, the same procedure for finding the variance and kernel weight is repeated for each speed center, resulting in a successive estimation process.

When the variances are determined, as expressed in (11), the equation should reach the maximum value with the variance value that makes the result of the equation greatest. A straightforward guaranteed approach is to perform a linear search; however, the analysis shows that more than 100,000 calculations are needed to find the correct value for each speed center's variance. Alternatively, the N-R method can determine the same variance values at a reasonable complexity. Although its calculation requires some computations as shown in the Appendix, it reaches sufficiently close variance value to the linear search method in less than 10 iterations of calculation; generally just 5 or 6 iterations are sufficient for the convergence. Therefore, the N-R method reaches the value faster than the linear search method. The overall algorithm is given below.

Algorithm 1 Summary of the algorithm for parameter estimation

```
load the data
obtain PDF with KDE algorithm as given by (3)
obtain empirical CDF with "ecdf" function as formulated in (6)
take the derivative of eCDF and then smooth it
find peak values of smoothed PDF to get mean values with PDA
for each mean values
    determine corresponding F values of speed center from eCDF
    Use Newton–Raphson method:
    while error between two variance values is bigger than a threshold
        calculate H values via (12)
        maximize (11) by changing variance values
        decide on variance values with (A.9)
    end
    Getting variance value that maximizes (11)
    Finding the kernel weight with (15)
    OR use linear search method:
    for σ^2 from 0 to 100 with step size 0.001
        Calculate H values via (12)
        calculate (11) by changing variance values
    end
    determine variance value that makes (11) maximum
    find kernel weight via (15)
end
```

3. Numerical calculations

In this section, we will examine the proposed approach for an assumed system and observe the advantages and drawbacks of the system. A traffic scenario with 5 speed centers is assumed. We will assess the performance of the approach for estimated mean, variance, and kernel weight values. The assumed scenario has the following typical parameters that are thought to form in a highway passing through a city:

$$\mu_1 = 40 \quad \mu_2 = 70 \quad \mu_3 = 80 \quad \mu_4 = 100 \quad \mu_5 = 115$$

$$\sigma_1^2 = 7 \quad \sigma_2^2 = 6 \quad \sigma_3^2 = 5 \quad \sigma_4^2 = 6 \quad \sigma_5^2 = 7$$

$$\alpha_1 = 0.1 \quad \alpha_2 = 0.2 \quad \alpha_3 = 0.3 \quad \alpha_4 = 0.25 \quad \alpha_5 = 0.15$$

The PDF of the assumed system is given in Figure 3, while the CDF plot is given in Figure 4. The mean values of the speed centers estimated via a peak detection algorithm are as follows:

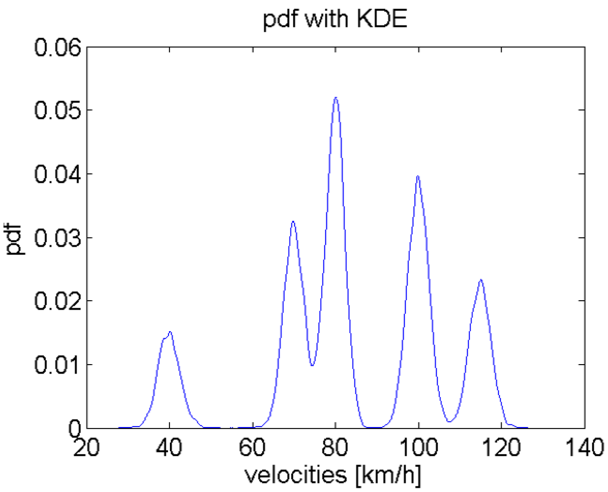

Figure 3. PDF plot with KDE.

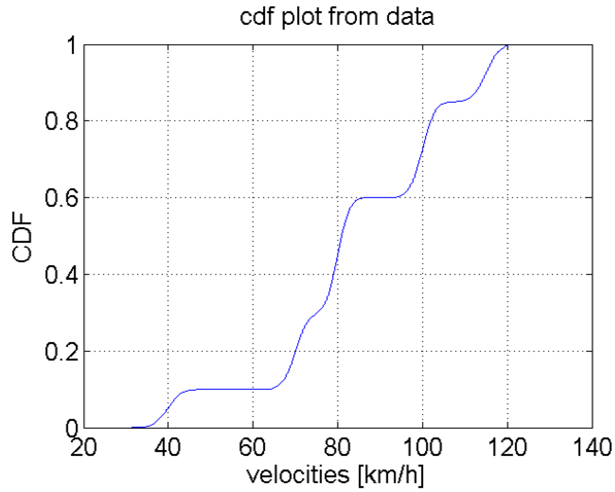

Figure 4. CDF plot for the example system.

$$\hat{\mu}_1 = 39.8921 \quad \hat{\mu}_2 = 69.7892 \quad \hat{\mu}_3 = 79.9655 \quad \hat{\mu}_4 = 100.0481 \quad \hat{\mu}_5 = 115.0541$$

We can see that the results are very close to real values as the MMSE (minimum mean square error) is 0.0125. The variances and kernel weights are estimated by using two methods as explained below. For the linear search method, which takes a long time but provides more accurate results, the estimated variances are

$$\hat{\sigma}_1^2 = 7.1350 \quad \hat{\sigma}_2^2 = 5.5230 \quad \hat{\sigma}_3^2 = 5.6700 \quad \hat{\sigma}_4^2 = 5.8980 \quad \hat{\sigma}_5^2 = 6.2970$$

$$\hat{\alpha}_1 = 0.0983 \quad \hat{\alpha}_2 = 0.1956 \quad \hat{\alpha}_3 = 0.3102 \quad \hat{\alpha}_4 = 0.2475 \quad \hat{\alpha}_5 = 0.1569$$

When the error amounts are analyzed, it is observed that kernel weights and speed centers have less error when compared to variance values. However, the estimation of variances is an intermediate step before the estimation of the kernel weights. Although variance estimation provides useful information about traffic density, the speed centers and kernel weights are more critical in assessing traffic density. We observe that the proposed approach can estimate the mean values and the kernel weights very close to the actual values. The MMSE for variance is 0.2399 and for kernel weights is 3.6030×10^{-5}. For the N-R method, which reaches the result quickly, the results are as follows:

$$\hat{\sigma}_1^2 = 7.1350 \quad \hat{\sigma}_2^2 = 5.5232 \quad \hat{\sigma}_3^2 = 5.6695 \quad \hat{\sigma}_4^2 = 5.8983 \quad \hat{\sigma}_5^2 = 6.2963$$

$$\hat{\alpha}_1 = 0.0983 \quad \hat{\alpha}_2 = 0.1956 \quad \hat{\alpha}_3 = 0.3102 \quad \hat{\alpha}_4 = 0.2475 \quad \hat{\alpha}_5 = 0.1472$$

The MMSE for variance values found by using the N-R method is 0.2399 and for kernel weights it is 2.8076×10^{-5}.

As can be seen from the estimated values, the proposed approach can accurately estimate the targeted parameters, as MMSE values are very small for the traffic density estimation. Since the model adopts a successive approach, it is expected that an increase in error (error propagation) occurs when the latter parameters in the order are estimated. However, for the sample system, the difference between α_5 and its real value is less than the difference between α_2 and α_3 and their real values. This is due to the fact that for every speed center first the variance is found and then its kernel weight is estimated, thereby eliminating the error propagation.

As a second example, we examine the system with real data of New York City's traffic (https://data.cityofnewyork.us/Transportation/Real-Time-Traffic-Speed-Data/xsat-x5sa), which is obtained via sensor feeds at the major arterials and highways. Unfortunately the data have only one kernel weight. The PDF plot of the data is shown in Figure 5, while its empirical CDF plot and smoothed version of the CDF are seen in Figure 6. The estimated speed center is 51.17875 by using the peak detection algorithm. For the linear search method, the variance value is 325.6390 and the kernel weights are 1, as expected. For N-R method, the variance value is 325.6430 and the kernel weight is 1. The variance value of the system is too large, because there is no other speed center and all the vehicles are around a single cluster.

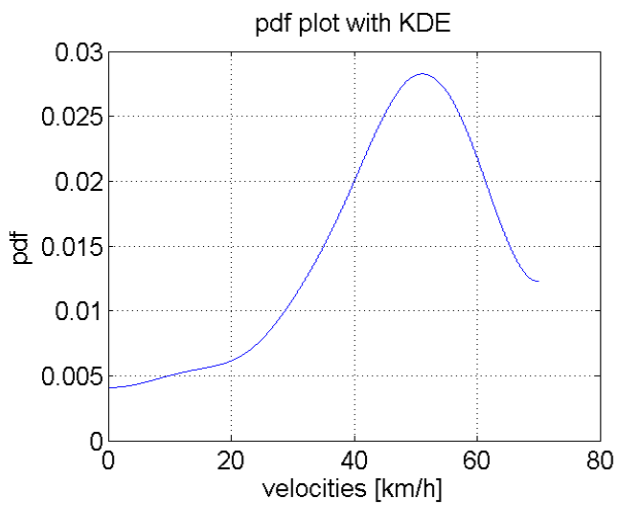

Figure 5. PDF plot for the real-time speed data.

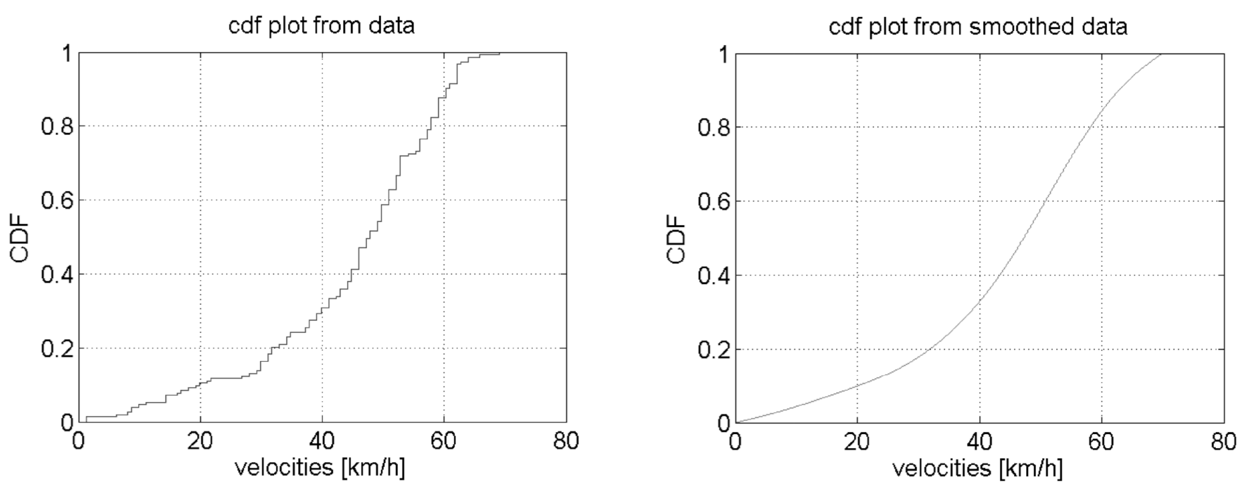

Figure 6. CDF plot and its smoothed version for the real-time speed data.

As seen in Figure 5, the speed center estimation matches the peak velocity values in the PDF plot. Since there is only one kernel, its weight is equal to 1. We could not find real data with more speed centers as the companies working in this area are mostly willing to show their final products like maps, and they are reluctant to share their data.

4. Assessment

The system is examined with several scenarios and it achieves accurate results in all. Even when it is tested for low and high-speed values, it performs well for the estimations. In contrast to [4], which only estimates kernel weights, this work deals with not only kernel weights but also speed centers and variance values. While estimating variance, the N-R method reaches the results very quickly and it does not have much difference between linear search method results.

The model is also tested for systems that have different kernel numbers like 1, 3, 5, and 7 and the results are shown in the Table. The estimations of mean values are found by peak detection. The weights and bandwidths are estimated either by N-R or linear search methods. The results show that the performance of the estimation process is accurate when compared with the existing studies. The error for mean estimation is less than 4%, while for kernel weights it is less than 1.1×10^{-4} when both methods are used. As for the variance, the MMSE error is higher, but less than 0.7. The variance estimates are exploited for the estimation of kernel weights, which are more critical in the overall process. As seen in the Table, for different scenarios, the proposed approach performs consistently well.

Table. MMSE of traffic density estimation with different kernel numbers.

Used method	Peak detection	Newton–Raphson		Linear search	
# of kernels	Mean	Kernel weights	Variance	Kernel weights	Variance
1	0.013572	0.000000	0.324786	0.000000	0.324900
3	0.034949	0.000072	0.223417	0.000107	0.159526
5	0.012502	0.000028	0.239866	0.000036	0.239853
7	0.032502	0.000030	0.638211	0.000030	0.631461

A criticism of this study could be about the error propagation due to its successive approach. Yet, as explained above and seen in the results of the example, the successive system model does not pave the way for error propagation since the mean values are estimated first; then for each mean value, its variance value is estimated, followed by its kernel weight estimation from CDF of KDE.

5. Conclusion and future work

In traffic management, one of the most important factors and indicators is the performance of the transportation networks, typically measured via the mobility on the roads. Therefore, traffic density estimation and its prediction play a crucial role in traffic-related decisions. In this study, traffic density estimation was resolved by estimating the mean values, bandwidths, and kernel weights of clusters, which represent a group of moving vehicles. The prediction part of the work is left for a future study.

In the study, first probability densities are modeled by using kernel density estimation. Secondly, empirical CDF is found by using the Kolmogorov–Smirnov test. Then speed centers are found by using a peak detection algorithm. In the last step, variance values and kernel weights are found successively by using the separation of parameters property of the nonlinear least square method. Here the variances or the kernel bandwidths are determined in two ways: linear search method and Newton–Raphson method. The proposed model and the estimation procedure are then tested with a sample system and it is observed that the proposed estimator achieves a good performance. Moreover, real-time traffic data are used to examine the system, and they fit the normal distribution hypothesis, and also the speed center is found accurately. Due to the lack of more than one kernel property of the data, we were not able to check the estimation of kernel weights and variance values for the real data. In addition, the model performed very well for simulated systems that have different numbers of speed centers, variance values, and kernel weights. The current proposal can be extended to include the prediction of traffic density for a given time interval. By the development of sequential tracking algorithms, estimation and prediction of the desired parameters can be accomplished using smaller amounts of data. The overall approach can then be benchmarked with applications like Yandex Navi.

Acknowledgment

This project was partly supported by TÜBİTAK, Project #113C037 (Bideb-2232).

References

[1] Scott DW. Multivariate Density Estimation: Theory, Practice, and Visualization. Hoboken, NJ, USA: John Wiley & Sons, 1992.

[2] Laxhammar L, Falkman G, Sviestins E. Anomaly detection in sea traffic - a comparison of the Gaussian mixture model and the kernel density estimator. In: Information Fusion 2009; 6–9 July 2009; Seattle, WA, USA. pp. 756-763.

[3] Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.

[4] Tabibiazar A, Basir O. Kernel-based optimization for traffic density estimation in ITS. In: IEEE 2011 Vehicular Technology Conference; 5–8 Sept 2011; New York, NY, USA: IEEE. pp. 1-5.

[5] Jäntschi L, Bolboaca SD. Distribution fitting 2. Pearson-Fisher, Kolmogorov-Smirnov, Anderson-Darling, Wilks-Shapiro, Kramer-von-Misses and Jarque-Bera statistics. Bull Univ Agric Sci 2009; 66: 691-697.

[6] Yılan M, Ozdemir M. A simple approach to traffic density estimation by using kernel density estimation. In: 23rd Signal Processing and Communications Applications Conference; 16–19 May 2015; Malatya, Turkey. pp. 1865-1868 (original article in Turkish with an abstract in English).

[7] Kay SM. Fundamentals of Statistical Signal Processing: Estimation Theory, vol. 1. Upper Saddle River, NJ, USA: Pearson Education, 1993.

[8] Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. Ann Stat 2010; 38: 2916-2957.

[9] Xie Z, Yan J. Kernel density estimation of traffic accidents in a network space. Comput Environ Urban 2008; 32: 396–406.

[10] Djuric P, Miguez J. Model assessment with Kolmogorov-Smirnov statistics. In: IEEE International Conference on Acoustics, Speech and Signal Processing 2009; 19–21 April 2009; Taipei, Taiwan. pp. 2973-2976.

[11] Chapra SC, Canale RP. Numerical Methods for Engineers. 6th ed. New York, NY, USA: McGraw-Hill, 2010.

[12] Bodewig E. Matrix Calculus. 2nd ed. Amsterdam, Netherlands: North-Holland Publishing Company, 2014.

[13] Meyer CD. Matrix Analysis and Applied Linear Algebra. Philadelphia, PA, USA: SIAM, 2000.

[14] Weir MD, Hass J, Thomas GB. Thomas' Calculus. 12th ed. Boston, MA, USA: Pearson Education, 2010.

Appendix

The Newton–Raphson method is derived to reach the zero crossing point and its general formula is as [11]

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \tag{A.1}$$

where i is the iteration number.

However, in this model, we deal with maximization problem of (11); thus, in order to apply the N-R method, we need to take the derivative of the (11) first, and then we can implement N-R to find zero crossing point. Therefore, the first and second derivatives of (11) have to be found.

Here we perform the search on variance and we have two different components, H and F, as given in (12) and (14). Since F is found from empirical CDF, we concentrate only on H. Let us use s instead of σ^2 in the equations.

If we use short notation to identify (11), it can be written as

$$B = H\left(H^T H\right)^{-1} H^T \quad \text{and} \quad A = F^T B F \tag{A.2}$$

Then the first and second derivatives of A become

$$\frac{\partial A}{\partial s} = F^T B' F \tag{A.3}$$

$$\frac{\partial^2 A}{\partial s^2} = F^T B'' F, \tag{A.4}$$

whereas the first and second derivatives of B are [12,13]

$$\begin{aligned} \frac{\partial B}{\partial s} &= H'\left(H^T H\right)^{-1} H^T + H\left(H^T H\right)^{-1} H'^T \\ &\quad - H\left(H^T H\right)^{-1}\left(H'^T H + H^T H'\right)\left(H^T H\right)^{-1} H^T \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} \frac{\partial^2 B}{\partial s^2} &= H''\left(H^T H\right)^{-1} H^T + 2H'\left(H^T H\right)^{-1} H'^T + H\left(H^T H\right)^{-1} H''^T \\ &\quad - H'\left(H^T H\right)^{-1}\left(H'^T H + H^T H'\right)\left(H^T H\right)^{-1} H^T \\ &\quad - H\left(H^T H\right)^{-1}\left(H'^T H + H^T H'\right)\left(H^T H\right)^{-1} H'^T \\ &\quad - H'\left(H^T H\right)^{-1}\left(H'^T H\right)\left(H^T H\right)^{-1} H^T - H'\left(H^T H\right)^{-1}\left(H^T H'\right)\left(H^T H\right)^{-1} H^T \\ &\quad - H\left(H^T H\right)^{-1}\left(H'^T H\right)\left(H^T H\right)^{-1} H'^T - H\left(H^T H\right)^{-1}\left(H^T H'\right)\left(H^T H\right)^{-1} H'^T \\ &\quad - H\left(H^T H\right)^{-1}\left(H''^T H + H'^T H'\right)\left(H^T H\right)^{-1} H^T \\ &\quad - H\left(H^T H\right)^{-1}\left(H'^T H' + H^T H''\right)\left(H^T H\right)^{-1} H^T \\ &\quad + 2H\left(H^T H\right)^{-1}\left\{\left(H'^T H + H^T H'\right)\left(H^T H\right)^{-1}\right\}^2 H^T \end{aligned} \tag{A.6}$$

Here the first and second derivatives of H, given in (12), are given by [14]

$$\frac{\mathrm{d}H}{\mathrm{d}s} = -\frac{(x-\mu)\, e^{-\frac{x^2-2\mu x+\mu^2}{2s}}}{2^{\frac{3}{2}}\sqrt{\pi} s^{\frac{3}{2}}} \tag{A.7}$$

$$\frac{\mathrm{d}^2H}{\mathrm{d}s^2} = -\frac{(x-\mu)\left(3s - x^2 + 2\mu x - \mu^2\right) e^{-\frac{x^2-2\mu x+\mu^2}{2s}}}{2^{\frac{5}{2}}\sqrt{\pi} s^{\frac{7}{2}}} \tag{A.8}$$

Then the variance estimation becomes

$$s_{i+1} = s_i - \frac{A^{'}}{A^{''}} \tag{A.9}$$