A METHODOLOGY TO DISAGGREGATE SPEED DATA COLLECTED BY ROAD DETECTORS

Original

Availability:
This version is available at: 11583/2663806 since: 2017-01-26T08:57:37Z

Publisher:
Transportation Research Board

Published
DOI:

Terms of use:
openAccess
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

(Article begins on next page)
A METHODOLOGY TO DISAGGREGATE SPEED DATA COLLECTED BY ROAD DETECTORS

Catani L., Tremblay J.-M., Bassani M., Cirillo C.

BACKGROUND, MOTIVATIONS AND OBJECTIVES

- Speed is the fundamental parameter in road design, traffic operation and control, and road safety.
- Speed data acquisition is performed by several methods (magnetic, pneumatic, laser, video, etc.) that may collect data in aggregate or disaggregate form; each method has different functionalities, costs, and accuracies.
- Aggregate data conceal information about the real trend of discrete values, that are more powerful for operating purposes.

Algorithm that computes disaggregate values from aggregate speed samples by considering different statistical distributions. Such individual speeds are necessary to operate with continuous distribution functions and to derive basic descriptive measures.

METHODOLOGY

The algorithm has been provided for three statistical distributions (Normal, Lognormal, and Gamma) usually employed for speed modeling. It was developed by using R software v3.1.1.

Load database (speed counts) → Define speed thresholds $T_i$ → From intervals to distribution: estimates $\theta$ with log-likelihood function (MLE) and optimization process (BFGS) → Generation of speed sample according to counts (inverse transform sampling)

Definition of counts. Assuming that $Y_i$ are observed speeds that follow a particular distribution $f_i$, characterized by a parameter $\theta_i$ $Y_1, Y_2, ..., Y_n \sim f(\theta)$ for $j = 1, 2, ..., n$

Defining the speed thresholds $T_i$, the counts of vehicles $c_i$ are defined as:

$c_i = \sum_{j=1}^{n} I(Y_j \in [T_{i-1}, T_i])$

for $i = 1, 2, ..., h$

$I(Y_j) = \left\{ \begin{array}{ll} 1 & \text{if } Y_j \in [T_{i-1}, T_i) \\ 0 & \text{if } Y_j \notin [T_{i-1}, T_i] \end{array} \right.$

Maximum Likelihood Estimation method. The Log-Likelihood (LL) function and the estimated parameter $\hat{\theta}_MLE$ are respectively:

$LL(\theta) = \sum_{i=1}^{h} c_i \cdot \log f(T_i; \theta) - F(T_i; \theta)$

$\hat{\theta}_{MLE} = \arg \max \, \sum_{i=1}^{h} c_i \cdot \log f(T_i; \theta) - F(T_i; \theta)$

Inverse Transform Sampling technique to compute individual speeds according to the distribution parameter and the sample size.

$V = F^{-1}(u)$ where $u = \frac{1}{h} \cdot \frac{1}{n} \cdot \sum_{i=1}^{h} c_i$

Example. Speed sample of 350 data.

Parameters estimated by Salter and Hounsell (1) methodology

<table>
<thead>
<tr>
<th>Speed class</th>
<th>mean, $\theta$</th>
<th>std.dev, s</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>79.9 km/h</td>
<td>11.5 km/h</td>
</tr>
<tr>
<td>5</td>
<td>80.3 km/h</td>
<td>13.4 km/h</td>
</tr>
</tbody>
</table>


ACKNOWLEDGEMENTS

The work included in this paper has been supported by the Compagnia di San Paolo (Italy) and the Politecnico di Torino under the grant “Studio per Finanziamento di Progetti di InterNazionalizzazione della Ricerca”, approved with the Rectoral Decree n. 208 of the 24th of May, 2013.

The authors wish to acknowledge the Department of Civil and Environmental Engineering at University of Maryland, College Park (US) for hosting Lorenzo Catani as a visiting student. This international cooperation agreement and the support from the Politecnico di Torino and the National Transportation Center at the University of Maryland has made this research possible. Speed data was collected by Alessio Bertola who is gratefully acknowledged for his contribution.

CONCLUSIONS

- The modeled distribution parameters are similar to those of original data; the mean is generally not affected by aggregation, while some discrepancies are noticed in the standard deviation;
- The choice of the statistical distribution which best interprets the field observations may be made on the basis of accuracy measures such as those considered in this manuscript;
- The proposed algorithm facilitates the construction of databases of individual data that are essential for conducting a variety of investigations which deal with speed analysis.

CASE STUDY

Speed surveys were performed in six stations on rural roads in the Province of Turin (North West of Italy). Data were collected at the same time both with loop detectors (counts) and cross-registration (individual), during off-peak hours to avoid traffic congestion.

Double-loop detectors: No. of classes: 7, thresholds: 30, 50, 70, 90, 110, and 130 km/h, aggregation period: 1 hour.

Cross-registrations: recording period: 15 minutes, but comparable observation frequencies inside each speed interval (difference lower than 10%).

Collected data were also analyzed to estimate the sample estimates $\hat{\theta}_i$, $s$ or $\alpha_i$ and to define if they are prone to follow a specific distribution by using the Kolmogorov-Smirnov test.

- estimated parameters are consistent (max difference equal to 12%)
- no evidence to suggest that one function distribution (Normal, Lognormal or Gamma) is superior to the others

Algorithm application. Distribution means are similar to those computed in the preliminary analysis; standard deviations are influenced by data aggregation (only seven classes) and by the sample composition process (inverse transform sampling). Due to different observation periods, to compare observed and modeled speeds, reduced samples were created through a random extraction from the outcome vector, with the actual number of values inside each interval.

The range in variability is small in the central part of the distribution, while it tends to increase towards the tails of the functions, leading to a greater variability in modeled speeds. Although the maximum and minimum values can deviate from the equality line, the data close to the average are very close to it. As expected, the algorithm works reasonably well and sometimes very well for the speed located across the central speed classes.

Some measures of accuracy (MPR, MAD, SDE, MPE) were used to provide information on the goodness-of-fit between observed and modeled data. These indexes confirm the results of statistical analysis performed on observed samples, or rather there is no one distribution that fits better with respect to the others.

Download the poster here!