Data fusion of instantaneous speeds and point-to-point travel times from probe vehicle data

Erik Jenelius
erik.jenelius@abe.kth.se

KTH Royal Institute of Technology
Dept. of Transport Science
Div. of Traffic and Logistics

1 Introduction

GPS devices in vehicles and smartphones have emerged as a new type of traffic sensor. These probe vehicle (also called FCD) sensors are opportunistic in the sense that their original purpose is not to collect traffic data, but have a great potential for cost efficient traffic monitoring and management applications.

However, a number of limitations mean that new sophisticated methods are needed to process the data and generate useful information, compared to traditional sensors. Sampling frequencies are often low (less than one per minute), so that vehicles may have traversed significant distances between reports. This means that that the spatial resolution of the individual travel time observations is typically lower than the resolution of the travel time model to be estimated. A general approach to overcome this problem is to use a large number of probe vehicle observations, which, through randomness in the sampled positions, cover different, partially overlapping sections of the network.

In the literature, there are two main ways to use probe vehicle data for travel time estimation:

1. *Point-to-point travel times*: Derive the average travel speed from the traversed distance and the difference in the time stamps between consecutive sampled locations (Jenelius and Koutsopoulos, 2013b; Hofleitner et al., 2012).

2. *Point speeds*: Use the reported instantaneous speed at each sampled location. This information is not available for all probe vehicle systems (Work et al., 2010).

A feature of probe vehicle data is the protocol that is used to determine the times and locations where vehicles are sampled. The two main principles are:

1. *Space-based sampling*: The vehicle trajectory is sampled at a certain location or at a certain distance from the previous sample point (Westgate et al., 2013).

2. *Time-based sampling*: The vehicle trajectory is sampled at a certain point in time or at a certain time interval since the previous sample point (Hofleitner et al., 2012).

It is well-known that time-based sampling means that locations are sampled more densely in parts where the speed is low; with space-based sampling, on the other hand, the sampled locations are independent of the speed (e.g., Liu et al., 2007; Westgate et al., 2013). Jenelius and Koutsopoulos (2013a) show that the sampling principle determines the proper specification of the likelihood function for point-to-point travel time data, differing in the probability distribution of how the travel time of the vehicle is allocated along the path between consecutive reports.

Using only point speed data, time-based sampling will lead to biased estimates of link travel time distributions since low-speed observations are over-represented. As far as we are aware, no rigorous correction for this problem has been proposed previously. Furthermore, few papers have considered how to combine point speed and point-to-point travel time information from vehicle traces in a consistent way.

The purpose of this paper is first to highlight the role of the sampling protocol in the estimation of travel time models with instantaneous speeds from probe vehicle data by MLE. It is shown that using only instantaneous speeds when sampling is time-based leads to biased parameter estimates. Second, a data fusion approach of instantaneous speeds and point-to-point travel times is proposed. It is shown that this resolves the problem of biased parameter estimates.

2 Travel time model

A path is partitioned into segments (links), along which the travel speed of each vehicle is assumed to be constant. Furthermore, the travel speed distributions are assumed to be independent between segments. Positions along the path are defined by the distance x from the start point of the path. The reciprocal of the travel speed is referred to as the travel time rate. The travel time rate on each segment j is a stochastic variable with pdf $p_{\mathrm{tr},j}(u \,|\, \theta_j)$.

2.1 Using only point speed information

Consider K independent samples of vehicle trajectories. Vehicle trajectory k is sampled at position x_1^k on segment m_1^k, reporting instantaneous speed v_1^k, $k = 1, \ldots, K$. Using this information only, a likelihood function is formulated as

$$\mathcal{L}(\theta) = \prod_{k=1}^{K} p(v_1^k \,|\, \theta) = \prod_{k=1}^{K} \frac{1}{(v_1^k)^2} p_{\mathrm{tr},m_1^k}\left(\frac{1}{v_1^k} \,\middle|\, \theta_{m^k}\right). \tag{1}$$

The likelihood is separable in the segments. The factor $1/\prod_{k=1}^{K}(v_1^k)^2$ does not influence estimation.

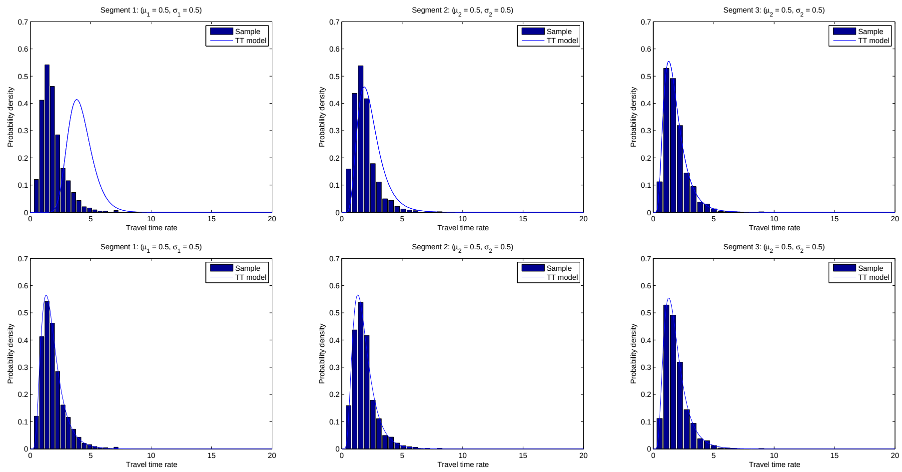

Figure 1: Time-based sampling, estimated (lines) versus empirical (bars) segment travel time rate distributions. Top: Using only point speeds. Bottom: Using both point speeds and point-to-point travel times.

2.2 Fusing point speed and point-to-point travel time information

Assume now that the preceding report of each vehicle is given, with location x_0^k on segment m_0^k and instantaneous speed v_0^k. With time-based sampling, the time τ^k between the reports is sampled according to the pdf $p_{\text{st}}(\tau^k \mid \beta_{\text{st}})$. Given positions x_0^k and x_1^k, the total travel distance $d^k = x_1^k - x_0^k$ and the distance traversed on each segment $j = m_0^k, \ldots, m_1^k$, denoted ℓ_j^k, are known. Using this information a likelihood function is formulated as

$$\mathcal{L}(\theta) = \prod_{k=1}^{K} p(d^k, v_1^k \mid x_0^k, \tau^k, v_0^k, \theta) = \prod_{k=1}^{K} p(v_1^k \mid \theta) \cdot p(d^k \mid x_0^k, \tau^k, v_0^k, v_1^k, \theta) \tag{2}$$

Given instantaneous speeds v_0^k and v_1^k, the times spent on segments m_0^k and m_1^k are $\tau_0^k = \ell_{m_0^k}^k / v_{m_0^k}$ and $\tau_1^k = \ell_{m_1^k}^k / v_{m_1^k}$. The second factor in (2) is thus the pdf of traversing intermediate segments $m_0^k+1, \ldots, m_1^k-1$ in time $\tau^k - \tau_0^k - \tau_1^k$; this pdf is derived by (Jenelius and Koutsopoulos, 2013a). Let $N^k = m_1^k - m_0^k - 1$ denote the number of intermediate segments, let f_j^k denote the unobserved fraction of the travel time spent on segment j, and let Δ_{N^k} denote the N^k-dimensional unit simplex. Thus,

$$\mathcal{L}(\theta) \propto \prod_{k=1}^{K} p_{\text{tr},m_1^k}\left(\frac{1}{v_1^k} \,\middle|\, \theta_{m_1^k}\right) \cdot \int_{\Delta_{N^k}} \prod_{j=m_0^k+1}^{m_1^k-1} f_j^k \, p_{\text{tr},j}\left(\frac{\tau^k - \tau_0^k - \tau_1^k}{\ell_j^k} f_j^k \,\middle|\, \theta_j\right) dF^k \tag{3}$$

3 Numerical experiments

The full paper describes the practical implementation of (2) and (3) and analyzes the impacts of using point-speed information in isolation and in combination with point-to-point travel time information for both time-based and space-based sampling protocols. A simple path with three segments and lognormal travel time distributions is used to illustrate the results. As an example, Figure 1 shows the estimation results using only point-speed information (top) and both information sources (bottom) with time-based sampling. As can be seen, using point speeds only may give highly biased estimates, while using both data sources leads to no discernable bias.

References

Hofleitner, A., Herring, R., Abbeel, P. and Bayen, A. (2012) Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. *IEEE Transactions on Intelligent Transportation Systems* 13, 1679–1693.

Jenelius, E. and Koutsopoulos, H. N. (2013a) Impact of sampling protocol on bias and consistency in travel time estimation on probe vehicle data. Submitted to *IEEE Transactions on Intelligent Transportation Systems.*

Jenelius, E. and Koutsopoulos, H. N. (2013b) Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B*, accepted for publication.

Liu, K., Yamamoto, T. and Morikawa, T. (2007) Comparison of time/space polling schemes for a probe vehicle system. Proceedings of the 14th World Conference on Intelligent Transport Systems, Beijing, China.

Westgate, B. S., Woodard, D. B., Matteson, D. S. and Henderson, S. G. (2013) Travel time estimation for ambulances using Bayesian data augmentation. *Annals of Applied Statistics*, forthcoming.

Work, D. B., Blandin, S., Tossavainen, O.-P., Piccoli, B. and Bayen, A. M. (2010) A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress* 2010, 1–35.