Stochastic Modeling for Vehicle Platoons (II): Statistical Characteristics

Baibing Li

School of Business & Economics
Loughborough University
Loughborough LE11 3TU, United Kingdom
E-mail: b.li2@lboro.ac.uk

ABSTRACT

This two-part paper presents a new approach to stochastic dynamic modeling for vehicle platoons. Part I develops a vehicle platoon model to capture the dynamics of vehicles' grouping behavior and proposes an online platoon recognition algorithm. On the basis of the developed platoon model, Part II investigates various important characteristics of vehicle platoons and derives their statistical distribution models, including platoon size, within-platoon headway, between-platoon headway and platoon speed. It is shown that the derived statistical distributions include some important existing models in the literature as their special cases. These statistical distribution models are crucial for us to understand the traffic platooning phenomenon. In practice, they can be used as the inputs for the design of traffic management and control algorithms for traffic with a platoon structure. Real traffic data is used to illustrate the obtained theoretical results.

Keywords: Between-platoon headway, Platoon size, Platoon speed, Within-platoon headway.

1. Introduction

Traffic platooning is an important traffic phenomenon. Vehicle platoons increase the capacity of roads and hence, when the traffic platoon structure is taken into consideration, the efficiency of traffic management can substantially be enhanced.

In Part I of this two-part paper in Li (2016), we develop a stochastic model to describe the dynamic behavior of vehicle platooning. We characterize vehicle platoons by both vehicle speeds and vehicle time headways so that the dynamic nature of the platoon-to-platoon transition process and within-platoon movements can be captured.

In part II, we turn to investigate various statistical properties of vehicle platoons. Traffic platoons are usually described by several important characteristics, including platoon size, platoon speed, within-platoon headway and between-platoon headway. These characteristics play an important role in traffic management and control. For example, between-platoon headway and platoon size are crucial inputs to traffic signal control algorithms (see, e.g., Jiang et al. 2006).

In the traffic literature, not much attention has been paid to date to investigating these platoon characteristics. Here we outline some important existing studies in this area.

The first strand of research is mostly based on queueing theory, aiming to develop some new theoretical distribution models. This includes the Borel-Tanner distribution (Tanner 1961; Haight and Breuer, 1960) and Miller distribution (Miller, 1961) for platoon sizes.

The other line of research takes an empirical approach, aiming to find which existing statistical distributions are suitable to describe the collected traffic data. Most of the studies on car-followers' headways fall into this category (e.g. Buckley, 1962, 1968; Ashton, 1971; Branston, 1976).

The recent research continues this line of research and uses an empirical approach to measuring the important platoon characteristics. Based on a large sample of measurements on the real traffic, Jiang et al. (2006) explored the platoon characteristics by using some common statistical distributions. They found that the exponential distribution was suitable for platoon size data and the normal distribution was a good approximation for platoon speed measurements. With respect to headways, their results suggested that the normal and log-normal distributions fitted well to within-platoon

headway data and platoons' inter-arrival times respectively. On the other hand, Ramezani et al. (2010) suggested the shifted exponential distribution for platoon sizes. They also found that the normal distribution was suitable for within-platoon headways and the shifted exponential distribution fitted between-platoon headways well. In addition, Sun and Benekohal (2005) also suggested the shifted exponential distribution for platoon sizes.

Li (2016) in part I of this paper has developed a stochastic model to describe the dynamic behavior of vehicle platooning. Here in part II we will investigate theoretical properties of the model for some important platoon characteristics (platoon size and speed, and within and between platoon headways) and derive their statistical distributions. The derived statistical distributions provide a probabilistic approach to understanding the statistical properties of the platoon characteristics by taking into consideration vehicles' dynamic grouping behavior that varies at different velocity levels. We note that the existing researches on vehicle platoons in the literature are solely based on vehicle time headways, and the dependence of the platoon characteristics on velocity is not captured (see, e.g., Tanner 1961; Miller, 1961; Cowan, 1975; Jiang et al. 2006).

This paper is structured as follows. In the next section we briefly summarize the dynamic platoon model proposed in Li (2016). Then we investigate various statistical properties of vehicle platoons' characteristics and use a practical example to illustrate the resulting statistical distributions in Section 3. Section 4 is devoted to model comparison and assessment. Here the derived statistical distributions are regarded as stand-alone models. We explore the relationships of these models with the existing distribution models in the literature. We also compare and assess the performances of the proposed distribution models. Finally, we draw conclusions in Section 5. The proofs of the theorems are offered in Appendix A. Some computational issues are discussed in Appendix B.

2. Summary of the stochastic dynamic model for vehicle platoons

In this section, we briefly summarize the stochastic platoon model developed in Part I of this paper; see Li (2016) for further details.

In the traffic literature, a vehicle platoon is defined to be a group of vehicles traveling together at approximately the same speed. Consider a traffic flow consisting of a number of consecutive vehicles indexed by $n = 1,2 \ldots \ldots$ Each individual vehicle n is characterized by two microscopic traffic variables, i.e. vehicle speed v_n and vehicle time headway h_n.

In Li (2016), a Markov regime-switching stochastic process is used to model the dynamic behavior of platoon-to-platoon transitions, and a state-space model is employed to describe individual vehicles' dynamic movements within each vehicle platoon.

For this end, a vehicle platoon indicator $G_n \in \mathcal{M}_P = \{1,2, \ldots, 2M\}$ for vehicle n is used to describe platoon-to-platoon transitions that follows a Markov switching process with a transition matrix Q:

$$Q = \begin{bmatrix} q_{11} & \cdots & q_{1\times(2M)} \\ \vdots & \ddots & \vdots \\ q_{(2M)\times 1} & \cdots & q_{(2M)\times(2M)} \end{bmatrix}, \quad (1)$$

where the entry in the *i*th row and *j*th column of matrix Q is the transition probability $q_{ij}(h_n) = \Pr\{G_n = i | G_{n-1} = j, h_n\}$.

A vehicle platoon $\mathbb{P}_m(j)$, which is indexed by m and of size L_m and is associated with a velocity mode j ($j \in \mathcal{M}_V$), is defined to be a number of consecutive vehicles $m_1 + 1, \ldots,$ and $m_1 + L_m$ such that the following conditions (C1)-(C3) are met (Li, 2016):

(C1) either $\{G_{m_1} \neq j\} \cap \{G_{m_1} \neq j + M\} \cap \{G_{m_1+1} = j\}$ or $\{G_{m_1+1} = j + M\}$;

(C2) $G_{m_1+L_m+1} \neq j$;

(C3) $G_n = j$ for all $m_1 + 2 \leq n \leq m_1 + L_m$.

The within-platoon movements of vehicles are described by a state-space model. Each speed measurement v_n is assumed to follow a normal distribution with mean $\bar{v}_n$ and standard deviation σ_0:

$$v_n = \bar{v}_n + \sigma_0 \varepsilon_n \quad \text{with} \quad \varepsilon_n \sim N(0,1). \quad (2)$$

For a vehicle n belonging to vehicle platoon $\mathbb{P}_m(j)$, its mean speed $\bar{v}_n$ of is equal to the mean speed level of the platoon (denoted as μ_j), plus a speed adjustment w_n made by the individual driver, i.e. $\bar{v}_n = \mu_j + w_n$. To reflect the dynamics of the within-platoon movements of vehicles, the speed drift w_n is assumed to follow an auto-regressive AP(p) model:

$$w_n = \sum_{k=1}^{p} \gamma_k w_{n-k} + \sigma_j e_n, \quad \text{with} \quad e_n \sim N(0,1) \tag{3}$$

where the standard deviation σ_j characterizes the magnitude of the adjustment and γ_k are coefficients.

In Li (2016), the platoon indicator G_n is characterized by two traffic indicators, headway mode R_n and velocity mode S_n, and the platoon indicator is expressed as $G_n = S_n + MR_n$. The headway mode $R_n \in \mathcal{M}_H = \{0,1\}$ indicates the status of the headway of a vehicle n: $R_n = 0$ representing the 'car-following' status and $R_n = 1$ representing the 'free-speed' status. On the other hand, velocity mode $S_n \in \mathcal{M}_V = \{1, \dots, M\}$ represents the velocity level of a vehicle n; A velocity mode j is associated with a mean speed level μ_j (with $\mu_1 < \cdots < \mu_M$) and standard deviation σ_j. Both the traffic indicators R_n and S_n are modeled by the Markov processes with probability transition matrices P_H and P_V respectively. The probability transition matrix P_H with entry in the *i*th row and *j*th column $\Pr\{R_n = i | R_{n-1} = j, h_n\}$ is given by:

$$P_H = \begin{bmatrix} r_0(h_n) & r_0(h_n) \\ 1 - r_0(h_n) & 1 - r_0(h_n) \end{bmatrix}. \tag{4}$$

Conditional on headway h_n, the probability transition matrix $P_V = [p_{ij}(h_n)]$ is modeled as

$$p_{ij}(h_n) = \Pr\{S_n = i | S_{n-1} = j, h_n\} = \frac{a_{ij}(h_n - \tau)^{b_{ij}}}{1 + \sum_{k \neq i} a_{kj}(h_n - \tau)^{b_{kj}}}, \tag{5}$$

where $a_{kj} \geq 0$ and $b_{kj} \geq 0$ for $k \neq j$, and $\tilde{a}_{ii} = 0$ and $b_{ii} = 0$. Both of $r_0(h_n)$ and $p_{ij}(h_n)$ in (4) and (5) are dependent on the current headway h_n.

The probability transition matrix of the platoon indicator G_n is given by $Q = P_H \otimes P_V$, where $\otimes$ denotes the Kronecker product of the two matrices P_H and P_V. We therefore have

$$q_{ij}(h_n) = q_{i(j+M)}(h_n) \quad \text{and} \quad q_{(i+M)j}(h_n) = q_{(i+M)(j+M)}(h_n) \quad \text{for any } i,j \in \mathcal{M}_V . \tag{6}$$

3. Statistical models for vehicle platoons' characteristics

In this section, we investigate some important characteristics of vehicle platoons $\mathbb{P}_m(j)$ $(j \in \mathcal{M}_V)$, including platoon size, within-platoon headway, between-platoon headway, and platoon speed.

Following the traffic literature (e.g., Breiman et al., 1968; Cowan, 1975), we assume that headways h_n $(n = 1,2, \dots)$ across a traffic stream are approximately independent of each other, and

each follows a mixture distribution with two components: a 'car-following' component $g_0(h)$ that is associated with the vehicles following its lead vehicle, and a 'free-speed' component $g_1(h)$ that is associated with the vehicles traveling at a free speed:

$$g(h) = \theta g_0(h) + (1-\theta) g_1(h), \tag{7}$$

where θ is the mixing probability. Following Cowan (1975) and Li (2016), we use the following gamma distributions to describe the headway components:

$$g_i(h) = (h-\tau)^{\alpha-1} \exp(-(h-\tau)/\lambda_i)/[\lambda_i^{\alpha}\Gamma(\alpha)] \quad (\text{for } h \geq \tau) \qquad i \in \mathcal{M}_H, \tag{8}$$

where τ is the minimum time headway, $\alpha \geq 1$ is the common shape parameter, and λ_i is the scale parameter of the distribution $g_i(h)$ with $\lambda_1 > \lambda_0$. Given the headway h_n of a vehicle n, we can explicitly write out the expression for the probability in (4):

$$r_0(h_n) = \Pr\{R_n = 0 | h_n\} = \frac{\theta g_0(h_n)}{\theta g_0(h_n) + (1-\theta) g_1(h_n)}. \tag{9}$$

In addition, based on the headway model (7) and noting $Q = P_H \otimes P_V$, we can work out the expected transition probabilities $\bar{q}_{ij} \coloneqq E\{q_{ij}(h)\}$:

$$\begin{aligned} \bar{q}_{ij} = \bar{q}_{i(j+M)} &= \int_{\tau}^{+\infty} \frac{a_{ij}(h_n-\tau)^{b_{ij}}}{1+\sum_{k\neq i} a_{kj}(h_n-\tau)^{b_{kj}}} \, r_0(h) g(h) dh \\ &= \theta \int_{\tau}^{+\infty} \frac{a_{ij}(h_n-\tau)^{b_{ij}}}{1+\sum_{k\neq i} a_{kj}(h_n-\tau)^{b_{kj}}} \, g_0(h) dh \qquad i,j \in \mathcal{M}_V. \end{aligned} \tag{10}$$

$$\begin{aligned} \bar{q}_{(i+M)j} = \bar{q}_{(i+M)(j+M)} &= \int_{\tau}^{+\infty} \frac{a_{ij}(h_n-\tau)^{b_{ij}}}{1+\sum_{k\neq i} a_{kj}(h_n-\tau)^{b_{kj}}} \, [1-r_0(h)] g(h) dh \\ &= (1-\theta) \int_{\tau}^{+\infty} \frac{a_{ij}(h_n-\tau)^{b_{ij}}}{1+\sum_{k\neq i} a_{kj}(h_n-\tau)^{b_{kj}}} \, g_1(h) dh \qquad i,j \in \mathcal{M}_V. \end{aligned}$$

The expected transition probabilities $\bar{q}_{ij}$ $(i,j = 1,\ldots,2M)$ describe dynamic transition behavior of platoons averaged across the entire traffic stream. It plays a very important role in the following analysis. A detailed discussion on the calculation of (10) is given in Appendix B.

3.1. Platoon size distributions

We now consider the statistical distribution of platoon sizes. We will first develop a theoretical model and then use the results in Li (2016) to illustrate the proposed model.

3.1.1. The model

The size of a platoon $\mathbb{P}_m(j)$ ($j \in \mathcal{M}_V$) is defined to be the number of the vehicles within the platoon. In this subsection, we denote the platoon size of $\mathbb{P}_m(j)$ as $L_m(j)$ to emphasize the dependence of the platoon size on the velocity level j.

We first note that, given that vehicle $m_1 + 1$ is the platoon leader with $S_{m_1+1} = j$, we have either $G_{m_1+1} = j \le M$ (hence $R_{m_1+1} = 0$) or $G_{m_1+1} = j + M$ (hence $R_{m_1+1} = 1$). We therefore consider $G_{m_1+1} = l$ with $l = j$ or $j + M$. We calculate the conditional probability that the vehicle platoon has only one vehicle:

$$\Pr\{L_m(j) = 1 | h_{m_1+2}\}$$
$$= \Pr\{G_{m_1+2} \neq j \big| G_{m_1+1} = l, h_{m_1+2}\} = 1 - q_{jl}(h_{m_1+2}) = 1 - q_{jj}(h_{m_1+2}),$$

where the last equality in the above equation is obtained from equation (6). In general, the conditional probability that the vehicle platoon includes k vehicles given the relevant headways is:

$$\Pr\{L_m(j) = k | h_{m_1+i}, i = 2, \dots, k+1\}$$
$$= \Pr\{G_{m_1+k+1} \neq j, G_{m_1+i} = j, i = 2, \dots, k, \big| G_{m_1+1} = l, h_{m_1+i}, i = 2, \dots, k+1\}$$
$$= \left[1 - q_{jj}(h_{m_1+k+1})\right] \prod_{i=2}^{k} q_{jj}(h_{m_1+i}). \quad (11)$$

This distribution is a generalized geometric distribution that is conditional on the headways.

The dependence on the headways in (11) makes theoretical analysis difficult. We hence consider the marginal probabilities for the platoon size by integrating out the headways:

$$l_{jk} := E\left[\Pr\{L_m(j) = k | h_{m_1+i}, i = 2, \dots, k+1\}\right]$$
$$= \int_{\tau}^{+\infty} (1 - q_{jj}(h_{m_1+k+1})) g(h_{m_1+k+1}) dh_{m_1+k+1} \prod_{i=2}^{k} \int_{\tau}^{+\infty} q_{jj}(h_{m_1+i})\, g(h_{m_1+i})\, dh_{m_1+i}$$
$$= (1 - \bar{q}_{jj})(\bar{q}_{jj})^{k-1} \quad \text{for} \quad k = 1,2, \dots \text{and } j \in \mathcal{M}_V, \quad (12)$$

with $\bar{q}_{ij}$ given by (10). We can see from (12) that, after integrating out the headways, the generalized geometric distribution (11) collapses to the ordinary geometric distribution $Geo(1 - \bar{q}_{jj})$. It is important to note that the platoon size distributions $Geo(1 - \bar{q}_{jj})$ at the different velocity modes $j \in \mathcal{M}_V$ are different, and consequently vehicle platoons traveling at different velocity levels have

different average sizes. As a whole, the platoon size over the entire traffic stream is a mixture distribution of these individual geometric distributions.

From equation (12), we can calculate the average platoon size for $j \in \mathcal{M}_V$:

$$\bar{L}(j) = \sum_{k=1}^{+\infty} k l_{jk} = \sum_{k=1}^{+\infty} k(1-\bar{q}_{jj})(\bar{q}_{jj})^{k-1} = 1/(1-\bar{q}_{jj}). \quad (13)$$

This quantity is important in traffic management.

In the literature, Dunne et al. (1968) investigated a geometric distribution, $l_k=(1-q)q^{k-1}$, to describe platoon sizes. The parameter q in the model of Dunne et al. (1968), however, was uniform across all vehicle velocities because Dunne et al. (1968) did not take into account speed variability and their platoon classification was solely based on time headways. In the numerical example investigated in the next subsection, we will see that this assumption is restrictive and unrealistic in practice.

3.1.2. Numerical illustration

We use the results obtained in the empirical study in Li (2016) to illustrate the platoon size distributions. The expected transition probability matrix obtained in Li (2016) is:

$$EQ = \bar{Q} = \begin{bmatrix} 0.081 & 0.102 & 0.081 & 0.102 \\ 0.390 & 0.369 & 0.390 & 0.369 \\ 0.082 & 0.123 & 0.082 & 0.123 \\ 0.447 & 0.406 & 0.447 & 0.406 \end{bmatrix},$$

where each element $\bar{q}_{ij}$ of $\bar{Q} = [\bar{q}_{ij}]$ is the probability that a vehicle platoon indicator $G_n = S_n + MR_n$ is equal to i, given that the platoon indicator of its lead vehicle is j. Here $R_n \in \{1,2\}$ and $S_n \in \mathcal{M}_V = \{1,2\}$, with $S_n = 1$ for the lower speed mode and $S_n = 2$ for the higher speed mode. For example, $\bar{q}_{21} = 0.390$ is the transition probability that platoon indicator G_n is 2 (and hence is of the car-following status and traveling at the higher speed level), given that the platoon indicator of its lead vehicle is 1 (and hence the lead vehicle is of the car-following status and traveling at the lower speed level).

On the basis of the expected transition matrix $\bar{Q}$, we can work out the average vehicle platoon size $E\{L_m(j)\}$ for $j \in \mathcal{M}_V$. For the vehicle platoons associated with the lower velocity mode, we use

equation (13) to obtain the average platoon size $1/(1-0.081) = 1.088$ and the corresponding variance 0.096. This shows that vehicles at the lower velocity mode tended to travel alone rather than to form a vehicle platoon. In other words, the following vehicle tended not to form a platoon with the lead vehicle when the lead vehicle traveled at a low speed. On the other hand, for vehicle platoons associated with the higher velocity mode, the average platoon size is equal to $1/(1-0.369) = 1.584$ and the corresponding variance is 0.927. Hence, on average, platoon sizes at the higher velocity mode tended to be larger. It is therefore unrealistic to assume the same platoon size distribution across the entire traffic stream as it did in Dunne et al. (1968).

From equation (12), the distribution for platoon size at the lower velocity level is

$$l_{1k} = 0.919\ (0.081)^{k-1} \qquad \text{for} \qquad k = 1,2,\dots, \tag{14}$$

whereas the distribution for platoon size at the higher velocity level is

$$l_{2k} = 0.631(0.369)^{k-1} \qquad \text{for} \qquad k = 1,2,\dots \tag{15}$$

The upper left panel in Figure 1 shows a grouped bar plot for the lower velocity mode, where the actual proportion of the platoon sizes (represented as the left white bar) and their corresponding theoretical value (the right black bar) calculated using distribution (14) are clustered around the same location on the x-axis. Likewise, for the higher velocity mode, the upper right panel shows the actual proportion of the platoon sizes (the left white bar) and the corresponding theoretical value (the right black bar) calculated using distribution (15) clustered at each location on the x-axis.

Overall, the two graphs show that the theoretical values calculated using the derived probability mass functions (14)-(15) are close to the actual values. In addition, in each case, the frequency decreases rapidly as the platoon size increases. It can also be seen that a majority of vehicle platoons at the lower velocity mode (see the upper left panel) has a size of 1, whereas more platoons at the higher velocity mode (the upper right panel) have a size larger than 1. This clearly shows that vehicles' platooning behavior can differ from each other substantially at the different velocity levels.

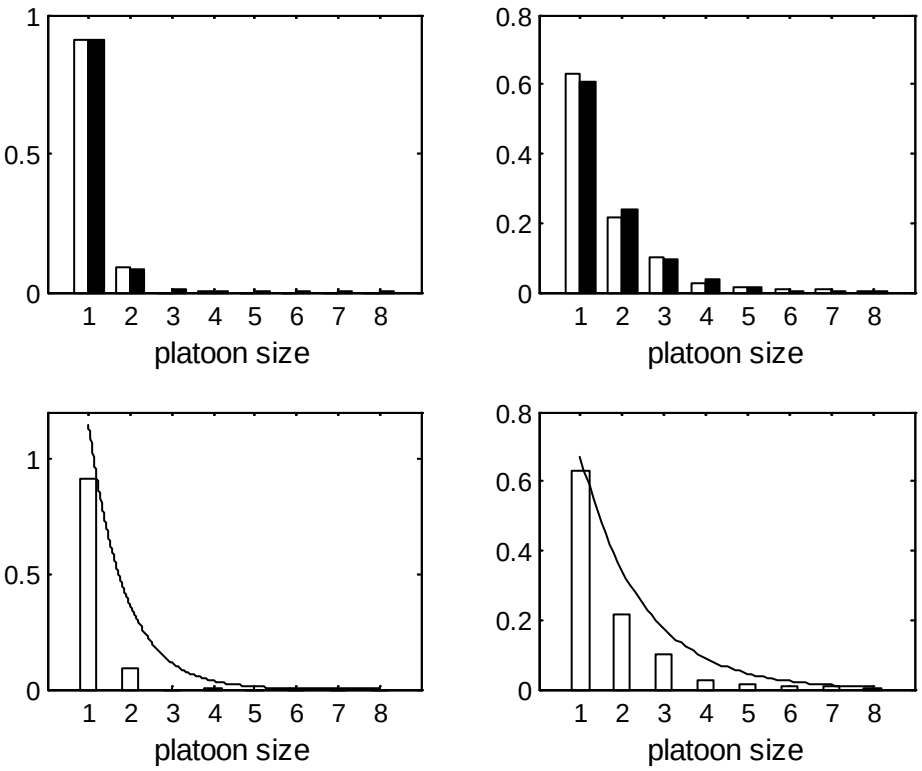

Figure 1. Actual proportions of platoon sizes (the left white bars) and the corresponding probabilities (the right black bars) for the lower velocity mode (the upper left panel) and for the higher velocity mode (the upper right panel) respectively; Actual proportions of platoon sizes and the corresponding shifted exponential approximations for the lower velocity mode (the lower left panel) and for higher velocity mode (the lower right panel) respectively.

3.2. Distributions of within-platoon headways

Next, we turn to consider the within-platoon headway distributions. We first develop a theoretical model and then use the results in Li (2016) to illustrate the proposed model.

3.2.1. The model

Consider any two consecutive vehicles (indexed as $n-1$ and n respectively) in a platoon $\mathbb{P}_m(j)$ ($j \in \mathcal{M}_V$). Let h_n denote the headway of vehicle n. Then, given $G_{n-1} = l$ with either $l = j$ or $l = j + M$, the conditional probability that vehicle n stays within the platoon $\mathbb{P}_m(j)$ is

$$\Pr\{n \in \mathbb{P}_m(j) | n-1 \in \mathbb{P}_m(j), h_n\} = \Pr\{ G_n = j | G_{n-1} = l, h_n\} = q_{jl}(h_n) = q_{jj}(h_n),$$

where the last equality in the above equation is from equation (6).

For any two consecutive vehicles $n-1$ and $n \in \mathbb{P}_m(j)$ ($j \in \mathcal{M}_V$), the probability density function for within-platoon headways for mode j is given by

$$f_{WP}(h_n; j) := f(h_n | n-1 \text{ and } n \in \mathbb{P}_m(j)).$$

From Bayes' rule, we can obtain

$$f(h_n | n-1 \text{ and } n \in \mathbb{P}_m(j)) = \frac{\Pr\{n \in \mathbb{P}_m(j) | n-1 \in \mathbb{P}_m(j), h_n\} f(h_n | n-1 \in \mathbb{P}_m(j))}{\int_\tau^{+\infty} \Pr\{n \in \mathbb{P}_m(j) | n-1 \in \mathbb{P}_m(j), h_n\} f(h_n h | n-1 \in \mathbb{P}_m(j)) dh_n}.$$

We note that $\Pr\{n \in \mathbb{P}_m(j) | n-1 \in \mathbb{P}_m(j), h_n\} = q_{jj}(h_n)$. In addition, $f(h_n | n-1 \in \mathbb{P}_m(j)) = g(h_n)$ as $\{h_n, n = 1,2, \dots\}$ are mutually independent. Hence, from equation (7), we obtain

$$f_{WP}(h_n; j) = \frac{q_{jj}(h_n) g(h_n)}{\int_\tau^{+\infty} q_{jj}(h_n) g(h_n) dh_n} = \theta p_{jj}(h_n) g_0(h_n) / \bar{q}_{jj}, \qquad \text{for } j \in \mathcal{M}_V \tag{16}$$

with $\bar{q}_{jj}$ given by (10). We can now calculate the average headway over $f_{WP}(h_n; j)$:

$$\rho_{WP}(j) = \int_\tau^{+\infty} h \, f_{WP}(h; j) dh \quad \text{for } j \in \mathcal{M}_V.$$

See Appendix B for a detailed discussion on the numerical computation of $\rho_{WP}(j)$. The following theorem discusses some theoretical properties of $f_{WP}(h; j)$. See Appendix A for proof.

Theorem 1. Let $f_{WP}(h; j)$ (for $j \in \mathcal{M}_V$) be given by (16) and $g_0(h)$ be given by (8). We have

(i) $\rho_{WP}(j) < \int_\tau^{+\infty} h \, g_0(h) dh < \int_\tau^{+\infty} h g(h) dh$;

(ii) $\lim_{h \to \infty} \frac{f_{WP}(h; j)}{g_0(h)} = 0$;

(iii) $f_{WP}(h; j)$ has the mode at $h^* \in (\tau, \tau + (\alpha - 1)\lambda_0)$ and then it strictly decreases for $h > h^*$ if $\alpha > 1$; if $\alpha = 1$, then $f_{WP}(h; j)$ strictly decreases for all $h \geq \tau$.

By definition (16) and from Theorem 1, $f_{WP}(h; j)$ is the probability density function of within-platoon headway at velocity mode $j \in \mathcal{M}_V$. We can see that $f_{WP}(h; j)$ essentially depends on the product of two components, i.e. $p_{jj}(h)$ and $g_0(h)$. Consequently, because the transition probability $p_{jj}(h_n)$ is different at the different velocity modes, the within-platoon headway behaves differently for the different velocity modes.

The above theorem also shows that the overall shape of $f_{WP}(h; j)$ is similar to $g_0(h)$. The right tail of $f_{WP}(h; j)$, however, decreases more rapidly than that of $g_0(h)$. In empirical analysis for within-

platoon headways, therefore, it is quite often that we only see a histogram around its mode h^*, beyond which the histogram vanishes rapidly.

In addition, Theorem 1 demonstrates that: (a) the average within-platoon headway is less than the average headway of the 'car following' component $\int_{\tau}^{+\infty} hg_0(h)dh$; (b) which in turn is less than the average headway of the entire traffic stream, i.e. $\int_{\tau}^{+\infty} hg(h)dh$.

Finally, we point out that, in Jiang et al. (2006), the platoons are defined using a deterministic rule, that is, a vehicle and its lead vehicle belong to the same platoon if its time headway is less than 2.5 s. Hence, the within-headway distribution used in Jiang et al. (2006) is $f_{WP}(h;j) \propto I(h \leq 2.5)g(h)$, i.e. it is the traffic headway distribution $g(h)$ truncated over the interval of $[\tau, 2.5]$, where $I(A)$ is an indicator function of set A. Clearly, the derived distribution in equation (16) is not suitable to those vehicle platoons recognised using the deterministic rule.

3.2.2. Numerical illustration

Based on the numerical results in the empirical study in Li (2016), the corresponding within-platoon headway distribution at the lower velocity level is:

$$f_{WP}(h;1)$$
$$= 23.783(h-0.490)^{1.320}\exp(-(h-0.490)/0.507)[1+4.842(h-0.490)^{0.093}]^{-1}.$$

Similarly, the obtained within-platoon headway distribution at the higher velocity level is:

$$f_{WP}(h;2)$$
$$= 5.236\,(h-0.490)^{1.320}\exp(-(h-0.490)/\,0.507)[1+0.279(h-0.490)^{0.061}]^{-1}.$$

The upper left and upper right panels of Figure 2 display the above probability density functions for the lower velocity mode and for the higher velocity mode respectively. For comparison purposes, we also display the histograms of the corresponding within-platoon headways on the same graph. Overall, we can see that the within-platoon headways are fairly homogeneous: the probability densities of within-platoon headways, as shown in Theorem 1, vanish rapidly beyond their modes.

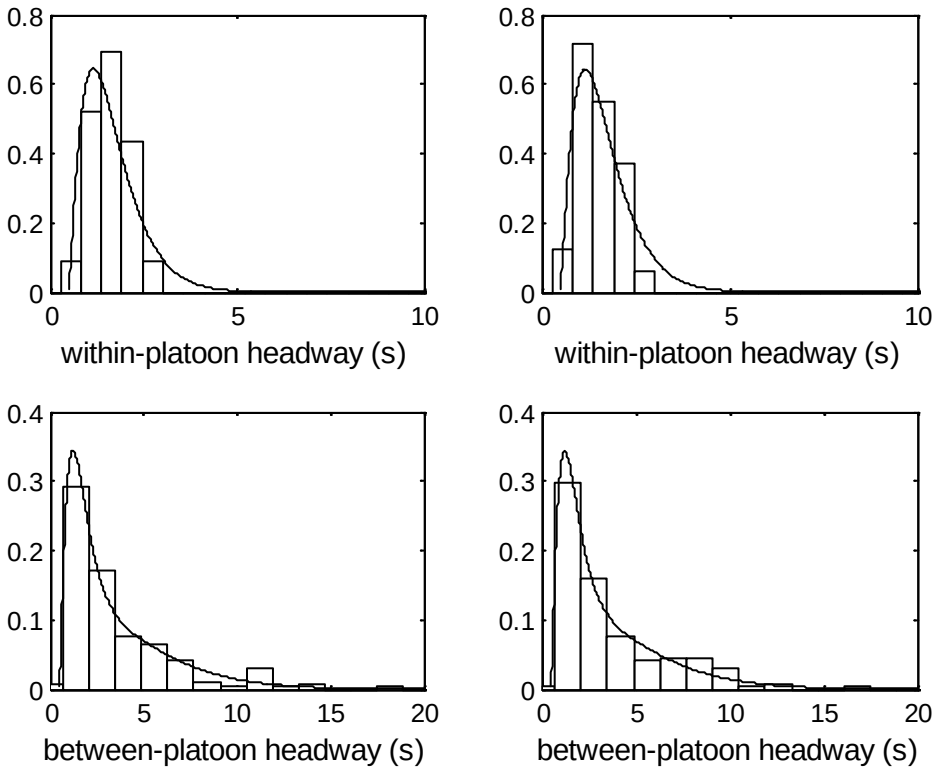

Figure 2. Distribution of within-platoon headway $f_{WP}(h;1)$ for the lower velocity mode (upper left) and distribution of the higher velocity mode $f_{WP}(h;2)$ (upper right) respectively; Distribution of between-platoon headway $f_{BP}(h;1,2)$ (lower left) and distribution of between-platoon headway distribution $f_{BP}(h;2,1)$ (lower right) respectively.

3.3. Distributions of between-platoon headways

Next, we consider the distributions of between-platoon headways.

3.3.1. The model

The between-platoon headway of any two consecutive platoons is defined to be the elapsed time between the front of the last vehicle of a platoon passing a point on the roadway and the front of the platoon leader of the following platoon passing the same point.

We note that there are two different scenarios where two consecutive vehicle platoons are separated from each other: (a) the two platoons belong to the same velocity mode and there is a large temporal gap between the two platoons; (b) the two platoons do not belong to the same velocity mode, and hence the temporal gap between them can be large and can be small.

First, we consider case (a) where the two consecutive vehicles under investigation belong to two different platoons which have the same velocity mode: $n-1 \in \mathbb{P}_{m-1}(i)$ and $n \in \mathbb{P}_m(i)$ for any $i \in \mathcal{M}_V$. In this case, the two platoons must have a large temporal gap between them.

Specifically, given $G_{n-1} = l$ with either $l = i$ (hence $R_{n-1} = 0$) or $l = i + M$ (so $R_{n-1} = 1$), the conditional probability that vehicle n is the platoon leader of $\mathbb{P}_m(i)$ is

$$\Pr\{n \in \mathbb{P}_m(i) | n-1 \in \mathbb{P}_{m-1}(i), h_n\} = \Pr\{ G_n = i + M | G_{n-1} = l, h_n\} = q_{(i+M)i}(h_n) ,$$

where the last equality in the above equation is from equation (6).

The probability density function for between-platoon headways, $f_{BP}(h_n|i,i)$, is given by

$$f_{BP}(h_n|i,i) \coloneqq f(h_n | n-1 \in \mathbb{P}_{m-1}(i), n \in \mathbb{P}_m(i), R_n).$$

Now we apply Bayes' rule to obtain:

$$f(h_n | n-1 \in \mathbb{P}_{m-1}(i), n \in \mathbb{P}_m(i), R_n) = \frac{q_{(i+M)i}(h_n) g(h_n)}{\int_\tau^{+\infty} q_{(i+M)i}(h_n) g(h_n) dh_n}$$

$$= (1-\theta) p_{ii}(h_n) g_1(h_n) / \bar{q}_{(i+M)i} \qquad \text{for } i \in \mathcal{M}_V . \tag{17}$$

On the basis of (17), we can calculate the average headway over $f_{BP}(h_n|i,i)$:

$$\rho_{BP}(i,i) = \int_\tau^{+\infty} h \, f_{BP}(h;i,i) dh \qquad \text{for } i \in \mathcal{M}_V .$$

See Appendix B for the numerical computation of $\rho_{BP}(i,i)$. Similar to Theorem 1, we have

Theorem 2. Let $f_{BP}(h|i,i)$ (for $i \in \mathcal{M}_V$) be given by (17) and $g_1(h)$ be given by (8). We have

(i) $\rho_{WP}(i) < \rho_{BP}(i,i) < \int_\tau^{+\infty} h \, g_1(h) dh$;

(ii) $\lim_{h\to\infty} \frac{f_{BP}(h|i,i)}{g_1(h)} = 0$;

(iii) $f_{BP}(h|i,i)$ has the mode at $h^* \in (\tau, \tau + (\alpha-1)\lambda_1)$ and then it strictly decreases for $h > h^*$ if $\alpha > 1$; if $\alpha = 1$, then $f_{BP}(h|i,i)$ strictly decreases for all $h \geq \tau$.

By definition (17) and from Theorem 2, $f_{BP}(h|i,i)$ is the probability density function of between-platoon headway when two consecutive platoons are both at the same velocity mode $i \in \mathcal{M}_V$. Essentially the probability density function in (17) depends on $p_{ii}(h_n)$ and $g_1(h_n)$. Hence, between-platoon headway behaviors differently at the different velocity levels. Theorem 2 (i) also shows that

the average between-platoon headway $\rho_{BP}(i,i)$ is larger than the average within-platoon headway $\rho_{WP}(i)$ for each $i \in \mathcal{M}_V$.

Now we turn to case (b) and consider two consecutive vehicles, $n-1 \in \mathbb{P}_{m-1}(i)$ and $n \in \mathbb{P}_m(j)$ ($i \neq j$ and $i,j \in \mathcal{M}_V$). Clearly the platoon indicator G_n of vehicle n can be either j or $j+M$; In addition, G_{n-1} can be equal to either i or $i+M$. For given $G_{n-1} = l$ with $l = i$ or $i+M$, we have

$$\begin{aligned}
&\Pr\{n \in \mathbb{P}_m(j) | n-1 \in \mathbb{P}_{m-1}(i), h_n\} \\
&= \Pr\{ (G_n = j) \cup ((G_n = j+M) | G_{n-1} = l, h_n\} \\
&= \Pr\{ G_n = j | G_{n-1} = l, h_n\} + \Pr\{ G_n = j+M | G_{n-1} = l, h_n\} \\
&= q_{jl}(h_n) + q_{(j+M)l}(h_n) = p_{ji}(h_n),
\end{aligned}$$

where the last equality in the above equation is obtained from equation (6). Clearly, the results are identical for either $l = i$ or $i+M$.

For two consecutive vehicles, $n-1 \in \mathbb{P}_{m-1}(i)$ and $n \in \mathbb{P}_m(j)$ ($i \neq j$ and $i,j \in \mathcal{M}_V$), the probability density function for between-platoon headways, $f_{BP}(h_n; j, i)$, is given by

$$f_{BP}(h_n; j, i) \coloneqq f(h_n | n \in \mathbb{P}_m(j), n-1 \in \mathbb{P}_{m-1}(i)).$$

We apply Bayes' rule to obtain

$$\begin{aligned}
&f(h_n | n \in \mathbb{P}_m(j), n-1 \in \mathbb{P}_{m-1}(i)) = \frac{p_{ji}(h_n) g(h_n)}{\int_\tau^{+\infty} p_{ji}(h_n) g(h_n) dh_n} \\
&= [\bar{q}_{ji} + \bar{q}_{(j+M)i}]^{-1} p_{ji}(h_n) g(h_n), \qquad \text{for } i \neq j
\end{aligned} \tag{18}$$

where $\bar{q}_{ij}$ is given by (10) and $g(h)$ is the headway distribution in (7). We calculate the average headway over $f_{BP}(h_n; j, i)$:

$$\rho_{BP}(j,i) = \int_\tau^{+\infty} h \, f_{BP}(h; j, i) dh \qquad \text{for } i,j \in \mathcal{M}_V \text{ and for } i \neq j.$$

See Appendix B for the numerical computation of $\rho_{BP}(j,i)$. Similar to Theorems 1 and 2, we have

Theorem 3. Let $f_{BP}(h; j, i)$ (for $i,j \in \mathcal{M}_V$ and $i \neq j$) be given by (18) and $g(h)$ be given by (7). We have $\rho_{BP}(j,i) \geq \int_\tau^{+\infty} h \, g(h) dh$.

By definition (18) and from Theorem 3, $f_{BP}(h|j,i)$ (for $i,j \in \mathcal{M}_V$ and $i \neq j$) is the probability density function of between-platoon headway when two consecutive platoons travel at different

velocity levels $i, j \in \mathcal{M}_V$. Theorem 3 shows that the average between-platoon headway is larger than the average headway across the entire traffic stream. In addition, noting that $g(h)$ in (7) is a mixture distribution, $f_{BP}(h|j,i)$ in equation (18) is also a mixture distribution,

$$f_{BP}(h_n; j, i) = \theta[\bar{q}_{ji} + \bar{q}_{(j+M)i}]^{-1} p_{ji}(h_n) g_0(h) + (1-\theta)[\bar{q}_{ji} + \bar{q}_{(j+M)i}]^{-1} p_{ji}(h_n) g_1(h),$$

which indicates that the between-platoon headway in case (b) can be small or large. In addition, from Theorem 1 (i) and Theorem 3, we can see that the average between-platoon headway is larger than the average within-platoon headway.

3.3.2. Numerical illustration

Based on the numerical results of the empirical study in Li (2016), the corresponding between-platoon headway distribution are:

$$f_{BP}(h; 2,1) = \frac{(h-0.490)^{1.413}}{1+4.842(h-0.490)^{0.093}} \{11.168 \exp\left(-\frac{h-0.490}{0.507}\right) + 0.535 \exp\left(-\frac{h-0.490}{1.974}\right)\}$$

and

$$f_{BP}(h; 1,2) = \frac{(h-0.490)^{1.381}}{1+0.279(h-0.490)^{0.061}} \{2.392 \exp\left(-\frac{h-0.490}{0.507}\right) + 0.115 \exp\left(-\frac{h-0.490}{1.974}\right)\}.$$

The lower left and lower right panels of Figure 2 display the above probability density functions of between-platoon headway respectively. For comparison purposes, we also display the histograms of the corresponding between-platoon headways on the same graph.

Overall, we can see from Figure 2 that the between-platoon headways have a heavier tail. In addition, the average between-platoon headways are much larger than that of the within-platoon headways, as asserted in Theorems 1 and 3.

From Figure 2 (the lower left and right panels), there are rather a lot of vehicles with a headway of less than 2.5 seconds that are considered to be in separate platoons. This is because, by the platoon definition, two consecutive sets of vehicles travelling at different speed levels will be classified into two separate platoons, even if their temporal gap is relatively small.

3.4. Platoon speed distribution

3.4.1. The model

The speed of a platoon can be characterized by two quantities: its speed level and speed dispersion. First, we define the platoon speed $\bar{V}_m$ of $\mathbb{P}_m(j)$ ($j \in \mathcal{M}_V$) to be

$$\bar{V}_m(j) = \sum_{l=1}^{L_m} \bar{v}_{m_1+l} / L_m = \mu_j + \delta_m,$$

with $\delta_m = \sum_{l=1}^{L_m} w_{m_1+l} / L_m$, i.e. $\bar{V}_m$ is the average speed of the vehicles within the platoon.

From equation (2), it is straightforward to conclude that, for given δ_m, the observed platoon speed $V_m = \sum_{l=1}^{L_m} v_{m_1+l} / L_m$ follows a normal distribution with a mean of $\bar{V}_m(j)$ and variance of σ_0^2 / L_m:

$$V_m = \mu_j + \delta_m + \tilde{\varepsilon}_m \qquad \text{for } m = 1,2 \ldots$$

where $\tilde{\varepsilon}_m \coloneqq \sum_{l=1}^{L_m} e_{m_1+l} / L_m$. In addition, $\tilde{\varepsilon}_m$ and $\tilde{\varepsilon}_l$ for two platoons m and l ($m \neq l$) are mutually independent. Also note the heterogeneity of $\tilde{\varepsilon}_m$; the variances $var(\tilde{\varepsilon}_m) = \sigma_0^2 / L_m$ are different when the platoons have different sizes. Clearly, given δ_m, the conditional distribution of the observed platoon speeds $\{V_m\}$ ($m = 1,2 \ldots$) may be approximated by a normal mixture with a sufficiently large number of components, K.

Since $\delta_m = \sum_{l=1}^{L_m} w_{m_1+l} / L_m$ also follows a normal distribution, the unconditional univariate distribution of V_m is also normal. Note, however, because δ_m and δ_l for two platoons m and l ($m \neq l$) are not mutually independent, V_m and V_l are correlated. Therefore, the unconditional distribution of a vector of observed platoon speeds $[V_1, \ldots, V_m]^T$ is an m-dimensional multivariate normal distribution.

Next, we turn to consider platoons' speed dispersion. We note that the vehicles that follow the platoon leader in $\mathbb{P}_m(j)$ usually travel at different speeds, as indicated by equation (3). However, although the speeds of the vehicles within the platoon may vary from vehicle to vehicle, the variation of the individual speeds may largely cancel out each other; it is the difference, $\bar{d}_m(j) = \bar{v}_{m_1+L_m} - \bar{v}_{m_1+1}$ (for $L_m > 1$), i.e. vehicle speeds between the last vehicle within the platoon and the platoon leader, that reflects the overall change in the vehicle speeds. Hence, we use $\bar{d}_m(j)$ to characterize the dispersion of platoon speeds.

Let $d_m(j) = v_{m_1+L_m} - v_{m_1+1}$ (for $L_m > 1$) denote the observed platoon speed dispersion. Note that both $d_m(j)$ and $\bar{d}_m(j)$ are conditional on the platoon length L_m. Here we suppress the

dependence for notational simplicity. Under the assumptions for equations (2) and (3), it can be seen that, for given w_{m_1+l} $(l = 1, \dots, L_m)$, $d_m(j)$ follows a normal distribution with a mean of $\Delta w_m = w_{m_1+L_m} - w_{m_1+1}$ and variance of $2\sigma_0^2$.

3.4.2. Numerical illustration

We display the histogram of the platoon speeds for the lower velocity mode (higher velocity mode) in the upper left panel (upper right panel) of Figure 3. We also superimpose the probability density function of a mixture normal with two components for each case. It seems that the normal mixture with two components does not provide a brilliant fit.

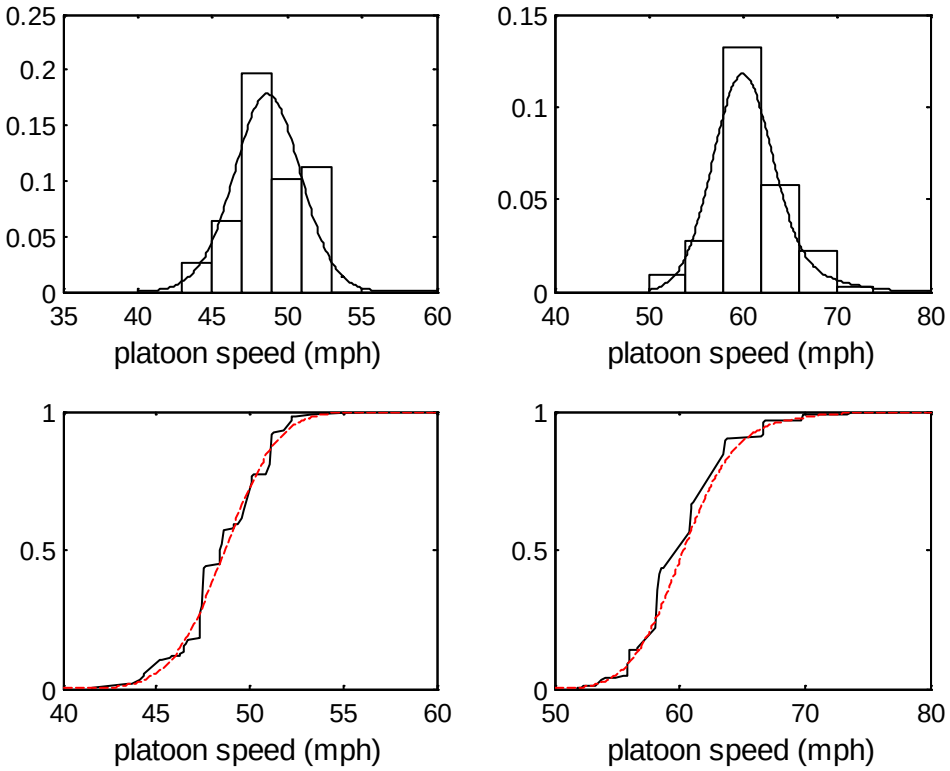

Figure 3. Histograms and the corresponding normal mixture density functions of platoon speeds for the lower velocity mode (upper left panel) and for the higher velocity mode (upper right panel); Empirical cumulative distributions (real line) and the corresponding normal mixture distribution functions (broken line) of platoon speeds for the lower velocity mode (lower left panel) and for the higher velocity mode (lower right panel).

To better assess the goodness of fit, we also plot the empirical cumulative distributions of the platoon speeds and the corresponding normal mixture distribution for the lower velocity mode (the lower left panel) and for the higher velocity mode (the lower right panel) respectively. We can see a

clear gap between the empirical cumulative distribution and the theoretical values. For example, from the lower right graph we can see that the real line (the empirical cumulative distribution) is higher than the broken line (the theoretical values calculated from the normal mixture with two components) in the middle part of the graph.

This is not surprising: due to the heterogeneity of the platoon speeds, a sufficiently large number of normal mixture components is expected to fit to the platoon speeds. In Section 4.4, we will explore the quality of fit for the normal mixture with more than two components.

4. Model comparison and assessment

In the previous section, several statistical distributions were derived based on a state space model with a Markov regime-switching process. In practice, these distributions can be regarded as stand-alone models for the platoon characteristics and used to fit to vehicle platoon data.

In this section, we will investigate the relationships between the developed distribution models for the platoon characteristics and the existing statistical distribution models in the literature. In addition, we will also assess the numerical performances of the derived statistical distributions.

In the following numerical comparison, we use the vehicle platoons recognized in Li (2016). In total there were 1057 vehicles under investigation and 712 platoons were recognized in Li (2016). During the platoon recognition, the probability of the platoon indicator for each vehicle was first calculated, and then the platoon classification was undertaken using Algorithm B in Li (2016) that was based on a state space model with a Markov switching process; see Li (2016) for the details.

We choose two widely used model selection criteria, i.e. the deviance and Akaike information criterion (AIC), to compare and contrast the models (see, e.g., Lunn et al., 2012).

Deviance in statistics is a quality-of-fit statistic for a model. It is a generalization of the least square criterion in multiple linear regression analysis to cases where model-fitting is achieved by maximum likelihood (McCullagh and Nelder, 1989). Specifically, the deviance for a model with a parameter vector ξ is defined to be the log maximum likelihood multiplied by -2 (e.g., Lunn et al., 2012):

$$D(\hat{\xi}) = -2log\mathcal{L}(\hat{\xi}),$$

where $\mathcal{L}(\xi)$ is the likelihood function of a model and $\hat{\xi}$ is the maximum likelihood estimate of ξ. When the underlying model is a normal distribution, $D(\hat{\xi})$ reduces to the least square criterion.

AIC is also a commonly used criterion for comparison among several models. Unlike the deviance that only takes into account quality of fit, AIC is a trade-off between quality of fit and model complexity. AIC is defined to be the deviance plus 2 times the number of the parameters in the model:

$$\text{AIC}= D(\hat{\xi}) + 2\dim(\xi),$$

where $\dim(\xi)$ denotes the number of the parameters. Note that only the differences in AIC between models are important; hence, of any two models for comparison, the model with a smaller AIC is more favorable. Clearly, AIC penalizes more complicated models.

Lunn et al. (2012, pp 166-167) have used the following guideline for model selection: (a) differences in AIC between two models more than 10 might definitely rule out the model with the higher AIC; (b) differences between 5 and 10 are substantial; and (c) there is uncertainty about the choice of model for differences less than 5. In this section, we will follow this guideline for model selection.

4.1. Platoon size distributions

We first consider the following geometric distribution $Geo(1-q_j)$ for platoon sizes derived in Section 3.1:

$$l_{jk}=(1-q_j)q_j^{k-1} \qquad \text{for} \qquad k = 1,2,\ldots, \tag{19}$$

where the parameter q_j is the probability that a vehicle remains in the same platoon traveling at speed level j.

In the literature, there are several other existing distribution models for platoon sizes, including: (a) the Borel-Tanner distribution (Tanner, 1961; Miller, 1961); (b) the Miller's model (Miller, 1961); and (c) the shifted exponential distribution (Sun and Benekohal, 2005; Jiang et al., 2006; Ramezani et al., 2010). See also Cowan, (1975), Galin (1980), and Gartner et al. (2001) for further discussions about these models.

The Borel-Tanner distribution, $l_k = (k\alpha e^{-\alpha})^{k-1} e^{-\alpha}/k!$, is a discrete distribution with a single parameter α. On the other hand, the Miller distribution, $l_k = (m+1)(m+s+1)!\,(s+k-1)!/[s!\,(m+s+k+1)!]$, is a discrete distribution with two parameters m and s. Both models are derived under the assumptions of Poisson arrivals and constant service time in queueing theory.

It is of interest to compare model (19) with the Borel-Tanner distribution and Miller distribution. First, the geometric distribution (19) has a simpler mathematical expression. Its parameter q_j has a clear interpretation in practice. In addition, distribution (19) is a velocity-specific model, assuming the statistical characteristics of platoon sizes are different at different speed modes. This assumption has support from empirical analysis (see the upper left and upper right graphs in Figure 1). On the other hand, the Borel-Tanner distribution and Miller distribution assume the same average platoon size across the entire traffic stream, regardless of the traveling speeds of vehicles under investigation.

In terms of the connections among these distributions, we point out that the geometric distribution (19), Borel-Tanner distribution, and Miller distribution have similar mathematical forms for their means; they can all be written as

$$expected\ platoon\ size = 1/(1-\varphi),$$

where we have $\varphi = q_j$ for model (19), $\varphi = \alpha$ for the Borel-Tanner distribution, and $\varphi = (s+1)/(m+s+1)$ for Miller distribution. Hence, when the corresponding parameters are carefully calibrated, these three distributions will have the same estimate of the average platoon size.

Next, we turn to consider the exponential distribution model for platoon sizes. There are several recent studies (e.g., Sun and Benekohal, 2005; Jiang et al., 2006; Ramezani et al., 2010) using the exponential distribution to approximate the platoon size distribution. In theory, the geometric distribution (19) is closely related to the exponential distribution, as shown in the following lemma:

Lemma (Johnson, 2011, pp.159). If a random variable X follows a geometric distribution $Geo(p)$ with $p = 1/(\lambda n)$, then the distribution of the random variable $Y = X/n$ approaches to the exponential distribution with mean λ when n becomes large.

Hence, an exponential distribution is the limiting case of the corresponding geometric distribution under some mild conditions. To illustrate the above theoretical result, we display the actual proportions of the platoon sizes and the exponential approximation at the lower (or higher) velocity mode in the lower left (or right) panel of Figure 1. It can be seen that the shifted exponential distribution is a good approximation to the actual platoon sizes but the gap between them is substantial.

Finally, based on the deviance and AIC, we compare the numerical performances of these models. Note that although the Borel-Tanner distribution and Miller distribution model assume the identical parameters across the entire traffic stream, we apply them to each of the different velocity modes in the following numerical comparison. This is because, as shown in Figure 1, the statistical characteristics of the platoon sizes at the two velocity models are very different, and hence a pooled analysis would not make much sense.

Table 1. The deviance and AIC for the shifted exponential distribution with continuity correction, Borel-Tanner distribution, Miller distribution, and geometric distribution (19).

	Exponential distribution	Borel-Tanner distribution	Miller distribution	Geometric distribution (19)
		At the lower	speed mode	
Deviance	228.4	140.0	140.5	140.0
AIC	232.4	142.0	144.5	142.0
		At the higher	speed mode	
Deviance	1198.8	1118.9	1108.0	1110.4
AIC	1202.8	1120.9	1112.0	1112.4

We calculate the maximum likelihood estimate for each model, and hence work out the deviance and AIC values. The results are displayed in Table 1. It can be seen from Table 1 that overall, the shifted exponential distribution performs poorly. For both the lower speed mode and higher speed mode, the differences in AIC between the shifted exponential distribution and the other models are much larger than 10. This indicates that this model should be ruled out in this analysis.

On the other hand, the other three models have similar performances: the differences in AIC among them are small. This suggests that the Borel-Tanner distribution, Miller distribution, and geometric distribution perform equally well for this analysis.

4.2. Within-platoon headway distributions

For the case where there are only two velocity modes, the within-platoon distribution for one velocity mode developed in the previous section can be written as:

$$f_{WP}(h) = c(h-\tau)^{\alpha-1}\exp(-(h-\tau)/\lambda)/[1+a(h-\tau)^{b}], \tag{20}$$

where a, b, α, λ, and τ are parameters and c is the normalization constant.

In the literature, the normal distribution is also used to describe within-platoon headways. For example, Jiang et al. (2006) and Ramezani et al. (2010) found that a normal distribution fitted their within platoon headway data well. The following theorem shows that the model (20) and normal distribution have a close relationship.

Theorem 4. If a random variable X follows distribution (20), then the distribution of the random variable $Y = (X-\alpha\lambda)/(\alpha^{1/2}\lambda)$ approaches to the normal distribution with mean of τ and unit variance for any fixed b when $a \to 0$ and $\alpha \to +\infty$.

See Appendix A for proof. Theorem 4 shows that normal distributions are a special case of the model (20) for within-platoon headways.

There are, however, two important differences between the two models. First, within-platoon headways are usually asymmetrical; the skewness of the headways cannot be captured by the normal approximation. Hence, the normal approximation is a good choice only when within-platoon headways are more or less symmetric. Secondly, model (20) has a clear practical interpretation: it is the headway distribution of the car-following mode, corrected by the probability that a vehicle will remain in the same platoon. There is no such an interpretation when the normal approximation is used.

Now we consider the issue of model selection. We calculate the deviances and AIC values to compare the two within-platoon headway models, i.e. the normal distribution and developed model (20), as displayed in Table 2.

For the lower speed mode, it can be seen from Table 2 that the proposed distribution model (20) provides a better fit in terms of quality-of-fit, with deviance of 23.2, when comparing to the normal approximation with deviance of 32.7. However, their difference in AIC is small (less than 5), suggesting the two models perform equally well. This is due to the fact that model (20) has more parameters and AIC penalises a more complicated model.

At the higher speed mode, on the other hand, the difference in AIC between the two models is large. Hence, the proposed distribution model (20) is a better choice for this case.

Table 2. The deviance and AIC for the normal distribution and the proposed model (20) for within-platoon headways.

	Normal distribution	Proposed distribution (20)
	At the lower speed mode	
Deviance	32.7	23.2
AIC	36.7	33.2
	At the higher speed mode	
Deviance	483.0	340.6
AIC	487.0	350.6

4.3. Between-platoon headway distributions

We now consider the between-platoon headway models. For the case where there are only two velocity modes and two platoons travel at different speed levels, the between-platoon distribution developed in the previous section is given by:

$$f_{BP}(h) = ca(h-\tau)^{b+\alpha-1}\{\theta\lambda_0^{-\alpha}\exp(-(h-\tau)/\lambda_0) + (1-\theta)\lambda_1^{-\alpha}\exp(-(h-\tau)/\lambda_1)\}/[1+a(h-\tau)^b] \quad (21)$$

where α, τ, λ_0, λ_1, θ, a and b are parameters and c is the normalization constant. Model (21) has a clear practical interpretation: it is the vehicle time headway distribution, corrected by the probability that a vehicle will switch to the other speed mode.

In the literature, the log-normal distribution (Jiang et al., 2006) and shifted exponential distribution (Ramezani et al., 2010) are used to describe between-platoon headways.

Theoretically, there is no directly connection between model (21) and the lognormal distribution. We note however, with $\theta = 0$ and $a \neq 0$, the distribution in (21) approaches to a three-parameter gamma distribution $g_1(h) = (h-\tau)^{\alpha-1}\exp(-(h-\tau)/\lambda_1)/[\lambda_1^{\alpha}\Gamma(\alpha)]$ when $b \to +\infty$. Therefore, model (21) includes the gamma distribution as its special case. In the literature, the gamma distribution and lognormal distribution are often used interchangeably (see, e.g., Wiens, 1999). Firth (1988) proves that analyzing log-normal data assuming a gamma distribution is more efficient than analysing gamma data assuming log-normality. Because the model (21) is much more flexible than the gamma distribution, we would expect that model (21) usually provides a better fit than the lognormal distribution.

On the other hand, model (21) and the shifted exponential distribution are closely related, as shown in the following theorem.

Theorem 5. Suppose a random variable X follows distribution (21). If $a \neq 0$ is fixed with $\theta = 0$ and $\alpha = 1$, then the distribution of the random variable X approaches to a shifted exponential distribution with mean of $\tau + \lambda_1$ and variance of λ_1 when $b \to +\infty$.

The proof of Theorem 5 is trivial. Theorem 5 shows that as a special case of (21), the shifted exponential distribution may fit between-platoon headways well in some applications.

Table 3. The deviance and AIC for the lognormal distribution, shifted exponential distribution and the proposed model (21) for between-platoon headways.

	Lognormal distribution	Shifted exponential distribution	Proposed distribution (21)
	At the lower speed mode		
Deviance	787.5	786.3	519.3
AIC	791.5	790.3	533.3
	At the higher speed mode		
Deviance	805.6	779.2	525.0
AIC	809.6	783.2	539.0

We now consider the numerical performances of these models. We calculate the deviance and AIC values to compare the three between-platoon headway models. The results are displayed in Table 3. It can be seen that the proposed distribution model (21) for between-platoon headways has a much lower values of deviance and AIC than the other two competitors, and hence it is a better choice than the lognormal distribution and shifted exponential distribution for both the lower and higher speed modes.

4.4. Platoon speed distributions

Finally, we turn to consider platoon speed distributions. In the literature, the normal distribution is used to approximate platoon speed distribution (see, e.g. Jiang et al., 2006). In Section 3.4, it was suggested that a normal mixture model with a sufficiently large number of components K be used to approximate the platoon speed distribution. Clearly, the normal mixture model includes the normal distribution used in Jiang et al. (2006) as its special case with $K = 1$.

To quantitatively evaluate the models, we calculate the deviance and AIC values based on the vehicle platoons obtained in Li (2016). In particular, we choose $K = 2, 3, 4$, and compare the normal mixture with different numbers of components to the normal approximation. The results are displayed in Table 4.

Table 4. The deviance and AIC for the normal distribution and the normal mixture for platoon speeds.

	Normal distribution	Normal mixture (2 components)	Normal mixture (3 components)	Normal mixture (4 components)
		At the lower speed mode		
Deviance	916.1	914.9	581.7	548.5
AIC	920.1	924.9	597.7	570.5
		At the higher speed mode		
Deviance	2733.6	2682.2	2232.8	1789.3
AIC	2737.6	2692.2	2248.8	1811.3

It can be seen from Table 4 that the normal mixture with two components and the normal approximation have similar performances in terms of deviance and AIC for the lower velocity mode, where their difference in AIC is less than 5. For the higher velocity mode, however, the normal mixture with two components performs better than the normal approximation.

When the number of the components is taken higher than 2, the normal mixture model greatly outperforms the normal approximation for both the lower and higher speed modes. This suggests that a normal mixture does a better job to accommodate the heteroscedasticity in platoon speeds due to the speed drifts within platoons.

5. Concluding remarks

In this paper, on the basis of the stochastic dynamic model for vehicle platoons developed in Li (2016), we have proposed several statistical distribution models for some important platoon characteristics, including platoon size, within-platoon headway, between-platoon headway and platoon speed.

These statistical distributions can be regarded as stand-alone models for platoon characteristics. We have explored the relationships between the proposed models and the existing models in the literature. It is shown that the statistical models investigated in this paper include some important existing models as special cases. It is therefore not surprising that the derived statistical distributions fitted the vehicle platoon data equally well or outperformed the existing models in the numerical analysis.

In practice, the developed statistical distribution models can help us better understand traffic platooning behavior. These statistical characteristics also provide crucial inputs into the traffic management and control algorithms for traffic with a platoon structure.

Finally, we would like to point out that the vehicle platoons used in the numerical analysis of this paper were recognized based on the vehicle classification algorithm in Li (2016). Consequently, the probability distributions arising from the developed model tend to provide a good fit to the platoon data; therefore care must be taken when interpreting the numerical results. More real traffic data are

needed in future research to test the proposed distribution models for the platoon characteristics. We note, however, given the fact that the derived distributions include some important existing distribution models as their special cases, the quality of fit of the proposed distribution models will not be worse than these existing models.

Acknowledgements

The author would like to thank the reviewers for their constructive comments that helped improve the quality of the earlier versions of the paper. The author also thanks Dr. Patikhom Cheevarunothai, Professor Yinhai Wang, and Professor Nancy Nihan for kindly providing the traffic data used in this paper.

Appendix A. Proofs of theorems

We provide the proofs of the Theorems in this appendix.

Proof of Theorem 1.

To show (i), we note that the function $p_{jj}(h)$ is strictly decreasing. From Chebyshev's integral inequality (see, e.g. Niculescu and Persson, 2006), we obtain

$$\int_{\tau}^{+\infty} h\, p_{jj}(h) g_0(h) dh < \int_{\tau}^{+\infty} p_{jj}(h) g_0(h) dh \int_{\tau}^{+\infty} h g_0(h) dh = (\bar{q}_{jj}/\theta) \int_{\tau}^{+\infty} h g_0(h) dh.$$

This indicates that $(\theta/\bar{q}_{jj}) \int_{\tau}^{+\infty} h\, p_{jj}(h) g_0(h) dh < \int_{\tau}^{+\infty} h g_0(h) dh$. In addition, $\int_{\tau}^{+\infty} h g_0(h) dh < \int_{\tau}^{+\infty} h g(h) dh$ is obvious by noting $\lambda_0 < \lambda_1$.

The proof of (ii) is trivial. We turn to (iii) and focus on $\alpha > 1$. We note that for any $h \geq \tau + (\alpha - 1)\lambda_0$, both $p_{jj}(h)$ and $g_0(h)$ are decreasing, and hence $f_{WP}(h; j)$ is a decreasing function of h. In addition, it is easy to verify that $df_{WP}(h; j)/dh$ has a unique root in $(\tau, \tau + (\alpha - 1)\lambda_0)$ and there exists a point $h^* \in (\tau, \tau + (\alpha - 1)\lambda_0)$ such that $p_{jj}(h) g_0(h)$ attains its maximum.

Proof of Theorem 2.

We focus on part (i) and show $\rho_{WP}(i) < \rho_{BP}(i, i)$; the other parts of theorem can be shown in a similar way as done in the proof of Theorem 1.

From equations (16) and (17), $\rho_{WP}(i) < \rho_{BP}(i,i)$ holds if and only if

$$\frac{\int_{\tau}^{+\infty} h p_{ii}(h) g_0(h) dh}{\int_{\tau}^{+\infty} p_{ii}(h) g_0(h) dh} < \frac{\int_{\tau}^{+\infty} h p_{ii}(h) g_1(h) dh}{\int_{\tau}^{+\infty} p_{ii}(h) g_1(h) dh}.$$

This can equivalently be written as

$$\int_{\tau}^{+\infty} p_{ii}(h) g_0(h) dh \int_{\tau}^{+\infty} h p_{ii}(h) g_1(h) dh > \int_{\tau}^{+\infty} p_{ii}(h) g_1(h) dh \int_{\tau}^{+\infty} h p_{ii}(h) g_0(h) dh.$$

We note $p_{ii}(h) g_0(h) / [p_{ii}(h) g_1(h)] = (\lambda_1/\lambda_0)^{\alpha} \exp\{-h[1/\lambda_0 - 1/\lambda_1]\}$ and $\lambda_1 > \lambda_0$. Hence, $p_{ii}(h) g_0(h) / [p_{ii}(h) g_1(h)]$ strictly decreasing. From Theorem 7 in Liu et al. (2009), we conclude that the above inequality holds.

Proof of Theorem 3.

We note that the function $p_{ji}(h)$ is strictly increasing. From Chebyshev's integral inequality (see, Niculescu and Persson, 2006), we obtain that

$$\int_{\tau}^{+\infty} h p_{ji}(h) g(h) dh > \int_{\tau}^{+\infty} p_{ji}(h) g(h) dh \int_{\tau}^{+\infty} h g(h) dh.$$

This indicates that $\rho_{BP}(j,i) = \int_{\tau}^{+\infty} h p_{ji}(h) g(h) dh \,/\, \int_{\tau}^{+\infty} p_{ji}(h) g(h) dh > \int_{\tau}^{+\infty} h g(h) dh.$

Proof of Theorem 4.

Clearly for any fixed b and as $a \to 0$, model (20) collapses to a gamma distribution $f_{WP}(h) = c(h-\tau)^{\alpha-1} \exp(-(h-\tau)/\lambda)$ which approaches to a normal distribution when α becomes large (see, e.g., Patel and Reed, 1996).

Appendix B. Computational issues on platoons' characteristics

In this appendix, we discuss the numerical evaluations of various numerical characteristics of vehicle platoons using a Monte Carlo method. Consider a random variable X that follows a mixture distribution $g(x)$ consisting of two components, $g_0(x)$ and $g_1(x)$, with a weight function $w(x) \geq 0$:

$$g(x) = c w(x) [\pi g_0(x) + (1-\pi) g_1(x)],$$

where the parameter π $(0 \leq \pi \leq 1)$ is the mixing probability. The parameter c is a normalizing constant. We wish to evaluate an integral of the following form for a given $A(x)$:

$$E[A(X)] = c\int_{-\infty}^{+\infty} A(x)w(x)[\pi g_0(x) + (1-\pi)g_1(x)]dx.$$

Analytically this may not be evaluated easily for some choice of the functional forms of $w(x)$, $g_0(x)$, $g_1(x)$, and $A(x)$. A quick approach is to use a Monte Carlo method:

$$E[A(X)] \approx c\sum_{k=1}^{K} A(x_k)w(x_k),$$

where x_k $(k = 1., \ldots, K)$ are independently drawn from $\pi g_0(x) + (1-\pi)g_1(x)$. K is a sufficiently large number. This is summarized in the following Algorithm.

Algorithm C.

Given: c, $\pi g_0(x) + (1-\pi)g_1(x)$, $w(x)$ and $A(x)$.

Set $E = 0$.

For $k = 1{:}\,K$

Draw q from the uniform distribution on $[0, 1]$.

If $q < \pi$, then draw x_k from $g_0(x)$; otherwise draw x_k from $g_1(x)$.

Calculate $E = E + A(x_k)w(x_k)$.

End of for loop

Return $E = cE/K$.

We can apply Algorithm C to calculate $\bar{q}_{ij}$ by setting

$$w(x) = \frac{a_{ij}(x-\tau)^{b_{ij}}}{1+\sum_{k\neq i} a_{kj}(x-\tau)^{b_{kj}}} \quad \text{and} \quad A(x) = 1,$$

with $c = \theta$ and $\pi = 1$ if $i \leq M$, and with $c = 1-\theta$ and $\pi = 0$ otherwise.

In addition, we can calculate the average within-platoon headway $\rho_{WP}(j)$ by setting

$$w(x) = \frac{1}{1+\sum_{k\neq i} a_{kj}(x-\tau)^{b_{kj}}}, \quad A(x) = x, \quad c = \theta/\bar{q}_{jj}, \text{ and } \pi = 1.$$

We can calculate the average between-platoon headway $\rho_{BP}(j, j)$ by setting

$$w(x) = \frac{1}{1+\sum_{k\neq i} a_{kj}(x-\tau)^{b_{kj}}}, \quad A(x) = x, \quad c = (1-\theta)/\bar{q}_{(j+M)j}, \text{ and } \pi = 0.$$

Finally, in order to calculate the average between-platoon headway $\rho_{BP}(j, i)$ $(i \neq j)$, we set

$$w(x) = \frac{a_{ji}(x-\tau)^{b_{ji}}}{1+\sum_{k\neq j} a_{ki}(x-\tau)^{b_{ki}}},\quad A(x) = x,\quad c = 1/(\bar{q}_{ji} + \bar{q}_{(j+M)i}) \text{ and } \pi = \theta.$$

References

Ashton, W.D., (1971). Distributions for gaps in road traffic. IMA Journal of Applied Mathematics 7(1), 37-46.

Branston, D., (1976). Models of single lane time headway distributions. Transportation Science 10(2), 125-148.

Breiman L., Gafarian A. V., Lawrence R. L., Murthy V. K., (1968). An experimental analysis of single-lane time headways in freely flowing traffic. 4th Int. Sympos. Traffic. Karlsruhe.

Buckley, D.J., (1962). Road traffic headway distributions. In Australian Road Research Board (ARRB) Conference, 1st, 1962, Canberra (Vol. 1, No. 1).

Buckley, D.J., (1968). A semi-poisson model of traffic flow. Transportation Science 2(2), 107-133.

Cowan, R. J., (1975). Useful headway models. Transportation Research 9(6), 371-375.

Dunne, M.C., Rothery, R.W., Potts, R.B., (1968). A discrete Markov model of vehicular traffic. Transportation Science 2(3), 233-251.

Firth, D. (1988). Multiplicative Errors: Log-Normal or Gamma? Journal of the Royal Statistical Society, Ser. B 2, 266-268.

Galin, D., (1980). A comparison of some bunch-size models for two-lane rural roads—an Israeli experience. Transportation Research Part A: General 14(1), 51-56.

Gartner, N., Messer, C. J., Rathi, A. K., (2001). Traffic Flow Theory: A State-of-the-Art Report. Transportation Research Board, Washington, DC.

Haight, F. A., Breuer, M. A., (1960). The Borel-Tanner distribution. Biometrika 47(1), 143-150.

Jiang, Y., Li, S., Shamo, D. E., (2006). A platoon-based traffic signal timing algorithm for major–minor intersection types. Transportation Research Part B: Methodological 40(7), 543-562.

Johnson, J. L., (2011). Probability and Statistics for Computer Science. New York: Wiley.

Li, B., (2016). Stochastic modeling for vehicle platoons (I): Dynamic grouping behavior and online platoon recognition. Submitted to Transportation Research Part B: Methodological.

Liu, W.-J., Ngô, Q.-A., Huy, V. N., (2009). Several interesting integral inequalities, J. Math. Inequal. 3 (2), 201–212.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., (2012). The BUGS book: A practical introduction to Bayesian analysis. London: Chapman & Hall/CRC.

McCullagh, P., Nelder, J., (1989). Generalized Linear Models, Second Edition. London: Chapman & Hall/CRC.

Miller, A.J., (1961). A queueing model for road traffic flow. Journal of the Royal Statistical Society. Series B (Methodological), 23 (1), 64-90.

Niculescu, C., Persson, L. E., (2006). Convex Functions and Their Applications: A Contemporary Approach. New York: Springer Science & Business Media.

Patel, J.K., Read, C.B., (1996). Handbook of the normal distribution. 2nd ed. London: Chapman & Hall/CRC.

Ramezani, H., Benekohal, R.F., Avrenli, K.A., (2010). Statistical distribution for inter platoon gaps, intra-platoon headways and platoon size using field data from highway bottlenecks. In Traffic Flow Theory and Characteristics Committee: Summer Meeting and Conference, Annecy, France.

Sun, D., Benekohal, R.F., (2005). Analysis of work zone gaps and rear-end collision probability. Journal of Transportation and Statistics 8(2), 71.

Tanner, J. C., (1961). A derivation of the Borel distribution. Biometrika 48(1), 222-224.

Wiens, B.L., (1999). When log-normal and gamma models give different results: a case study. The American Statistician 53(2), 89-93.