Loughborough
University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

A Model of Pedestrians' Intended Waiting Times

for Street Crossings at Signalized Intersections

Baibing Li

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

ABSTRACT

For the purposes of both traffic-light control and the design of roadway layouts, it is important to understand pedestrian street-crossing behavior because it is not only crucial for improving pedestrian safety but also helps to optimize vehicle flow. This paper explores the mechanism of pedestrian street crossings during the red-man phase of traffic light signals and proposes a model for pedestrians' waiting times at signalized intersections. We start from a simplified scenario for a particular pedestrian under specific traffic conditions. Then we take into account the interaction between vehicles and pedestrians via statistical unconditioning. We show that this in general leads to a U-shaped distribution of the pedestrians' intended waiting time. This U-shaped distribution characterizes the nature of pedestrian street-crossing behavior, showing that in general there are a large proportion of pedestrians who cross the street immediately after arriving at the crossing point, and a large proportion of pedestrians who are willing to wait for the entire red-man phase. The U-shaped distribution is shown to reduce to a J-shaped or L-shaped distribution for certain traffic scenarios. The proposed statistical model was applied to analyze real field data.

Keywords: Intended waiting time; Pedestrian street crossing; Signalized intersection; Vehicle time headway

1. Introduction

In the urban areas of most large cities around the world, there is intensive interaction between pedestrians and vehicles. It is hence important to understand pedestrian street-crossing behavior because properly designed and placed pedestrian facilities encourage pedestrians to follow traffic regulations and to cross streets at safe locations during safe periods. However, in most of the current traffic systems in metropolitan cities, pedestrians usually receive a lower priority than motorized vehicles. Traffic-signal control has normally sought to optimize vehicle flow, and pedestrians have been fitted around that flow (Ahuja et al., 2005; Tiwari et al., 2007; Ishaque and Noland, 2008). Recently, research into pedestrian behavior has become increasingly important because traffic accidents involving pedestrians have become common in many cities of emerging-economy countries due to their rapid motorization and urbanization. According to the WHO reports, the number of road traffic deaths is about one hundred thousand per year in India and China respectively, among which pedestrian fatalities account for a large proportion. Take Delhi, for example, the percentage of road traffic fatalities involving pedestrians has reached 40% to 50% of the total fatalities (Tiwari et al., 2007). In more developed countries, the total road traffic fatalities are low but as vulnerable road users, the percentages of traffic fatalities involving pedestrians are also high. For instance, there were a total of 1850 fatalities in road accidents in 2010 across the U.K., among which 21.9% involved pedestrian fatalities (DfT, 2012). In London, pedestrian-involved-fatality accounted for 46.0% in 2010, i.e., 58 out of a total of 126 road traffic fatalities (TfL, 2012).

In response to the rapid research advances, two comprehensive reviews of studies of pedestrian behavior in urban areas have recently been published, Ishaque and Noland (2008) and Papadimitriou et al. (2009). The two articles review existing studies and discuss various issues that are encountered in the modeling of pedestrian behavior including: (a) pedestrian micro-simulation methods; (b) pedestrian speeds at crossings and the associated behavior such

as gap-acceptance behavior and compliance rates; and (c) pedestrian speeds on pavements. It is also noteworthy that apart from the empirical studies reviewed in Ishaque and Noland (2008) and Papadimitriou et al. (2009), there is attention in the recent literature paid to the methodological research on pedestrian behavior, such as pedestrian walking behavior based on discrete choice (Antonini et al., 2006; Robin et al., 2009), pedestrian flow modeling (Huang et al., 2009), pedestrians route choice and activity scheduling via dynamic programming (Hoogendoorn and Bovy, 2004a, b), etc.

The literature review, however, suggests that although a lot of attention has been paid to the research on pedestrian's behavior in recent years, very little has been done on the methodological development of pedestrian waiting time models at signalized intersections. In recent years there have been only a few empirical studies on investigating pedestrian waiting times. For instance, Keegan and O'Mahony (2003) considered the impact on pedestrians' waiting times at a signalized crossing in Dublin when a new type of countdown timer was deployed. More recently Lipovac et al. (2012) have compared pedestrians' behavior with and without pedestrian countdown displays at two signalized pedestrian crossings in Bosnia and Herzegovina. Keegan and O'Mahony (2003) and Yang et al. (2006) used questionnaires/interviews to investigate the factors that influence pedestrian-crossing behavior during a red-man phase. In addition, Hamed (2001), Tiwari et al. (2007), and Wang et al., (2011) applied the Cox proportional hazard model to identify the factors that may lead to unsafe crossing in Jordan, India, and China respectively.

The purpose of this paper is to investigate the statistical modeling of pedestrians' intended waiting times at signalized intersections. Currently, the commonly used model of waiting time in queueing theory is the general-purpose exponential distribution (see, e.g. Bocharov et al., 2003). In traffic studies, the exponential distribution was also used to model pedestrians' waiting times for street crossings (e.g. Zhuang and Wu, 2011). In addition, in safety analysis

on pedestrians' crossings, survival analysis was often applied (e.g. Hamed 2001; Tiwari et al. 2007; Wang et al., 2011) where the exponential distribution and its extensions (such as Weibull distribution, gamma distribution, etc.) were explicitly or implicitly assumed. The exponential distribution and its extensions, however, are in general not suitable for the investigation of pedestrians' waiting times at signalized intersections because in theory the waiting time assumed in these models can be infinitely long. It is therefore unable to capture the characteristics of the *bounded* waiting time at signalized intersections (in particular the tail area closer to the upper bound of the red-man phase). This is shown in Figure 1 with a U-shaped distribution for the field data on intended waiting times prior to unsafe crossings during the red-man phase; these data were collected at a signalized intersection in a Chinese city, Kunming. A detailed data description and analysis will be provided in Section 4. Similar U-shaped patterns were also documented in the literature (e.g. Lipovac et al., 2012).

(Figure 1 is here)

Figure 1. Histogram (left) and empirical distribution function (right) of pedestrians' intended waiting times at an intersection in Kunming.

The data displayed in Figure 1 challenge the exponential distribution model, indicating that pedestrians' waiting times at signalized intersections can greatly differ from the exponential distribution and its extensions. This has some important practical implications. First, this shows that the exponential-distribution-based microscopic simulations for pedestrians' intended waiting times at signalized intersections are in general inadequate because the exponential distribution is in general unable to model pedestrian behavior in the red-man phase of traffic lights. Secondly, the U-shape of the distribution suggests that the arithmetic mean and standard deviation are in general not suitable summary statistics for measuring the centrality and variability of pedestrians' intended waiting times at signalized intersections. In addition, the

commonly used statistical testing methods such as the z-test and t-test may perform poorly because the data are in general far from being normally distributed. Lachenbruch (2002), for instance, discussed the difficulties in hypothesis testing for bounded outcomes, in particular when the data are L-shaped, J-shaped, or U-shaped. Lachenbruch showed that the efficiency of statistical inference is in general low when normality-based methods are used to analyze bounded outcomes.

This paper explores the mechanism of pedestrian street crossings during the red-man phase of traffic light signals and proposes a new distributional model for pedestrians' waiting times at signalized intersections. We investigate pedestrians' intended waiting times during the red-man phase of traffic lights, and show that the U-shaped distribution exhibited in Figure 1 is the outcome of the interaction between vehicles and pedestrians.

This paper is structured as follows. Section 2 discusses the modeling of pedestrians' intended waiting times at signalized intersections. In Section 3 we investigate the properties of the proposed model and statistical inference for the model. To illustrate this model, an empirical study is conducted in Section 4. Concluding remarks are offered in Section 5. Finally, the proofs of the theorems are given in the Appendix.

2. Modeling for pedestrian s' intended waiting times

In this section we first briefly review some general-purpose models for waiting time. Then we develop a statistical model for pedestrians' intended waiting times at signalized intersections. The statistical modeling process is undertaken in three steps. We start from a simplified scenario for a particular pedestrian under specific traffic conditions. We then use a generalized linear model to capture the relationship between the intended waiting time and the traffic conditions. Finally, we derive the distribution of interest by taking into account the interaction between vehicles and pedestrians.

2.1 General-purpose models for waiting time

Let random variable ξ denote the waiting time required for a service. The most commonly used model for waiting time is the general-purpose exponential distribution $\Pr\{\xi < w\} = 1 - \exp(-w/\gamma)$ with probability density function (p.d.f.) $f_{EXP}(w) = \gamma^{-1}\exp(-w/\gamma)$ $(w \geq 0)$, where $\gamma > 0$ is the average waiting time. This model is widely used in applications of queuing theory (see, e.g. Bocharov, 2003), and many other areas in operations research and statistics. Harris (1968), however, has found that in practice the average waiting time γ may be affected by a number of factors, and it varies as a gamma variate. Harris shows that a more suitable model for those problems is the following Pareto distribution:

$$f_P(w) = \begin{cases} \theta C^\theta w^{-(\theta+1)} & w \geq C \\ 0 & w < C \end{cases},$$

With parameters $\theta > 0$ and $C \geq 0$. The Pareto distribution is frequently used in applications where the waiting-time measurements have a heavier right-tail.

2.2 Model for bounded waiting time

The aforementioned general-purpose models for waiting time do not impose any restrictions on the duration of the waiting time. In some applications, however, the waiting time has an upper bound denoted by $C > 0$. To apply the Pareto distribution to model bounded waiting times, we must transform it so that the support of the waiting time is on the bounded period $[0, C]$. To do this, we define a mapping $\eta = f(\xi)$ from $[C, +\infty)$ to $[0, C]$: $\eta = C(1 - C/\xi)$. The transformed random variable η follows the distribution below:

$$f_{BP}(w; \theta, C) = (\theta/C)(1 - w/C)^{\theta-1} \qquad \text{for } 0 \leq w \leq C. \qquad (1)$$

Distribution (1) has a bounded support on $[0, C]$, and hence it is termed bounded Pareto distribution (denoted $BP(\theta, C)$ in this paper). Clearly the p.d.f. $f_{BP}(w; \theta, C)$ is strictly

decreasing (or increasing) if $\theta > 1$ (or $0 < \theta < 1$); and it is constant (i.e., a uniform distribution on $[0, C]$) if $\theta = 1$, as illustrated in Figure 2.

(Figure 2 is here)

Figure 2. Illustration of $BP(\theta, C)$ with $C = 80$ and $\theta = 4$ (real line), $\theta = 0.5$ (broken line), and $\theta = 1$ (dotted line).

For pedestrian crossings at signalized intersections, the pedestrians' intended waiting times are bounded by the duration of the red-man phase of traffic lights. This fact has an enormous impact on pedestrian waiting times. Distribution $BP(\theta, C)$ with $\theta > 1$ can thus be used to model the scenario where a pedestrian highly values his/her time and becomes more impatient as he/she waits longer. This type of pedestrian tends to ignore the red-man signal and to seek every opportunity to cross the street. Consequently, the probability that he/she waits decreases rapidly as time increases. The empirical study in Kaiser (1994) found that pedestrian impatience and risk-taking behavior increase after twenty seconds of delay. Keegan and O'Mahony (2003) show that the pedestrians in their sample who did not wait at the crossing overestimated the length of time they would have had to wait by 200% on average.

On the other hand, $BP(\theta, C)$ with $\theta < 1$ can be used to describe the scenario where a pedestrian tends to be law-abiding and not to risk his/her safety. Keegan and O'Mahony (2003) show that law-abiding pedestrians are aware that the light will soon change, making a safe crossing possible. Consequently, the longer they wait, the less likely that they will cross the street during the red-man phase. They are also more realistic about the waiting time: the law-abiding pedestrians in Keegan and O'Mahony (2003) overestimated the length of time they would have had to wait by only 69% on average, as opposed to 200% in the previous case.

Finally, the distribution $BP(\theta, C)$ with $\theta = 1$ can be used to model the scenario where a pedestrian's attitude to risk and time saving is neutral, and the probability that he/she crosses the street during the red-man phase is constant over the entire duration of the phase.

Although $BP(\theta, C)$ has the potential to model pedestrians' intended waiting times, it is in general not flexible enough to accommodate various traffic conditions and the heterogeneity of pedestrians. For instance, it is unable to capture the features exhibited in Figure 1.

2.3 Vehicle time headway

Pedestrian street-crossing behavior is in general the outcome of interaction between pedestrians and vehicles. Although many traffic parameters may affect pedestrian crossings such as the vehicular speed, the most important factor is the vehicle time headway that characterizes the gap between two consecutive vehicles because the vehicular speed is capped by a relatively low speed limit in urban areas, especially in city/town centers. The vehicle time headway is the elapsed time between the front of the lead vehicle passing a crossing point on the street and the front of the following vehicle passing the same point. Because pedestrians normally seek an appropriate gap between vehicles for street crossings, the vehicle time headway provides a measure of a pedestrian's opportunity to cross the street during the red-man phase.

We define the effective critical headway (ECH) to be the minimum vehicle time headway required by a pedestrian to cross safely. In other words, a pedestrian will cross a street only if the current vehicle time headway is greater than his/her ECH. Empirical studies have provided evidence on the relationship between pedestrian-crossing behavior and the gap between vehicles. In an early study, for instance, Cohen et al. (1955) found that 92% of pedestrians crossed the road when the available gap was 7 s; no one crossed when the gap was shorter than

1.5 s; and everyone crossed when the gap was 10.5 s or more. More empirical evidence has been provided by later research; see Ishaque and Noland (2008) and the references therein.

Now, let μ denote the ECH of a pedestrian and h the vehicle time headway. Following the previous analysis, the intended waiting time W of a pedestrian P for a given vehicle time headway h during the red-man phase is modeled as:

$$W|(h, P) \sim \begin{cases} f_{BP}(w; \theta, C) & if\ h > \mu \\ \delta(w - C) & if\ h \leq \mu \end{cases}, \tag{2}$$

where $\delta(w)$ is the Dirac delta function. The above model simply says that the pedestrian's intended waiting time W during the red-man phase follows $BP(\theta, C)$ if the vehicle time headway is greater than the ECH; otherwise the pedestrian has to be prepared to wait for up to the entire red-man phase. The mathematical expectation of the intended waiting time is:

$$E[W|(h, P)] = \begin{cases} C/(1 + \theta) & if\ h > \mu \\ C & if\ h \leq \mu \end{cases}. \tag{3}$$

We now model the relationship between the intended waiting time and the vehicle time headway. We first define the excess headway as the difference between the actual vehicle time headway and the ECH, i.e., $h - \mu$. As mentioned earlier, this quantity plays an important role in explaining the variability in waiting time: the higher the level of the excess headway, the more likely that the pedestrian will cross the street (Cohen et al., 1955). To establish a quantitative link between the excess headway and the intended waiting time, we note that $BP(\theta, C)$ belongs to the exponential distribution family (see, e.g., McCullagh and Nelder, 1989):

$$f_{BP}(w; \theta, C) = (1/C)\exp\{\theta log\left(1 - \frac{w}{C}\right) + \log(\theta) - log\left(1 - \frac{w}{C}\right)\} \quad for\ 0 \leq w \leq C,$$

with the mean function $\partial\{-\log(\theta)\}/\partial\ \theta = -1/\theta$. We thus use a generalized linear model to approximate the complicated relationship and relate the parameter θ to the attribute, the excess headway $h - \mu$, via the natural link function so that the variability of the waiting time can be explained using the excess headway: $\theta = \alpha + \beta \max(h - \mu, 0)$, where α is the intercept and β

is the coefficient of the attribute. In this paper, the coefficient β is termed the sensitivity coefficient for the excess headway $h - \mu$. Clearly, it takes only a positive value. Furthermore, from Eq. (3) we find that $\alpha = 0$ as $E[W|(h, P)] = C$ for $h \leq \mu$. Therefore, the above generalized linear model reduces to $\theta = \beta\max(h - \mu, 0)$. Model (2) is thus rewritten as

$$f_W(w|(h, P)) = \begin{cases} [\beta\max(h - \mu, 0)/C](1 - w/C)^{\beta\max(h-\mu,0)-1} & if\ h > \mu \\ \delta(w - C) & if\ h \leq \mu \end{cases} \quad (4)$$

for $0 \leq w \leq C$. The conditional expectation of the intended waiting time can be written as a unified form: $E[W|(h, P)] = C/\{1 + \beta \max(h - \mu, 0)\}$. As expected, it decreases as the excess headway increases. We also note that the model $\theta = \beta\max(h - \mu, 0)$ indicates that θ can be greater (or less) than unity, and hence the probability density (1) is strictly decreasing (increasing) if: (a) the sensitivity coefficient is large (small) for a given excess headway; or (b) the excess headway level is high (low) for a given pedestrian.

There are many sophisticated statistical models for vehicle time headway. In particular, Cowan (1975) proposed a widely used model for the vehicle time headway in which the vehicle headway H is assumed to include two components, $H = \tau + T$. The constant τ is the "tracking or following" component, representing the minimum time difference between the lead vehicle and the following vehicle passing the same point. The random variable T is the "free" component, assumed to follow an exponential distribution: $T \sim Exp(\lambda, \tau)$. The p.d.f. of the vehicle time headway in Cowan's model is

$$f_H(h) = p\delta(h - \tau) + (1 - p)\lambda^{-1}\exp\{-(h - \tau)/\lambda\} \qquad for\ h \geq \tau, \quad (5)$$

where p is the proportion of vehicles associated with the "tracking" component.

2.4   *Waiting time model for given pedestrian*

Now we turn to consider pedestrians. The existing studies in the literature classify pedestrians into categories; the pedestrians in different categories behave differently for street crossings during the red-man phase.

In this paper we follow this approach and classify pedestrians into two broad categories, risk averse and risk taking, according to whether or not the ECH is greater than the minimum headway. Risk-averse pedestrians have a higher average level of ECH, so they tend to wait until they are sure it is safe to cross. This category includes those who tend not to trade safety with time and those who have less mobility. Keegan and O'Mahony (2003) found that 20% of the pedestrians in their survey always waited. Hamed (2001) showed that a pedestrian's past involvement in a traffic accident seems to prevent him/her from accepting higher risk. Ahuja et al. (2005) noted that pedestrians are more law-abiding at traffic signals if they are accompanied by children or their mobility is impaired or they have heavy luggage. On the other hand, risk-taking pedestrians have a lower average level of ECH, so they tend to cross the street whenever possible. This category includes individuals who value their time highly and thus tend to take risks during street crossings. For instance, most commuters fall into this category. Hamed (2001) reported that pedestrians who frequently use a certain crossing and who live nearby are likely to reduce their waiting times by accepting higher risk. Younger and/or male pedestrians also tend to be risk-takers (Oxley et al., 1997; Hamed, 2001; Tiwari et al., 2007). Note that, as shown in Cohen et al. (1955), depending on the nature of traffic and in particular the minimum headway, a pedestrian can be risk-averse in one scenario and risk-taking in another. We assume that pedestrians from different categories also have different levels of the sensitivity coefficient β for a given level of excess headway: risk-taking pedestrians are more sensitive than risk-averse pedestrians and thus are more likely to cross the street unsafely.

Now we take into account both the traffic conditions and the pedestrian characteristics to model the intended waiting time. We first note that model (2) is a distributional model of the

waiting time, conditional on given time headway and given pedestrian type. Hence, we will employ a useful method in statistics, statistical unconditioning, to work out the unconditional distribution from conditional distribution (4) so that the resulting distribution applies to all time headways and pedestrians. Technically, we integrate the waiting-time distribution (2) over the distribution of headway distribution (5). This leads to the following waiting-time distribution:

$$f_W(w|P) = \int_\tau^{+\infty} f_W(w|(h,P))f_H(h)dh, \tag{6}$$

where P is an indicator of the pedestrian type with two nominal levels, RT (risk-taking) and RA (risk-averse). Let μ_{RT} and μ_{RA} denote the ECH of risk-taking and risk-averse pedestrians, and β_{RT} and β_{RA} the corresponding sensitivity coefficients for the excess headway. The following theorem gives a model of intended waiting time for risk-taking pedestrians.

Theorem 1. Let τ denote the minimum vehicle time headway. For risk-taking pedestrians with the ECH $\mu_{RT} \leq \tau$ and sensitivity coefficient β_{RT}, the cumulative distribution function (c.d.f.) of the intended waiting time is a mixture of two component distributions:

$$F_W(w|P = RT) = (1-p)F_1(w) + pF_2(w) \qquad \text{for } 0 \leq w \leq C,$$

where $F_2(w) = 1 - (1 - w/C)^{\beta_{RT}(\tau-\mu_{RT})}$, $F_1(w) = 1 - \left(1 - \frac{w}{C}\right)^{\beta_{RT}(\tau-\mu_{RT})}\{1 - \lambda\beta_{RT}\ln\left(1 - \frac{w}{C}\right)\}^{-1}$, and p is the proportion of vehicles associated with the "tracking" component and λ is the average headway of the "free" component in the headway model (5).

See the Appendix for proof. As it can be seen from the proof of Theorem 1, $F_2(w)$ and $F_1(w)$ in Theorem 1 describe the probabilities that risk-taking pedestrians cross the street when the headway is at the minimum level τ and when the gap is larger respectively.

Likewise, for risk-averse pedestrians, we can obtain:

Theorem 2. Let τ denote the minimum vehicle time headway. For risk-averse pedestrians with the ECH $\mu_{RA} > \tau$ and sensitivity coefficient β_{RA}, the cumulative distribution function (c.d.f.) of the intended waiting time is a mixture of two component distributions:

$$F_W(w|P = RA) = (1 - q)F_3(w) + qF_4(w) \qquad \text{for } 0 \le w \le C,$$

with $1 - q = (1 - p)\exp\{-(\mu_{RA} - \tau)/\lambda\}$, $F_3(w) = \left\{1 - \lambda\beta_{RA}\ln\left(1 - \frac{w}{C}\right)\right\}^{-1}$, and $F_4(w) = I([C, +\infty))$, where $I(S)$ is an indicator function for a set S that is equal to 1 if $w \in S$ and 0 otherwise. p and λ are the proportion of vehicles associated with the "tracking" component and the average headway of the "free" component in the headway model (5) respectively.

The derivation of equation in Theorem 2 is given in the Appendix. Theorem 2 shows that the first component of the distribution, $F_3(w)$, characterizes the probability that risk-averse pedestrians seek an opportunity to cross the street, whereas the second component, $F_4(w)$, describes the probability that risk-averse pedestrians are willing to wait for the entire red-man phase. Hence, for risk-averse pedestrians, $F_W(w|P = RA)$ is in general a J-shaped distribution.

2.5   *Model for intended waiting time*

In the previous subsection we investigated street-crossing behavior for different pedestrian types. We now consider a general population that consists of both risk-taking and risk-averse pedestrians. Let $\pi = \Pr\{risk\ taking\}$ denote the proportion of risk-taking pedestrians, and thus $1 - \pi = \Pr\{risk\ averse\}$ is the proportion of risk-averse pedestrians. From the total probability theorem, we obtain the following c.d.f. of intended waiting time that is a mixture of four component distributions:

$$F_W(w) = \Pr\{W < w\} = \pi F_W(w|P = RT) + (1 - \pi)F_W(w|P = RA)$$

$$= r_1 F_1(w) + r_2 F_2(w) + r_3 F_3(w) + r_4 F_4(w) \tag{7}$$

with $r_1 = \pi(1 - p)$, $r_2 = \pi p$, $r_3 = (1 - \pi)(1 - q)$, and $r_4 = (1 - \pi)q$.

In practice, the traffic conditions and the layout of intersections vary, and hence it is not unusual that one or more components are absent in model (7). For example, if risk-averse pedestrians are very cautious and μ_{RA} is relatively large compared to the minimum headway τ, then $q = 1 - (1 - p)\exp\{-(\mu_{RA} - \tau)/\lambda\}$ is close to 1. Consequently, the third component is absent and all risk-averse pedestrians tend to wait for the green-man signal.

We note that in the literature the classification of pedestrians varies from study to study. For instance, Keegan and O'Mahony (2003) classified pedestrians into two categories, "waiter" and "walker," according to whether or not they wait for the green-man signal. On the other hand, Liu et al. (2000) and Yang et al. (2006) classified pedestrians into "law-obeying" and "opportunistic". Clearly, the category of "waiter" or "law-obeying" corresponds to only the fourth component of model (7), and the category of "walker" or "opportunistic" includes the remaining three components.

On the basis of the previous analysis, we see that the four components of model (7) result from the interaction between pedestrians and vehicles, where each component describes the probability that a risk-taking or risk-averse pedestrian crosses the street during the red-man phase in the scenario of either the minimum headway or a larger gap, as illustrated by the following 2×2 classification table.

(Table 1 is here)

Hence, model (7) is able to differentiate different sub-types of "walker." This is of practical interest because different pedestrian types have different implications for pedestrian safety.

3. Model properties and statistical inference

3.1 *General class of model for pedestrians' waiting times*

The model developed in the previous section belongs to a more general class of model for bounded waiting times. To show this, we define a distribution family having a c.d.f. of $G(w; A, B, C)$ with parameters $A \geq 0$, $B \geq 0$:

$$G(w; A, B, C) = 1 - \left(1 - \frac{w}{C}\right)^A \left\{1 - B \ln\left(1 - \frac{w}{C}\right)\right\}^{-1} \quad \text{if} \quad w \in [0, C].$$

This distribution family includes the bounded Pareto distribution as a special case when $B = 0$. It is easy to show the following limiting cases for this distribution family.

Theorem 3. The c.d.f $G(w; A, B, C)$ approaches

(i) the degenerate distribution at $w = 0$ if either parameter A becomes sufficiently large but B is fixed at a given value, or B becomes sufficiently large but A is fixed at a given value;

(ii) the degenerate distribution at $w = C$ if A becomes sufficiently small and $B = 0$.

From part (ii) of Theorem 3, $G(w; A, 0, C)$ approaches the fourth component of (7) as parameter A becomes sufficiently small. Hence, the distribution of intended waiting time developed in the previous section can be written as a more general mixture of four component distributions that belong to the same distribution family:

$$F_W(w)$$
$$= r_1 G(w; A_{RT}, B_{RT}, C) + r_2 G(w; A_{RT}, 0, C) + r_3 G(w; 0, B_{RA}, C) + r_4 G(w; A_{RA}, 0, C), \quad (8)$$

where $A_{RT} = \beta_{RT}(\tau - \mu_{RT})$, $B_{RT} = \lambda \beta_{RT}$ and $B_{RA} = \lambda \beta_{RA}$. A_{RA} is a parameter to ensure the fourth component is a continuous distribution. These parameters satisfy the constraints $B_{RT} > B_{RA}$ and $A_{RT} > A_{RA}$.

Theorem 3 also leads to a couple of other important scenarios that are of interest in practice. First, from Theorem 3(i), when A_{RT} becomes large, the waiting-time distribution $F_W(w|P = RT)$ for risk-taking pedestrians reduces to the degenerate distribution at $w = 0$. In practice, this corresponds to the scenario where the minimum headway is large and almost all

risk-taking pedestrians cross the street immediately after arriving. Secondly, when B_{RT} becomes large but A_{RT} is fixed, the intended waiting-time distribution $F_W(w|P = RT)$ reduces to a mixture of the degenerate distribution at $w = 0$ and a bounded Pareto distribution $G(w; A_{RT}, 0, C)$. This corresponds to the scenario where the average vehicle time headway of the "free" component is large and/or risk-taking pedestrians are very sensitive to larger gaps between vehicles. In this case the pedestrians associated with the first component $F_1(w)$ tend to cross the street immediately after arriving. Consequently, $F_W(w|P = RT)$ is likely an L-shaped distribution.

3.2 Summary statistics

In practice, it usually helps to use a few numerical values to summarize the information contained in a distribution. For single-modal symmetric distributions such as normal distributions, the mean and standard deviation are the most commonly used summary statistics. For U-shaped distributions, however, these summary statistics could be misleading sometimes (see, e.g., Lachenbruch, 2002).

For the mixture distribution (8), we advocate providing summary information for each individual component because each component represents a specific group of pedestrians. We suggest using the median $w_{0.5}$ as a measure of centrality and the interquartile range $w_{0.75} - w_{0.25}$ as a measure of dispersion, where for a continuous c.d.f. $H(w)$, the quartile w_s is defined to be the value such that $H(w_s) = s$ for a value of $s \in (0,1)$. Both the median and the interquartile range are robust statistics and are particularly useful for skewed distributions.

We now consider the c.d.f. $G(w; A, B, C)$. To solve the equation $G(w_s; A, B, C) = s$ for w_s, let $y = \left(1 - \frac{w_s}{C}\right)^A$. Then we obtain

$$y = (1 - s)[1 - (B/A) \ln(y)]. \tag{9}$$

It is easy to show that (9) has a unique root $y_s \in [1-s, 1]$ that can be calculated using any simple numerical method. The quartile is thus given by $w_s = C(1 - y_s^{1/A})$.

Some special cases follow immediately. First, for the second component $G(w; A_{RT}, 0, C)$, it is straightforward to obtain the quartile $w_s^{(2)} = C[1 - (1-s)^{1/A_{RT}}]$. Likewise, for the fourth component we have $w_s^{(4)} = C[1 - (1-s)^{1/A_{RA}}]$. In particular, $w_s^{(4)} = C$ as A_{RA} becomes sufficiently small. For the third component $G(w; 0, B_{RA}, C)$ we can obtain $w_s^{(3)} = C\left[1 - \exp\left\{-\frac{s}{B_{RA}(1-s)}\right\}\right]$. The following theorem shows the relationships between the medians of the individual components.

Theorem 4. For the mixture distribution (8) with $A_{RT} > A_{RA}$ and $B_{RT} \geq B_{RA}$, the medians $w_{0.5}^{(i)}$ ($i=1,\ldots,4$) of the components satisfy: (i) $0 \leq w_{0.5}^{(1)} \leq w_{0.5}^{(2)} \leq w_{0.5}^{(4)} \leq C$; (ii) $0 \leq w_{0.5}^{(1)} \leq w_{0.5}^{(3)} \leq C$.

From Theorem 4 we have $w_{0.5}^{(1)} + w_{0.5}^{(2)} \leq w_{0.5}^{(3)} + w_{0.5}^{(4)}$. Hence, on average, risk-taking pedestrians have a smaller intended waiting time than do risk-averse pedestrians. We define the overall average waiting time to be $w_{0.5} = \sum_{i=1}^{4} r_i w_{0.5}^{(i)}$.

3.3 *Model estimation and selection*

We note that model (7) is a mixture of continuous and discrete components that can cause problems in statistical inference. Two simple solutions are available to circumvent the problems. First, we can replace (7) with the more general model, Eq. (8), for which all the components are continuous. This approach is useful for theoretical analysis.

An alternative approach is to follow a commonly used approach in data analysis and to use a discretized version of Eq. (7) (see, e.g., Zucchini and MacDonald, 2009, pp. 10). This is particularly preferable when the limiting cases occur and some components converge to a degenerate distribution. For instance, when B_{RT} is large and A_{RA} is small, the first component reduces to a degenerate distribution (see Theorem 3). We can thus use a discretized probability mass function for the intended waiting time:

$$\Pr(W = t_j)$$

$$= r_1 I(\{t_j = 0\}) + r_2 p(t_j; A_{RT}, 0, C) + r_3 p(t_j; 0, B_{RA}, C) + r_4 I(\{t_j = C\}) \quad (j{=}1,\dots,K),$$

$$(10)$$

where $p(t_j; A, B, C) = G(t_{j+1}; A, B, C) - G(t_j; A, B, C)$ is the probability mass function that is discretized from $G(w; A, B, C)$ with $t_{K+1} = +\infty$. $I(\{t = a\})$ is the indicator function at $t = a$, i.e., it is equal to 1 if $t = a$; and zero otherwise.

Usually parameter identifiability can be an issue for mixture distributions of discrete components. We note, however, the second and third components in model (10) are not discrete distributions but rather they are continuous distributions discretized at pre-specified intervals. The major difference between a genuine discrete distribution and the discretized continuous distributions lies in the fact that the latter depend on only a few parameters rather than a number of point mass probabilities. In addition, the constraints $B_{RT} > B_{RA}$ and $A_{RT} > A_{RA}$ can help address the issue of identification.

We next consider the data for statistical inference. The data collected usually contain a substantial number of censored values because the observation of intended waiting times is often interrupted by the green-man signal. Specifically, if this signal appears before a pedestrian crosses, the observation of the intended waiting time is considered censored (Tiwari et al., 2007), and the only information available is that the waiting time is longer than the observed duration. Hence, a random sample of n pedestrians includes n observations, and each

is characterized by a data pair $\{(w_j, c_j)\}$, where w_j is the observed *actual* waiting time of pedestrian j, defined to be the time difference between arriving at the crossing point during the red-man phase and leaving the curb. c_j is the corresponding indicator, with $c_j = 0$ if the time that pedestrian j is willing to wait is observed, and $c_j = 1$ if the observation is interrupted by the green-man signal.

Because of the nature of the mixture model, it is convenient to employ the EM algorithm for statistical inference. The EM algorithm has long been used in the literature of transport and traffic studies. An early application is the estimation of intersection origin-destination matrices (Nihan and Davis, 1989). Li (2005) investigated the estimation of general transport-network origin-destination matrices using the EM algorithm.

Specifically, consider a mixture model with m component distributions, $f(w) = \sum_{i=1}^{m} r_i f_i(w, \boldsymbol{\phi})$, where each component has a p.d.f. $f_i(w; \boldsymbol{\phi})$ and c.d.f. $F_i(w; \boldsymbol{\phi})$ ($i=1,\ldots,m$), and vector $\boldsymbol{\phi}$ includes all the parameters of the component distributions. Let $\mathbf{r} = [r_1, \ldots, r_m]^T$ be the vector of weights. For model (8), for instance, $m = 4$ and $\boldsymbol{\phi} = [A_{RT}, B_{RT}, A_{RA}, B_{RA}]^T$. We introduce the following *unobserved* data to represent which particular sub-population that each observation $\{(w_j, c_j)\}$ arose from: $\mathbf{z}_j = [z_{j1}, \ldots, z_{jm}]^T$, where z_{ji} is one (or zero) according as whether the corresponding waiting-time observation $\{(w_j, c_j)\}$ arose (or did not arise) from the ith component of the mixture ($i=1,\ldots,m$). The complete data $\mathbf{x}_j$ include both the observed data $\{(w_j, c_j)\}$ (also termed incomplete data) and the unobserved data $\mathbf{z}_j^T$, i.e. $\mathbf{x}_j = \{w_j, c_j, \mathbf{z}_j^T\}$ ($j=1,\ldots,n$). The complete-data log likelihood for $\boldsymbol{\phi}$ and $\mathbf{r}$ is:

$$logL(\boldsymbol{\phi}, \mathbf{r}) = \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ji} R_i(w_j, \boldsymbol{\phi}),$$

where $R_i(w_j; \boldsymbol{\phi}) = (1 - c_j)\log f_i(w_j; \boldsymbol{\phi}) + c_j \log[1 - F_i(w_j; \boldsymbol{\phi})]$.

The EM algorithm includes a number of iterations, each consisting of two steps, the Expectation step (E-step) and Maximum step (M-step). Following McLachlan and Krishnan

(2008), it is easy to show that the E-step simply requires replacing z_{ji} by their corresponding

conditional expectations, $z_{ji}^{(k)} = \frac{r_i^{(k)} R_i(w_j; \boldsymbol{\phi}^{(k)})}{\sum_{l=1}^{4} r_l^{(k)} R_l(w_j; \boldsymbol{\phi}^{(k)})}$, where $\boldsymbol{\phi}^{(k)}$ and $\mathbf{r}^{(k)}$ denote the values of

$\boldsymbol{\phi}$ and $\mathbf{r}$ at the kth iteration. We use a variant of the EM algorithm in the paper, ECM

(Expectation Conditional Maximization), in which the M-step includes a couple of conditional

maximization (CM) steps: each block of parameters, $\boldsymbol{\phi}$ or $\mathbf{r}$, is maximized individually,

conditionally on the other parameters remaining fixed. See McLachlan and Krishnan (2008) for

a detailed description of the EM algorithm and its extensions.

There are several methods that can be used to derive the standard errors of the estimated

parameters in the mixture model. One commonly used method is to approximate the standard

errors using the inverse of the observed information matrix related to the incomplete data

likelihood. This likelihood can derived directly using the mixture distribution $f(w) = \sum_{i=1}^{m} r_i f_i(w, \boldsymbol{\phi})$ based on the incomplete data $\{(w_j, c_j)\}$. There are also some other methods

available such as the supplemented EM algorithm or using a bootstrap approach. See, e.g.,

Zucchini and MacDonald (2009, Section 3.6), for a detailed discussion.

In practice, there may exist several competing models for which one or more components

are absent, and we want to choose one model among the others. Two commonly used criterions

for model selection are AIC and BIC. The BIC criterion penalizes models with a higher

number of parameters more severely than AIC. In addition, the BIC will select with probability

1 the true model (assuming it is in the class of models considered) as sample size becomes

large, while AIC will tend to choose more complex models (see, e.g. Hastie et al., 2009). We

use BIC for model selection in this paper.

4. Empirical study

To illustrate the proposed model, we return to the field data discussed in Section 1 and undertake an empirical analysis using the EM algorithm discussed in Section 3.

The intersection chosen for the analysis is located in a busy area of a Chinese metropolitan city, Kunming. It has four arms, with two major roads (Renmin Road East and Baita Road) crossing. Around the intersection are a theater, a number of small shops, and several residential areas. The pedestrian-crossing behavior was observed at a crossing point of the north arm of the intersection. The intersection was signalized with the standard three-phase cycle.

Data were collected on two consecutive weekend days from 10:00 a.m. to 12:00 noon and 2:00 p.m. to 4:00 p.m. During the observation period the total duration of the traffic-light cycle was 110 s, and the duration of the red-man phase was 75 s. For each pedestrian in the sample, we recorded the actual waiting time during the red-man phase, and whether the pedestrian crossed the street in the red-man phase or waited until the green signal showed. We also recorded other information about the pedestrians, including the gender and age group (young, middle-aged, and elderly). The separation into age groups was based on best-effort guesses on the part of the observer. In total, 283 valid observations were included in the following analysis. Table 2 displays the percentages of the pedestrians in the different categories.

(Table 2 is here)

Initially, we considered the general 4-component model (8), but we found that the 3-component model below would suffice since it had the same likelihood value of -567.3 but a smaller BIC of 1162.9 (as opposed to1174.1 for the 4-component model):

$$F_W(w) = r_1 G(w; A_{RT}, B_{RT}, C) + r_2 G(w; A_{RT}, 0, C) + r_4 G(w; A_{RA}, 0, C). \tag{11}$$

Two observations immediately follow from model (11). First, because the third component of model (8) is absent here, it suggests that the risk-averse pedestrians' ECH was large so that

almost all of the risk-averse pedestrians were willing to wait for the full red-man phase. This is in line with our observation during the data collection that the traffic was heavy and the vehicle time headway was relatively small. Second, during the data analysis it was observed that the estimate of parameter B_{RT} was very large and the estimate of parameter A_{RA} was very small, suggesting that the first and the last components in (11) converged to the degenerate distribution at $w = 0$ and the degenerate distribution at $w = C$ respectively. Hence, we considered a discretized 3-component model in the subsequent analysis:

$$\Pr(W = j) = r_1 I(j = 0) + r_2 p(j; A_{RT}, 0, C) + r_4 I(j = C) \quad (j=0,\dots,75). \tag{12}$$

The results are summarized in Table 3 (the first panel). It can be seen from Table 3 that about 13.8% (with standard error of 5.7%) of the pedestrians intended to cross the street immediately after arriving at the crossing point during the red-man phase. About 50.6% (with standard error of 6.2%) of the pedestrians were willing to wait for the entire phase. The remaining 35.6% (with standard error of 3.3%) of the pedestrians initially waited and then found an acceptable gap and crossed. The median intended waiting time for this pedestrian type was 28.8 s. The overall average intended waiting time including all three components was 48.2 s.

(Table 3 is here)

To assess the performance of the model in terms of fitting the data, we conducted a goodness-of-fit test. For this end, the duration of the red-man phase was split into seven sub-intervals as displayed in Figure 1, i.e. [0, 10), [10, 20), [20, 30), [30, 40), [50, 60), and [60, 75]. The observed frequencies O_i and the expected frequencies E_i ($i=1,\dots,7$) obtained using model (12) in the sub-intervals were compared:

$$X^2 = \sum_{i=1}^{7} (O_i - E_i)^2 / E_i.$$

The calculated test statistic was $X^2 = 3.51$. Asymptotically X^2 follows the chi-square distribution with the degrees of freedom 7-3-1=3, so the critical value at the 5% significance level is 7.81. Hence model (12) was not rejected at the 5% level.

Next, to investigate how different pedestrian types perceived the risk associated with street crossings, we undertook data analysis for each age group using model (12). The results are displayed in Table 3 (the second to fourth panels). It can be seen that:

(a) The estimate of r_4 increases with age. This suggests that senior pedestrians are more likely to wait for the entire red-man phase.

(b) The estimate of r_1 decreases with age. This suggests that more young pedestrians intended to cross the street immediately after arriving at the crossing point during the red-man phase.

(c) The overall average intended waiting time increased with age.

These findings suggest that young pedestrians tend to be risk-takers whereas senior pedestrians are more risk-averse. This is understandable because a pedestrian's decision on street crossings is a trade-off between safety and time-saving, subject to his/her mobility. Most senior pedestrians in Chinese metropolitan cities are retired, so their safety is usually more important than time saving. Moreover, elderly pedestrians are often less mobile, so they tend not to accept smaller gaps between vehicles.

In addition, the data analysis was also performed separately for male and female pedestrians, and the results are displayed in Table 3 (the last two panels). It can be seen that more females (52.7% with standard error of 8.9%) than males (46.9% with standard error of 8.6%) were willing to wait for the entire red-man phase. In addition, fewer females (11.5% with standard error of 8.8%) than males (15.7% with standard error of 7.7%) intended to cross the street immediately after arriving at the crossing point during the red-man phase. The overall

average intended waiting time of the females (54.7 s) was longer than that of the males (43.0 s), indicating that the crossing behavior of females was less aggressive.

The above analyses for different gender groups were not adjusted for the factor 'age', hence potentially the differences exhibited for different sexes could be due to the distribution for 'age', i.e., the more aggressive behavior of the male pedestrians might be explained purely in terms of 'age' if the females were older than the males. Referring to Table 2, however, there is no evidence that the female pedestrians in the data were older than the males.

5. Concluding remarks

This paper has investigated the statistical modeling of pedestrians' intended waiting times for street crossings at signalized intersections. In the literature, the exponential distribution is the most commonly used model for waiting time. For pedestrian crossings at signalized intersections, however, the pedestrians' intended waiting times are bounded by the duration of the red-man phase of the traffic lights. This fact has an enormous impact on pedestrians' waiting times because for risk-averse pedestrians, empirical studies show that the longer they wait, the less likely that they will cross the street during the red-man phase. Clearly the exponential distribution is unable to capture such a nature for pedestrian street crossings.

To accommodate the complex nature of the distribution of intended waiting times, the statistical modeling in this paper initially started with a simplified scenario for a particular pedestrian under specific traffic conditions. Then we took into account the interaction of vehicles and pedestrians via statistical unconditioning. This led to a mixture distribution with four components. When the weights of the first and fourth components are large, this distribution is U-shaped, showing that in general there are a large proportion of pedestrians who cross the street immediately after arriving at the crossing point, and a large proportion of pedestrians who are willing to wait for the entire red-man phase. The weights of the

components are in general determined by the traffic conditions and pedestrian types, so the general U-shaped distribution can sometimes reduce to a J-shaped or L-shaped distribution.

The proposed model can be used in simulation-based experiments to gain a better understanding of the interaction between vehicles and pedestrians, and to evaluate the impact of new traffic-light control schemes and roadway-layout redesigns on pedestrian-crossing behavior. The model also has important implications for statistical analyses of pedestrian crossings. Care must be taken in the analysis of waiting-time related problems because the distributions of intended waiting times are in general far from normal.

In this paper we have focused on univariate analysis in which only one variable, pedestrian waiting time, is explicitly modeled. In some risk analyses, a multivariate approach is required so that a number of risk factors can be taken into account simultaneously in the analyses. We note, however, most existing multivariate approaches are based on the assumption of normal distribution or exponential distribution (or its extensions), hence they are not suitable methods to analyze the problem here. A fruitful direction for future research would be to develop a multivariate modeling approach, on the basis of the model proposed in this paper, for pedestrians' waiting times at signalized intersections.

Acknowledgements

Appendix: proofs of theorems

To simplify the notation, the dependence of μ and β on the pedestrian type is suppressed in the proofs of Theorems 1 and 2 below.

Proof of Theorem 1. From equation (6) we obtain

$$f_W(w|P = RT)$$

$$= (1-p)\int_\tau^{+\infty} f_W(w|(h,P))\lambda^{-1}\exp\{-(h-\tau)/\lambda\}dh + p\int_\tau^{+\infty} f_W(w|(h,P))\delta(h-\tau)dh.$$

Let $f_1(w)$ and $f_2(w)$ denote the integrals of the first and second terms on the right-hand-side (RHS) of the above equation respectively. By the definition of function $\delta(h-\tau)$, it is straightforward to obtain $f_2(w) = [\frac{\beta(\tau-\mu)}{C}](1-w/C)^{\beta(\tau-\mu)-1}$, hence it is a bounded Pareto distribution, $BP(\beta(\tau-\mu), C)$. Now we consider $f_1(w)$. Noting $\mu \leq \tau$, it can be shown by some algebra that

$$f_1(w) = (\beta(\tau-\mu)/C)\left(1-\frac{w}{C}\right)^{\beta(\tau-\mu)-1}\{1-\lambda\beta\ln\left(1-\frac{w}{C}\right)\}^{-1}$$

$$+(\lambda\beta/C)\left(1-\frac{w}{C}\right)^{\beta(\tau-\mu)-1}\left\{1-\lambda\beta\ln\left(1-\frac{w}{C}\right)\right\}^{-2}.$$

The proof is completed by noting that the c.d.f. corresponding to the above density function is

$$F_1(w) = 1-\left(1-\frac{w}{C}\right)^{\beta(\tau-\mu)}\{1-\lambda\beta\ln\left(1-\frac{w}{C}\right)\}^{-1}.$$

Remark: Recalling headway distribution (5), we can see that $f_2(w)$ describes the probability that risk-taking pedestrians cross the street when the headway is at the minimum level τ, whereas $f_1(w)$ characterizes the probability that risk-taking pedestrians wait for a larger gap to cross during the red-man phase.

Proof of Theorem 2. From equation (6) we obtain

$$f_W(w|P = RA) = \int_\tau^\mu f_W(w|(h,P))[(1-p)\lambda^{-1}\exp\{-(h-\tau)/\lambda\} + p\delta(h-\tau)]dh$$

$$+ \int_\mu^{+\infty} f_W(w|(h,P))(1-p)\lambda^{-1}\exp\{-(h-\tau)/\lambda\}dh.$$

Note that from equation (4) we have $f_W(w|(h,P)) = \delta(w-C)$ for $h \leq \mu$. Hence, the first term of the RHS is the degenerate distribution at $w = C$ multiplied by a weight which can be shown to be $q = 1-(1-p)\exp\{-(\mu-\tau)/\lambda\}$.

On the other hand, by some algebra it can be shown that the second term of the RHS is

$$f_3(w) = (\lambda\beta/C)\left(1 - \frac{w}{C}\right)^{-1}\left\{1 - \lambda\beta\ln\left(1 - \frac{w}{C}\right)\right\}^{-2}$$

multiplied by a weight of $1 - q$. It is easy to verify that $f_3(w)$ is a probability density function

with the c.d.f. of $F_3(w) = \left\{1 - \lambda\beta\ln\left(1 - \frac{w}{C}\right)\right\}^{-1}$. This completes the proof.

Next, we focus on the proof of Theorem 4 because the proof of Theorem 3 is immediate.

Proof of Theorem 4. To show part (i), we note that $y_{0.5} \geq 0.5$ from (9). Hence $w_{0.5}^{(1)} =$

$C\left(1 - y_{0.5}^{1/A_{RT}}\right) \leq C\left(1 - (0.5)^{1/A_{RT}}\right) = w_{0.5}^{(2)}$. The result of $w_{0.5}^{(2)} \leq w_{0.5}^{(4)}$ is obvious from

$A_{RT} > A_{RA}$. To show part (ii), we note that from $B_{RT} \geq B_{RA}$ we have $(1 - 2y)/B_{RT} \geq$

$-1/B_{RA}$. Since the left-hand side is equal to $\left(\frac{1}{A_{RT}}\right)\ln(y)$ from (9), we obtain $\left(\frac{1}{A_{RT}}\right)\ln(y) \geq$

$-1/B_{RA}$. This completes the proof by exponenting both sides of the inequality.

References

Ahuja, S., Hao, X., MacDonald, M., Bell, M., Phull, S., 2008. Pedestrian crossing behaviour at signalised crossings. In: Proceedings of European Transport Conference, Leiden, The Netherlands.

Antonini, G., Bierlaire, M., Weber, M., 2006. Discrete choice models of pedestrian walking behavior. Transportation Research Part B 40 (8), 667-687.

Bocharov, P. P., D'Apice, C., Pechinkin, A. V., 2003. Queueing Theory. Boston: VSP.

Cohen, J., Dearnaley, E. J., Hansel, C. E. M., 1955. The risk taken in crossing a road. OR: Journal of the Operational Research Society 6 (3), 120-128.

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

DfT, 2012. Reported Road Causalities in Great Britain: 2010 Annual Report. https://www.gov.uk/government/publications.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statisctical Learning. New York: Springer.

Hamed, M. M., 2001. Analysis of pedestrians' behavior at pedestrian crossings. Safety science 38 (1), 63-82.

Harris, C. M., 1968. The Pareto distribution as a queue service discipline. Operations Research 16 (2), 307-313.

Hoogendoorn, S. P., Bovy, P. H. L., 2004a. Pedestrian route-choice and activity scheduling theory and models. Transportation Research Part B 38 (2), 169–190.

Hoogendoorn, S. P., Bovy, P. H. L., 2004b. Dynamic user-optimal assignment in continuous time and space. Transportation Research Part B 38 (7), 571-592.

Huang, L., Wong, S. C., Zhang, M., Shu, C.-W., Lam, W. H. K., 2009. Revisiting Hughes' dynamic continuum model for pedestrian flow and the development of an efficient solution algorithm. Transportation Research Part B 43 (1), 127-141.

Ishaque, M.M., Noland, R., 2008. Behavioural issues in pedestrian speed choice and street crossing behavior: a review. Transport Reviews 28 (1), 61-85.

Kaiser, S. H., 1994. Urban intersections that work: a new definition for level of service. Transport Research Record 1438. Washington DC: TRB National Research Council.

Keegan, O., O'Mahony, M., 2003. Modifying pedestrian behavior. Transportation Research Part A 37 (10), 889-901.

Lachenbruch, P. A., 2002. Analysis of data with excess zeros. Statistical Methods in Medical Research 11 (4), 297-302.

Li, B., 2005. Bayesian inference for origin-destination matrices of transport networks using the EM algorithm. Technometrics 47 (4), 399-408.

Lipovac, K., Vujanic, M., Maric, B., Nesic, M., 2012. Pedestrians' behavior at signalized pedestrian crossings. Journal of Transportation Engineering 139 (2), 165-172.

Liu, R., Cruz da Silva, J. P., da Maia Seco, A. J., 2000. A bi-modal micro-simulation tool for the assessment of pedestrian delays and traffic management. In: Proceedings of the ninth international association of travel behaviour research conference, Gold Coast, Australia.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. 2nd ed. London: Chapman & Hall/CRC.

McLachlan, G. J., Krishnan, T., 2008. The EM Algorithm and Extensions. 2nd ed. New York: Wiley.

Nihan, N. L., Davis, G. A., 1989. Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts. Transportation Science 23 (2), 77-90.

Oxley, J., Fildes, B., Ihsen, E., Charlton, J., Days, R., 1997. Differences in traffic judgements between young and old adult pedestrians. Accident Analysis and Prevention 29 (6), 839-847.

Paradimitriou, E., Yannis, G., Golias, J., 2009. A critical assessment of pedestrian behavior models. Transportation Research Part F 12 (3), 242-255.

Robin, T., Antonini, G., Bierlaire, M., Cruz, J., 2009. Specification, estimation and validation of a pedestrian walking behavior model. Transportation Research Part B 43 (1), 36-56.

TfL, 2012. Towards the year 2010: monitoring casualties in Greater London. http://www.tfl.gov.uk/corporate/projectsandschemes/23293.aspx.

Tiwari, G., Bangdiwala, S., Saraswat, A., Gaurav, S., 2007. Survival analysis: pedestrian risk exposure at signalized intersections. Transportation Research Part F 10 (2), 77-89.

Wang, W., Guo, H., Gao, Z., Bubb, H., 2011. Individual differences of pedestrian behaviour in midblock crosswalk and intersection. International Journal of Crashworthiness 16 (1), 1-9.

Yang, J., Deng, W., Wang, J., Li, Q., Wang, Z., 2006. Modeling pedestian' road crossing behavior in traffic systems micro-simulation in China. Transportation Research Part A 40 (3), 280-290.

Zhuang, X., Wu, C., 2011. Pedestrians' crossing behaviors and safety at unmarked roadway in China. Accident Analysis and Prevention 43 (6), 1927-1936.

Zucchini, W., MacDonald, I. L., 2009. Hidden Markov Models for Time Series. London: Chapman & Hall/CRC.

Table 1. Pedestrians' crossing probability distributions

for different pedestrian types and traffic conditions

Pedestrian type	Traffic condition	
	Minimum headway	Larger headway
Risk taking	$F_2(w)$	$F_1(w)$
Risk averse	$F_4(w)$	$F_3(w)$

Table 2. Percentages of the pedestrians in different age and gender groups

	Young	Middle-aged	Elderly
Male	15.9%	25.8%	11.0%
Female	19.1%	20.4%	7.8%

Table 3. Parameter estimates of the three-component pedestrian waiting-time model (12)

Parameter	All age/gender groups	(average waiting time=48.2 s)		
	A_{RT}	r_1	r_2	r_4
Estimate	1.429	0.138	0.356	0.506
Standard error	0.155	0.057	0.033	0.062

Parameter	Young pedestrians	(average waiting time =44.3 s)		
	A_{RT}	r_1	r_2	r_4
Estimate	1.294	0.196	0.364	0.440
Standard error	0.216	0.103	0.054	0.108

Parameter	Middle-aged pedestrians	(average waiting time =47.5 s)		
	A_{RT}	r_1	r_2	r_4
Estimate	1.357	0.122	0.407	0.471
Standard error	0.218	0.083	0.055	0.090

Parameter	Elderly pedestrians	(average waiting time =57.1 s)		
	A_{RT}	r_1	r_2	r_4
Estimate	2.018	0.069	0.240	0.691
Standard error	0.613	0.128	0.063	0.139

Parameter	Male pedestrians	(average waiting time =43.0 s)		
	A_{RT}	r_1	r_2	r_4
Estimate	2.112	0.157	0.374	0.469
Standard error	0.278	0.077	0.044	0.086

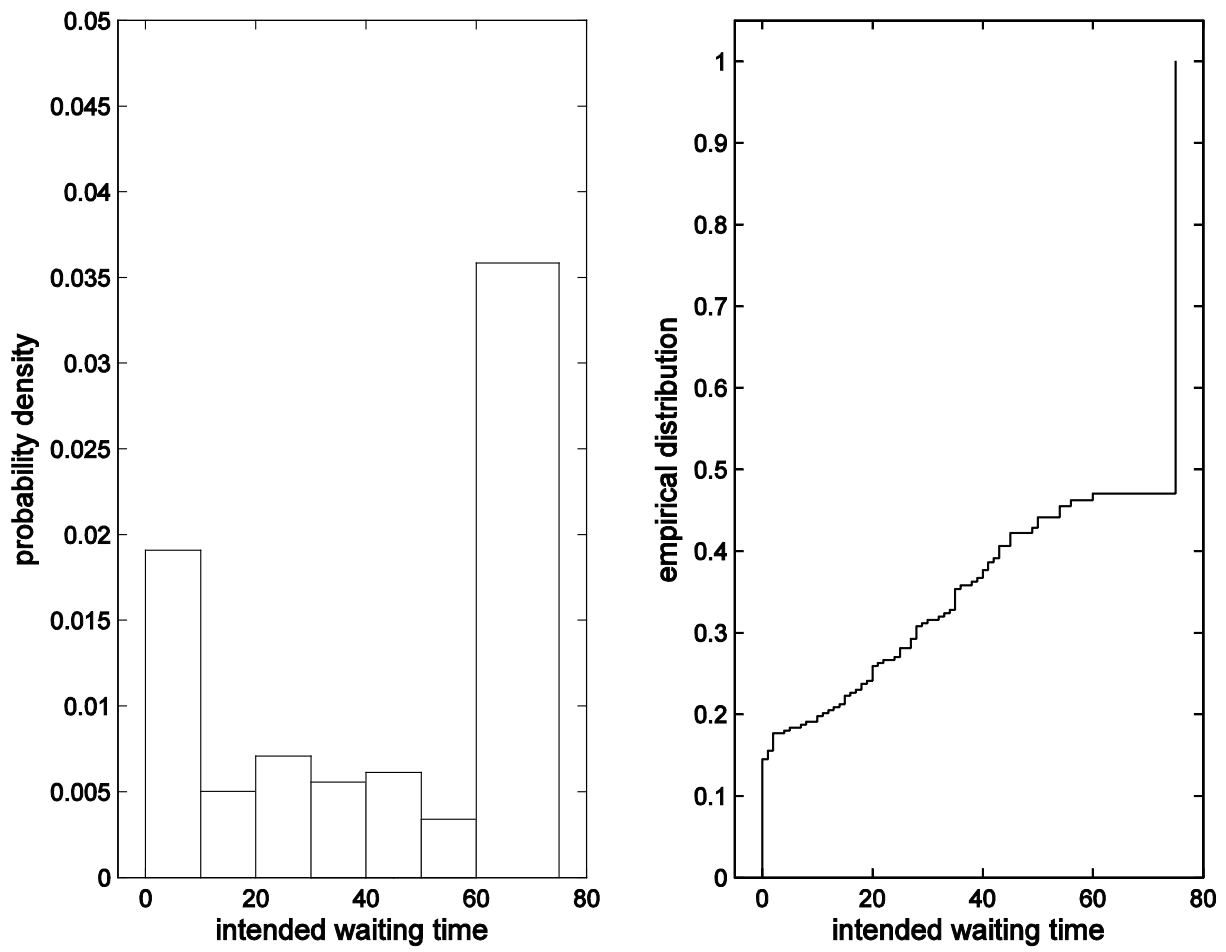| Parameter	Female pedestrians	(average waiting time =54.7 s)		
	A_{RT}	r_1	r_2	r_4
Estimate	0.832	0.115	0.358	0.527
Standard error	0.152	0.088	0.055	0.089

Figure 1. Histogram (left) and empirical distribution function (right) of pedestrians'

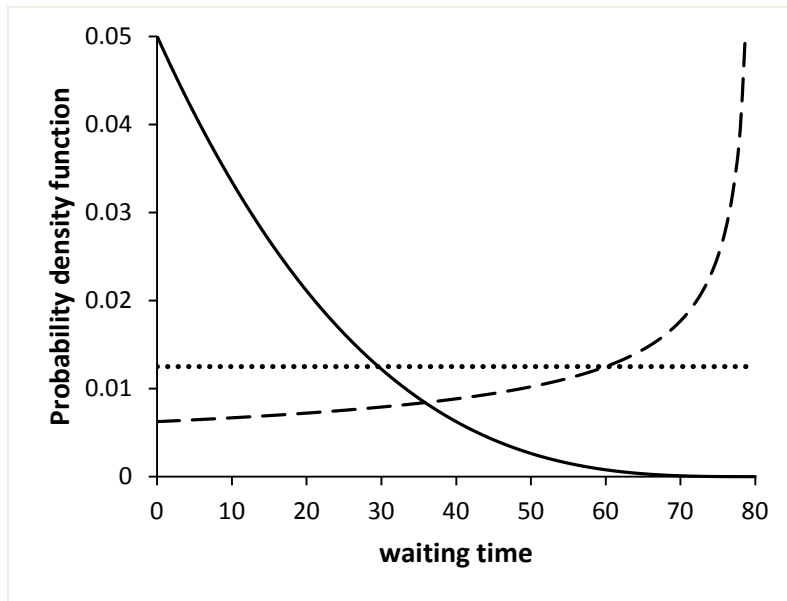intended waiting times at an intersection in Kunming.

Figure 2. Illustration of $BP(\theta, C)$ with $C = 80$ and $\theta = 4$ (real line), $\theta = 0.5$ (broken line),

and $\theta = 1$ (dotted line).