Safety Science 64 (2014) 155–162

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes

Lai Zheng [a,*,1], Karim Ismail [b], Xianghai Meng [a]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, 202 Haihe Road, Nangang District, Harbin, Heilongjiang, China
[b] Department of Civil and Environmental Engineering, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 13 July 2013
Received in revised form 5 November 2013
Accepted 6 December 2013
Available online 14 January 2014

Keywords:
Shifted Gamma-GPD model
Safety continuum
Bayesian approach
Crash estimation

ABSTRACT

Safety continuum plays an important role in the development of traffic conflict techniques. This study proposes a shifted Gamma-Generalized Pareto Distribution (GPD) model to map the whole safety continuum and then estimate crashes. Two important model parameters, the threshold and shifted value, are discussed in detail. The threshold, which is mapped as the boundary to distinguish conflicts and normal events, is estimated simultaneously with other four Gamma-GPD parameters by Bayesian approach. The shifted value, which is introduced by shifted reciprocal mapping and mapped as the boundary to distinguish conflicts and crashes, is determined by a crash-based approach. The proposed model is applied to estimate crashes related to lane change maneuvers on freeways, and the Bayesian approach is also compared with the classical maximum likelihood estimation approach. More accurate and less uncertain estimated crashes are obtained through the Bayesian approach, and this also shows the superiority of shifted reciprocal mapping approach over the linear mapping approaches. Meanwhile, the estimated model parameters show that the boundary to distinguish conflicts and crashes is consistent while the boundaries to distinguish conflicts and normal events might be varied across different segments.

1. Introduction

In the development of traffic conflict techniques, the safety continuum plays an important role in extending road safety analysis from crashes to serious conflicts and from serious conflicts to normal traffic events. Different from the use of crashes and serious conflicts, the primary focus of which is exceptional and unsuccessful events, analyzing the whole safety continuum provides "a complete picture" of road safety situations and can improve the understanding of mechanisms leading to crashes (Svensson and Hydén, 2006). As stated by Archer (2005), the safety continuum of traffic events is:

> *Theoretical concept inferred in relation to the use of proximal safety indicators whereby all interactions are placed on the same scale with safe passages at one extreme and accidents involving fatalities at the other.*

In order to describe the safety continuum, many hierarchy models have been proposed to represent the relationship between frequency and severity of the traffic events, and examples include distribution function in terms of nearness to a collision (Glauz and Migletz, 1980), safety pyramid model (Hydén, 1987), and diamond-shaped severity hierarchy model (Svensson, 1998). In these models, traffic events are ordered from the safest to most dangerous as: normal events (or undisturbed passages), potential conflicts, slight conflicts, serious conflicts, and crashes. Many studies suggested that if a stable relationship between levels of safety hierarchy could be established then the observations of non-crash events would be used to estimate crashes, and the relationship usually takes the form (Hauer and Garder, 1986):

$$\lambda = \sum \pi_i \cdot c_i \qquad (1)$$

where λ is the expected number of crashes; π_i is the crash-to-conflict ratio for conflicts of severity level i; c_i is the number of conflicts of severity level i.

Although Eq. (1) provides a theoretical framework, the application of this idea is limited because the required stability of crash-to-conflict ratio is difficult to ensure. Moreover, some thresholds to distinguish different severity levels are difficult to be objectively determined. To overcome these limitations, Songchitruksa and Tarko (2006) introduced the extreme value theory (EVT) for road safety estimation. The EVT provides a single dimension to measure the severity of traffic events, which fits within the safety continuum framework and abandons the assumption of fixed crash-to-conflict ratio. However, the proposed generalized extreme value distribution mainly focused on the extreme events and the threshold issue was still not well investigated.

---

[*] Corresponding author. Tel.: +1 613 796 8488.
E-mail addresses: Lai.Zheng@carleton.ca, zhenglai@hit.edu.cn (L. Zheng).
[1] Now at Carleton University, Canada as a visiting student.

0925-7535/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ssci.2013.12.003

In this study, a shifted Gamma-Generalized Pareto Distribution (hereinafter, shifted Gamma-GPD) is proposed to map the whole safety continuum. According to the authors' knowledge, it is the first attempt that different severity levels of events are incorporated into a unique distribution, and the crashes estimated directly from the distribution rather than relying on the correlation between crashes and event counts. During the modeling process, the subjectivity of threshold determination is also eliminated.

2. Theoretical framework

2.1. Generalized Pareto distribution

Let $X_1, X_2, \ldots, X_n$ are independently and identically distributed random variables with unknown distribution function $F(x) = \Pr(X_i \leqslant x)$, then the distribution function of X over a threshold u is:

$$F_u(x) = \Pr(X - u \leqslant x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{2}$$

Pickands (1975) showed that, for sufficiently high threshold u, the conditional distribution function $F_u(x)$ could be approximated by a generalized Pareto distribution (GPD), and the form is as follows:

$$G(x; u, \sigma, \xi) = 1 - \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-1/\xi} \tag{3}$$

where u is the predetermined threshold; $\sigma > 0$ is the scale parameter; $-\infty < \xi < \infty$ is the shape parameter; the right hand is taken to be $1 - \exp(-(x-u)/\sigma)$ when $\xi = 0$.

2.2. Model extension and the safety implication

The application of extreme value distributions to road safety analysis is based on safety continuum. This continuum is usually mapped onto the continuum of separation time and/or space during an encounter ("a simultaneous arrival in a certain limited area", Laureshyn et al., 2010) between two road users. Taking the post encroachment time (PET) as an example, the smaller the PET the more potential the event ends up with a crash, and a crash happens when the PET $\leqslant 0$.

Since the GPD distribution fits the observations over a threshold, it needs to map the condition for crash occurring from the left-hand tail (i.e., PET $\leqslant 0$) to the right-hand tail. Negated mapping, which transforms the PET to negated PET as used in the study of Songchitruksa and Tarko (2006), is the most direct way. As shown in Fig. 1(a), after the mapping, negated PET $\geqslant 0$ indicates a crash occurs. However, an issue on this linear transformation is that severity differences between different PET levels are considered as the same. This is somewhat inappropriate because the severity differences should be more sensitive with the decrease of the PET, especially when the PET is less than a critical value which distinguishes the normal events and the conflicts. In this context, a non-linear mapping approach which can amplify the differences between small PET values is needed. There are many non-linear transformation forms (e.g., reciprocal, exponential, and logarithmic form), and the reciprocal form is used in this study because it can significantly enlarge the difference between small PET values than others. Corresponding to the reciprocal form, the shifted reciprocal mapping approach is proposed (see Fig. 2(b)). The PETs at first are shifted by a value ($\delta > 0$) before the reciprocal mapping, and then $1/(\text{PET} + \delta) \geqslant 1/\delta$, which is PET $\leqslant 0$, indicates a crash occurs.

With the shifted reciprocal mapping approach, the risk of collision as well as the estimated crashes can be obtained. The risk of collision R is defined as the probability of observing an event with the shifted reciprocal PET equals to or is greater than $1/\delta$, that is:

$$R = \Pr\left(Z \geqslant \frac{1}{\delta}\right) = 1 - G\left(\frac{1}{\delta}\right) = \left(1 + \frac{\xi}{\sigma}\left(\frac{1}{\delta} - u\right)\right)^{-1/\xi} \tag{4}$$

where Z is the shifted reciprocal PET; $G(\cdot)$ is the generalized Pareto distribution. Therefore, for a long period of T, the estimated crashes based on the observation period t is:

$$C_T = \frac{T}{t} \cdot R \tag{5}$$

where C_T is the estimated number of crashes for the period of T.

2.3. Model parameters interpretation

It is noted that a new parameter shifted value δ is introduced besides the GPD parameter threshold u, scale parameter σ, and shape parameter ξ. Of these 4 parameters, threshold u and shifted value δ are of importance, not only for the model estimation but also for their safety implications. Recalling the diamond shaped safety hierarchy, as shown in Fig. 2, the full distribution of PETs embodies the same information. The threshold u is mapped as the boundary to distinguish conflicts and normal events, and the conflicts are then taken as extremes to be fitted with GPD. $1/\delta$, which defines the tail region of GPD, is the boundary to distinguish crashes and conflicts. The determination of threshold u and shifted value δ will be discussed in next section.

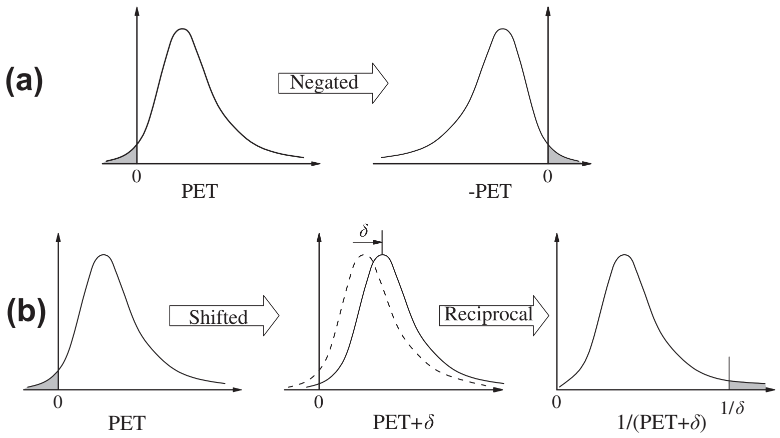

Fig. 1. Mapping approach: (a) negated mapping; (b) shifted reciprocal mapping.