APPLICATION OF THE LOMAX DISTRIBUTION TO ESTIMATE THE CONDITIONAL PROBABILITY OF CRASH

Andrew Tarko and
Purdue University
Lyles School of Civil Engineering, 550 Stadium Mall Drive, West Lafayette, IN 47907
Phone: + 1 (765) 413 1946 E-mail: tarko@purdue.edu

1. INTRODUCTION

Although crash frequency and severity are conventional and most direct measures of road safety, crash data have limitations that include low quality, lack of reliable insight into the events leading to the crash, and long data collection times (Tarko et al., 2009). Valid and practical surrogate measures for modeling and estimating safety are needed to allow preventive improvements of safety before human lives and health are lost.

To be applicable in transportation safety, surrogate events should be observable and logically connected with crashes. The introduction of new technologies for detecting and tracking road users opened new opportunities for quantifying and measuring road user’s motion leading to more precise measurements of non-crash events. More precise data collection methods might help validate surrogate measures by converting non-crash events into a corresponding crash frequency or even severity.

This paper presents the recent development in estimating the probability of crash conditional on traffic conflict. This development increases the traffic method practicality and also provides the theoretical and statistical bridge between conflicts and crashes. A case of study is introduced for road departures studied in a driving simulator to demonstrate the applicability of the method. Application of traffic conflicts at these facilities will provide a proactive approach to safety estimation in driving simulators, naturalistic driving, site observations, and when autonomous vehicles are in the mix of traffic.

2. THE METHOD

Spatial or temporal proximity of crash, or separation between road users, is applied for identification of traffic conflicts. In this case conflict is defined as “…an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged.” (Amundsen & Hyden, 1977). Various separation measures reported in the literature include Time to Collision (TTC), post-encroachment time (PET), Time to Accident (TA), range rate and other (Zheng et al., 2014; Babu et al. 2017). In the presented here study, an interaction between road users is considered a traffic conflict if the separation between road users is too small to be acceptable while the speeds warrant a serious outcome.

Glauz and Migletz (1980) proposed to connect conflicts and crashes with a tail of a suitable distribution of separations between road users during these events. The authors did not develop this concept into an estimation method. Campbell et al. (1996) in their unpublished report and then Songchitrunka and Tarko (AAP 2004) applied the extreme value theory (EVT) to traffic conflicts to estimate the probability of crash. Detailed reports of the latter study (Songchitrunka, 2004; Songchitrunka and Tarko, 2006) described sampling schemes of the extreme value applied to model the probability of right-angle crash. They recommended further study to improve the estimation efficiency. Tarko and Songchitrunka (2005) and Tarko (2012) continued this line of research by applying the Generalized Pareto (GP) distribution.
This study proposes the Lomax distribution, which is a special case of the GP with the zero shift parameter. The Lomax distribution is estimated with exceedances $x$ calculated from the observed separations between road users $s$:

$$x = s_c - s$$

where $x$ is the exceedance over threshold separation $s_c$ and corresponding to separation $s$. The exceedance $0 < x < s_c$ indicates a traffic conflict while its zero or negative value (not measurable) corresponds to a crash.

The Lomax distribution, as any GP, is the Gamma mix of exponential distributions. The relevance of this statistical representation of a traffic conflict and the corresponding crash is discussed in another work (Tarko, 2018). This relevance is supported here based on assumption of a traffic conflict caused by human or other failure leading to initial unawareness of the danger with the following additional elements:

1. Heyward’s definition of a traffic conflict with separation $s$ observed during the conflict smaller than threshold value $s_c$. The threshold value $s_c$ is sufficiently short to claim a conflict,
2. Time-invariant response to the conflict during the initial unawareness of the conflict (exponential distribution),
3. Heterogeneity of the response represented with the Gamma-distributed response rates.

The probability of a crash given a conflict $P(C|N)$ is calculated as:

$$P(C|N) = (1 + \theta s_c)^{-k}$$

where $\theta$ and $k$ are parameters of the Lomax distribution estimated with the observed exceedances $x$. The maximum-likelihood estimation method (MLE) and the probability-weighted moment-based estimation method (PWM) were applied (Greenwood et al., 1979). A new method called Single-Parameter Estimate (SPE) method is proposed to overcome certain weaknesses of the existing methods (Tarko, 2018).

The theory implies that conflicts claimed based on the violation of the threshold separation should yield an unbiased estimate of the expected number of crashes during the observation period regardless of the threshold value as long as this value is sufficiently small. This property is used to identify the proper threshold separation which is the largest among those yielding statistically equal estimates of the expected number crash.

3. DATA COLLECTION

Two female and two male undergraduate and graduate students at Purdue University between the ages of 22 and 24 subjects were driving in a driving simulator 19 times along the same road. The simulated road was 26.7 miles long with four 12-ft lanes wide divided by a 26-ft wide median. The shoulders were deliberately narrow (4 ft) and without guardrails in order to add difficulty to staying in the traveled way (see Figure 1). The position of the car and other data were measured at 13.7 Hz frequency. The driving schedule in afternoons before going home set a realistic commute-type situation that encouraged the subjects to drive reasonably fast to save time and leave for home earlier. There was additional monetary and time loss for casing a crash.

The lateral distance of the lead tire to the edge of a traveled way (lateral clearance) was used as a measure of the nearness to road departure. The traveled way included the traffic lanes and a well-maintained paved shoulder. In the driving simulation discussed, a departure occurred when the lead tire crossed the outer edge of the 4-ft shoulder. Subject were instructed to driving only in the right lane; only departures to the right were possible. A near-departure period started when the lateral clearance became
smaller than the threshold, and the near-departure period ended when the driver regained the lateral clearance above the threshold. The smallest nearness to the departure during the event represented the event. Figure 2 shows the number of near-departures in each of the 19 simulation runs and for four thresholds: 1, 2, 3, and 4 ft.

4. RESULTS

The Lomax-based analysis was applied to the lateral clearances divided into eight sequences of runs (two sequences for each of the four subjects). The single-parameter estimation method (SPE) was applied to estimate the $k$ parameter with an assumed $\theta = 1/s_c$, where $s_c$ is the threshold clearance. Figure 3 presents the estimates of the expected number of departures for the various thresholds standardized into departure rates per 100,000 miles. All the results exhibit the patterns implied by the theory. The expected number of crashes estimated based on large threshold clearances reduced gradually when the threshold was reduced. The declining trend flattened and became a subject of random fluctuation when the threshold clearances were small. It is easy to identify the largest threshold that could comfortably be assumed suitable. The suitable thresholds are marked with filled circles and the corresponding number of conflict are highlighted with a bold font. The estimates of the expected number of crashes based on these thresholds are shown in Table 1.

![Figure 1](image1.png)  
**Figure 1** Example renderings of the simulated road with the simulator

![Figure 2](image2.png)  
**Figure 2** Number of conflicts in individual runs of an example driver 385 (threshold clearances: 1, 2, 3, and 4 ft.)

![Figure 3](image3a.png)  
(a) Figure 3 The expected departure rates for driver 385: (a) runs 1-9, (b) runs 10-19.
The safety-related performance of Subject 385 during the first and second sequence of runs (1-9 and 10-19 runs) might be considered comparable due to the similar numbers of encroachments into the right-hand shoulder. One would classify these encroachments as near-departures and compared the average number of shoulder encroachments per run: 244/9=27.1 vs. 250/10=25.0 to conclude that the rates are quite similar with a weak indication that the performance might be slightly better in the second sequence of runs.

The applied Lomax-based analysis lead to a quite different conclusion. The driver improved its safety performance remarkably. First of all, the assumed threshold of 4 feet was clearly found as too large. The obvious decreasing trend of the expected number of crash estimates flattened around 3.7 ft in the first sequence of runs and around 3.2 ft in the second sequence. This indicates the growing confidence but also increasing skills of the driver. This improvement is particularly obvious when comparing the probabilities of departure conditional on near-departure event. This probability became much smaller in the second sequence of runs. It reduced the expected departure rate by factor 28. Subject 385, who did not have previous experience with the road and the simulator, turned out to be a very efficient learner.

### Table 1

Estimated expected number and rates of departures for the eight sequences of runs (two sequences of runs for each of the four subjects).

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Runs</th>
<th>Threshold $s_c$ (ft)</th>
<th>Conflicts $Q_N$ (events/all runs)</th>
<th>Departures $Q_C$ (events/all runs)</th>
<th>Departure Rate $R_C$ (events/100,000 mi)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Expected</td>
<td>$P_{10}^*$</td>
<td>$P_{90}^*$</td>
<td>Expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$P_{10}^*$</td>
</tr>
<tr>
<td>385</td>
<td>1-10</td>
<td>3.7</td>
<td>134</td>
<td>0.04428</td>
<td>0.01732</td>
</tr>
<tr>
<td>385</td>
<td>11-19</td>
<td>3.2</td>
<td>34</td>
<td>0.00141</td>
<td>0.00022</td>
</tr>
</tbody>
</table>

Note: The 80-percent confidence interval estimated with the MATLAB `bootstrap` and `prctile` based on 500,000 bootstrap data samples (MATLAB, 2016)

5. **CONCLUSION**

The analysis of the traffic conflicts data produced results that closely followed the anticipated trends prompted by the theory. The most encouraging is the flattening trend in the estimated expected number of crashes when the threshold separation (lateral clearance) falls below a certain value. The point where the trend changes is easy to identify in most cases.

The results confirm that the number of claimed traffic conflicts carry useful information about traffic safety. Nevertheless, relying only on traffic conflicts to compare various scenarios is insufficient in cases where the conditional probability of crash changes between the scenarios.

The largest suitable threshold clearance below which the expected crash estimates do not exhibit any trend can be loosely interpreted as the boundary between the lateral positions comfortable to the driver and the lateral positions that are not comfortable. In other words, lateral clearances smaller than the largest suitable threshold are observed during a driver’s attempt to increase the clearance to a comfortable level.

The element in the results useful to evaluating driving performance is the probability of crash given a conflict. This probability depends on the threshold clearance adopted by the driver and how effective the driver is in responding to undesirable crash nearness. A small threshold and large variability in the exceedances (or separations) increase the probability of crash.
REFERENCES


