Probability approach for ranking high-accident locations

A. S. Al-Ghamdi
Department of Civil Engineering, King Saud University, Saudi Arabia

Abstract

Several methods for ranking high accident locations have been considered. Although the budget constrains is the primary factor affecting the number of locations to be considered in future improvement programs, some techniques based on either accident frequency and/or accident rate are typically used. This paper presents a statistical technique employing the theoretical relationship concept of Poisson and binomial distributions. It is known that counting distributions, in particular Poisson and Binomial distributions, have been used in traffic research for a long time. Based on a certain definition for a hazardous location, the paper gives an illustration on how to rank such locations by computing a list of probabilities. The analyst can use one of these probabilities in order to select some locations to be candidates for future improvement. The paper suggests a generic format for the technique described in the study.

1 Introduction

Poisson distribution has been used in traffic research for a long time. Kinzer in 1933 discussed the possible use of the Poisson distribution for traffic [1]. Three years later, 1936, Adams published numerical examples. Greeshields in 1947 used the Poisson distribution in the analysis of classic work on traffic intersections [1]. In 1955 the Eno Foundation published Poisson and Traffic, which dealt with the applications of the Poisson distribution to the problems of street and highway traffic [2]. Gerlough and Barnes [2] published “Poisson and Other Distributions in Traffic” to deal with some of the relationships which have
been found useful in handling the random properties of traffic. They stated that “One principal tool is the Poisson distribution”.

Many statistical models have been developed to establish the empirical relationships between vehicle accidents and the geometric design of highways for different roadway classes and accident severity types. However, most of these models were developed on the basis of the conventional multiple linear regression approach, and have been shown to be lacking the distributional property to adequately describe discrete, nonnegative, and typically sporadic, vehicle accident events on the road. These unsatisfactory properties of the linear regression models have led to the investigation of the Poisson regression and negative binomial regression models in recent studies.

Miauo and Lum [3] illustrated how the Poisson regression model can be used to evaluate the effects of key highway geometric design elements on truck accident involvement rates. Jovanis and Chang [4] applied a Poisson regression as a means to predict accidents. They argued that the statistical properties of Poisson regression are superior to those of linear regression for applications regarding highway safety. Al-Ghamdi [5] used Poisson distribution to describe the distribution of traffic accident occurrences. He developed a test statistic to enable traffic analysts to compare traffic accident rates in various transportation facilities (i.e., intersections and road sections). Following is a brief background about binomial and Poisson distributions.

1.1 Binomial distribution

Binomial distribution is formed from a sequence of independent Bernoulli trials, in which the number of successes of a certain number of trials is the quantity of interest. The expansion of the binomial \((q + p)^n\) forms the basis of the binomial distribution function. If \(n\) is a positive integer, the \((x+1)\)th term in this expansion is

\[
\binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}
\]

\(p\), the binomial distribution of a random variable \(X\) is defined as

\[
\Pr[X = k] = \begin{cases} 
\binom{n}{x} p^x q^{n-x} & (x = 0,1,2,\ldots,n) \\
0 & \text{otherwise}
\end{cases}
\]
where 0 < p < 1 and q = 1 - p. The mean and variance of X are np and npq, respectively.

This distribution has been used in several traffic applications. In congested traffic as in the case where the ratio of the observed variance/mean is substantially less than 1, for instance, the binomial distribution can be used to describe the distribution of traffic arrivals. When n is very large, and p is very small the binomial distribution is approximated by the Poisson distribution.

### 1.2 Poisson distribution

A random variable X is said to have a Poisson distribution with parameter $\lambda$ if it has discrete pdf of the form

$$
Pr[X = k] = \frac{e^{-\lambda} \lambda^k}{k!} \quad (k = 0, 1, 2, \ldots; \lambda > 0)
$$

The random variable X has the same mean and variance. Along with exponential distribution, the pdf of which is defined below, the Poisson distribution has been applied in traffic studies, particularly in those studies which involve simulation applications. A continuous random variable X has the exponential distribution with parameter $\lambda$; $\lambda > 0$ if its pdf has the following form:

$$
f(x, \lambda) = \begin{cases} 
\frac{1}{\lambda} e^{-x/\lambda} & x > 0 \\
0 & \text{otherwise}
\end{cases}
$$

The application of the Poisson distribution to traffic studies has been in existence since the 1930's [4]. This distribution has been used in fitting traffic accidents and vehicle arrivals at certain locations.

### 2 Methodology

Given that an intersection is considered hazardous when at least k accidents occur during time $t$, the question is: What is the probability that there will be at least m intersections in a study area, which have more than k accidents? One can use this probability to decide about the number of intersections that should be candidates for future improvement, and hence, the budget can be estimated.

Computing the probability in the above question, one should consider how to relate the concept of Poisson distribution and the concept of binomial
distribution. In other words, we know that the occurrence of accidents at a certain location within a period of time follows Poisson distribution. In the same context, the number of hazardous intersections among a finite number of intersections is said to be binomially distributed. Following is a description in mathematical terms for this relationship.

Let \( Y \) denote the number of accidents in a given intersection during a period of time. It is known that \( Y \) is Poisson distributed. Then the probability that a given intersection will contain more than \( k \) accidents is:

\[
p = \Pr (Y > K) = \sum_{i=k+1}^{\infty} f(i / \lambda) = \sum_{i=k+1}^{\infty} \frac{e^{-\lambda} \lambda^i}{i!} \tag{1}
\]

Therefore,

\[
1 - p = \sum_{i=0}^{K} f(i / \lambda) = \sum_{i=0}^{K} \frac{e^{-\lambda} \lambda^i}{i!} \tag{2}
\]

Now let \( X \) denote the number of intersections, among the \( n \) intersections in the city, in which there are more than \( k \) accidents at each. It can be said that \( X \) follows binomial distribution with parameter \( p \) (this parameter is the Poisson probability above). Then, for \( x = 0, 1, \ldots, n \),

\[
P_x (X = x) = \binom{n}{x} p^x (1 - p)^{n-x}
\]

and

\[
P_x (X \geq m) = \sum_{x=m}^{n} \binom{n}{x} p^x (1 - p)^{n-x} \tag{3}
\]

This is the probability that there will be at least \( m \) intersections with more than \( k \) accidents.

Now we can substitute eqn (1) and eqn (2) into eqn (3) and calculate the probability under question. A “probability list” can be developed for \( m \) as a variable. This probability list can be used to determine how many intersections are hazardous, as will be illustrated in the case study below.

3 Case Study

Let’s suppose to have the following list of intersections with accident occurrences:
### # of accidents in 1 year

<table>
<thead>
<tr>
<th>Intersection</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

\[
\text{n}=10 \quad \Sigma = 32 \\
\hat{\lambda} = 3.2
\]

The overall mean (accident rate) is 3.2 accidents/intersection. The above data was obtained from Traffic Department in Riyadh (Capital of Saudi Arabia). The data represents accident occurrence at ten intersections in the northern part of the city during a one-year period. The goal is to decide how many intersections from the above list are to be considered as candidates for improvement. For this purpose, the criterion for an intersection to be hazardous is defined by the analyst as: any intersection with accident frequency above 3 during a one-year period.

Hence, \( k = 3 \), in eqn (2) and eqn (3):

\[
p = Pr(Y > 3) = \sum_{i=0}^{3} \frac{e^{-3.2} 3.2^i}{i!} = 0.6025 \approx 0.603
\]

and,

\[
1 - p = 1 - 0.6025 = 0.3975 \approx 0.397
\]

The probability (0.603) will be used as the parameter for binomial distribution to compute the following probability list, where \( m \) is variable and \( k = 3 \):
<table>
<thead>
<tr>
<th>$m$</th>
<th>$p(X \geq m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.994</td>
</tr>
<tr>
<td>2</td>
<td>0.952</td>
</tr>
<tr>
<td>3</td>
<td>0.828</td>
</tr>
<tr>
<td>4</td>
<td>0.61</td>
</tr>
<tr>
<td>5</td>
<td>0.361</td>
</tr>
<tr>
<td>6</td>
<td>0.161</td>
</tr>
<tr>
<td>7</td>
<td>0.053</td>
</tr>
<tr>
<td>8</td>
<td>0.012</td>
</tr>
<tr>
<td>9</td>
<td>0.002</td>
</tr>
<tr>
<td>10</td>
<td>\approx 0</td>
</tr>
</tbody>
</table>

Thus, for example, when $k$ is greater than 3 accidents, the probability that 4 intersections are hazardous is 0.61. One may use this probability as a criterion for determining how many intersections are to be considered hazardous. For example, if the probability selected (criterion) is at least 0.80, only 3 intersections, based on the above probability list, are considered to be hazardous and hence to be treated. That is, the three highest ranked sites in the Table should be treated.

### 4 Generic approach

Based on the above derivation and illustration, the approach can be generalized as described below.

Firstly, sites with accident frequencies should be listed decreasingly as below:

<table>
<thead>
<tr>
<th>Site.</th>
<th>No. of accidents in period $t$ for site $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_1$</td>
</tr>
<tr>
<td>2</td>
<td>$y_2$</td>
</tr>
<tr>
<td>3</td>
<td>$y_3$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$n-1$</td>
<td>$y_{n-1}$</td>
</tr>
<tr>
<td>$n$</td>
<td>$y_n$</td>
</tr>
<tr>
<td>$n=10$</td>
<td>$\Sigma y_i$</td>
</tr>
</tbody>
</table>

where,

$y_1 \geq y_2 \geq \ldots \geq y_{n-1} \geq y_n$

and
where,

\[ n = \text{Number of sites} \quad \quad i = 1, \ldots, n \]
\[ \hat{\lambda} = \text{Accident rate (accidents/site).} \]

From eqn (2):

\[ \hat{p} = \Pr(Y > K) = 1 - \sum_{i=0}^{K} \frac{e^{-\hat{\lambda}} \hat{\lambda}^i}{i!} \]  
\[ \text{(5)} \]

Therefore,

\[ 1 - \hat{p} = \sum_{i=1}^{k} \frac{\sum_{i=0}^{n} y_i \left( \sum_{j=0}^{n} y_j \right)^i}{i!} \]  
\[ \text{(6)} \]

Substituting into (3), the probability under question is:

\[ P = \Pr(X \geq m) = \]

\[ \sum_{x=m}^{n} \binom{n}{x} \left( 1 - \sum_{i=1}^{k} \frac{\sum_{i=0}^{n} y_i \left( \sum_{j=0}^{n} y_j \right)^i}{i!} \right)^x \left( \sum_{i=1}^{k} \frac{\sum_{i=0}^{n} y_i \left( \sum_{j=0}^{n} y_j \right)^i}{i!} \right)^{n-x} \]  
\[ \text{(7)} \]

Now a list of probabilities can be tabulated as:
530 Urban Transport and the Environment for the 21st Century

\[ m \quad P(X \geq m) \]

\[
\begin{align*}
1 & \quad p_1 \\
2 & \quad p_2 \\
3 & \quad p_3 \\
. & \quad . \\
. & \quad . \\
i & \quad p_i \approx 0
\end{align*}
\]

\( m \) can be selected according the following criterion:

\[ P \geq \varphi \]

for a selected criterion level of probability \( \varphi \) (e.g., 0.50), the corresponding \( m \) in Eq.(7) gives a probability equal to or greater than \( \varphi \) and represents the hazardous sites that should be candidates for treatment. Hence, the analyst can go to the ranked list of sites and choose the first \( m \) sites, which satisfy the criterion above. These \( m \) sites are the hazardous locations and will be subject to improvement.

5 Conclusion

This primary finding of this paper is a statistical approach to estimate the probability of having a certain number of sites which can be classified as hazardous locations (a location with at least \( k \) accidents in a defined period of time) within a certain region. It is known from traffic safety literature that accident occurrences follow Poisson distribution. In addition, the number of hazardous sites within a finite number of sites can be described as a binomial distribution. The study developed a relationship between the two variables with Poisson and binomial distributions, as just described, to obtain the probability of having a certain number of hazardous sites. This probability is useful in planning for future safety improvement programs. That is, the analyst can get better idea about how many sites at least should be treated and, hence, how much budget should be allocated.

The approach is illustrated with an application. All that the analyst needs is to know the accident rate for the sites under consideration (accidents per site) and to define a hazardous location (number of accidents occur within certain time to consider the location as hazardous).

References


