Modelling on Pedestrian Accidents in Malaysia

NOORIZAM DAUD and HASLINDA AB. MALEK
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam
MALAYSIA
noorizam@tmsk.uitm.edu.my, linda_gurl04@yahoo.com
http://www.uitm.edu.my

Abstract: - Malaysia has been experiencing rapid growth in population, economy and motorization. This increase in population and motorization has led to the increase in the number of pedestrian accidents on road networks. Recently, road safety has become a major concern for many countries including Malaysia. Many researches pertaining to the pedestrian accident data has been conducted in aiming to reduce the rate of pedestrian accidents in Malaysia. The main aim of this study is to identify which distribution is the best fitted for pedestrian accident data obtained from 30 locations in Malaysia from year 2006 to 2008. It is found that Negative Binomial distribution provides a better fit for daytime and night-time of pedestrian accident data as compared to the Poisson distribution. This study also emphasized on the identification of pedestrian accident blackspots using hierarchical Bayesian approach based on the correlation between the mean number of pedestrian accidents during night-time and daytime. This study reveals that Tapah emerged as the top ranked pedestrian accident location in this period.

Key-Words: - Pedestrian Accidents, Negative Binomial, Poisson, Blackspots, Hierarchical Bayesian, correlation.

1 Introduction

The pedestrians are known as the most vulnerable of all road users. Malaysia has been experiencing rapid growth in population, economy and motorization. This increase in population and motorization has led to the increase in the number of pedestrian accidents on road networks. The children and elderly people are considered to be at a higher risk regarding pedestrian crashes [9]. The pedestrian accidents contribute to the health and social problem in this country. Moreover, road accidents are one of the major contributors of human deaths in Malaysia. Pedestrian exposure research has been conducted for decades; however, the crashes between pedestrians and motor vehicles become more serious problem. Therefore, this paper concentrate on finding the best distribution model of pedestrian accidents and use it in identifying the potential pedestrian accident blackspots. The identification of pedestrian accident prone locations could help the road authority to make the decision regarding the road safety especially in order to improve the pedestrians' safety and hence, it could reduce the number of such pedestrian accidents in the near future.

Many research related to the accident data has been conducted in order to help in reducing the rate of accidents. Among the studies conducted on the distribution related to the accident data are Poisson distribution, Binomial distribution, Negative Binomial distribution, Truncated Negative Binomial distribution and Log Series distribution. Besides, many researchers claimed that among the most popular distributions applied in road safety analysis are the standard Poisson and Negative Binomial (NB) distributions [1, 6, 8]. Furthermore, some of the researchers have proved that the accident data can be modelled by a Poisson distribution, for example Zahavi [13] has shown that the Poisson distribution is the best fitted model for the monthly accident data at Israel from 1959 until 1961. The same result found by Chapman [3] for daily accident occurred at New South Wales, Australia from 1958 until 1963. On the other hand, Weber [12] has proved that the Negative Binomial distribution could fit the accident data for 148,000 drivers in California from 1961 until 1963. He concluded that the number of accidents for a certain period of time can be modelled by a Negative Binomial distribution. The same result found by Wan and Kamarulzaman [11]. They have fitted a Poisson and Negative Binomial distributions for accident data obtained from 25 locations in United Kingdom from year 1974 to 1992. Based on the comparison of the fitted model, they concluded that Negative Binomial distribution provides a better fit for the accident data as compared to the Poisson distribution. However, Senn and Collie [10] have fitted the Bivariate Negative Binomial distribution to the accident data for two periods where the first period covers from year 1979 until 1980 and the second period covers from year 1981 until 1982. They concluded that the Bivariate Negative Binomial distribution is appropriate to model accident data with such features. In addition to that, Maher [5] also suggested that this model could be extended to the data on two types of accident (slight and severe) and data on two period of time (night-time and daytime).

More recently, Bayesian techniques are used to tackle problems in traffic safety. A number of authors have used the empirical Bayesian approach, combined with multivariate regression models to estimate the safety at various types of facilities. This approach was first proposed by Bonneson *et al.*[2] to estimate the safety at two way stop controlled intersections on rural highways. On the other hand, Noorizam and Kamarulzaman [7] have used a hierarchical Bayesian approach in order to rank the accident blackspots based on the correlation between the mean number of fatal and serious injury motor vehicle accidents. Furthermore, Higle and Hecht [4] have conducted a controlled experiment to compare the efficiency of different Bayesian and conventional hazardous location identification methods. It was concluded that Bayesian identification methods generally perform better than the conventional methods in correctly identifying hazardous locations.

2 Methodology

This section will discuss about data used in the study and also the method used in searching the suitable distribution to fit this pedestrian accident data. In addition, the discussion also covers on the hierarchical Bayesian approach by considering the correlation exists between the mean number of night-time and daytime pedestrian accidents.

2.1 Data Collection

The data used in this study was obtained from the Malaysian Institute of Road Safety Research (MIROS). The number of pedestrian accidents occurred at 30 different locations in Malaysia from the year 2006 until 2008. There are three variables measured:

1) The number of pedestrian accidents in each location.
2) The location of pedestrian accidents.
3) The time of pedestrian accidents (night-time or daytime).

2.2 FITTING DISTRIBUTION TO THE PEDESTRIAN ACCIDENT DATA

Here, Poisson and Negative Binomial distributions would be considered in order to identify the appropriate distribution to model the night-time and daytime pedestrian accident data. The parameters for both distributions are estimated using the method of maximum likelihood. The Chi-square goodness of fit test will be used to determine the appropriate distribution to the data.

2.3 HIERARCHICAL BAYESIAN APPROACH BASED ON THE CORRELATION

Consider two discrete random variables X_i and Y_i each representing the number of pedestrian accidents in night-time and the number of pedestrian accidents in daytime occurring at a certain location. Since both X_i and Y_i satisfy the characteristics of a Poisson process, it is reasonable to assume that both variables are following Poisson distribution with mean λ_1 and λ_2, respectively. Assuming λ_1 follows Gamma distribution with parameter $\alpha+\beta_1$ and γ

whereas λ_2 follows Gamma distribution with parameter $\alpha+\beta_2$ and γ. The λ_1 and λ_2 each were

having means $\mu_1 = \dfrac{\alpha+\beta_1}{\gamma}$ and $\mu_2 = \dfrac{\alpha+\beta_2}{\gamma}$,

According to Maher [5], λ_1 and λ_2 can be assumed to be

$$\lambda_1 = U + V$$
$$\lambda_2 = U + W$$

where;
$U \sim Gamma(\alpha, \gamma)$, $V \sim Gamma(\beta_1, \gamma)$ and
$W \sim Gamma(\beta_2, \gamma)$.

Assuming U and V, and also U and W are independent and because of U is common to both, λ_1 and λ_2 are positively correlated with correlation coefficient denoted as ρ. This value of correlation coefficient, ρ can be obtained by using the following formula:

$$\rho = \frac{\alpha}{\sqrt{(\alpha+\beta_1)(\alpha+\beta_2)}} \tag{1}$$

Maher [5] also has suggested the application of Bivariate Negative Binomial distribution to be used to analyze traffic accidents during night-time and daytime. The application of this distribution is based on the assumption that the positive correlation exists

ISBN: 978-1-61804-020-6

between the mean number of pedestrian accidents during night-time and daytime.

The Bivariate Negative Binomial distribution as suggested by Maher [5] is given below:

$$P(x_i, y_i) = \iiint f_{U}(u) f_{V}(v) f_{W}(w) e^{-(u+v)} \frac{(u+v)^{x}}{x!} e^{-(u+w)} \frac{(u+w)^{y}}{y!} du dv dw \quad (2)$$

in which $f_U(u), f_V(v), f_W(w)$ are the probability density functions of three gamma variables *U*, *V*, and *W*.

3 Analysis and Findings

Based on the parameter estimations, the corresponding estimated values for Poisson and Negative Binomial distributions using method of maximum likelihood are given in Table 1.

Table 1: Summary of the parameter estimates

Time of Accidents	Distribution	Estimated Value
Night-time	$X \sim Poi(\lambda_1)$	$\hat{\lambda}_1 = 1.99$
Daytime	$Y \sim Poi(\lambda_2)$	$\hat{\lambda}_2 = 3.53$
Night-time	$X \sim NB(k, w)$	$\hat{k} = 1.187$ $\hat{w} = 0.374$
Daytime	$Y \sim NB(k, w)$	$\hat{k} = 2.654$ $\hat{w} = 0.429$

Table 2 shows the summary of the Chi-square goodness of fit test. Based on the comparison of the fitted model, it is found that Negative Binomial distribution provides a better fit for night-time and daytime of pedestrian accident data as compared to the Poisson distribution.

In this study, the correlation between the mean number of pedestrian accidents during night-time and daytime was found to be 0.41. Based on the p-value (0.024), at the 5% level of significance, we could conclude that there is a significant correlation between the mean number of pedestrian accidents during night-time and daytime. Hence, the Bivariate Negative Binomial is appropriate to be used in analyzing this data where it parameters need to be estimated. By using the value of correlation coefficient, $\rho = 0.41$, mean of night-time accidents, $\mu_1 = 2$, mean of daytime accidents, $\mu_2 = 3.5$ and γ=1, the parameters of Bivariate Negative Binomial which are α, β_1, β_2 can be found using the simultaneous equations technique. Each posterior mean of λ_1 and λ_2 are obtained using a written MATLAB program. The top ten pedestrian accident locations identified to be potential hazardous pedestrian accident locations is given in Table 3.

Table 2: Summary of the chi-square goodness of fit test

Distribution	Chi-square Value	Chi-square Table
X~Poi(1.99)	29.186	$(\chi^2_{0.05,3} = 7.813)$
Y~Poi(3.53)	47.761	$(\chi^2_{0.05,5} = 9.236)$
X~NB(1.187,0.374)	3.420	$(\chi^2_{0.05,3} = 7.813)$
*Y~NB(*2.654, 0.429*)*	11.659	$(\chi^2_{0.05,6} = 12.592)$

Table 3: Posterior mean with the correlation coefficient 0.41

Location	X_i	Y_i	Posterior Mean	Ranking
Tapah	17	17	10.358	R(1)
Mersing	14	19	10.061	R(2)
Barat daya	9	17	8.081	R(3)
Kuala Lipis	7	18	7.753	R(4)
Langkawi	7	16	7.237	R(5.5)
Pekan	6	17	7.202	R(7)
P. Besar	11	10	6.763	R(8)
Kulai Jaya	13	8	6.730	R(9)
Machang	8	13	6.742	R(10)

4 Conclusion

Negative Binomial distribution provides a better fit for night-time and daytime of pedestrian accident data as compared to the Poisson distribution. Since the mean number of pedestrian accidents during night-time and daytime are found to have a significant positive correlation, a Bivariate Negative Binomial distribution is used to model the night-time and daytime pedestrian accidents. Based on the overall posterior mean value, Tapah emerged as the top ranked pedestrian accident location in this study.

References:

[1] Austin, R.D. & Carson, J.L. An Alternative Accident Prediction Model for Highway Rail Interfaces. *Accident Analysis and Prevention* 34, 2002, pp. 31–42.

[2] Bonneson, J.A. & McCoy, P.T. Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways. *Transportation Research Record 1401*, 1993, pp. 83 - 89.

[3] Chapman, R.A. Numbers of Accidents per Day and Times between Accidents. *Traffic Engineering and Control,* 12(2), 1971, pp. *82-84.*

[4] Higle, J.L. & Hecht, M.B. A Comparison of Techniques for the Identification of Hazardous Locations. *Transportation Research Record*, 1988, pp. 10 – 19.

[5] Maher, M.J. A Bivariate Negative Binomial Model To Explain Traffic Accident Migration. *Accident, Analysis & Prevention.* Vol. 22. No. 5, 1990, pp. *487-498.*

[6] Miaou, S.P. The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions. *Accident Analysis and Prevention* 26-4, 1994, pp. 471–482.

[7] Noorizam, D. & Kamarulzaman, I. *Hierarchical Bayesian Approach for Ranking of Accident Blackspots.* Faculty of Science, Technology. Universiti Kebangsaan Malaysia, 2007.

[8] Saccomanno, F., Fu, L. & Miranda-Moreno, L.F. Risk-Based Model for Identifying Highway–Rail Grade Crossing Black spots. *Transportation Reearch Record 1862*, 2004, pp. 127–135.

[9] Satiennam, T. & Tanaboriboon Y. Road Accidents in Thailand. *Journal of the International Association of Traffic and Safety Sciences, IATSS Research*. Vol. 29, 2005, pp. 88-100.

[10]Senn, S.J. & Collie, G.S. Accident Black spots and The Bivariate Negative Binomial. *Journal of Traffic Engineering Control,* 1988.

[11]Wan, F.W.M. & Kamarulzaman, I. *Fit Distributed Accident data.* Faculty of Science Technology, Universiti Kebangsaan Malaysia, 2001.

[12]Weber. Accident Rate Potential: An Application of Multiple Regression Analysis of a Poisson Process.*Journal of the American Statistical Association*, Vol. 66, No. 334, 1971, pp. 285-288.

[13]Zahavi, Y. The Problem of Accident Distribution. *Traffic Quarterly,* 16(4), 1962, pp. 540-548.