AN INTRODUCTORY GUIDELINE FOR THE USE OF BAYESIAN STATISTICAL METHODS IN THE ANALYSIS OF ROAD TRAFFIC ACCIDENT DATA

CJ Mollett Pr Eng
Chief Engineer : Traffic Engineering
Department of Economic Affairs , Agriculture and Tourism : Branch - Transport
25 Alfred Street
Cape Town
8001

Dr CJ Bester
Department of Civil Engineering
University of Stellenbosch
Private Bag X1
Matieland
7602

1. INTRODUCTION

Engineers often have to analyse accident data to estimate the level of safety at different road infrastructure elements (segments and intersections) in order to identify hazardous (unsafe) locations and to evaluate the effectiveness of road safety countermeasures. By 'safety' is meant either the 'true' underlying accident rate or the 'true' underlying accident frequency at a location.

There are a variety of methods available to analyse road traffic accident data. These methods can be classified into two categories : 'Conventional' and Bayesian.

Perhaps because of its perceived complexity Bayesian methods are not often used to analyse accident data even though the general consensus is that Bayesian methods are superior to conventional methods (1). The objective of this paper is to provide guidelines on how the Bayesian approach can be used to estimate safety (accident rate/frequency) at any location (segment or intersection) or a group of locations, and how these estimates can then be used to identify Accident Prone Locations (APL's) and to evaluate the effectiveness of remedial measures.

The paper will commence with an overview of the theoretical principles underlying the Bayesian approach and in doing so relevant comparisons will be made with the 'conventional' approach. This will be followed by a literature review to show how the Bayesian approach has been applied in practice.

The main body of the paper is divided into three sections – dealing with the estimation of safety, the identification of hazardous locations and the evaluation of road safety remedial measures respectively.

2. THE BAYESIAN APPROACH

2.1 THEORETICAL FRAMEWORK

As in other disciplines of statistics, the analysis of accident data is concerned with the determination of parameters and constants which has great practical importance – such as the true accident frequency|rate (m) at a location – but whose values are and cannot be precisely known. Methods for dealing with this problem can be divided into two categories – a) 'Conventional' methods and b) 'Bayesian' methods.

Bayesian methods are distinguished from ‘conventional’ methods in that any parameter in a problem (such as the true accident frequency|rate at a location) is regarded as a random variable with a probability distribution. Whereas in ‘conventional’ methods the observed accident experience *x* at a site is considered to be an unbiased estimate of the true level of safety at a site, Bayesian assumes that the observed accident experience *x* is a variable and that it is Poisson distributed about *m* – the true level of safety.

$$P(x/m) = \frac{m^x e^{-m}}{x!} \qquad \ldots[1]$$

It is further assumed that *m* is constant over time and that the observed accident experience in different years are random variables that are Poisson distributed about *m* .

The Bayesian approach further assumes that *m* varies between different sites and that the exact value for any particular site is unknown and is regarded as a Gamma variable with the following probability density function:

$$f(m) = \frac{\alpha^{\beta} m^{\beta-1} e^{-\alpha m}}{\Gamma(\beta)} \qquad \ldots[2]$$

Where

α - The shape parameter.
β - The scale parameter.

$$E(m) = \frac{\alpha}{\beta} \qquad \ldots[3] \qquad\qquad VAR(m) = \frac{\alpha}{\beta^2} = \frac{E(m)^2}{\beta} \qquad \ldots[4]$$

The distribution of *x* between different sites is Negatively Binomially distributed about $\bar{x}$ as follows:

$$P(x_i) = \frac{\Gamma(x_i + k)}{x_i!\Gamma(k)} \left(\frac{k^{-1}\bar{x}}{1 + k^{-1}\bar{x}} \right)^{x_i} \left(\frac{1}{1 + k^{-1}\bar{x}} \right)^{k} \qquad \ldots[5]$$

Where *k* is the dispersion parameter and the mean and variance are :

$$E(x_i) = \bar{x} \qquad \ldots[6] \qquad\qquad VAR(x_i) = \bar{x} + \frac{(\bar{x})^2}{k} \qquad \ldots[7]$$

The larger the value of *k* the less the dispersion. For very large values of k: $VAR(\bar{x}) \rightarrow \bar{x}$. In this case $E(x_i) = VAR(x_i)$ and the between-site variation can be described by the Poisson distribution. The more similar the sites are within a group the smaller the level of dispersion in accident frequencies, as measured by *k*, will be.

According the Al-Masaeid (4), the Bayesian approach is a probabilistic method capable of augmenting the most recent information with the available historical data or prior knowledge to achieve better estimates.

According to Abbess et al. (1) the Bayesian approach assumes that a probability distribution can be found before any data become available – this distribution is called the *prior* distribution of the parameter. Once information becomes available *Bayes theorem* can be used to convert the prior distribution into a *posterior* distribution. When even more information becomes available the *posterior* distribution, using *Bayes theorem*, can be updated to obtain even more accurate estimates of the parameter.

In the analysis of accident data one of the primary objectives is the estimation of the true level of safety at a particular location or a group of locations. The first step in applying the Bayesian approach in determining the level of safety at a location is to assume that the level of safety at all the study locations are the same and that it is estimated by E(m). The *prior* distribution of the true level of safety is therefor given by Equation 2. The next step is to use the observed accident experience (*x*) at a site to convert the *prior* distribution to a *posterior* distribution. This allow the prior estimate to be updated to a more accurate estimate of safety denoted as *E(m|x).*

2.2 REGRESSION-TO-MEAN

Any observation consist of three components : its true value (T), a random error component (R) and a systematic error component (S) such that X = T + R + S.

The regression-to-mean effect (RTM) is related to the random error component (R). Assuming S = 0 then RTM = X – T. This concept is graphically illustrated in Figure 1 where T = m|x. The regression-to-mean effect is especially significant when sites are selected for treatment on the basis of a high observed accident frequency/rate.

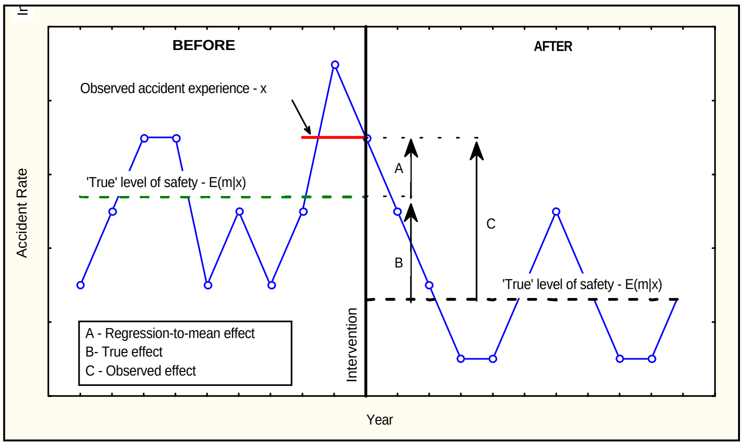

Figure 1 : Regression-to-mean effect

Conventional methods assume that the observed accident experience (X) is an appropriate measure of the true level of safety (T) i.e. X=T. This is a valid assumption only if the random error is equal to zero. Bayesian methods use the observed accident experience to estimate the true level of safety (T = m|x). The random error component (R) is therefor largely eliminated in the estimation procedure.

Conventional methods attempt to deal with the regression-to-mean effect in the following different ways:

a) Use accident data collected over longer periods – preferably over 5 years or more. The reliability of a safety estimate using accident data collected over such a long period of time is questionable considering the many site-specific external factors that could effect the level of safety over time. External factors which will not be accounted for by using a reference group of sites.

b) Using control groups. According to Al-Masaied (2) the 'before and after with control group' method is theoretically sound provided that comparison groups have geometric, operational characteristics <u>and</u> levels of safety similar to that of the treatment locations.

c) Using Bayesian methods to calculate the regression-to-mean effect and then to combine this with the results of the 'conventional' analysis. This approach is based on the potentially flawed assumption that the regression-to-mean effect is equal at all sites under consideration.

2.3 REFERENCE GROUPS

Both Bayesian and 'conventional' methods require a reference group comprising of a sufficient sample of locations similar to the study location/s. In some instances however the Bayesian approach can be applied to obtain reliable estimates of safety without the use of a reference group.

Classical methods require a reference group to account for time dependant trends in the accident rate and to eliminate the regression-to-mean effect. In order for the regression-to-mean effect to be eliminated the reference group sites should have levels of safety similar to that of the study site. Failure to select control sites that have a similar level of safety will not account for the regression-to-mean effect, even if the sites are similar to the study site in all other respects. This could obviously present practical problems in obtaining a sufficiently large sample of control sites.

The Bayesian approach require a reference group to determine the *prior* distribution of *m*. In selecting the reference group for the Bayesian approach locations should only have similar geometric and operational characteristics. Accident histories should not be considered in the selection process otherwise a biased estimate of *m* will be obtained.

Since the effect of the *prior* distribution diminishes as it is updated with observed data to form the *posterior* distribution the selection criteria for suitable reference group sites can be relaxed somewhat in order to ensure a sufficient sample size of sites.

The Bayesian approach allows for the use of accident models to determine the *prior* parameters.

2.4 INFORMATION REQUIREMENTS

To obtain reliable estimates of safety conventional methods require a minimum of 3 years, preferably 5 years worth of accident data. In order therefor to conduct a reliable before-and-after analysis at least 6 – 10 years' worth of accident data is required.

With the Bayesian approach reliable estimates of safety can be obtained using only 1 year's worth of accident data, providing the size of the reference group is sufficiently large. This has the advantage that only 'fresh' data, untarnished by the effect of external influences, is used in estimating safety at a location. Each additional year's worth of accident data means a new updated posterior distribution with an incremental increase in the accuracy and reliability of the estimate.

2.5 ANALYSIS AND INTERPRETATION

The end 'products' of the Bayesian approach is :

a) an estimate of the 'true' level of safety at a location or a group of locations [E(m|x)] and its variance [VAR(m|x)], and
b) the parameters of the probability density function of E(m|x).

This allow statistical inferences about E(m|x) to be made eg. confidence intervals and hypothesis testing.

The same flexibility to make statistical inferences is not generally forthcoming from the conventional approach to accident analysis.

3. EXAMPLES OF THE APPLICATION OF THE BAYESIAN METHOD

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 Hauer[1]. Bonneson et al. (6) and Belanger (5) applied Hauer's method to estimate the safety at two way stop controlled intersections on rural highways. Hauer (7) used his method to estimated the safety of signalised intersections.

Al-Masaeid et al (3) showed how to evaluate the safety impact of highway projects using the empirical Bayesian approach. He described a methodology to estimate the safety on road segments and groups of road segments with or without the use of exposure information, during the 'before' and 'after' periods. An equation is proposed to determine the probability of an improvement in safety between the 'before' and 'after' periods. Al-Masaeid et al. (4) used a similar approach to determine the accident reduction potential of pavement markings.

Hauer et al. (8) proposed a Bayesian method to identify hazardous locations and to evaluate the efficiency of an identification method. Their method however did not make provision for the inclusion of exposure data. Higle and Witowski (10) complemented Hauer et al.'s (8) work by proposing a method that would allow the incorporation of exposure data. The work by Higle and Witowski (10) was found to contain certain errors, which was later corrected by Morris. (11)

[1] Hauer E. Empirical Bayes Approach to the estimation of "Unsafety" : the Multivariate regression method. *Accident Analysis and Prevention*, Vol 24, No. 5 1992, pp 457 – 477.

Higle and Hecht (9) 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 and not identifying non-hazardous locations.

Al-Masaeid (4) conducted a performance evaluation of different safety evaluation methods. The Bayesian approach was compared to the simple before-and-after methodology as well as the 'before and after with control group' methodology. He concluded that the simple before-and-after methodology overestimates the effectiveness of remedial measures and that this method should not be used. He found that the empirical Bayesian method to be comparable with the before and after with control group method, and recommend that the Bayesian method be used if there is any difficulty in identifying a suitable and large number of comparison locations.

4. THE MEASUREMENT OF SAFETY

The following sections will describe in more detail the use of Bayesian methods to determine the level of safety for single sites and groups of sites with or without the use of exposure (traffic volume) information. The use of accident data without the use of exposure information should be handled with the utmost and is not recommended.

4.1 ACCIDENT NUMBER METHODOLOGY: SITE LEVEL

The prior parameters of the Gamma function (α, β) can be estimated form the reference group data as follows :

$$\hat{\alpha} = \frac{\bar{x}}{(s^2 - \bar{x})} \qquad \ldots[9] \qquad\qquad \hat{\beta} = \frac{\bar{x}^2}{(s^2 - \bar{x})} \qquad \ldots[10]$$

The values of E(m) and VAR(m) can be determined from Equations 4 and 5.

If *x* is the number of accidents at a site then the posterior distribution of *m* is of gamma type with parameters:

$$\beta' = x + \hat{\beta} \qquad \ldots[11] \qquad\qquad \alpha' = 1 + \hat{\alpha} \qquad \ldots[12]$$

Where

$$E(m \mid x) = \frac{\beta'}{\alpha'} \qquad \ldots[13] \qquad\qquad VAR(m \mid x) = \frac{E(m \mid x)^2}{\beta'} \qquad \ldots[14]$$

Hauer et al. (7) proposed the following simplified equations to determine *E(m|x)* and *VAR(m|x)* :

$$E(m \mid x) = aE(m) + (1 - a)x \qquad \ldots[15]$$

$$VAR(m \mid x) = a(1 - a)E(m) + (1 - a)^2 x \qquad \ldots[16]$$

Where

$$a = \frac{E(m)}{E(m) + VAR(m)} \qquad \ldots[17]$$

4.2 ACCIDENT NUMBER METHODOLOGY: GROUP-OF-SITES LEVEL

The expected number of accidents at a group of *n* similar locations is obtained by using the convolution principle as follows:

$$m_t = \sum_{i=1}^{n} m_i \qquad \text{…[18]}$$

Where m_t has a gamma probability density function with parameters $\Sigma\beta'_i$ and α'.

$$\sum \beta'_i = n\beta + \sum x_i \qquad \text{…[19]}$$

The expected mean and variance of m_t are :

$$E(m_t) = \frac{\sum \beta'_i}{\alpha'} \qquad \text{…[20]} \qquad VAR(m_t) = \frac{\sum \beta'_i}{(\alpha')^2} \qquad \text{…[21]}$$

4.3 ACCIDENT RATE METHODOLOGY : SINGLE SITE LEVEL

Let r be the accident rate at a location and x the observed accident frequency :-

$$r = \frac{x}{V} \qquad \text{…[22]}$$

Where V = Annual traffic in million vehicles (AADT*365/10^6).

The estimated prior parameters of the gamma distribution is as follows:

$$\hat{\alpha} = \frac{V^* \bar{r}}{V^* s^2 - \bar{r}} \qquad \text{…[23]} \qquad \hat{\beta} = \bar{r}\hat{\alpha} \qquad \text{…[24]}$$

Where V* is the harmonic mean of all the normalised traffic volumes and can be determined as follows

$$\frac{1}{V^*} = \frac{1}{n}\sum_{i=0}^{n}\frac{1}{V_i} \qquad \text{…[25]}$$

Once the parameters of prior distribution have been determined, the next step is to combine the prior information with the site-specific data to obtain the posterior distribution.

$$\alpha' = \alpha + V \qquad \text{…[26]} \qquad \beta' = \beta + x \qquad \text{…[27]}$$

4.4 ACCIDENT RATE METHODOLOGY : GROUP-OF-SITES

At the group of sites level, the total expected accident rate is given by the sum of the individual accident rates. The total expected accident rate (r_t) for a group of *n* sites is given by:

$$r_t = \sum_{i=1}^{n} r_i \qquad \ldots[28]$$

The expected value and variance of r_t is given by :

$$E(r_t) = \sum_{i=1}^{n} \frac{\beta'_i}{\alpha'_i} \qquad \ldots[29] \qquad\qquad VAR(r_t) = \sum_{i=1}^{n} \frac{\beta'_i}{\alpha'^{2}_i} \qquad \ldots[30]$$

The Gamma parameters of r_t can be estimated as follows:

$$\alpha_t = \frac{E(r_t)}{VAR(r_t)} \qquad \ldots[31] \qquad\qquad \beta_t = \frac{[E(r_t)]^2}{VAR(r_t)} \qquad \ldots[32]$$

The probability density function of r_t is then as follows :

$$f(r_t) = \frac{\alpha_t^{\beta_t} r_t^{\beta_t - 1} e^{-\alpha_t r_t}}{\Gamma(\beta_t)} \qquad \ldots[33]$$

4.5 USING PREDICTION MODELS

An alternative approach to estimate safety is to use accident prediction models for sites similar to the site in question. The simplest type of accident model relate accident frequency to traffic flows for each site category while in the more sophisticated models the accident frequency can be related to traffic flows as well as geometric variables.

The general form of the models generally used for segments and intersections are as follows :

For segments: $y = b_0 F^{b_1}$

For intersections: $y = b_0 F_1^{b_1} F_2^{b_2}$

Where

F – *ADT (Average Daily Traffic)*
F_1 – *Major crossroad ADT*
F_2 – *Minor crossroad ADT*

For example, Bonneson and McCoy (6) developed the following model for two-way stop controlled intersections on rural highways.

$$E(m) = 0.69\left(\frac{F_1}{1000}\right)^{0.256}\left(\frac{F_2}{1000}\right)^{0.831} \qquad \ldots[34]$$

E(m) - *Expected number of accidents per 3 year period..*

Using a prediction model as in Equation 34 will ensure that the resulting estimates will reflect the effect of traffic flow at a particular site.

The true underlying accident frequency E(m|x) and its variance is determined as follows :

$$E(m \mid x) = aE(m) + (1-a)x \qquad \ldots[35]$$

$$VAR(m \mid x) = a(1-a)E(m) + (1-a)^2 x \qquad \ldots[36]$$

where

$$a = \frac{E(m)}{E(m) + VAR(m)} \qquad \ldots[37] \qquad\qquad VAR(m) = \frac{[E(m)]^2}{k} \qquad \ldots[38]$$

If sufficient data are available to develop an appropriate prediction model then *k* can be estimated using this model. This requires fitting a multivariate Negative Binomial model to the observed data. The process of coefficient determination is iterative. Initially a value of *k* is assumed for the first round of regression. From the results of the first round a new value of *k* is determined. This new value is then fed into the second round of regression and so on until the value of *k* converge.

<u>Example</u>

The k value Bonneson an McCoy (6) calculated in the calibration of the model shown in Equation 34 = 4.

Assume one of the intersections in the study by Bonneson and McCoy had the following details :

- *F_1 = 4000 and F_2 = 2200*
- *Observed accident frequency at intersection over a 3 year period (x) = 4*

From Eqn. 34: $E(m) = 0.692\left(\frac{4000}{1000}\right)^{0.256}\left(\frac{2200}{1000}\right)^{0.831} = 2.3$

From Eqn 38 : $VAR(m) = \frac{2.3^2}{4} = 1.32$

From Eqn 37 : $a = \frac{2.3}{2.3 + 1.32} = 0.64$

From Eqn 35: $E(m \mid x) = 0.64(2.3) * 4(1 - 0.64) = 2.9$

From Eqn 36: $VAR(m \mid x) = 0.64(1 - 0.64)2.3 + (1 - 0.64)^2 4 = 1.05$

If however existing (already developed) models are used then *k* cannot be estimated directly. In such a case there are two approaches that could be considered.

a) Assume an appropriate range of values for k. This approach does not require a reference group of sites to determine *E(m).*

According to Mountain et al. (13) a number of authors have fitted negative binomial distributions to observed accident frequencies and obtained *k* values in the range 0.5 to 2.8.

b) If there are insufficient data to determine a reliable *k* for sites similar to the study site then a broader grouping of sites can be used to generate a sufficient number of sites.

According to Mountain et al. (13), if a broader grouping of sites is used then *k* can be estimated as follows: -

$$k = \phi(1 + \frac{1}{\theta}) \qquad \ldots[39]$$

Where ϕ is determined from fitting a negative binomial distribution to the observed accident frequencies of each site in the reference group. The value of ϕ can be estimated from the method of moments as follows :

$$\phi = \frac{\bar{x}^2}{{s_x}^2 - \bar{x}} \qquad \ldots[40]$$

The value of θ can be determined from fitting a Gamma distribution to the values of y for each site. Where y is the predicted number of accidents at a site using the prediction model.

$$\theta = \frac{\bar{y}^2}{{s_y}^2 - \bar{y}} \qquad \ldots[41]$$

5. THE IDENTIFICATION OF ACCIDENT PRONE LOCATIONS (APL'S)

The procedure for the identification of APL's are as follows :

a) Determine the values of *E(m|x)* and the posterior parameters of each study site using one of the methodologies described under section 3.

b) Determine the probability that *E(m|x)* exceeds *E(m).*

$$P(m \mid x > m) = 1 - \int_0^{E(m)} \frac{\alpha^\beta (m \mid x)^{\beta-1} e^{-\alpha(m|x)}}{\Gamma(\beta)} d(m \mid x) \qquad \ldots[42]$$

If *P(m|x > m)* > δ, where δ is the chosen level of confidence (eg. 95 %), then the location can be considered hazardous.

P(m|x > m) can be determined using the *GAMMADIST(x, alpha, beta, cumulative)* function of Microsoft Excel97®. Where x = E(m), alpha = β, beta = α and cumulative = TRUE.

Example

The reference group statistics: r = 2.0 acc/mvkm ; s^2 = 14 ; V = 1.20 mvkm.*

The site statistics: x = 4 per year/km ; AADT = 3850 ; V = 1.4 mvkm

Step 1 : Determine the prior parameters from Eqn's 23 and 24.

$$\hat{\alpha} = 0.16 \qquad \hat{\beta} = 0.32$$

Step 2 : Determine posterior parameters from Eqn's 26 and 27.

$$\alpha' = 0.16 + 1.4 = 1.56 \qquad \beta' = 0.32 + 4 = 4.32$$

$$E(m \mid x) = {}^{4.32}\!/_{1.56} = 2.77 \qquad VAR(m \mid x) = {}^{2.77^2}\!/_{4.32} = 1.77$$

Step 3 : Determine P(m|x) > m)

$$P(m \mid x > m) = 1 - GAMMADIST(2.0, 4.32, 1.56, TRUE) = 0.97$$

Since P(m|x > m) > 0.95 the site can be considered hazardous.

6. Evaluation of remedial measures

The assessment of the effectiveness of engineering measures in improving safety a site requires comparing the level of safety after treatment with the expected level of safety had no improvements taken place.

a) Determine the true accident frequency/rate – $E(m_b|x_b)$ – for the before period as well as its posterior parameters from any of the methodologies described in Section 3.

b) Determine the true accident frequency/rate – $E(m_a|x_a)$ – for the after period as well as its posterior parameters from any of the methodologies described in Section 3.

c) Determine the accident reduction factor as follows

$$E = \frac{100(m_b \mid x_b - m_a \mid x_a)}{m_b \mid x_b} \qquad \ldots[43]$$

d) Determine the joint probability function of m_a and m_b

$$f(m_b, m_a) = \frac{\alpha_a{}^{\beta_a} m_a{}^{\beta_a - 1} e^{-\alpha_a m_a}}{\Gamma(\beta_a)} \cdot \frac{\alpha_b{}^{\beta_b} m_b{}^{\beta_b - 1} e^{-\alpha_b m_b}}{\Gamma(\beta_b)} \qquad \ldots[44]$$

This joint probability function is based on the assumption that m_a and m_b are independent of each other.

e) Compute the probability that the level of safety in the after period (m_a) is less than the level of safety in the before period (m_b).

According to Al-Masaeid (2) $p(m_a < m_b)$ can be estimated as follows :

$$p(m_a < m_b) = 1 - \sum_{j=0}^{\beta_a - 1} \left[\frac{\alpha_a}{\alpha_b}\right] \left[\frac{\alpha_b}{\alpha_a + \alpha_b}\right]^{\beta_b + j} \cdot \frac{\Gamma(\beta_b + j)}{\Gamma(j+1)\Gamma(\beta_b)} \qquad \ldots[45]$$

The equation can be solved using a spreadsheet. E.g. the gamma function $\Gamma(\beta_b)$ can be solved using a combination of the EXP and GAMMALN functions in Microsoft Excel®. $\Gamma(\beta_b) = EXP(GAMMALN(\beta_b))$

Example

A site has the following parameters:

Before : $\alpha_b = 3.4$, $\beta_b = 1.2$, $E(m_b|x_b) = 2.80$
After : $\alpha_a = 4.0$, $\beta_a = 1.9$, $E(m_a|x_a) = 2.10$

ARF = 25 %

If $$D_j = \left[\frac{\alpha_a}{\alpha_b}\right]\left[\frac{\alpha_b}{\alpha_a + \alpha_b}\right]^{\beta_b + j} \cdot \frac{\Gamma(\beta_b + j)}{\Gamma(j+1)\Gamma(\beta_b)}$$

Then $D_0 = 0.46$ *and* $D_1 = 0.25$

$$\sum_{j=1}^{0.9} D_j = D_0 + 0.9D_1 = 0.46 + 0.9(0.25) = 0.685$$

$P(m_a < m_b) = 1 - 0.685 = 0.315$

The conclusion is therefor that for this particular site there has been no significant improvement in safety.

7. Conclusion

It has been shown that the Bayesian approach to estimation of safety has certain distinct advantages over the conventional methods. It can be utilised effectively in situations where sample sizes are inadequate. Provided sufficient prior information is available it can provide more reliable estimates of safety and it provides results that enable statistical inferences to be made.

The Bayesian approach has been applied successfully by renowned researchers to measure safety at various types of locations and to use these estimates to identify accident prone locations and to evaluate the effectiveness of road safety remedial measures at the *site* level as well as at *the group-of-sites level*.

Guidelines and procedures have been provided to simplify the use of a Bayesian approach to the analysis of accident data. It has been shown how the true level of safety at a site or a group of sites can be determined by combining the accident statistics of a group of reference sites with the accidents at a particular site. It has also been shown how accident prediction models can be used in conjunction with the Bayesian approach to estimate safety.

It has been shown how the results of the safety estimation process in which Bayesian methods were used can be applied to identify accident prone locations and whether there has been a significant change in the level of safety as a result of some remedial measure.

8. Recommendation

In South Africa greater emphasis should be placed on the use of Bayesian methods in road safety research and more research should be conducted into developing appropriate accident prediction models to support the application of the Bayesian approach.

8. LIST OF REFERENCES

1. Abbess C, Jarret D and Wright CC ; Accidents at blackspots : estimating the effectiveness of remedial treatment, with special reference to the 'regression-to-mean' effect ; Traffic Engineering and Control ; October 1981; Pages 535 – 542.

2. Al-Masaeid HR ; Performance of safety evaluation methods ; Journal of Transportation Engineering ; September/October 1997 ; Pages 364 – 369.

3. Al-Masaeid HR, Sinha KC ; Analysis of accident reduction potentials of pavement markings ; Journal of Transportation Engineering ; September/October 1994 ; Pages 723 – 735.

4. Al-Masaeid HR, Sinha KC and Kuczek T ; Evaluation of safety impact of highway projects ; Transportation Research Record 1401 ; 1993 ; Pages 9 – 16.

5. Belanger C ; Estimation of safety of four-legged unsignalized intersections ; Transportation Research Record 1467 ; 1994 ; Pages 23 – 29.

6. Bonneson JA and McCoy PT ; Estimation of safety at two-way stop-controlled intersections on rural highways ; Transportation Research Record 1401 ; 1993 ; Pages 83 - 89.

7. Hauer E, Ng JCN and Lovell J ; Estimation of safety at signalised intersections ; Transportation Research Record 1185 ; 1988 ; Pages 48 – 61.

8. Hauer E and Persaud BN ; Problem of identifying hazardous locations using accident data ; Transportation Research Record 975 ; 1984 ; Pages 36 – 43.

9. Higle JL and Hecht MB ; A comparison of techniques for the identification of hazardous locations ; Transportation Research Record ; Pages 10 – 19.

10. Higle JL and Witkowski JM ; Bayesian identification of hazardous locations ; Transportation Research Record 1185 ; 1988 ; Pages 24 – 31.

11. Morris CH ; Discussion on Higle and Witkowski's paper – Bayesian identification of hazardous locations ; Transportation Research Record 1185 ; 1988 ; Pages 31 – 32.

12. Mountain L, Fawaz B and Sineng L ; The assessment of changes in accident frequencies at treated intersections : a comparison of four methods ; Traffic Engineering and Control ; Volume 33 ; Number 2 ; February 1992 ; pages 85 – 87.

13. Mountain L and Fawaz B ; The accuracy of estimates of expected accident frequencies obtained using an Empirical Bayes approach ; Traffic Engineering and Control ; Volume 32 Number 5; May 1991 ; Pages 246 – 251.

14. Nembhard DA and Young M ; Parametric emperical Bayes estimates of truck accident rates ; Journal of Transportation Engineering ; July/August 1995 ; Pages 359 – 375.

AN INTRODUCTORY GUIDELINE FOR THE USE OF BAYESIAN STATISTICAL METHODS IN THE ANALYSIS OF ROAD TRAFFIC ACCIDENT DATA

CJ Mollett Pr Eng
Chief Engineer : Traffic Engineering
Department of Economic Affairs , Agriculture and Tourism : Branch - Transport
25 Alfred Street
Cape Town
8001

Dr CJ Bester
Department of Civil Engineering
University of Stellenbosch
Private Bag X1
Matieland
7602

CURRICULUM VITAE

CJ Mollett Pr Eng

Mr CJ Mollett currently holds the position of Chief Engineer : Traffic Engineering in the Transport Branch of the Provincial Administration of the Western Cape. He has been in the employ of the Provincial Administration : Western Cape since 1990. He has been working in the Traffic Engineering division of the Branch since 1994.

Mr Mollett completed his B.Sc Civil Engineering degree at UCT in 1989 and is currently busy with his masters degree in Traffic and Transportation Engineering at the University of Stellenbosch. Mr Mollett has completed a number of courses towards a B.Com degree at UNISA, amongst others majoring in Statistics and Quantitative Management.