October 25, 2004

FROM EMPIRICAL BAYES TO FULL BAYES: METHODS FOR ANALYZING TRAFFIC SAFETY DATA

Alicia L. Carriquiry, PhD [1], Iowa State University
Michael Pawlovich, PhD, PE [2], Iowa Department of Transportation [3]

Summary

Traffic safety engineers are among the early adopters of Bayesian statistical tools for analyzing crash data. As in many other areas of application, empirical Bayes methods were their first choice, perhaps because they represent an intuitively appealing, yet relatively easy to implement alternative to purely classical approaches. With the enormous progress in numerical methods made in recent years and with the availability of free, easy to use software that permits implementing a fully Bayesian approach, however, there is now ample justification to progress towards fully Bayesian analyses of crash data.

The fully Bayesian approach, in particular as implemented via multi-level hierarchical models, has many advantages over the empirical Bayes approach. In a full Bayesian analysis, prior information and all available data are seamlessly integrated into posterior distributions on which practitioners can base their inferences. All uncertainties are thus accounted for in the analyses and there is no need to pre-process data to obtain Safety Performance Functions and other such prior estimates of the effect of covariates on the outcome of interest. In this light, fully Bayesian methods may well be less costly to implement and may result in safety estimates with more realistic standard errors.

In this manuscript, we present the full Bayesian approach to analyzing traffic safety data and focus on highlighting the differences between the empirical Bayes and the full Bayes approaches. We use an illustrative example to discuss a step-by-step Bayesian analysis of the data and to show some of the types of inferences that are possible within the full Bayesian framework.

1. Introduction

State Departments of Transportation and engineers engaged in research have been collecting and analyzing traffic accident information for decades. Of particular interest to practitioners and researchers alike are ranking of hazardous sites, evaluation of the effectiveness of site improvements and prediction of the effect of potential modifications to a set of sites.

[1]Dr. Carriquiry is Professor, Department of Statistics, Iowa State University, Ames, Iowa
[2]Dr. Pawlovich is Traffic Safety and Crash Analysis Engineer, Office of Traffic and Safety, Iowa Department of Transportation
[3]This work was partially funded through a contract between the Department of Statistics at Iowa State University and the Iowa Department of Transportation.

In all of these cases, it is important to obtain a reliable estimate of the expected number of crashes at a specific site or group of sites in order to compare with actual occurrences. Often times, those estimates and the policy decisions that rely on them are based on relatively scarce information about the site or group of sites, either because traffic volumes at those particular locations are comparatively low, or because only a few years (or perhaps even a single year) of crash data are available for the locations of interest.

When data at a particular location span only a few years, the naive statistical analysis that relies on information from that site alone fails to capture the true (yet unobservable) long-term behavior at that site. The estimated long-term crash rate obtained by averaging observed crash rates over a few years can be unduly influenced by a single year with an unusually high (or low) number of crashes. A similar bias is introduced when crash rates are estimated for sites with very low traffic volumes, where a small number of crashes can result in a very high estimated crash rate at the site if rate is estimated using only information from that site. This is known as the *regression to the mean* effect and has been discussed extensively in the literature (e.g., Hauer, 1986; Hauer et al., 1986; Persaud, 1988).

In recent years, interest in the Bayesian approach to data analysis has increased significantly in many areas of application, including traffic safety. Perhaps the most influential first step towards the application of Bayesian methods in traffic safety is the work by Hauer and colleagues (Hauer, 1986; Hauer et al., 1986; Persaud, 1988; Hauer, 1996a; Hauer, 1996b; Harwood et al., 2002; Hauer et al., 2002; Higle et al., 1988; Higle et al., 1989; HSIS, 2001; Pendleton, 1991; Pendleton et al., 1991; Persaud et al., 1997; Persaud et al., 1998; Persaud et al., 2002a; Persaud et al., 2002b)), who have been instrumental proponents of the empirical Bayes (EB) approach to analyzing crash data. These researchers have eloquently and effectively argued in favor of EB methods, by pointing out the advantages of the EB approach over the classical approach to statistical analysis. In particular, the manuscript entitled *Estimating Safety by the Empirical Bayes Method: A Tutorial* (2002a) has provided an accessible step-by-step description of the EB approach to statistical analysis in various situations of interest to traffic safety engineers and as a consequence, the EB approach is now widely accepted in the field.

While demonstrably better suited to this type of application than naive statistical methods, the EB approach suffers several serious drawbacks. Perhaps most unfortunate among those is the need to spend time, resources and effort on the estimation of what are known as Safety Performance Functions (SPFs) required for implementation of the EB method. We argue later in this work that it is possible to improve on the prediction of the expected number of crashes at a site while at the same time avoiding the need to obtain estimates of SPFs and other quantities such as Accident Modification Factors (AMFs). Indeed, as will become clear in later sections, the need to use estimated SPFs and AMFs is one of the critical components of the EB approach to analyzing crash data. Except in very exceptional situations, the need to estimate the SPFs is also one of its largest limitations.

In this manuscript, we attempt to provide an accessible discussion of what we will call fully Bayesian (FB) methods for analyzing traffic safety data. Fully Bayesian methods are not new to traffic safety engineering; several authors have proposed FB methods for estimating crash rates (Davis and Gao, 1993; Davis and Guan, 1996; Davis, 2000; Davis and Yang, 2001; Pawlovich, 2003), for ranking sites (Schluter et al., 1997; Pawlovich, 2003) and for identifying black spots in road segments (Saccomanno et al., 2001; Flahaut et al., 2004).

We focus here on the comparison of FB with EB and in particular on their differences and comparative advantages. We argue that FB approaches present significant advantages over EB methods, both in terms of the inferences that are possible (output) and also in terms of the information that is required to implement either approach (input). EB is a special case of FB that arises when an FB analysis is simplified by making certain assumptions. For example, an FB model can be reduced to an EB model if the investigator is willing to assume that the effect of covariates such as control signals and shoulder width on number of crashes is known without any uncertainty and can be represented via an SPF or an AMF. But associations between covariates and safety are typically estimated from crash data and thus are subject to uncertainty. This uncertainty is more fully accounted by the FB approach than by the EB approach so sometimes EB results can be unrealistically optimistic relative to FB results. In an FB approach, as will be argued later in this manuscript, it is unnecessary to carry out any out-of-sample estimation, since information available for all relevant sites is combined into a single analysis; that is, pre-processing some of the information into estimating equations such as SPFs is not a necessary step and in fact so doing can limit the type of inference that is possible.

The rest of the manuscript is organized as follows. In Section 2, we introduce a motivating example using simulated data and describe the steps that would be needed in order to implement an EB and an FB analysis of the data. In Section 3 we describe more precisely the differences between the EB and the FB approaches and focus on the advantages and disadvantages of each approach. Finally, we draw some conclusions and provide recommendations in Section 4. Two appendices are included in the manuscript. Appendix A is a technical appendix, with the formal statistical formulation for the approaches addressed in the manuscript. In Appendix B, we discuss computational resources available to implement an FB analysis of traffic safety data.

2. A hypothetical example

Engineers at the Department of Transportation in some state have assembled two years of crash data on a set of 20 intersections in an urban area. In addition to the number of crashes per intersection per year, engineers have also collected data on daily entering vehicles (DEV) at the intersections as well as on the presence of traffic control signals (controls or no controls). The questions of interest might be the following:

- How many crashes can be expected per 1,000 DEV at the intersection labeled 1?
- Which are the two intersections with highest crash rates among those without traffic controls and among those with traffic controls?
- Other factors being equal, are intersection controls associated to lower crash rates?

ID	**Controls**	DEV_1	DEV_2	y_1	y_2
1	0	5766	5581	10	2
2	0	4307	4743	8	6
3	0	2623	1941	6	2
4	0	4240	4835	2	5
5	0	4844	4625	2	9
6	0	3178	2422	0	2
7	0	5162	5315	6	5
8	0	5420	5250	7	3
9	0	5392	4876	2	4
10	0	2987	3169	3	6
11	1	5726	4581	6	0
12	1	4207	4844	4	2
13	1	2520	1981	1	0
14	1	4270	4235	1	4
15	1	3844	3625	2	3
16	1	4178	4422	0	2
17	1	4160	4355	2	2
18	1	6420	5850	4	1
19	1	6390	5856	4	3
20	1	2980	3148	2	1

Table 1: *Number of crashes (y) and DEV in each of two years.*

The crash data for the 20 intersections over the two years are given in Table 1. In the table, we also give an indicator for traffic controls. Note that intersection 1, for which we wish to obtain an estimate of expected number of crashes, is one of 10 intersections in the study area with no traffic control signals.

We use y_{ij} to denote the number of crashes at the ith intersection in the jth year and use θ_i to denote the expected number of crashes for the ith intersection. Note that the number of crashes can be calculated as the product of *crash rate* (e.g., number of crashes per 1,000 DEV) times the DEV (in thousands) at the intersection. Thus, $\theta_i = \lambda_i \times DEV_i$, where λ_i is the crash rate at the ith intersection. In the hypothetical example introduced above, $i = 1, ..., 20$, and $j = 1, 2$.

How do we estimate the expected number of crashes at the intersection of interest?

The easiest (and most unreliable!) approach is to use only the two years of crash data available for intersection 1. In this case, an estimate of the expected number of crashes per 1,000 DEV (or crash rate) is just

$$\hat{\lambda}_i = \frac{y_{i1} \times 1,000/DEV_1 + y_{i2} \times 1,000/DEV_2}{2}$$

For intersection 1 above, the estimate would be 1.05 crashes/1,000 DEV, with a coefficient of variation of about 67%. Note that regardless of the distribution from which we sample the y_{ij}, the mean $\hat{\lambda}_i$ has an asymptotic normal distribution with variance equal to the year-to-year variability in crashes (the within-intersection variance σ_w^2) for the ith intersection divided by the number of years of data available $\sigma_{wi}^2/2$. If σ_{wi}^2 is large, that is, the number of observed crashes at the site varies greatly from year to year, then the estimate of the expected number of crashes will have low reliability (high variance).

An approach that under some conditions results in a more reliable estimate of λ_i is to use the two years of data for the intersection of interest *in combination with* information about similar intersections. In this case, the underlying assumption is that our intersection is sampled from a population of similar intersections, with mean number of crashes μ. In other words, the implicit model here is:

$$\begin{aligned} y_{ij} &\sim (\theta_i, \sigma_w^2), \qquad \theta_{ij} = \lambda_i \times DEV_{ij} \\ \lambda_i &\sim (\mu, \sigma_b^2), \end{aligned}$$

where σ_b^2 denotes the intersection-to-intersection variability in crash numbers (the between-intersection variance) when considering only groups of similar intersections. In such a model, one can obtain the Best Linear Predictor (BLP) of θ_i, which is given by

$$\tilde{\lambda}_i = \mu + \phi(\hat{\lambda}_i - \mu).$$

for

$$\phi = \frac{\sigma_b}{\sqrt{\sigma_b^2 + \frac{\sigma_w^2}{n}}},$$

where n is the number of years of data available for the intersection of interest. If we think of ϕ as a weighting factor, then we can rewrite the expression above as

$$\tilde{\lambda}_i = \mu(1 - \phi) + \hat{\lambda}_i \phi \tag{1}$$

as in Hauer et al. (2002). That is, the estimated number of crashes at the ith intersection is given by the weighted average of the information collected for that intersection and the mean number of crashes at intersections just like the one that interests us. Note that as the number of data points n available for the ith intersection increases, the within-intersection variance component decreases and as a consequence, ϕ increases. Thus, when there is abundant information available for the intersection of interest, the estimator of its expected number of crashes is based almost exclusively on the data available for that intersection (see expression (1)). In contrast, as σ_w^2/n increases, ϕ decreases and then $\tilde{\lambda}_i$ tends to μ. This is intuitively appealing; as more data for the intersection of interest becomes available, we rely less and less on information about 'similar' interesections. The estimator above is called Best because it has smallest mean squared error among all linear estimators.

Assume for a moment that we let intersections 2 - 10 in Table 1 (all with no traffic controls) represent the "'similar"' cohort of intersection 1 that is of interest. How do we obtain the estimate $\tilde{\lambda}_1$ for intersection 1? If the 10 intersections without controls in Table

1 had identical DEV, we could estimate μ simply as the mean (over intersections and over years) number of crashes per 1,000 DEV. Similarly, σ_b^2 is estimated as the variance of the intersection means and σ_w^2 is estimated as the average (over the nine intersections) of the year to year variance in number of crashes at each intersection. Using the hypothetical data given in Table 1, we obtain $\hat{\mu} = 1.07$ crashes per 1,000 DEV, $\sigma_b^2 = 0.19$ and $\sigma_w^2 = 0.38$. The weighting factor ϕ is then estimated to be 0.69. Ignoring differences in DEV across intersections 1 and its cohort, the BLP of λ_1 is

$$\tilde{\lambda}_1 = 1.07(1 - 0.69) + 1.04 \times 0.69 = 1.05 \text{ crashes per 1,000 DEV.} \tag{2}$$

In the example, the average variance within intersections divided into the two years available for each intersection is about the same as the variability across intersections and thus the weighted average estimator is close to the mean of the two years of crash data for intersection 1. In other words, the nine additional intersections in the similar cohort do not contribute much additional information over what is contributed by intersection 1 itself. Typically, μ will be estimated from a larger cohort and thus the intersection mean will be shrunken more noticeably towards the cohort mean when scarce information about the intersection of interest is available. The estimator in (1), computed as a weighted average, has been called an *abridged EB* estimator (Hauer et al., 2002a). A *full* EB approach can also be implemented and will be discussed in the next section.

Essentially, implementing the EB method requires two steps. First, engineers identify a set of intersections that can be considered to be similar enough to the intersection of interest, and from data available for those intersections, estimate μ. In the example above, μ was computed just as a mean, but in real applications engineers estimate an SPF. The SPF is then used to estimate the expected number of crashes at intersections just like the one of interest. In the example above, the SPF might be a very simple regression function (linear or nonlinear) such as:

$$\log(\mu) = \alpha_0 + \alpha_1 \times \log(DEV), \tag{3}$$

which can also be written as

$$\mu = \exp(\alpha_0) \times DEV^{\alpha_1} \tag{4}$$

as in Hauer et al. (2002a). The values of the parameters (α_0, α_1) are estimated by fitting the linear regression model in (3) to data collected at the similar intersections. For example, using the two years of data available on the nine intersections that are similar to intersection 1 we obtain $\hat{\alpha}_0 = -2.7381$ and $\hat{\alpha}_1 = 0.497$. Thus, the SPF that might be used to calculate the expected mean number of crashes at intersections just like 1 is

$$\mu = \exp(-2.739) DEV^{0.497}. \tag{5}$$

In a year in which intersection 1 experiences volume equal to about 5,600 DEV, we estimate that the expected number of crashes is about 0.9 crashes per 1,000 DEV.

One challenge in applying expression (1) for estimating λ_i is that the 'population' parameter μ is also unknown (we assume for now that the within and between-intersection variances are known). In order to compute the estimator in (1) we need to either introduce an external (to the analysis at hand) estimator for μ (such as that resulting from an SPF calculation) or else must estimate μ together with the expected number of crashes at the intersection of interest.

It is at this juncture where the EB and FB approaches differ. An analyst implementing an EB approach would obtain a point estimate for μ which we denote $\hat{\mu}$ and would estimate λ_i as in (1) (e.g., Hauer et al., 2002a). Once the point estimate $\hat{\mu}$ is obtained, the EB analyst proceeds as if the estimate represents the *true* value of μ, with no uncertainty remaining. In contrast, an analyst implementing a FB approach would account for the uncertainty about the true value of μ by estimating a distribution of *likely* values of μ and then averaging the possible values of λ_i relative to that distribution. We explain this further in the following section.

In the next section, we briefly describe the fundamentals of the FB approach and show that the estimator in (1), an example of an EB estimator when a point estimate $\hat{\mu}$ is used in place of the unknown μ, arises from an FB analysis when certain simplifying assumptions are made.

3. The Bayesian paradigm

Before proceeding, it is useful to present the fundamental ideas behind the Bayesian approach to statistical analysis. All of the concepts presented in this section apply equally to the EB and the FB approaches. In fact, as a preview of coming attractions, we state that the major difference between EB and FB will be in the treatment of the prior parameters μ of Section 2 or α and β mentioned later in this section. Please read on.

As before, let y_i denote an observation, and suppose that the observation is drawn from some distribution (perhaps a Poisson distribution or a Normal distribution, or some other probabilistic distribution) with likelihood function $l(y|\theta)$, where θ is a parameter or vector of parameters indexing the function. For example, if the y are normally distributed, then $\theta = (\mu, \sigma^2)$ (the mean and the variance), whereas if $y \sim$ Poisson then θ represents the usual rate parameter.

Classical statistical inferences are based on the likelihood function. For example, the maximum likelihood estimator (MLE) of θ is the value $\hat{\theta}$ in the parameter space that maximizes the likelihood function. In the classical approach, parameters are *fixed* quantities and the emphasis is on obtaining point estimators for θ and standard errors for $\hat{\theta}$.

A Bayesian model has an additional component: a *prior* distribution that summarizes any knowledge about θ that might be available before observing any data. In the case of traffic safety, when the parameter of interest is the expected number of crashes at an intersection, we might have information to indicate that at certain types of intersections (e.g., those without controls and with high DEV) we can expect more crashes than at others. We use $p(\theta)$ to denote the prior distribution of θ and note that the prior will depend on its own parameters, which we denote μ. In the Bayesian approach, inferences about θ are based on the *posterior* distribution $p(\theta|y)$ of θ, obtained by combining the likelihood with the prior via Bayes' rule:

$$p(\theta|y) = \frac{p(\theta)l(y|\theta)}{C}, \tag{6}$$

where C is just a normalizing constant that does not depend on θ. A possible point estimator for θ is the mean of the posterior distribution $E(\theta|y)$. In the Bayesian context, however, emphasis is not placed on point estimators but on summaries of various kinds of the posterior distribution of the parameter.

An example of interest to traffic engineers is one where data are counts distributed as Poisson variables, a situation similar to the hypothetical example described in Section 2. As in the preceding section, we might wish to estimate the expected number of crashes at an intersection. A reasonable probabilistic model might be the following:

$$\begin{aligned} y &\sim \text{Poisson } (\lambda e), \text{ the likelihood function} \\ \lambda &\sim \text{Gamma } (\alpha, \beta), \text{ the prior distribution for } \lambda. \end{aligned}$$

Here, e denotes *exposure*. For example, if we are interested in number of crashes per 1,000 DEV, the exposure of an intersections with DEV equal to 2,000 is 2 and that corresponding to an intersections with DEV = 500 is 0.5. Expressions for the Poisson and the Gamma distributions are the following:

$$\begin{aligned} l(y|\lambda, e) &\propto \Pi_i \lambda_i^{(d\bar{y}_i)} \exp\{-\lambda_i d\bar{e}_i\} \\ p(\lambda|\alpha, \beta) &\propto \lambda^{\alpha-1} \exp\{-\beta\lambda\}, \end{aligned}$$

where the symbol '$\propto$' stands for 'is proportional to' and $\bar{y}_i$ is the observed mean crash count for intersections i over d years. In our hypothetical example, $d = 2$.

It is useful to recall that the MLE of λ_i is just the observed mean crashes $\bar{y}_i$ at the intersection and that the mean of a Gamma random variable is given by α/β. The quantity $\bar{y}$ is the naive estimator of λ_i based only on the years of data available for the intersection of interest. The prior mean α/β represents our 'best guess' for λ_i *prior* to observing any crash data for the intersection. If we apply Bayes' rule to derive the posterior distribution for λ we obtain:

$$\begin{aligned} p(\lambda|y) &\propto l(y|e, \lambda) \times p(\lambda|\alpha, \beta) \\ &\propto \lambda^{(d\bar{y})} \exp\{-d\bar{e}\lambda\} \lambda^{\alpha-1} \exp\{-\beta\lambda\} \\ &\propto \lambda^{(d\bar{y})+\alpha-1} \exp\{-(d\bar{e} + \beta)\lambda\}. \end{aligned}$$

The posterior distribution $p(\lambda|y)$ has the same general form as the prior distribution, so we conclude that *a posteriori* (after observing data y), the expected number of crashes at the intersection is distributed as a Gamma variable with parameters $(\alpha + d\bar{y}, \beta + d\bar{e})$, and thus can be estimated by the mean of the posterior distribution:

$$\tilde{\lambda} = \frac{\alpha + d\bar{y}}{\beta + d\bar{e}}. \tag{7}$$

Note that $\tilde{\lambda}$ is like a weighted average of $\bar{y}$ and α/β, the sample and prior means, respectively. In fact, it is easy to show that the Bayes estimator of λ will take on a value somewhere between the sample mean and the prior mean, as would be intuitively expected. When the number of years of data available for the ith intersection of interest increases, the prior parameters in (7) get 'swamped' by the term depending on $\bar{y}$ and as a consequence, $\tilde{\lambda} \to \bar{y}$. At the other extreme, if no data were available for the intersection, then $\tilde{\lambda} = \alpha/\beta$, the prior mean.

As we explain in the next section, the prior parameters (α, β) (or some function thereoff) play the role of μ, the parameter estimated from the cohort of similar intersections in Section 2. Notice that in order to implement the Bayesian approach described above, in principle it appears that we need to set values for the two parameters of the prior distribution (α, β). To choose adequate values for these parameters, we might proceed as in Hauer et al. (2002a) and others, and obtain values $(\hat{\alpha}, \hat{\beta})$ perhaps by fitting a model to data of the general form of (3). Thus, an EB approach (abridged or full, see below) is defined in general as the approach in which the parameters of the prior distribution are estimated from existing data.

4. The EB and FB approaches to analyzing traffic safety data

We have already introduced the Poisson model for crash counts. In fact, the Poisson model is a standard probability model for counts y_{ij} such as number of crashes. As is well known, however, the Poisson is somewhat limiting in that the mean and the variance of crash counts are modeled by a single parameter. Typically, crash data exhibit overdispersion (variance is larger than the mean) and thus a more flexible Poisson formulation would include an additional parameter to accommodate the extra variability observed in the sample. A very natural way to formulate such a model is the following. Consider the following two-tier model known as a Poisson-Gamma model already introduced in earlier sections:

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\lambda_i e_{ij}) \\ \lambda_i &\sim \text{Gamma}(\alpha, \beta). \end{aligned} \tag{8}$$

It can be shown (see Appendix A) that while the mean crash number continues to be equal to the expected value (mean) of $\lambda_i \times e$, the variance of crash numbers is now the sum of two components, one that arises from the usual Poisson variability and another one that arises from the second tier model imposed on λ_i. More formally,

$$\begin{aligned} E(y) &= eE(\lambda) \\ Var(y) &= eE(\lambda) + e^2 Var(\lambda). \end{aligned}$$

It can further be shown that the probability distribution of crash numbers y_{ij} computed from the two-tier model presented above is equal to the distribution of a negative binomial (NB) random variable (see Appendix A). Thus, the popular NB model used in traffic safety arises from the Poisson-Gamma model that was just introduced in (8).

Some recently published discussions on statistical modeling of crash data have proposed extensions of the model above for the case of roadway segments. One such extension is proposed for those cases in which the number of segments with crash counts equal to zero in the dataset exceeds the number that would be expected under the Poisson-Gamma model. The general class of models is known as zero-inflated models, and consist essentially of mixtures of Poisson processes with different intensities (e.g., Lee and Mannering, 2002; Shankar et al., 2003; Qin et al., 2004). We do not discuss those models here, but note that except for the added complexity that arises because of the need to estimate the mixing proportions, fitting these zero-inflated models does not differ significantly from fitting regular Poisson-Gamma models of the type addressed in this manuscript.

The Poisson-Gamma model above implies the following:

- Crash counts y_{ij} differ within intersections from year to year.
- Each intersection, at least as its characteristics do not change, has an expected crash rate λ_i.
- The expected number of crashes is different for different intersections, *even if their characteristics are very similar.*
- The expected number of crashes at similar intersections are thought of as draws from a common *population distribution* with its own mean and variance. This is interpreted as follows: the expected number of crashes at a group of similar intersections are deviations from an overall mean expected crashes for intersections with those specific characteristics.

The components of the model shown in (8) accommodate the four statements above as follows. At the highest level of the hierarchy, we assume that the parameters (α, β) index the distribution of the expected crash counts λ_i across intersections and determine the overall expected number of crashes at intersections with a certain set of characteristics. Thus, intersections that are similar in some relevant way differ in their expected crash counts but are assumed to have a common overall mean (in other words, are assumed to be generated by the same process). Each intersection with its own expected crash count will exhibit different number of crashes from one year to the next.

In model (8) we have chosen a Gamma distribution to represent the variability of the individual expected crash counts. If all intersections in our dataset are similar in relevant ways (e.g., have the same geometry, signalization, DEV), then model (8) might well provide an adequate representation of the data. That is, since all intersections belong to the same 'class', then an overall mean expected number of crashes can be conceptualized as representing all intersections.

If, however, our study dataset consists of a set of sites with differences in characteristics that are likely to affect the number of crashes at the site, then it is necessary to add more structure to the model so as to account for the potential systematic effects of covariates such as shoulder width and traffic controls on number of crashes. We now think of groups of similar intersections, where the groups are defined by relevant intersection characteristics.

Consider model (8) but with the following modification:

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\lambda_i e_{ij}) \\ \log(\lambda_i) &= \beta_0 + \log(\mathbf{x}_i)'\beta + c_i, \end{aligned} \tag{9}$$

where now β_0 is the intercept in an ordinary regression model (sometimes fixed at a value called an offset), $\mathbf{x}_i$ is a p-dimensional vector of covariates of safety associated to the ith intersection, β is a p-dimensional vector of unknown regression coefficients, and c_i is a random effect associated to intersection i, that is assumed to have mean zero and some variance σ_c^2. Model (9) is similar to (8) except that at the second level, the implicit model for the expected number of crashes is

$$\log \lambda_i \sim \text{N}(\beta_0 + \log(\mathbf{x}_i)'\beta, \sigma_c^2). \tag{10}$$

An equivalent representation of (10) is

$$\begin{aligned} \lambda_i &= \exp\{\beta_0 + \log(\mathbf{x}_i)'\beta + c_i\} \\ &= \exp\{\beta_0\}\Pi_{k=1}^{p} \exp\{x_{ik}^{\beta_k}\} \exp\{c_i\}. \end{aligned} \tag{11}$$

Note that the expression in (11) is similar to the SPF used for illustration in Hauer et al. (2002a).

The regression coefficients $\beta_1, \beta_2, ..., \beta_p$ indicate the expected change in expected number of crashes at the ith intersection when crash predictors $x_{i1}, ..., x_{ip}$ change. As an example, suppose that the model includes only one predictor, DEV, expressed in the logarithmic scale, where DEV is measured in total number of vehicles entering the intersection during a 24 hour period. Further, for illustration suppose that the regression coefficient associated to $\log DEV$ is equal to 0.5, that the intercept β_0 equals $\log 0.02$ and that the intersection of interest has an average of 4,000 entering vehicles during a day (these numbers are similar to those used for illustration in Hauer et al. (2002a). Then:

$$\begin{aligned} \log \lambda &= \log 0.02 + 0.5 \times \log 4,000 \\ &= 3.91 + 4.15 \\ &= 0.2349. \end{aligned}$$

Therefore, given those values of β_0, β_1, an intersection with 4,000 daily entering vehicles can expect to have about 1.3 crashes per 1,000 DEV.

The second tier in the hierarchical model shown in (10) depends on regression coefficients β, that are typically unknown. In the EB approach the effect of factors such as DEV and signalization on the expected number of crashes is estimated externally (to the study of interest) using SPFs. The estimated regression coefficients $\hat{\beta}$ that arise from fitting the SPFs are then used to determine the expected number of crashes at intersections with certain characteristics (e.g., certain values for the predictors $\mathbf{x}$). Once those 'class' estimates are obtained, they are in turn simply plugged into expressions similar to (1) in order to obtain an estimate of the expected number of crashes for the intersection of interest. In other words, to implement the full EB approach we would begin by formulating a two-tier model as before, and then we would plug in estimated values for the parameters of the second level in the model, where these parameters were estimated from data external to the study.

How would we implement an FB analysis? The difference between EB and FB resides only in the treatment of the hyperparameters (or the parameters of the second-level distributions). While the EB practitioner estimates the parameters from auxiliary data, the FB practitioner sets hyper-prior distributions on those parameters that depend themselves on a set of higher-level parameters. Consider, for example, the model in (9) and suppose that a priori, we have no information about the association between traffic controls, DEV and number of crashes. Since the regression coefficients associated to the various covariates in principle can take on any value on the real line, a reasonable choice for the hyperprior distribution of those parameters is a normal, with some fixed mean value and a large variance.

We refer again to the hypothetical data that was presented in Table 1. There are 20 intersections in the dataset, 10 with no signal controls and 10 with controls. We wish to estimate the expected mean number of crashes for some of the intersections (for example, for intersection 1) and we might also be interested in assessing the effect of traffic controls after controlling for differences in DEV across intersections. A possible model is

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\theta_{ij}) \\ \theta_{ij} &= \lambda_i \times e_{ij} \\ \log(\lambda_i) &= \beta_1 \text{ signal}_i + c_i \\ c_i &\sim \text{Normal}(0, \sigma_b^2). \end{aligned}$$

Priors for high-level parameters (β_1 and σ_b^2) need to be specified as well. Prior distributions can be *informative* or *non-informative*. For example, if we knew that the effect of traffic controls is to reduce the expected log crash rate by 0.3, we might consider choosing a prior distribution for β_1 that is centered at 0.3 and has some variance to reflect uncertainty about that value. On the contrary, if we have no information about the effect of traffic controls on crash rates prior to observing data, then we would choose a prior distribution with any mean and a very large variance, reflecting the fact that a priori, we have no preference for one value of β_1 over others. In the case of a variance, the prior distribution would be chosen from among the probability distributions with support on the positive real line. This is because variances cannot take on negative values. A standard prior distribution for a variance parameter is the Inverted Gamma (IG) distribution with parameters (δ, γ). The hyperparameters (δ, γ) are typically fixed at some numerical value chosen by the investigator to make the prior informative or uninformative, as appropriate in the specific application.

A full specification of the model is then the following:

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\theta_{ij}), \qquad \theta_{ij} = \lambda_i \times e_{ij} \\ \log(\lambda_i) &= \beta_1 \text{ signal}_i + c_i \\ c_i &\sim \text{Normal}(0, \sigma_b^2) \\ \beta_1 &\sim \text{Normal}(m, \tau^2) \end{aligned}$$

$$\sigma_b^2 \sim \text{Inv-Gamma}(\delta, \gamma) \tag{12}$$

Inferences in the Bayesian framework are based on the posterior distributions of the parameters in the model. In model (12) we can draw inferences about the following parameters: $(\lambda_1, ..., \lambda_n, \beta_1, \sigma_b^2)$ assuming that hyper-parameters $(m, \tau^2, \delta, \gamma)$ are fixed at some numeric values a priori. The joint posterior distribution of the model parameters is given by

$$p(\lambda_1, ..., \lambda_n, \beta_1, \sigma_b^2|y) \propto \text{ Normal}(m, \tau^2) \text{ IG}(\delta, \gamma)\Pi_i \text{ Poisson}(\lambda_i e_{ij}). \tag{13}$$

In our example, if $n = 20$ is the number of intersections in the dataset, the dimensionality of the posterior distribution is 22.

In principle it appears that working with a 22-dimensional probability distribution is analytically intractable and very difficult. Recent computational developments (Markov chain Monte Carlo methods, e.g., Gelfand and Smith, 1990; Casella and George, 1992; Gilks et al., 1998; Carlin and Louis, 2000; Gelman et al., 2004) however, permit drawing values of any parameter from its marginal posterior distribution in a very efficient manner. For example, it is very simple to 'draw' 1,000 likely values of λ_1 from its posterior distribution and then use Monte Carlo methods to obtain estimates for quantities such as the posterior mean, standard deviation and percentiles of λ_1. Thus, not only do we obtain point estimators for the parameters in the model, we also obtain the entire distribution of likely values for the parameter which we can then choose to summarize in different ways. We illustrate the implementation of these methods in the next section.

5. A FB analysis of the hypothetical data in Section 2

We now implement a FB analysis of the hypothetical data presented in Section 2. We fit a model like (12) to the data shown in Table 1 and estimate

- The expected number of crashes at each of the intersections.
- The regression coefficient associated to traffic controls.
- The probability that each of the intersections *within its traffic controls group* is ranked in the top three in terms of the expected number of crashes per 1,000 DEV.

We choose prior distributions that reflect ignorance about the values of the hyper-parameters (β_1, σ_b^2). (It is also be possible to incorporate information in this step if information is available.) Prior distributions are:

$$\begin{aligned} \beta_1 &\sim \text{Normal}(0, 1.0E06) \\ \sigma_b^2 &\sim \text{Inverse Gamma}(0.001, 0.001) \end{aligned}$$

Note that the prior variances associated to (β_1 and σ_b^2) are both very large. That means that a priori, we do not consider any possible value of β_1 or σ_b^2 to be more likely than others. Note too that the prior distribution used for σ_b^2 has support only on the positive portion of the real line, consistent with the fact that variances cannot take on negative values.

Intersection	**Mean**	**Std error**	**95% set**
1	1.018	0.1466	(0.757 , 1.352)
2	1.116	0.1819	(0.8481 , 1.58)
3	1.097	0.1956	(0.8157, 1.064)
4	0.9862	0.1465	(0.6985, 1.304)
5	1.04	0.1574	(0.7736, 1.401)
6	0.9288	0.1513	(0.5993, 1.219)
7	1.016	0.1455	(0.7526, 1.356)
8	0.9929	0.1415	(0.7173, 1.306)
9	0.9238	0.1438	(0.6218, 1.203)
10	1.077	0.1786	(0.8035, 1.512)
11	0.51	0.1091	(0.324, 0.7559)
12	0.5154	0.1091	(0.334, 0.7611)
13	0.4826	0.1095	(0.2885, 0.7179)
14	0.5083	0.1102	(0.3226, 0.7558)
15	0.5129	0.1128	(0.3278, 0.7691)
16	0.4723	0.1028	(0.285, 0.6896)
17	0.4951	0.1064	(0.3116, 0.7298)
18	0.4859	0.1019	(0.3056, 0.7055)
19	0.5097	0.1048	(0.3324, 0.7421)
20	0.4966	0.1071	(0.3113, 0.7342)

Table 2: *Posterior mean, standard deviation, and 95% credible set for crash rate* λ.

We fitted the model above using WinBUGS (BUGS Project, 2003). The code used to implement the example is given in Appendix B. Below we present and interpret some of the results. The program produces draws from the posterior distribution of each of the parameters in the model, and given those draws, Monte Carlo methods can then be used to approximate quantities such as the posterior mean and the posterior standard deviation of the parameter.

Consider for example, results presented in Table 2. Each row in the table represents one of the 20 intersections. The column labeled Mean is a point estimate of the crash rate per 1,000 DEV, which we have denoted λ. The second column gives an estimate of the posterior standard deviation of the crash rate per 1,000 DEV for each intersection, and the last column presents a 95% credible set (similar to the classical confidence interval) for the crash rate per 1,000 DEV. For intersection 1, for example, we estimate the crash rate per 1,000 DEV to be equal to 1.018 crashes per 1,000 DEV, with a standard deviation equal to 0.1466. With 95% probability, the crash rate per 1,000 DEV for intersection 1 is between 0.757 and 1.352. The posterior distribution of crash rate (number of crashes per 1,000 DEV) for intersections labeled 1 and 13 are shown in Figures 1 and 2, respectively.

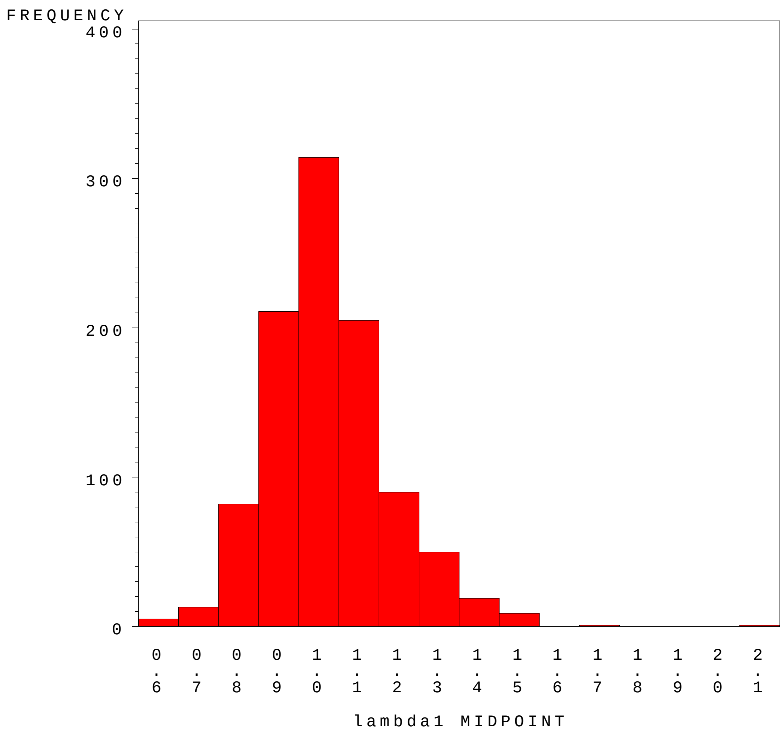

Figure 1: *Posterior distribution of the number of crashes per 1,000 DEV for intersection labeled 1.*

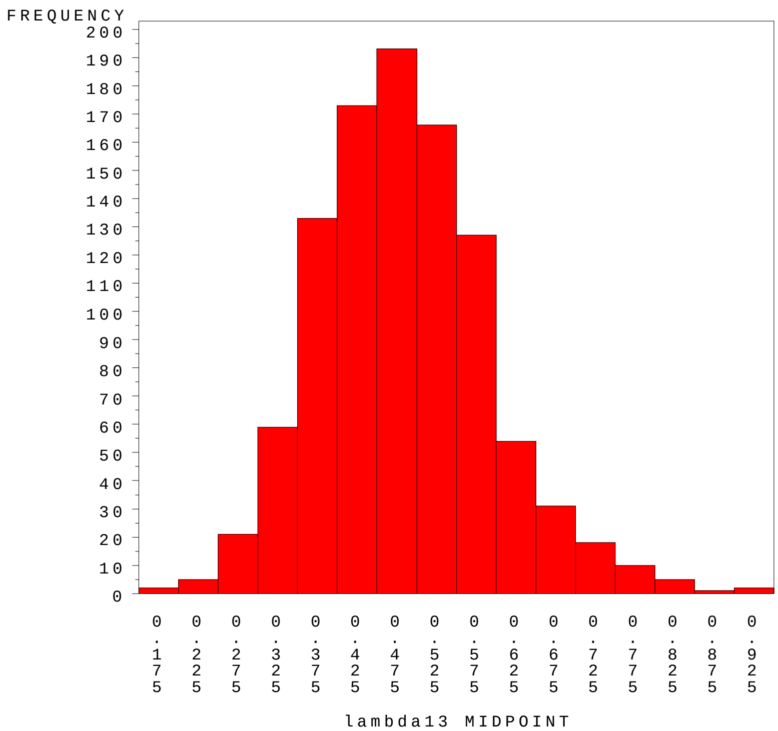

Figure 2: *Posterior distribution of the number of crashes per 1,000 DEV for intersection labeled 13.*

To assess which of the intersections has a worst safety record, we might wish to compute the probability that the crash rate at each intersection is the highest among similar intersections or perhaps can estimate the probability that each of the intersections is ranked in the top three regarding the number of crashes per 1,000 DEV. For example, when we compute those probabilities using the draws for each of the λ's we find that intersection 1 has only a 9.5% probability of being the most dangerous among intersections without controls, and that intersection 11 has an 11% probability of being the worst among intersections with traffic controls. We estimated the probability that each of the intersections in the group lacking traffic controls would rank in the top three in terms of crashes per 1,000 DEV and obtained the following estimates for intersections labeled 1 through 10, respectively: 0.28, 0.53, 0.43, 0.25, 0.31, 0.12, 0.28, 0.23, 0.14 and 0.43. Based on these results, we might conclude that intersections labeled 6 and 9 appear to have the lowest crash rates among intersections with no traffic controls.

The posterior distribution of the regression coefficient associated to traffic controls can also be estimated. For this example, the posterior mean of β_1 was -0.7914 with a standard deviation of 0.1619 and a 95% credible set of $(-1.042, -0.416)$. Since the probability that $\beta_1 \geq 0$ is negligibly small, we are quite confident that those intersections with some form of traffic control have significantly lower crash rates per 1,000 DEV than those without any controls. (As is the case with classical methods, however, cause-effect relations cannot be inferred from mere associations, regardless of the strength of those associations.)

6. Conclusions

We have presented, within the context of an example, the steps that are needed to implement a fully Bayesian (FB) approach for the statistical analyses of traffic safety data. Some similarities and differences between the two approaches are listed below.

- Both EB and FB recognize that similar intersections are likely to have similar (but not identical) expected number of crashes.

- A few years of information on an intersection would result in an unreliable estimate of that intersection's safety, so it is important to 'borrow' information from similar intersections to complement the scarce data. Both EB and FB combine information from all available *exchangeable* (or similar) intersections in order to better estimate intersection-level quantities. The combined estimate typically takes on the form of a weighted average, where the weights depend on the relative amount of information on the intersection of interest.

- In the EB approach, the mean expected number of crashes for a group of intersections with certain characteristics is estimated externally, using SPFs or similar statistical models. The estimated parameters that result from fitting the SPFs are then used as if they were true values. This implies that population-level estimates in the EB approach do not contribute to the uncertainty in the estimate of safety for a specific intersection. Often, EB analysts obtain unrealistically low standard errors for intersection-level estimates.

- In the FB approach, information about similar intersections is also incorporated into the estimate for a single intersection. In this case, however, it is explicitly recognized that the population-level estimates of safety are also uncertain and thus contribute to the variance of the intersection-level estimate λ.

While perhaps somewhat daunting at first sight, the FB approach is likely to be less costly than EB and can be even easier to implement. In addition, it has the potential for resulting in more reliable assessments. One important conclusion from the discussion above is that the significant efforts that are expended in estimating SPFs for a wide array of possible scenarios are not only not required but also not desirable in an FB approach to data analysis. Regardless of the quality of the parameter estimates obtained in the context of fitting an SPF, intersection-level estimates λ obtained via EB methods will in general appear to be more precise than they really are. This is because the uncertainties in the estimates of the parameters in SPFs are typically not accounted for in the estimator for λ. In the worst case, intersection-level estimates might also be biased. Thus our arguing that in the long run, FB methods may well prove to be, in addition, less costly as they do not require estimation of SPFs and other model parameters for inputting into the analysis. A comparison of the exact costs associated to one or the other approaches in the State of Iowa for example, would need to be undertaken within a carefully designed experiment. Here we offer some general reasons that appear to suggest that FB might well be cheaper than EB in most situations.

Implementing a FB approach requires no SPFs and this is one of the advantages of FB over EB. On the down-side, however, practitioners must choose prior distributions for the parameters in the model and ideally those ought to be based on information available about those parameters prior to observing the data. While choosing prior distributions may appear to be suspiciously similar to estimating SPFs, a FB estimate is typically more robust to poor specifications of the prior (as long as data are reasonable abundant) than EB estimates are to poorly estimated SPFs. In other words, as long as data on a reasonably large number of intersections (or road segments) are available over a few years, FB estimates of expected crash numbers or crash rates are often insensitive to different specifications of the prior. Nonetheless, a serious FB analysis would always include an assessment of the sensitivity of results to different prior specifications.

References

BUGS Project. (2003). BUGS Manual. http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml. MRC Biostatics Unit, Cambridge, UK. (date retrieved: November 27, 2003)

Carlin, Bradley P. and Thomas A. Louis. (2000). Bayes and Empirical Bayes Methods for Data Analysis, Second Edition. Chapman & Hall/CRC, Boca Raton, FL.

Casella, George and Edward I. George. (1992). Explaining the Gibbs Sampler. The American Statistician. Vol. 46, Issue 3, 167-174.

Davis, Gary A. (2000). Estimating Traffic Accident Rates While Accounting for Traffic-

Volume Estimation Error: A Gibbs Sampling Approach. Transportation Research Record 1717. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 94-101.

Davis, Gary A. and Shimin Yang. (2001). Bayesian Identification of High-Risk Intersections for Older Drivers via Gibbs Sampling, Transportation Research Record 1746. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 84-89.

Davis, Gary A. and Yihong Gao. (1993). Statistical Methods to Support Induced Exposure Analyses of Traffic Accident Data. Transportation Research Record 1401. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 43-49.

Davis, Gary A. and Yuzhe Guan. (1996). Bayesian Assignment of Coverage Count Locations to Factor Groups and Estimation of Mean Daily Traffic. Transportation Research Record 1542. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 30-37.

Gelfand, Alan, and Adrian S. M. Smith. (1990). "Sampling Based Approaches to Calculating Marginal Densities". Journal of the American Statistical Association, Vol. 85, 398-409..

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. (2004). Bayesian Data Analysis. 2nd edition. Chapman & Hall, London, England.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter. (1998). Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, Boca Raton, FL.

Harwood, D.W., K.M. Bauer, I.B. Potts, D. J. Torbic, K.R. Richard, E.R. Kohlman Rabbani, E. Hauer, and L. Elefteriadou. (2002). Safety Effectiveness of Intersection Left- and Right-Turn Lanes. Report No. FHWA-RD-02-089. Federal Highway Administration (FHWA), Washington, D.C.

Hauer, Ezra. (1986) On the Estimation of the Expected Number of Accidents. Accident Analysis and Prevention. Vol. 18, No. 1, Pergamon Press, Ltd., Great Britain, 1-12.

Hauer, Ezra. (1996a). Detection of Safety Deterioration in a Series of Accident Counts. Transportation Research Record 1542. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 38-43.

Hauer, Ezra. (1996b). Identification of Sites with Promise. Transportation Research Record 1542. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 54-60.

Hauer, Ezra, Douglas W. Harwood, Forrest M. Council, and Michael S. Griffith. (2002a).

Estimating Safety by the Empirical Bayes Method: A Tutorial. Transportation Research Record 1784. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 126-131.

Hauer, E., J. Lovell, and B.N. Persaud. (1986). New Directions for Learning About Safety Effectiveness. Publication No. FHWA-RD-86-015. United States Department of Transportation (US DOT)/Federal Highway Administration (FHWA)/Turner-Fairbank Highway Research Center (TFHRC), McLean, Virginia.

Hauer, Ezra, Jake Kononov, Bryan Allery, and Michael S. Griffith. (2002b). Screening the Road Network for Sites with Promise. Transportation Research Record 1784. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 27-32.

Highway Safety Information System (HSIS). (2001). The Safety Effectiveness of Intersection Left- and Right-Turn Lanes. Highway Safety Information System (HSIS) Summary Report. FHWA-RD-02-031. United States Department of Transportation (US DOT)/Federal Highway Administration (FHWA), McLean, VA.

Higle, Julia L. and James M. Witkowski. (1988). Bayesian Identification of Hazardous Locations. Transportation Research Record 1185. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 24-36.

Higle, Julie and Mari B. Hecht. (1989). A Comparison of Techniques for the Identification of Hazardous Locations. Transportation Research Record 1238. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 10-19.

Lee, J. and F. Mannering. (2002). "Impact of Roadside Features on the Frequency and Severity of Run-Off-Road Accidents: An Empirical Analysis". Accident Analysis and Prevention, Vol. 34, 349-161.

Pawlovich, Michael. (2003). "Evaluating Traffic Safety Network Screening: An Initial Framework Utilizing the Hierarchical Bayesian Philosophy". Unpublished PhD Dissertation, Department of Civil and Construction Engineering, Iowa State University, Ames, IA 50011-1210.

Persaud, Bhagwant N. (1988). Do Traffic Signals Affect Safety? Some Methodological Issues. Transportation Research Record 1185. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 37-47.

Persaud, Bhagwant, Dominique Lord, and Joseph Palmisano. (2002). Calibration and Transferability of Accident Prediction Models for Urban Intersections. Transportation Research Record 1784. Transportation Research Board (TRB)/National Research Council (NRC), Washington, D.C., 57-64.

Persaud, B., E. Hauer, R. Retting, R. Vallurupalli, and K. Mucsi. (1997). Crash Reductions Related to Traffic Signal Removal in Philadelphia. Accident Analysis and Prevention. Vol. 29:6. Elsevier Science, Ltd., Oxford, England, 803-810.

Persaud, Bhagwant and Thu Nguyen. (1998). Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. Transportation Research Record 1635. Transportation Research Board (TRB)/National Research Council, Washington, D.C., 113-120.

Pendleton, Olga J. (1991). Application of New Accident analysis Methodologies Volume I General Methodology. FHWA-RD-90-091. Federal Highway Administration (FHWA), McLean, VA.

Pendleton, Olga J, Ofelia Gonzalez, and Horacio Duarte. (1991). Application of New Accident analysis Methodologies Volume II A Users Manual for BEATS. FHWA-RD-91-014. Federal Highway Administration (FHWA), McLean, VA.

Qin, X., J.N. Ivan, and N. Ravishankar. (2004). "Selecting Exposure Measures in Crash Rate Prediction for Two-Lane Highway Segments". Accident Analysis and Prevention, Vol. 36, 183-191.

Saccomanno, F.F., R. Grosst, D. Green, and A. Mehmood. (2001). Identifying Black Spots Along Highway SS107 in Southern Italy Using Two Models. Journal of Transportation Engineering. Vol. 127, No. 6, November/December 2001, 515-522.

Shankar, V.N., G.F. Ulfarsson, R.M. Pendyala, and M.B. Nebergall. (2003). "Modelling Crashes Involving Pedestrians and Motorized Traffic". Safety Science, Vol 41, 627-640.

Schluter, P.J., J.J. Deely, and A.J. Nicholson. (1997). Ranking and Selecting Motor Vehicle Accident Sites by Using a Hierarchical Bayesian Model. The Statistician. Vol. 46, No. 3. Royal Statistical Society, England, 293-316.

Appendix A: Technical notes

The Negative Binomial model

An integer random variable y is distributed as a Negative Binomial (NB) variable with shaper parameter $\alpha > 0$ and inverse scale parameter $\beta > 0$ if its density function is

$$p(y) = \binom{y+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+y}\right)^{y}.$$

The density $p(y)$ is also the p.d.f. of the random variable Y, the number of failures in a sequence of Bernoulli trials before the occurrence of α successes, where the probability of success in each trial is equal to p. In the parametrization used here,

$$p = \frac{\beta}{\beta+1}$$

which leads to the more familiar representation of the NB density

$$p(y) = \binom{y+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y}.$$

Consider the Poisson-Gamma model described in an earlier section. For y denoting the number of crashes and λ denoting crash rate, let

$$\begin{aligned} y|\lambda &\sim \text{Poisson}(\lambda) \\ \lambda|\alpha,\beta &\sim \text{Gamma}(\alpha,\beta). \end{aligned}$$

The joint distribution of (y, λ) has p.d.f.

$$\begin{aligned} p(y,\lambda) &= p(y|\lambda)p(\lambda) \\ &= \frac{1}{y!}\lambda^{y}\exp\{-\lambda\}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp\{-\lambda\beta\} \\ &= \frac{1}{y!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{y+\alpha-1}\exp\{-\lambda(\beta+1)\}. \end{aligned}$$

The marginal distribution of the number of crashes can be obtained by integrating the joint distribution $p(y|\lambda)p(\lambda)$ with respect to λ to obtain:

$$\begin{aligned} p(y) &= \int_0^{\infty} p(y|\lambda)p(\lambda)d\lambda \\ &= \frac{1}{y!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\int \lambda^{y+\alpha-1}\exp\{-\lambda(\beta+1)\}d\lambda \\ &= \frac{1}{y!}\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+y}\right)^{y} \end{aligned}$$

Notice that for (α, y) integer variables, the following is true

$$\begin{aligned} \Gamma(y+\alpha) &= (y+\alpha-1)! \\ \Gamma(\alpha) &= (\alpha-1)!, \end{aligned}$$

and therefore

$$\begin{aligned}
\frac{1}{y!}\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)} &= \frac{(y+\alpha-1)!}{(\alpha-1)!y!} \\
&= \binom{y+\alpha-1}{\alpha-1.}
\end{aligned}$$

We have demonstrated that the NB model can be viewed as arising from a mixture of Poisson random variables where the mixing distribution reflects uncertainty about the value of the Poisson rate. In the case of traffic safety applications, the NB model makes perfect sense; it arises when crash counts (modeled as Poisson random variables) exhibit extra-Poisson variability represented in the model through a distribution for the Poisson rate λ.

Note that

$$\begin{aligned}
\mathrm{E}(y) &= \alpha\left(\frac{1-p}{p}\right) = \frac{\alpha}{\beta} = \mathrm{E}(\lambda) \\
\mathrm{Var}(y) &= \alpha\left(\frac{1-p}{p^2}\right).
\end{aligned}$$

Note too that the variance of y is larger than its mean since $p \in (0,1)$ and that it can be written as:

$$\begin{aligned}
\alpha\left(\frac{1-p}{p^2}\right) &= \frac{\alpha}{\beta^2}(\beta+1) \\
&= \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} \\
&= \mathrm{E}(\lambda) + \mathrm{Var}(\lambda).
\end{aligned}$$

Appendix B: Notes on computational issues

We used Markov chain Monte Carlo methods (MCMC) to draw samples from the posterior distributions of interest. We implemented the calculations using WinBUGS, a software freely available from http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml. The BUGS code used to obtain the samples of the 20 λ's and of the regression coefficient β_1 is listed below. The BUGS syntax is very similar to SPlus or R syntax.

We ran three independent Markov chains for each of the parameters in the model for 20,000 iterations. We monitored convergence using the Gelman-Rubin statistic $\sqrt{R}$ (Gelman et al., 2004) and also using visual approaches such as observing trace plots. We thinned the chains using a factor of 10 and discarded the first 10,000 iterations in each chain as burn-in runs. Thus, inferences are based on samples of size 3,000 for each of the parameters in the model.

```
model{
  for(i  in 1:N){
      for (j in 1:M){
         y[i,j] ~ dpois(theta[i,j])
         theta[i,j] <- lambda[i] * e[i,j]
         e[i,j] <- dev[i,j] / 1000
                          }
      log(lambda[i]) <- beta1*signal[i] + c[i]
      c[i] ~ dnorm(0.0, tau)
                                    }
  beta1 ~ dnorm(0.0, 1.0E-6)
  tau ~ dgamma(0.01, 0.01)
  sigma <- 1/ sqrt(tau)
             }
Data
list(N = 20, M = 2)

y[ ,1] y[ ,2] dev[ ,1] dev[ ,2]  signal[]
10  2     5766  5581       0
 8  6     4307  4743       0
 6  2     2623  1941       0
 2  6     4240  4835       0
 2  9     4844  4625       0
 0  2     3178  2422       0
 6  5     5162  5315       0
 7  3     5420  5250       0
 2  4     5392  4876       0
 3  6     2987  3169       0
 6  0     5725  4581       1
 4  2     4207  4834       1
```

```
 1  0     2520  1981       1
 1  4     4270  4235       1
 2  3     3844  3625       1
 0  2     4178  4422       1
 2  2     4160  4355       1
 4  1     6420  5850       1
 4  3     6390  5856       1
 2  1     2980  3248       1
END

Inits
list( beta1 = 0, tau = 0.1)
list( beta1 = 1, tau = 0.5)
list( beta1 = 2, tau = 1.0)
```

The posterior probabilities of rankings of intersections were obtained from the posterior draws in two steps. First, we estimated the ranking of each of the 20 intersections in each of the 3,000 draws. This was done using the following PERL script (T. Peterson, personal communication):

```
open(IN,"<c:\\lambdas.dat");
open(OUT,">c:\\lambdas.rank");

while(<IN>) {

     chop;
     @row = split(/ /,$_);
     $index = 1;
     foreach $value (sort @row) {
        $row_sort{$value} = $index++;
     }
     foreach $value ( @row) {
        printf OUT "%3d", $row_sort{$value}," ";
     }
     print OUT "\n";
}

close(IN);
close(OUT);
exit(0);
```

The ranks were then sorted and ranked themselves using standard functions in SAS to sort, rank and identify minima, maxima or percentiles in a numerical list.