Bayesian Analysis of Road Accidents: A general framework for the multinomial case*

Denis Bolduc† and Sylvie Bonin‡

March 3, 1998

Contents

---

*This research extends the work that we performed with the support of the *Programme d'Action concertée de soutien à la recherche en sécurité routière* jointly financed by the *ministère des Transports du Québec, la Société de l'assurance automobile du Québec* and *le Fonds pour la formation des chercheurs et l'aide à la recherche FCAR*. We would like to thank prof. Ben Heydecker for his input in the beginning of this project.

†GREEN, Département d'Économique, Université Laval, Ste-Foy, Canada, G1K 7P4.
‡GRIMES, Département d'Aménagement,Université Laval, Ste-Foy, Canada, G1K 7P4.

RÉSUMÉ

Habituellement, la détection des sites d'accidents routiers dangereux est effectuée à partir de méthodes de bayes empiriques appliquées à des taux d'accidents et/ou des proportions d'accidents qui se sont produits dans des conditions données. Ces méthodes comparent les taux et proportions observés avec ceux qui se produisent normalement dans un ensemble de sites routiers considérés comme semblables. Les approches existantes exploitent des lois de distribution binomiales. Dans le présent article, nous décrivons une méthodologie générale à information complète pour analyser le niveau de danger des sites routiers, qui fait appel à des distributions multinomiales. La technique proposée, de type bayésienne, permet de traiter simultanément les problèmes d'hétérogénéité déterministe et aléatoire ainsi que celui de la corrélation spatiale attribuable à la proximité ou l'environnement similaire caractérisant les sites à l'étude. Notre cadre méthodologique englobe des approches bayésiennes de pratique courante qui étudient les proportions d'accidents impliquant une caractéristique donnée. Les propriétés et l'intérêt de la nouvelle méthode sont démontrés à l'aide d'un exemple simple basé sur des données d'accidents de la ville de Québec.

ABSTRACT

The detection of dangerous road sites is usually performed using empirical methods which focus on observed accident frequencies and/or proportions of accidents with a given feature. The most widely used detection tools have an empirical Bayes (EB) background. The EB approaches rely on the comparison of frequencies and/or proportions of accidents at a given site with the amounts that would normally occur at similar sites. Currently, analytical techniques for accident proportions describe the number of accidents with a given feature using a binomial distribution. This paper extends to the multinomial case the general EB technique that we recently suggested to analyze road accident proportions. Our proposed approach is a full-information Bayes method that allows for both deterministic and random heterogeneity as well as spatial-correlation among the sites under investigation. The technique can also be used to analyze accident frequencies. An empirical example based on accident data taken from the Québec city database, will serve to demonstrate its usefulness.

1 Introduction

An important aspect in road safety research concerns the development of analytical tools to identify road sites with high risk. Within a context of optimization subject to financial constraints, decisions have to be taken as to which sites should be considered for treatment or safety improvement. The most economically reasonable selection criterion is to select those sites which had the highest accident rate in the preceding year. This is a bad procedure because of the well known regression to the mean problem. Even if no remedial treatment is made, the number of accidents recorded at the same site in the following year will naturally decrease toward its temporal mean. In other word, very high accident rates should be viewed as outliers.

Empirical Bayes (EB) techniques have gained in popularity because it accounts explicitly for the regression to the mean problem and also because it incorporates in the analysis the information about sites considered as similar to the one under investigation (see Hauer 1986; Higle and Witkowsky ,1989 and also Heydecker and Wu ,1991). To implement an EB approach, the analyst must put great care in defining the population of sites to include in the analysis. In order for the approach to make sense, sites should be rather homogeneous; i.e. comparable in terms of characteristics. At one extreme, if sites are selected according to a narrow concept of similarity, the referent population becomes too small to generate accurate estimates. On the other extreme, sites become so different that the amount of information carried for the analysis turns out to be small. The solution taken in Hauer (1992) is to define larger groups of sites and control for deterministic heterogeneity through a multivariate regression based on site specific characteristics, such as traffic flow, for example. After controlling for those differences, sites obviously become more comparable. This makes it possible to save important degrees of freedom in performing statistical analyses.

Although controlling for deterministic heterogeneity is very important, it may be the case that the modeler does not have access to all the important variables that would be required to perform a satisfactory investigation of the problem. Therefore, it is possible that the heterogeneity could not fully be explained deterministically which leaves a certain degree of heterogeneity. This type of heterogeneity is usually accounted for through the error term. Bolduc and Bonin (1997) suggested a full information EB approach which accounts for the presence of random and deterministic heterogeneity as well as spatial correlation. It's limitation is that it only applies within a binomial framework. The present paper extends their general framework to the multinomial case. In Section 2, we suggest a multinomial based approach which makes restrictive assumptions about heterogeneity and spatial autocorrelation. Section 3 describes the general multinomial approach where the assumptions just mentioned are relaxed. The last Section presents the results of a simple application to a Québec city accident database to demonstrate the usefulness of the approach.

2 Standard Analyses of Accident Proportions

In this section we describe a standard EB approach to study accident proportions. We call standard an approach that is correct and simple to implement, but which makes restrictive distributional assumptions about the process generating the data. The more general versions considered in Section 3, are more computer intensive but allow for a lot more flexibility and realism. In particular, two types of heterogeneity and spatial-correlation are assumed to be potentially present. For the convenience of the reader, we first review the original approach of Heydecker and Wu (1991) which is formulated for the binomial case. Then, we proceed with the extension that we propose for the multinomial case. This first extension will still be viewed as standard because of the assumptions made.

2.1 The Binomial Case

2.1.1 The Model

The most basic EB approach to study accident proportions was first formulated in Heydecker and Wu (1991). Their methodology examines proportions of accidents that occurred at a site with a given feature (e.g. proportion of accidents occurring at night, during weekends, head-on collisions,...). The implementation critically depends on the distributional assumptions made about the occurrence of an accident in a particular situation. The model is now presented. The observation at location i which registered x_i accidents with a given feature out of a total of n_i accidents at that site during a given period of time, is assumed to have a binomial distribution with mean parameter θ. We write it as:

$$f(x_i \,|\, n_i, \theta) = \begin{pmatrix} n_i \\ x_i \end{pmatrix} \theta^{x_i}(1-\theta)^{n_i - x_i}, \quad 0 \leq x_i \leq n_i. \tag{1}$$

To model variability among similar sites, the mean θ is postulated to be beta distributed with density:

$$g_b(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, 0 < \theta < 1, \tag{2}$$

where $B(\alpha, \beta) = \{\Gamma(\alpha)\,\Gamma(\beta)\} \;/\; \Gamma(\alpha+\beta)$ denotes the beta function with parameters α and β and $\Gamma(s) = \int_0^\infty e^{-z} z^{s-1} dz$, is the gamma function. The b subscript stands for *before* to emphasize the *a priori* nature of the distribution. A beta distribution is used because it can be mixed conveniently with the binomial. The mean and variance of the beta distribution are computed as follows:

$$E_b(\theta) = \frac{\alpha}{\alpha+\beta} \quad \text{and } V_b(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{3}$$

Combining the two preceding distributions lead to the unconditional binomial-beta distribution for x_i expressed in terms of α and β:

$$h(x_i | n_i, \alpha, \beta) = \begin{pmatrix} n_i \\ x_i \end{pmatrix} \frac{B(\alpha + x_i, \beta + n_i - x_i)}{B(\alpha, \beta)}. \tag{4}$$

To obtain the posterior distribution of θ, apply the Bayes theorem which, using equations (1),(2) and (4), is written as:

$$g_a(\theta \mid n_i, x_i, \alpha, \beta) = \frac{f(x_i | n_i, \theta) \cdot g_b(\theta \mid \alpha, \beta)}{h(x_i | n_i, \alpha, \beta)}.$$

This leads to the following adjusted beta distribution:

$$g_a(\theta \mid \alpha + x_i, \beta + n_i - x_i) = \frac{\theta^{\alpha + x_i - 1}(1-\theta)^{\beta + n_i - x_i - 1}}{B(\alpha + x_i, \beta + n_i - x_i)}, \quad 0 < \theta < 1, \tag{5}$$

where the a subscript stands for *after* or *a posteriori*. The empirical Bayes approach is implemented in two steps. First maximize the log-likelihood of the observed sample defined using equation (4), with respect to the parameters α and β in order to get $\hat{\alpha}$ and $\hat{\beta}$. In the second step, use the posterior distribution displayed in (5) to identify the most dangerous sites.

2.1.2 Bayesian Analysis

The Bayesian analysis is performed using the posterior distribution of θ evaluated at the maximum likelihood (ML) estimates $\hat{\alpha}$ and $\hat{\beta}$. The posterior distribution represents the state of knowledge concerning θ after the observations $(x_1, ..., x_I)$ have been combined with the prior information. The Bayesian estimator of the accident proportion at site i is given by the posterior mean:

$$E_b(\theta \mid i) = \frac{\alpha + x_i}{\alpha + \beta + n_i}. \tag{6}$$

Measures other than point estimates can also be computed to help assessing the degree of hazardousness of the sites under investigation. Let θ^m denote the median proportion of accidents associated with the prior distribution, that is, the state of knowledge before observing the sample $(x_1, ..., x_I)$. It can be found by solving the following integral:

$$\int_{\theta=\theta^m}^{1} g_b(\theta \mid \alpha, \beta)\ d\theta = 0.5. \tag{7}$$

In practice, the last function is evaluated at $\hat{\alpha}$ and $\hat{\beta}$. Given θ^m, two useful probabilities can be computed. The expression

$$\begin{aligned} B_{1i} &= \int_{\theta=\theta^m}^{1} g_a(\theta \mid \alpha + x_i,\ \beta + n_i - x_i)\ d\theta \\ &= \Pr(\theta > \theta^m), \end{aligned} \tag{8}$$

gives the probability that site i is more hazardous that normal, among a population of sites of the same kind. A site can then be viewed as dangerous if its B_{1i} value is larger than a critical bound such as 0.8, for example. The previous interpretation reiterates how important it is to define the population of similar sites properly. Another probability leading to a more conservative decision is:

$$\begin{aligned} B_{2i} &= \int_{\theta'=0}^{1} [\int_{\theta=\theta'}^{1} g_a(\theta \mid \alpha + x_i, \beta + n_i - x_i)\ d\theta\ ]\ g_b(\theta' \mid \alpha, \beta)\ d\theta', \\ &= E_{\theta'}[\Pr(\theta > \theta')]. \end{aligned} \tag{9}$$

In words, it can be interpreted as the average probability that the mean proportion of accidents at a given site is greater than at other sites of the same kind [1].

2.2 The Multinomial Case

We now proceed with an extension to the multinomial case of the approach just described. In the binomial approach, the data is assumed to be binomial while the mean parameter θ is assumed to be beta distributed. The extension involves the use of the multinomial distribution for the accident data and of the Dirichlet distribution for the parameter θ which in this case, is a vector. For convenience, we put in the Appendix the main distributional properties associated with those two distributions.

2.2.1 The Model

We now make the assumption that the accident feature we are focusing on can be of $K+1, K \geq 1$ different types. As an example, consider that one is interested in analyzing the seasonal variations in accident rates. The notation and the model are now introduced. The observation at location i registered x_{ik} accidents with a given feature and type k, $k = 1, ..., K+1$ out of a total of n_i accidents at that site. In our example, x_{ik} would be the number of accidents that occurred during each season k. We use $K+1$ to emphasize that given the total number of accident n_i at site i, the number of accidents for one of the $K+1$ type can be deduced from the K other values of x_{ik}. As a convention, we will always assume that x_{iK+1}, the number of accidents of the last type, will be determined from $x_{iK+1} = n_i - \sum_{k=1}^{K} x_{ik}$. To describe the data generating process, we assume that the K-dimensional vector $x_i = (x_{i1}, ..., x_{iK})$ has a multinomial distribution with mean parameter vector $\theta = (\theta_1, ..., \theta_K)$, $0 < \theta_k < 1$, $k = 1, ..., K$, and n_i, $n_i > 1$. We write it as:

$$f(x_i|\theta, n_i) = \frac{n_i!}{\prod_{k=1}^{K+1} x_{ik}!} \prod_{k=1}^{K+1} \theta_k^{x_{ik}}, \tag{10}$$

[1] The first probability, B_1, is computed using the median value of the prior distribution as the criterion, while B_2 may be viewed as an average B_1 measure with the different $B_1 = \Pr(\theta > \theta')$ being computed over all possible values of θ', not only the unique value $\theta' = \theta^m$. For this reason, B_2 is always closer to 0.5 than is B_1. This explains its more conservative character.

where $x_{iK+1} = n_i - \sum_{k=1}^{K} x_{ik}$ and $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$. This, of course, implies that $0 < \sum_{k=1}^{K} \theta_k < 1$. To model the variability of the accident proportions among sites, θ is assumed to be Dirichlet distributed. The Dirichlet distribution is the multivariate version of the beta distribution. Dirichlet and Multinomial are known to mix conveniently and this explains our choice of distribution for the proportions. Applied to our situation, we have a K-dimensional Dirichlet density with parameter vector $\alpha = (\alpha_1, ..., \alpha_{K+1}),\ \alpha_j > 0,\ j = 1, ..., K+1$, that we write as:

$$g_b(\theta|\alpha) = \mathbf{1}(0 < \textstyle\sum_{k=1}^{K} \theta_k < 1) \cdot d(\alpha) \cdot \displaystyle\prod_{k=1}^{K+1} \theta_k^{\alpha_k - 1}, \tag{11}$$

where $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$ and also, $d(\alpha) = \Gamma(\sum_{k=1}^{K+1} \alpha_k) \,/\, \prod_{k=1}^{K+1} \Gamma(\alpha_k)$. Again, $\Gamma(\cdot)$ denotes the gamma function. In the last equation, $\mathbf{1}(0 < \sum_{k=1}^{K} \theta_k < 1)$ is an indicator that is equal to 1 if the condition inside the parentheses is satisfied and 0 otherwise. To simplify the notation, we now write this indicator function as : $\mathbf{1}(\circ)$. The mean vector and covariance matrix of θ can be computed as:

$$E_b[\theta_k] = \frac{\alpha_k}{\sum_{j=1}^{K+1} \alpha_j}, \qquad V_b[\theta_k] = \frac{E[\theta_k](1 - E[\theta_k])}{1 + \sum_{j=1}^{K+1} \alpha_j},$$

$$C_b[\theta_k, \theta_l] = \frac{-E[\theta_k] E[\theta_l]}{1 + \sum_{j=1}^{K+1} \alpha_j}, \quad k, l = 1, ..., K. \tag{12}$$

As indicated in the Appendix, when both equations (10) and (11) are combined, one obtains the Multinomial-Dirichlet distribution. It corresponds to the unconditional distribution of the $x_i = (x_{i1}, ..., x_{iK})$ expressed in terms of only the parameters $\alpha = (\alpha_1, ..., \alpha_{K+1})$ and the total number of accidents n_i. We write is as:

$$h(x_i|\alpha, n_i) = \frac{n_i!}{\prod_{k=1}^{K+1} x_{ik}!} \cdot \frac{\Gamma(\sum_{k=1}^{K+1} \alpha_k)}{\Gamma(\sum_{k=1}^{K+1} \{\alpha_k + x_{ik}\})} \cdot \prod_{k=1}^{K+1} \frac{\Gamma(\alpha_k + x_{ik})}{\Gamma(\alpha_k)}. \tag{13}$$

Our empirical Bayes implementation suggests to retain the value $\hat{\alpha}$ of α vector which maximizes equation (13) given the I observations of the K-dimensional vector $x_i, i = 1, ..., I$. As starting values for α in this estimation process or even as an alternative estimation, one could use the solution of a method of moments (MM) applied on the following relationships (or a subset of it) :

$$E[x_{ik}] \;=\; n\, p_k, \quad p_k = \frac{\alpha_k}{\sum_{j=1}^{K+1} \alpha_j}$$

$$
\begin{aligned}
V[x_{ik}] &= \frac{n+\sum_{j=1}^{K+1}\alpha_j}{1+\sum_{j=1}^{K+1}\alpha_j}\, n\, p_k\,(1-p_k) \\
C[x_{ik},x_{il}] &= -\frac{n+\sum_{j=1}^{K+1}\alpha_j}{1+\sum_{j=1}^{K+1}\alpha_j}\, n\, p_k\, p_l, \qquad k,l=1,...,K. \qquad (14)
\end{aligned}
$$

As is well known, under general conditions, the MM provides consistent estimates. Given a value $\ddot{\alpha}_{K+1}$ of α_{K+1}, that can be found using one of the equations for $V[x_{ik}]$, a simple MM estimation $\tilde{\alpha}_1,...,\tilde{\alpha}_K$ of $\alpha_1,...,\alpha_K$ would exploit the K expressions for the means in (14) to give:

$$
\tilde{\alpha}_k = \ddot{\alpha}_{K+1}\,\frac{\bar{x}_k}{\bar{x}_{K+1}}, \quad k=1,...,K,
$$

where $\bar{x}_k$ denotes the average number of accidents of type k among the set of sites under study. Of course, this defines a non-linear recursion that needs to be solved iteratively. The advantage of the last relationship is that it produces a MM solution that can be obtained using a criterion concentrated with respect to the first K coefficients. The maximum likelihood (ML) estimation is certainly more involved than this easily implemented approach but it is known to lead to efficient estimates.

In order to derive the posterior distribution for θ, one can apply the Bayes theorem using equations (10), (11) and (13). It can be shown that it leads to the following adjusted Dirichlet distribution:

$$
g_a(\theta|\alpha+x_i,n_i) = \mathbf{1}(\circ)\,\cdot d(\alpha+x_i)\,\cdot\prod_{k=1}^{K+1}\theta_k^{\alpha_k+x_{ik}-1}, \qquad (15)
$$

where $d(\alpha+x_i)=\Gamma(\sum_{k=1}^{K+1}\{\alpha_k+x_{ik}\})\,/\,\prod_{k=1}^{K+1}\Gamma(\alpha_k+x_{ik})$. Therefore, the mean vector and covariance matrix associated with this posterior distribution can be computed as:

$$
\begin{aligned}
E_a[\theta_k|i] &= \frac{\alpha_k+x_{ik}}{\sum_{j=1}^{K+1}\{\alpha_j+x_{ij}\}},\; V_a[\theta_k|i] = \frac{E[\theta_k](1-E[\theta_k])}{1+\sum_{j=1}^{K+1}\{\alpha_j+x_{ij}\}}, \\
C_a[\theta_k,\theta_l|i] &= \frac{-E[\theta_k]E[\theta_l]}{1+\sum_{j=1}^{K+1}\{\alpha_j+x_{ij}\}}, \quad k,l=1,...,K. \qquad (16)
\end{aligned}
$$

2.2.2 Bayesian Analysis

The EB estimate of the accident proportion at site i and of type k is:

$$
E_a[\theta_k|i] = \frac{\alpha_k+x_{ik}}{\sum_{j=1}^{K+1}\{\alpha_j+x_{ij}\}} \qquad (17)
$$

evaluated at $\hat{\alpha}$, the ML estimate computed in the first step. As described in the previous section, measures that we called B_{1i} and B_{2i} probabilities in the previous section, can be used to assess the degree of hazardousness of a site. Before we provide the formulas required for the analysis, it is important to mention a very convenient property associated with the Dirichlet. According to property 2 in the Appendix, if $\theta = (\theta_1, ..., \theta_K)$ is Dirichlet with parameter $\alpha = (\alpha_1, ..., \alpha_{K+1})$, then the marginal distribution of $\theta^{(L)} = (\theta_1, ..., \theta_L)$, $L < K$, is Dirichlet with parameter $\alpha^{(L)} = (\alpha_1, ..., \alpha_L, \sum_{k=L+1}^{K+1} \alpha_k)$. This implies that, once the full parameter vector α is estimated using the multinomial approach, all kinds of analysis can be performed about θ and subsets of θ_k's. In particular, this result implies that a given proportion θ_k associated with a single type k will be Dirichlet distributed with parameter vector $(\alpha_k, [\sum_{j=1}^{K+1} \alpha_j] - \alpha_k)$. As indicated in property 3 of the Appendix, the Dirichlet then reduces to the beta distribution. This implies that the binomial setting described in Section 2 is covered as a special case of the current approach.

As seen in equation (7), the calculation of the B_{1i} and B_{2i} probabilities require that median proportions be evaluated. In the multinomial setting, this implies that K such values $\theta_k^m, k = 1, ..., K$ must be found. The last property about the marginals from the Dirichlet distribution makes it simple to do because each θ_k^m can be found by solving :

$$\int_{\theta_k=\theta_k^m}^{1} g_b(\theta_k \mid \alpha_k,\, \alpha_{-k})\; d\theta_k = 0.5, \tag{18}$$

where by convention, $\alpha_{-k} = [\sum_{j=1}^{K+1} \alpha_j] - \alpha_k$, i.e. the sum of all α's with the exception of α_k. Calculations are assumed to be made at the maximum likelihood value $\hat{\alpha}$ of α. Note also that the $g_b(\cdot)$ marginal density considered in (18) corresponds to the beta density function. Given the K median proportion values, the multinomial extension of equation (8) is:

$$\begin{aligned} B_{1i} &= \int_{\theta_1=\theta_1^m}^{1} \cdots \int_{\theta_K=\theta_K^m}^{1} \mathbf{1}(\circ)\, g_a(\theta \mid \alpha + x_i, n_i)\; d\theta_1 ... d\theta_K, \\ &= \Pr(\theta_1 > \theta_1^m, ..., \theta_K > \theta_K^m), \end{aligned} \tag{19}$$

which is the probability that the proportion of accident with a given feature and for each type $k, k = 1, ..., K$ is greater than normal in a population of similar sites. Recall that $\mathbf{1}(\circ)$ is an idicator function which is equal to 1 if $0 < \sum_{k=1}^{K} \theta_k < 1$ and 0 otherwise. In an example where the accident type denotes the season, the B_{1i} value would be large for sites with a large accident proportions in each of the seasons. This integral is of dimension K, and as long as the level of integration is not more than 4, it can be computed numerically, otherwise, one would have to simulate it. The result about the marginals implies that the analysis can easily be performed for one type at a time using the beta distribution and using low level integrals for subsets of θ_k's. In particular, the B_{1i} value focusing on the accident that occurred at site i during spring and summer would be evaluated as:

$$B_{1i} = \int_{\theta_2=\theta_2^m}^{1} \int_{\theta_3=\theta_3^m}^{1} \mathbf{1}(\circ)\, g_a(\theta \mid \alpha_2 + x_{i2}, \alpha_3 + x_{i3}, \{\alpha_1 + x_{i1} + \alpha_4 + x_{i4}\}\,)\; d\theta_2\, d\theta_3,$$

where the calculation would be performed at $\hat{\alpha}$. Another probability useful for the analysis is the more conservative B_{2i} probability. Using the prior and posterior densities in equations (11) and (15), this measure could be computed as:

$$\begin{aligned} B_{2i} &= \int_{\theta_1'=0}^{1} \cdots \int_{\theta_K'=0}^{1} [\int_{\theta_1=\theta_1'}^{1} \cdots \int_{\theta_K=\theta_K'}^{1} \mathbf{1}(\circ)\, g_a(\theta|\cdot)\; d\theta_1 ... d\theta_K]\; \mathbf{1}(\circ)\, g_b(\theta'|\alpha) d\theta' \;, \\ &= E_{\,\theta'}[\,\Pr(\theta_1 > \theta_1', ..., \theta_K > \theta_K')\,]\,, \end{aligned} \tag{20}$$

which corresponds to the average B_{1i} value obtained when using all possible values of θ', not only the median vector $(\theta_1^m, ..., \theta_K^m)$. Of course, except for cases with $K \leq 2$, this integral of dimension $2K$ would have to be simulated. The numerical complexities associated with the computation give B_{1i} an advantage over B_{2i}. This statement will be reinforced in the general version with heterogeneity and spatial correlation, that we now describe.

3 The General Approach

We now extend the model just presented to allow for deterministic and random heterogeneity as well as spatial correlation among the sites investigated. As it was the case in Bolduc and Bonin (1997) for the binomial case, those effects are being handled in the multinomial case through the parameters α_k,

$k = 1, .., K+1$ involved in the prior distribution (11). Recall from equation (12) that the first two moments of the Dirichlet distribution are simple transformation of the α parameter vector and it is through this channel that the generalities are introduced.

3.1 The Model

Deterministic and random heterogeneity are accounted for by allowing the α vector to become site specific. By assumption, each site is associated with a random parameter vector $\alpha_i = (\alpha_{i1}, ..., \alpha_{iK+1})$ with components defined as:

$$\alpha_{ik} = \exp(\mathbf{z}_{ik}\,\varphi_k + \sigma_k \varepsilon_{ik}), \qquad k = 1, ..., K+1, \tag{21}$$

where $\mathbf{z}_{ik}$ is a row vector with as many columns as there are elements in the φ_k vector of coefficients associated with type k. The component $\mathbf{z}_{ik}\,\varphi_k$ serves to explain the deterministic heterogeneity using site specific information. The other component ε_{ik}, is a normally distributed error term with zero mean intended to capture both the random heterogeneity (that is the heterogeneity that remains unexplained by the deterministic component) and the spatial correlation across the sites, while σ_k is a standard deviation term to control for scale effects present across the types considered. The $\exp(\cdot)$ transformation is used to insure the positivity of all the α_{ik}'s. By assumption, the ε_{ik} may be affected by spatial correlation among sites. The individual terms specific to a given type k are assumed to arise from the following first-order spatial autoregressive process:

$$\varepsilon_k = \rho_k W_k \varepsilon_k + \xi_k = (I_I - \rho_k W_k)^{-1} \xi_k, \qquad k = 1, ..., K+1, \tag{22}$$

where ε_k is a I-dimensional vector, ρ_k is a spatial correlation parameter such that $-1 \leq \rho_k \leq 1$, I_I is a $(I \times I)$ identity matrix and W_k is a $(I \times I)$ weighting matrix depicting the relationships between sites which is specific to type k. The component ξ_k is a $(I \times 1)$ vector of standard normal random variates. A very simple form for W_k is defined as: $w_{i,j} = 1$ for sites i and j that are neighbors, and $w_{i,j} = 0$, otherwise, for $i = 1, ..., I$, and $j = 1, ..., I$. More general versions are considered in Bolduc and Bonin (1997). The focus here is more on the multinomial aspect than the spatial correlation itself.

We now make some notational simplifications in order to produce the different formulas required for the analysis. To incorporate explicitly the spatial correlation among the I sites modeled with (22) into the α_{ik} of a given site i, we use the notation:

$$\alpha_{ik} = \alpha_{ik}(\mathbf{z}_{ik}\,, \varphi_k, \sigma_k, \rho_k, \xi_k). \tag{23}$$

In terms of a data generating process, this last equation implies that: 1) a ξ_k vector of standard normal variates is drawn; 2) given W_k and given a value of ρ_k, a $(I \times 1)$ vector ε_k arises from equation (22); 3) given the ε_{ik} that applies for site i, the α_{ik} value is computed from equation (21). As a final notational convention, we call ξ the $([K+1]I \times 1)$ vector obtained from the vertical concatenation of the $K+1$ different $(I \times 1)$ vectors ξ_k. We will denote the joint normal density of ξ as $n(\xi)$. By assumption, it is $N(0, I_{[K+1]I})$.

Adapting the notation to allow for site specific proportions θ_{ik}, involves rewriting the equations in Section 2.2 adding a i subscript to θ_k and α_k. Of course, conditional on given values of α_{ik}, all formulas in Section 2.2 continue to hold. In the following, we exploit this fact. With the assumptions made, the mean of the prior distribution of θ_{ik} can be computed as:

$$\begin{aligned} E_b[\theta_{ik}] &= E_\xi\ [E_b(\theta_{ik} \mid \xi)] \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} E_b(\theta_{ik} \mid \xi)\ n(\xi)\, d\xi \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\alpha_{ik}(\mathbf{z}_{ik}\,, \varphi_k, \sigma_k, \rho_k, \xi_k)}{\sum_{l=1}^{K+1} \alpha_{il}(\mathbf{z}_{il}\,, \varphi_l, \sigma_l, \rho_l, \xi_l)} n(\xi)\, d\xi. \end{aligned} \tag{24}$$

All kinds of situations are covered by this last equation which involves KI-dimensional integrals. In the situation where σ_k, $k = 1, ..., K+1$ are all zero, the integrals disappear from the last equation and the prior mean becomes:

$$E_b[\theta_{ik}] = \frac{\exp(\mathbf{z}_{ik}\, \varphi_k)}{\sum_{l=1}^{K+1} \exp(\mathbf{z}_{il}\, \varphi_l)}, \qquad k = 1, ..., K, \tag{25}$$

which refers to a model with only deterministic heterogeneity. Because the expression in equation (24) is an expectation, for practical purposes it will be evaluated using an average of $E_b(\theta_{ik} \mid \xi)$ taken over R independent draws ξ^r of ξ. This type of simulator has proved to be very reliable in many previous applications, even when R is rather small. The simulator for $E_b[\theta_{ik}]$ that we denote as $\bar{E}_b[\theta_{ik}]$ is calculated as:

$$\bar{E}_b[\theta_{ik}] = \frac{1}{R} \sum_{r=1}^{R} \frac{\alpha_{ik}(\mathbf{z}_{ik}\,, \varphi_k, \sigma_k, \rho_k, \xi_k^r)}{\sum_{l=1}^{K+1} \alpha_{il}(\mathbf{z}_{il}\,, \varphi_l, \sigma_l, \rho_l, \xi_l^r)}. \tag{26}$$

Of course, to implement the Bayesian analysis, one needs fitted values for the unknown parameters $\varphi_k, \sigma_k, \rho_k$, $k = 1, ...K+1$ that we incorporate in a joint vector γ of right dimension. The current empirical Bayes implementation involves selecting the value of the parameters that maximize the following likelihood function:

$$h(x_i|\gamma, n_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} h(x_i|\gamma, n_i, \xi)\, n(\xi)\, d\xi \tag{27}$$

$$= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} [\frac{n_i!}{\prod\limits_{k=1}^{K+1} x_{ik}!} \cdot \frac{\Gamma(\sum\limits_{k=1}^{K+1} \alpha_{ik}(\xi))}{\Gamma(\sum\limits_{k=1}^{K+1} \alpha_{ik}(\xi) + x_{ik})} \prod_{k=1}^{K+1} \frac{\Gamma(\alpha_{ik}(\xi) + x_{ik})}{\Gamma(\alpha_{ik}(\xi))}]\, n(\xi)\, d\xi,$$

where $\alpha_{ik}(\xi)$ is a short notation for $\alpha_{ik}(\mathbf{z}_{ik}, \varphi_k, \sigma_k, \rho_k, \xi_k)$. Because the integral is multidimensional, in practice, it is replaced with the empirical mean:

$$\bar{h}(x_i|\gamma, n_i) = \frac{1}{R} \sum_{r=1}^{R} h(x_i|\gamma, n_i, \xi^r).$$

To implement the maximum likelihood estimation with $h(x_i|\gamma, n_i)$ replaced with $\bar{h}(x_i|\gamma, n_i)$ is known as maximum simulated likelihood (MSL) estimation. MSL estimation has well known properties and it usually performs very well. As noted by a referee, one possibility would be to apply the current setting to the moment conditions in equation (14). This would obviously lead to a method of simulated moments (MSM) which also produces estimators with well known properties.

3.2 Bayesian Analysis

To account for the randomness introduced in the α_{ik}'s, the formulas for B_{1i} and B_{2i} in equations (19) and (20) have to be adjusted accordingly. For given values of ξ, the conditional posterior density function of θ_{ik} can be computed as:

$$g_a(\theta_{ik}|\alpha_i + x_i, n_i, \xi) = \mathbf{1}(\circ)\ d(\alpha_i(\xi) + x_i) \cdot \prod_{k=1}^{K+1} \theta_{ik}^{\alpha_{ik}(\xi)+x_{ik}-1}, \tag{28}$$

where $d(\alpha_i(\xi) + x_i) = \Gamma(\sum_{k=1}^{K+1}\{\alpha_{ik}(\xi) + x_{ik}\}) / \prod_{k=1}^{K+1} \Gamma(\alpha_{ik}(\xi) + x_{ik})$. The analysis should be performed using the unconditional density function:

$$g_a(\theta_{ik}|\alpha_i + x_i, n_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g_a(\theta_{ik}|\alpha_i + x_i, n_i, \xi)\ n(\xi)\, d\xi. \tag{29}$$

Also, the K median values θ_{ik}^{m} should be obtained for each site i and each type k, by solving the following equation:

$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} [\int_{\theta_{ik}=\theta_{ik}^{m}}^{1} g_b(\theta_{ik} \mid \alpha_{ik}(\xi),\, \alpha_{i,-k}(\xi))\; d\theta_{ik}]\, n(\xi)\, d\xi = 0.5,$$

which is just equation (18) adapted to account for the randomness of ξ. In practice, with high dimensional integrals, this is replaced with the simulated function computed as:

$$\frac{1}{R}\sum_{r=1}^{R} \int_{\theta_{ik}=\theta_{ik}^{m}}^{1} g_b(\theta_{ik} \mid \alpha_{ik}(\xi^r),\, \alpha_{i,-k}(\xi^r))\; d\theta_{ik} \; = 0.5.$$

Given the median values θ_{ik}^{m}, $k = 1, ..., K$, the two required probabilities can then be computed as:

$$B_{1i} \;=\; \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} B_{1i}(\cdot|\xi)\; n(\xi)\; d\xi = E_{\xi}[B_{1i}(\cdot|\xi)] \tag{30}$$

and

$$B_{2i} \;=\; \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} B_{2i}(\cdot|\xi)\; n(\xi)\; d\xi = E_{\xi}[B_{2i}(\cdot|\xi)], \tag{31}$$

where:

$$B_{1i}(\cdot|\xi) \;=\; \int_{\theta_{i1}=\theta_{i1}^{m}}^{1} ... \int_{\theta_{iK}=\theta_{iK}^{m}}^{1} \mathbf{1}(\circ)\; g_a(\theta_{ik}|\alpha_i + x_i, n_i, \xi)\; d\theta_{i1}...d\theta_{iK},$$

and

$$B_{2i}(\cdot|\xi) \;=\; \int_{\theta_1'=0}^{1} ... \int_{\theta_K'=0}^{1} [\int_{\theta_{i1}=\theta_1'}^{1} ... \int_{\theta_{iK}=\theta_K'}^{1} \mathbf{1}(\circ)\; g_a(\theta_{ik}|\cdot)\; d\theta_{i1}...d\theta_{iK}] \;\; \mathbf{1}(\circ)\;\; g_b(\theta'|\alpha)d\theta'\;,$$

where $g_a(\theta_{ik}|\cdot)$ is defined as in equation (28). Note that in practice, the expectations in (30) and (31) are replaced with empirical means computed using R values of $B_{1i}(\cdot|\xi^r)$ and $B_{2i}(\cdot|\xi^r)$. Usually, a R value of at least 50 is enough to get approximations of these integrals with good precision. Given the notation in equation (20), one can see that B_{2i} can be interpreted as:

$$B_{2i} = E_{\xi}\; E_{\theta'}\, [\Pr(\theta_{i1} > \theta_1', ..., \theta_{iK} > \theta_K')].$$

As it was the case with the standard multinomial analysis, since the marginals from the Dirichlet are also Dirichlet, the previous formulas can easily be adjusted to analyze the proportions for a given type or for a given subset of types.

4 Application

We now apply the proposed methodology to the Québec city road accident data base. We retained a subset of accidents occurring during a four year period 1990-1993 at 4 leg intersections of comparable roads, to ensure some homogeneity in the data. The reference group includes 90 intersections selected among a set of 224 sites having registered at least sixteen accidents during the period (a minimum of 4 accidents per year). This selection is based on the premise that it would be economically unsound to study sites with very few accidents as a means to improve a socially unacceptable situation. The application focuses on proportions of accidents that occurred during specific periods of the week. Type 1 concerns accidents that took place on a Monday, Tuesday or Wednesday (MTW), type 2 refers to accidents on a Thursday or a Friday (TF) and finally type 3 covers the weekend (SS). The accident database retained for estimation is displayed in Table 1. The accident frequencies reported in terms of proportions are presented in the first set of columns in Table 2. For simplicity, the application is limited to the standard multinomial approach where no heterogeneity and spatial correlation are assumed to be present. Therefore all formulas used are taken from Section 2.2.

Table 3 presents the estimation results obtained in maximizing the log-likelihood function using the unconditional distribution of the accident counts x_{ik} observed at site i for each type k considered. The probability density function used for this purpose is the Multinomial-Dirichlet displayed in equation (13). The α parameters are all significantly different from zero. The coefficient α_1 refers to type 1 which concerns MTW, the two other coefficients concern TF and SS. Given the $\hat{\alpha}_k$'s, the EB analysis can then be implemented. The EB estimator of the accident proportion at site i and of type k, corresponds to the posterior mean displayed in equation (17). Estimated values obtained are produced in the second set of columns in Table 2. We can clearly observe that the regression to the mean correction is effective. Then, the median values θ_1^m, θ_2^m and θ_3^m can be found solving (18). The values obtained are reported at the bottom of Table 2. Given the median values, all kinds of B_1 values can be produced. A first set of B_1 values focuses on the analysis of one type at a time. Those values are produced in the last three columns of Table 2. To be more specific, the last column reports $\Pr(\theta_3 > \theta_3^m \mid i)$, that is:

$$B_{1i} = \int_{\theta_3=\theta_3^m}^{1} g_a(\theta_3 \mid \alpha_3 + x_{i3}, \{\alpha_1 + x_{i1} + \alpha_2 + x_{i2}\}\ )\ d\theta_3,$$

where in this case, the Dirichlet density coincides with a Beta density. The values obtained in the last three columns of Table 2 are very similar to those produced using the binomial approach of Section 2.1. This is certainly an interesting results but the main advantage of the multinomial setting is the ability to produce B1 values for several types taken simultaneously.

In the last column of Table 4, we report the B_1 probability values for the event $\theta_1 > \theta_1^m$, $\theta_2 > \theta_2^m$ which covers the weekdays. Results from the analysis permit to identify the sites which show their highest accident rates during those days. A visual inspection in Table 1 permitted to identify site 20 as problematic. This is confirmed by the EB analysis. In column titled B1_1*B1_2, we compute the same probability assuming that the events occur independently across those two week periods. In other words, instead of exploiting the multinomial setting, the analysis would be performed using series of independent binomial based studies. From the note at the bottom of Table 4, we can see that the site ordering differs between the two approaches. Also the computed probability values are not numerically the same. This provides some indication that the multinomial approach could be preferred in this case. A statistical test to decide between the multinomial and the binomial settings could be devised. Still, since the multinomial setting contains the binomial one as a special case, we prefer to use the more general approach which is, by definition, more flexible and therefore more attractive.

5 Conclusion

In this paper, we describe a methodology to account for site specific heterogeneity and spatial autocorrelation in a full information empirical Bayes framework for road accident analyses using accidents distributed according to a multinomial probability. The generalizations suggested in the present paper are likely to be of great importance and can potentially contribute to reach better decisions regarding the identification of the most dangerous sites. We provide a simple empirical example using the Québec city accident database. The multinomial approach is demonstrated to be very flexible and useful. A more detailed empirical study is obviously required but the main purpose of the present paper was to suggest the technique. The example aims to provide some evidence that the approach is feasible.

TABLE 1: Accident Data

Site	Total	MTW	TF	SS	Site	Total	MTW	TF	SS
5	24	10	10	4	**98**	18	4	6	8
7	22	9	8	5	**100**	19	8	7	4
8	16	8	7	1	**101**	40	18	12	10
11	23	10	6	7	**104**	27	8	14	5
13	22	12	4	6	**107**	24	9	11	4
14	17	7	7	3	**108**	34	16	10	8
16	19	5	9	5	**109**	27	18	7	2
19	19	10	5	4	**110**	30	12	13	5
20	73	36	32	5	**115**	33	21	8	4
21	19	12	5	2	**118**	44	16	16	12
22	65	27	21	17	**120**	30	20	7	3
23	20	6	7	7	**122**	35	17	11	7
25	33	14	13	6	**125**	63	24	24	15
27	28	11	15	2	**126**	22	9	9	4
29	31	9	4	18	**129**	43	19	18	6
30	22	14	7	1	**136**	17	8	7	2
31	18	9	7	2	**137**	20	13	3	4
32	30	17	8	5	**142**	19	8	4	7
34	25	10	10	5	**144**	19	7	8	4
35	23	11	9	3	**146**	16	6	6	4
39	17	10	6	1	**148**	28	11	11	6
41	31	11	3	17	**149**	28	11	8	9
43	16	11	3	2	**150**	22	6	13	3
44	22	10	9	3	**153**	18	14	3	1
46	30	9	11	10	**156**	28	9	8	11
48	20	8	5	7	**157**	16	7	4	5
51	20	10	9	1	**162**	21	6	10	5
55	19	5	5	9	**167**	39	16	14	9
57	30	18	8	4	**172**	25	10	11	4
58	20	15	2	3	**178**	16	7	3	6
60	30	15	10	5	**182**	19	6	9	4
65	17	10	4	3	**185**	31	16	5	10
66	20	14	6	0	**187**	45	28	13	4
69	18	8	3	7	**190**	24	12	8	4
76	17	6	7	4	**191**	20	9	8	3
80	17	4	6	7	**193**	27	17	7	3
83	22	12	7	3	**195**	38	19	11	8
85	29	13	13	3	**196**	16	7	7	2
87	21	8	8	5	**197**	25	11	8	6
88	63	23	25	15	**202**	18	5	9	4
91	46	25	12	9	**203**	25	12	8	5
92	23	7	12	4	**207**	25	8	10	7
93	16	10	4	2	**208**	26	12	9	5
94	18	6	10	2	**214**	42	20	13	9
96	22	9	10	3	**224**	19	9	6	4
					Max:	73	36	32	18
					Avg:	26,38	11,92	8,99	5,47

TABLE 2 : Proportions

	Proportions			Posterior Means			B1 based on Marginals		
Site	**MTW**	**TF**	**SS**	**MTW**	**TF**	**SS**	**MTW**	**TF**	**SS**
5	0,417	0,417	0,167	0,442	0,363	0,195	0,425	0,668	0,405
7	0,409	0,364	0,227	0,441	0,347	0,212	0,415	0,551	0,560
8	0,500	0,438	0,063	0,463	0,362	0,175	0,571	0,650	0,253
11	0,435	0,261	0,304	0,448	0,318	0,234	0,464	0,334	0,738
13	0,545	0,182	0,273	0,478	0,297	0,224	0,678	0,203	0,665
14	0,412	0,412	0,176	0,443	0,357	0,199	0,436	0,618	0,448
16	0,263	0,474	0,263	0,406	0,374	0,220	0,203	0,730	0,627
19	0,526	0,263	0,211	0,471	0,322	0,207	0,627	0,363	0,517
20	0,493	0,438	0,068	0,475	0,395	0,129	0,701	0,907	0,012
21	0,632	0,263	0,105	0,497	0,322	0,181	0,783	0,363	0,296
22	0,415	0,323	0,262	0,433	0,332	0,235	0,332	0,420	0,801
23	0,300	0,350	0,350	0,414	0,343	0,243	0,243	0,521	0,793
25	0,424	0,394	0,182	0,442	0,360	0,197	0,423	0,655	0,428
27	0,393	0,536	0,071	0,433	0,404	0,162	0,359	0,893	0,153
29	0,290	0,129	0,581	0,396	0,267	0,337	0,140	0,067	0,998
30	0,636	0,318	0,045	0,503	0,335	0,162	0,819	0,456	0,159
31	0,500	0,389	0,111	0,464	0,352	0,184	0,579	0,586	0,315
32	0,567	0,267	0,167	0,492	0,316	0,193	0,770	0,310	0,385
34	0,400	0,400	0,200	0,437	0,359	0,204	0,386	0,638	0,492
35	0,478	0,391	0,130	0,460	0,355	0,185	0,554	0,611	0,320
39	0,588	0,353	0,059	0,483	0,344	0,173	0,705	0,523	0,235
41	0,355	0,097	0,548	0,419	0,256	0,325	0,259	0,042	0,996
43	0,688	0,188	0,125	0,503	0,308	0,189	0,811	0,273	0,357
44	0,455	0,409	0,136	0,453	0,360	0,187	0,505	0,642	0,339
46	0,300	0,367	0,333	0,401	0,350	0,250	0,162	0,574	0,846
48	0,400	0,250	0,350	0,439	0,318	0,243	0,406	0,333	0,793
51	0,500	0,450	0,050	0,465	0,369	0,166	0,587	0,702	0,187
55	0,263	0,263	0,474	0,406	0,322	0,272	0,203	0,363	0,921
57	0,600	0,267	0,133	0,503	0,316	0,181	0,830	0,310	0,286
58	0,750	0,100	0,150	0,529	0,279	0,192	0,913	0,121	0,381
60	0,500	0,333	0,167	0,469	0,338	0,193	0,621	0,484	0,385
65	0,588	0,235	0,176	0,483	0,317	0,199	0,705	0,331	0,448
66	0,700	0,300	0,000	0,516	0,331	0,153	0,871	0,425	0,117
69	0,444	0,167	0,389	0,451	0,300	0,249	0,487	0,222	0,826
76	0,353	0,412	0,235	0,430	0,357	0,213	0,347	0,618	0,563
80	0,235	0,353	0,412	0,403	0,344	0,253	0,194	0,523	0,841
83	0,545	0,318	0,136	0,478	0,335	0,187	0,678	0,456	0,339
85	0,448	0,448	0,103	0,451	0,377	0,172	0,490	0,762	0,215
87	0,381	0,381	0,238	0,434	0,352	0,215	0,367	0,583	0,582
88	0,365	0,397	0,238	0,407	0,370	0,223	0,157	0,756	0,691
91	0,543	0,261	0,196	0,493	0,306	0,202	0,797	0,225	0,468
92	0,304	0,522	0,174	0,411	0,392	0,197	0,222	0,836	0,426
93	0,625	0,250	0,125	0,490	0,321	0,189	0,742	0,361	0,357
94	0,333	0,556	0,111	0,424	0,392	0,184	0,310	0,826	0,315
96	0,409	0,455	0,136	0,441	0,372	0,187	0,415	0,725	0,339

98	0,222	0,333	0,444	0,398	0,339	0,263	0,168	0,490	0,887
100	0,421	0,368	0,211	0,445	0,348	0,207	0,446	0,553	0,517
101	0,450	0,300	0,250	0,452	0,324	0,224	0,493	0,368	0,681
104	0,296	0,519	0,185	0,403	0,397	0,200	0,177	0,864	0,448
107	0,375	0,458	0,167	0,430	0,375	0,195	0,340	0,747	0,405
108	0,471	0,294	0,235	0,459	0,324	0,217	0,552	0,367	0,612
109	0,667	0,259	0,074	0,521	0,315	0,164	0,897	0,308	0,166
110	0,400	0,433	0,167	0,435	0,373	0,193	0,368	0,737	0,385
115	0,636	0,242	0,121	0,519	0,305	0,175	0,900	0,237	0,236
118	0,364	0,364	0,273	0,414	0,351	0,235	0,218	0,589	0,773
120	0,667	0,233	0,100	0,526	0,304	0,170	0,916	0,234	0,200
122	0,486	0,314	0,200	0,465	0,331	0,204	0,597	0,424	0,490
125	0,381	0,381	0,238	0,415	0,362	0,223	0,205	0,693	0,691
126	0,409	0,409	0,182	0,441	0,360	0,200	0,415	0,642	0,448
129	0,442	0,419	0,140	0,448	0,374	0,178	0,465	0,763	0,245
136	0,471	0,412	0,118	0,457	0,357	0,186	0,529	0,618	0,336
137	0,650	0,150	0,200	0,503	0,292	0,205	0,817	0,178	0,494
142	0,421	0,211	0,368	0,445	0,309	0,246	0,446	0,275	0,810
144	0,368	0,421	0,211	0,432	0,361	0,207	0,357	0,646	0,517
146	0,375	0,375	0,250	0,436	0,348	0,216	0,386	0,556	0,586
148	0,393	0,393	0,214	0,433	0,358	0,209	0,359	0,634	0,534
149	0,393	0,286	0,321	0,433	0,323	0,244	0,359	0,365	0,809
150	0,273	0,591	0,136	0,403	0,410	0,187	0,185	0,902	0,339
153	0,778	0,167	0,056	0,530	0,300	0,171	0,912	0,222	0,218
156	0,321	0,286	0,393	0,410	0,323	0,267	0,212	0,365	0,917
157	0,438	0,250	0,313	0,449	0,321	0,229	0,478	0,361	0,693
162	0,286	0,476	0,238	0,408	0,377	0,215	0,213	0,752	0,582
167	0,410	0,359	0,231	0,436	0,348	0,216	0,369	0,566	0,608
172	0,400	0,440	0,160	0,437	0,371	0,192	0,386	0,720	0,383
178	0,438	0,188	0,375	0,449	0,308	0,243	0,478	0,273	0,784
182	0,316	0,474	0,211	0,419	0,374	0,207	0,275	0,730	0,517
185	0,516	0,161	0,323	0,475	0,278	0,247	0,664	0,103	0,832
187	0,622	0,289	0,089	0,527	0,318	0,155	0,935	0,317	0,096
190	0,500	0,333	0,167	0,467	0,339	0,195	0,601	0,487	0,405
191	0,450	0,400	0,150	0,452	0,356	0,192	0,496	0,614	0,381
193	0,630	0,259	0,111	0,509	0,315	0,176	0,852	0,308	0,247
195	0,500	0,289	0,211	0,471	0,321	0,208	0,646	0,340	0,530
196	0,438	0,438	0,125	0,449	0,362	0,189	0,478	0,650	0,357
197	0,440	0,320	0,240	0,449	0,335	0,216	0,473	0,455	0,600
202	0,278	0,500	0,222	0,411	0,379	0,210	0,233	0,758	0,540
203	0,480	0,320	0,200	0,461	0,335	0,204	0,561	0,455	0,492
207	0,320	0,400	0,280	0,413	0,359	0,228	0,231	0,638	0,699
208	0,462	0,346	0,192	0,455	0,343	0,202	0,521	0,516	0,470
214	0,476	0,310	0,214	0,463	0,328	0,210	0,581	0,395	0,548
224	0,474	0,316	0,211	0,458	0,335	0,207	0,538	0,457	0,517
Max:	0,778	0,591	0,581	0,530	0,410	0,337	0,935	0,907	0,998
Avg:	0,452	0,340	0,207	0,453	0,341	0,207	0,494	0,503	0,496
Med:	0,452	0,339	0,203						
Prior Mean:	0,453	0,341	0,206						

TABLE 3 : Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t-ratio
α_1	26.26	9.37	2.80
α_2	19.79	7.06	2.80
α_3	11.96	4.28	2.79
Log-likelihood function: -415.36			
Number of iterations : 29			

TABLE 4 : Bayesian Analysis for: MTW$>\theta_1^m$, **TF**$>\theta_2^m$.
(5 most dangerous sites on weekdays)

	B1			
Site	B1_1 = P (**MTW**$>\theta_1^m$)	B1_2 = P(**TF**$>\theta_2^m$)	Product B1_1*B1_2	Probability value P(**MTW**$>\theta_1^m$,**TF**$>\theta_2^m$)
20	0.701	0.907	0.636	0.609
51	0.587	0.702	0.412	0.320
66	0.871	0.425	0.370	0.311
30	0.819	0.456	0.374	0.300
85	0.490	0.762	0.374	0.287

Note : Sites selected based on B1_1*B1_2 values are: 20 51 85 30 8 66.

6 Appendix

In this appendix, we provide a review on the multivariate distributions that we use in the paper. A good reference is Bernardo and Smith (1994).

6.1 The Multinomial Distribution

Let $x = (x_1, ..., x_K)$ be a discrete random vector where, by convention we write $x_{K+1} = n - \sum_{k=1}^{K} x_k$. It has a multinomial distribution of dimension K, with parameters $\theta = (\theta_1, ..., \theta_K)$ and n $(0 < \theta_k < 1,\ \theta_{K+1} + \sum_{k=1}^{K} \theta_k = 1,\ n \geq 1)$ if its probability function can be written as:

$$f(x|\theta, n) = \frac{n!}{\prod_{k=1}^{K+1} x_k!} \prod_{k=1}^{K+1} \theta_k^{x_k} \quad . \tag{A.1}$$

Recall that by definition $x_{K+1} = n - \sum_{k=1}^{K} x_k$ and $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$. This is done for notational simplicity. Although $K + 1$ different x_k terms are involved, only the first K are free; the last one is an explicit function of $x_k, k = 1, ..., K$. The same comment also applies for the θ_{K+1} term. The mean vector and covariance matrix are given by:

$$E[x_k] = n\theta_k,\ V[x_k] = n\theta_k (1 - \theta_k),\ C[x_k, x_l] = -n\theta_k\theta_l, \quad k, l = 1, ..., K. \tag{A.2}$$

Property 1:

If $x_1, ..., x_K$ are K independent Poisson random quantities with densities $f(x_k|\lambda_k)$, then the joint distribution of $x = (x_1, ..., x_K)$ given $\sum_{k=1}^{K} x_k = n$ is multinomial $f(x|\theta, n)$ with parameters n and $\theta_k = \lambda_k / \sum_{l=1}^{K} \lambda_l$.

This property is very interesting, but we are not going to make use of it in this paper.

6.2 The Dirichlet Distribution

Let $\theta = (\theta_1, ..., \theta_K)$ be a continuous random vector where $0 < \theta_k < 1$,and where again, to simplify the notation we write $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$. This implies that $\theta_k, k = 1, ..., K$ are all free to vary in the $(0, 1)$ interval and θ_{K+1} is a term that fills the gap between $\sum_{k=1}^{K} \theta_k$ and 1. The θ random vector has a Dirichlet distribution of dimension K, with parameters $\alpha =$

$(\alpha_1, ..., \alpha_{K+1}), \alpha_j > 0, \;\; j = 1, ..., K+1$ if its probability density is written as:

$$g(\theta|\alpha) = d(\alpha) \cdot \prod_{k=1}^{K+1} \theta_k^{\alpha_k - 1}, \tag{A.3}$$

where by definition, $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$. Also, $d(\alpha) = \Gamma(\sum_{k=1}^{K+1} \alpha_k) \,/\, \prod_{k=1}^{K+1} \Gamma(\alpha_k)$ and $\Gamma(s)$ is the gamma function, computed as: $\Gamma(s) = \int_0^\infty e^{-z} z^{s-1} dz$. The mean vector and covariance matrix can be computed as:

$$E[\theta_k] = \frac{\alpha_k}{\sum_{j=1}^{K+1} \alpha_j}, \qquad V[\theta_k] = \frac{E[\theta_k](1 - E[\theta_k])}{1 + \sum_{j=1}^{K+1} \alpha_j},$$
$$C[\theta_k, \theta_l] = \frac{-E[\theta_k]E[\theta_l]}{1 + \sum_{j=1}^{K+1} \alpha_j}, \quad k, l = 1, ..., K. \tag{A.4}$$

More generally, the moments of the Dirichlet distribution (see Wilks, 1962) can be computed using the formula:

$$E(\prod_{k=1}^{K+1} \theta_k^{b_k} \mid \alpha) = \frac{d(\alpha_1, ..., \alpha_{K+1})}{d(\alpha_1 + b_1, ..., \alpha_{K+1} + b_{K+1})}, \tag{A.5}$$

where by assumption $b_{K+1} = 0$. According to our previous definition, we have:

$$d(\alpha_1 + b_1, ..., \alpha_{K+1} + b_{K+1}) = \Gamma(\sum_{k=1}^{K+1} \{\alpha_k + b_k\}) \,/\, \prod_{k=1}^{K+1} \Gamma(\alpha_k + b_k).$$

Property 2*:*
Marginal distributions:
The marginal distribution of $x^{(m)} = (x_1, ..., x_m)$, $m < K$ is Dirichlet with parameters $(\alpha_1, ..., \alpha_m, \sum_{j=m+1}^{K+1} \alpha_j)$.

Property 3:
If $K = 1$, the density in (A.3) reduces to a beta density:

$$g(\theta_1 \mid \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \theta_1^{\alpha_1 - 1}(1 - \theta_1)^{\alpha_2 - 1}, \quad 0 < \theta_1 < 1. \tag{A.6}$$

6.3 The Multinomial-Dirichlet Distribution

This distribution is generated by mixing the two previous distributions in equations (A.1) and (A.3) in the following way:

$$h(x|\alpha,n) = \int_0^1 \dots \int_0^1 f(x|\theta,n) \cdot g(\theta|\alpha)\ d\theta, \tag{A.7}$$

where the integral is of dimension K, the number of components in the θ vector. As a result of this integral which can be resolved analytically, one obtains the Multinomial-Dirichlet probability function (see also DeGroot, 1986):

$$\begin{aligned} h(x|\alpha,n) &= \frac{n!}{\prod_{k=1}^{K+1} x_k!} \frac{d(\alpha)}{d(\alpha+x)} \\ &= \frac{n!}{\prod_{k=1}^{K+1} x_k!} \cdot \frac{\Gamma(\sum_{k=1}^{K+1} \alpha_k)}{\Gamma(\sum_{k=1}^{K+1} \alpha_k + x_k)} \cdot \prod_{k=1}^{K+1} \frac{\Gamma(\alpha_k + x_k)}{\Gamma(\alpha_k)}. \end{aligned} \tag{A.8}$$

An alternative result which avoids the use of $\Gamma(s)$ functions is :

$$h(x|\alpha,n) = \frac{n!}{(\sum_{j=1}^{K+1} \alpha_j)^{[n]}} \prod_{k=1}^{K+1} \frac{\alpha_k^{[x_k]}}{x_k!}, \tag{A.9}$$

where $z^{[s]} = \prod_{j=1}^{s} (z+j-1)$.

The mean vector and covariance matrix are given by:

$$\begin{aligned} E[x_k] &= n\,p_k, \quad p_k = \frac{\alpha_k}{\sum_{j=1}^{K+1} \alpha_j} \\ V[x_k] &= \frac{n + \sum_{j=1}^{K+1} \alpha_j}{1 + \sum_{j=1}^{K+1} \alpha_j}\ n\,p_k\,(1-p_k) \\ C[x_k, x_l] &= -\frac{n + \sum_{j=1}^{K+1} \alpha_j}{1 + \sum_{j=1}^{K+1} \alpha_j}\ n\,p_k\ p_l, \qquad k,l = 1,\dots,K. \end{aligned} \tag{A.10}$$

Property 4:

If $K = 1$, the density in (A.8) reduces to a binomial-beta density.

References

[1] Bernardo, J.M. and A.F.M. Smith (1994), Bayesian Theory, John Wiley and sons.

[2] Bolduc, D. and S. Bonin (1997), « Modèle bayésien généralisé pour lidentification des sites routiers dangereux », L'Actualité économique/Revue d'analyse économique, vol.73, nos 1-2-3, 1997, pp. 81-98.

[3] Bolduc, D. and S. Bonin (1995), «Bayesian Analysis of Road Accidents: Accounting for Deterministic Heterogeneity», *Compte-rendus de la IXe Conférence canadienne multidisciplinaire sur la sécurité routière*, Montréal, Canada.

[4] DeGroot, M.H. (1986), «Probability and Statistics», Second Edition, Addison-Weslwy Publishing Company, United States.

[5] Hauer, E. (1992), «Empirical Bayes Approach to the Estimation of Unsafety: the Multivariate Regression Method», *Accident Analysis & Prevention*, Vol 24, No 5, 457-477.

[6] Hauer, E. (1986), «On the Estimation of the Expected Number of Accidents», *Accident Analysis & Prevention*, Vol 18, No 1, 1-12.

[7] Heydecker, B. and J. Wu (1991), «Using the Information in Road Accident Records», For presentation at the 19th Summer Annual Meeting of PTRC, University of Sussex, England.

[8] Higle, J.L. and J.M. Witkowski (1989), «Bayesian Identification of Hazardous Locations», *Transportation Research Record* 1185, 24-36.

[9] Loveday, J. and D. Jarrett (1992), «Spatial Modelling of Road Accident Data», *Mathematics in Transport Planning and Control*, Institute of Mathematics and its Applications Conference Series 38, 433-446.

[10] Maher, M.J. (1990) «A Bivariate Negative Binomial Model to Explain Traffic Accident Migration», *Accident Analysis & Prevention*, Vol. 22, No. 5, 487-498.

[11] Wilks, S. S. (1962) «Mathematical Statistics», New-York, John Wiley.