

Article

# Hamiltonian Monte Carlo with Random Effect for Analyzing Cyclist Crash Severity

Mahdi Rezapour \* and Khaled Ksaibati

Wyoming Technology Transfer Center, Laramie, WY 82071, USA; khaled@uwyo.edu

\* Correspondence: mrezapou@uwyo.edu

**Abstract:** Vulnerable traffic users, such as bikers and pedestrians, account for a significant number of fatalities on the roadways. Extensive research has been conducted in the literature review to identify factors to those crashes. Studying factors to those crashes is especially important in the Western state in the US, due to one of the highest fatality rates in the nation and its unique geographic conditions. The first step in identifying factors to the severity of cyclist crashes is to find the underlying factors to that type of crash, while accounting for the heterogeneity in the dataset. Various techniques such as mixed parameter or mixed effect models have been employed in the literature to account for the heterogeneity of the dataset. In the mixed effect model, often the random effect parameter has been assigned subjectively, and based on some attributes and engineering intuitions. Those assignments are expected to account for the heterogeneity in the dataset and enhancement of the model fit. However, a question might arise whether those factors could account for an optimum amount of the heterogeneity in the dataset. A more reasonable way might be to let the algorithm such as the finite mixture model (FMM) to identify those clusters based on parameters of the Gaussian model, means and covariance matrices of the dataset, and allocate each observation to the related clusters. Thus, in this study, to capture optimum amount of heterogeneity, first we implemented the finite mixture model in the context of maximum likelihood, due the label switching issue of the method in the context of the Bayesian method. After assignment of the parameters to the observation, the main method of Hamiltonian Monte Carlo (HMC) with random effect was implemented. The results highlighted a significant improvement in the model fit, in terms of Widely Applicable Information Criterion (WAIC). The results of this study highlighted factors such as older biker age, increased number of lanes, nighttime travelling, increased posted speed limit and driving while under emotional conditions are some factors contributing to an increased severity of bikers' crash severity. Extensive discussion has been made regarding the methodological algorithms and model parameters estimations.



**Citation:** Rezapour, M.; Ksaibati, K. Hamiltonian Monte Carlo with Random Effect for Analyzing Cyclist Crash Severity. *Signals* **2021**, *2*, 527–539. <https://doi.org/10.3390/signals2030032>

Academic Editor: Alexander Kocian

Received: 10 April 2021

Accepted: 3 August 2021

Published: 17 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** cyclist safety; Hamiltonian Monte Carlo; random effect; finite mixture model; pedestrian

## 1. Introduction

The benefits of cycling could be summarized as the wellbeing of the population through developing a healthy and sustainable environment, which could be achieved, in turn, by prevention of obesity and the consequent enhancement of cardiorespiratory fitness [1]. The benefits of shifting from car to bicycle have also been seen in terms of health effects through a decrease in air pollution emission, and greenhouse emissions, and an increase in physical activity [2].

However, safety is a significant barrier for active transportation activities. The situation would be direr as an increase in the rate of cycling would increase the risk of involvement in traffic crashes, as well as the severity of those crashes. More than half of the road fatalities worldwide are related to vulnerable road users, including bicyclists and pedestrians [3]. In addition, while about 10% of trips in the U.S. have been made by walking or bicycling [4], those types of commuting accounts for about 14% of all traffic fatalities [5].

It has been highlighted in the previous study that the proportion of biking and cycling are overrepresented in crash fatalities [6]. For instance, bicyclists are twelve times more likely to be killed in road crashes than motor vehicle drivers, and also in the US they are three times and twice as likely to be involved in fatal crashes compared with Dutch and Germany, respectively [7]. The following paragraphs would outline few studies related to cyclists' crashes.

The impact of the built environment was evaluated on bicyclist injury severity [8]. The generalized ordered logit and generalized additive models were employed. The results highlighted that employment density, reflective clothing, age and posted speed limit are some of the factors contributing to the severity of bikers' crash severity. The mixed logit analysis was used for analyzing bicyclist injury severity at intersection, and non-intersection locations [9]. The results highlighted that at the intersection when the bikers did not wear a helmet, or when the motorist was under the influence, the severity of crashes was increased. On the other hand, the results highlighted at non-intersection, when the motorist was under the influence, or a vehicle strikes the side of bicycle and the time that the heavy truck hit a bicycle, the severity of crashes increases.

Bicyclist crash types were investigated on national and local levels [10]. The results highlighted that while the majority of injuries occurred at intersections, most of the fatal and disabling injury bicyclist crashes occurred at the non-intersection area. The results also highlighted variation across contributory factors at the national and state levels.

Factors to injury severity of electric bikes were investigated in China [11]. The results highlighted that older riders (age > 25) are likelier to be involved in severe crashes. In another study, the impact of transportation infrastructure on bicycling injuries and crashes by reviewing the literature was evaluated [12]. The results showed that infrastructure impacts injury and the risk of crashes. In addition, factors such as a multi-lane roundabout significantly could increase the risk of bikers' crashes.

In another study, factors impacting the bikers' crash severity was evaluated considering alcohol, lighting and speed characteristics [13]. The results found that the severity of crashes would increase when the bikers or drivers drink alcohol, or when the drivers drive the vehicles above the posted speed limit.

Studies in the literature highlighted variations across determinants of bikers' crash severity. Those differences are linked based on mainly the study's area geography (e.g., urban vs. rural) [14]. For instance, it was found that factors affecting bicycle-motor vehicle crashes in rural areas are driver age, bicycle age, vehicle speed and divided roadway, while in urban areas factors include head-on crashes, single-unit truck and road conditions [14]. Thus, based on the above literature, it has been highlighted that the contributory factors to bikers' crash severity would vary based on various areas geographies characteristics [15]. As a result, it is important to have a vision regarding factors affecting bikers' severity in a mountainous area like Wyoming, with very low traffic and one of the highest fatality rates in the country.

To obtain less biased point estimates regarding the factors affecting the severity of bicyclist crashes, it is important to implement a method that could account for the heterogeneity in the dataset due to its structure. Not accounting for the heterogeneity in the dataset might result in biased or even erroneous model parameters' estimates. Thus, in this study we accounted for that heterogeneity by means of random effect, or hierarchical technique, in the context of the Hamiltonian Monte Carlo technique. Additionally, although the majority of the past studies in the context of traffic safety used a subjective hierarchy or random effect, in the current study we allocate an objective hierarchy based on the finite mixture model. Later, we allocate the identified clusters to the random effect in the main analysis.

The remainder of this study is as follows: methodology section details the methodological approach taken in this study. The data and results sections outline the employed data and the obtained results of the statistical analysis. The last section of discussion presents and discusses the identified results.

## 2. Method

This section contains three subsections. First the process of the finite mixture model is outlined, followed by the main model of Hamiltonian Monte Carlo (HMC). Finally, the methods for evaluation of the model fit are presented. Each section also discusses the model implementation related to the relevant mathematical equations and algorithms.

### 2.1. Finite Mixture Model

Here, instead of employing subjective assignment for our random effect, we implemented an objective method for assigning the random effect attribute. Thus, it is important to discuss the methodological approach of coming up with an objective attribute. The method of the finite mixture model (FMM) assumes that the dataset arises from data components. In the model-based method, instead of working with distance measures in K-means, for instance, we would work with a set of dimensional data.

A multivariate Gaussian distribution might be assumed for each component of  $f_k(x_i; \theta_k)$  by parameters of  $N(\mu_k, \Sigma_k)$ . The estimated clusters with multivariate Gaussian distribution would get the shape of ellipsoid with the means being centered at  $\mu_k$ , while volume, shape and orientation would be determined by the covariance matrix of  $\Sigma_k$ . The covariance matrices' parameters would be identified by Eigen-decomposition [16]:

$$\Sigma_k = \lambda_k D_k A_k D_k^T \tag{1}$$

where  $\lambda_k$  is scalar with the eigenvalue,  $A_k$  is the diagonal matrix, with the diagonal elements proportional to the corresponding eigenvalues, and  $D_k$  is an orthogonal matrix accommodating the eigenvector. These parameters determine volume, shape and orientation of the ellipsoid based on combinations of setting constrains on  $\Sigma_k$ , and based on various components of  $\Sigma_k$ .

The finite mixture model could be written as follows:

$$f(x_i; \varphi) = \sum_{k=1}^U \pi_k f_k(x_i; \varphi_k) = \sum_{k=1}^U \pi_k f_k(x_i; \mu_k, \Sigma_k), \text{ where } \pi_k > 0, \text{ and } \sum_{k=1}^U \pi_k = 1 \tag{2}$$

The parameters of the mixture model are  $\varphi = \{\mu_1, \dots, \mu_U; \Sigma_1, \dots, \Sigma_U\}$ , and  $f_k$  is the density for observation  $x_i$  with parameters vectors of  $\theta_k$ . We know that the density of multivariate normal density of  $f_k$  as  $\varphi_k$  would be written as:

$$\varphi_k(x_i | \mu_k, \Sigma_k) = \frac{\exp\left(-0.5(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)\right)}{\det(2\pi \Sigma_k)^{0.5}} \tag{3}$$

The expectation maximization (EM) algorithm is an iterative method for finding maximum likelihood, where the model depends on the unobserved latent variable. For the EM algorithm, the complete data are  $x_i = (y_i, z_i)$ , where  $z_i$  is an unobserved proportion of data, while the  $y_i$  is observed. Now, the E step is [17]:

$$\hat{z}_{ik} = \frac{\hat{\pi}_k f_k(y_i | \hat{\varphi}_k)}{\sum_{j=1}^U \hat{\pi}_j f_j(y_i | \hat{\varphi}_j)} \tag{4}$$

The M part would be employed to maximize the log likelihood equation by considering the fixed parameter of  $z_{ik}$  instead of  $\hat{z}_{ik}$  from above as

$$\mathcal{L}(\varphi_k, \tau_k, z_{ik} | x) = \sum_{i=1}^n \sum_{k=1}^U z_{ik} \log[\pi_k f_k(y_i | \varphi_k)] \tag{5}$$

### Model Parameters Estimation

The model implementations, in general, largely depend on Equations (4) and (5). First, to start the process, it is required to have initial values for cluster allocations and the covariance matrix. Those could be assigned by methods such as the K-means technique.

Now, the single matrix of covariance for all classes would be passed to the next algorithm for the finite mixture model process.

The objective function is to maximize the likelihood of  $f(x_i; \varphi)$  in Equation (2) by EM algorithm. The EM algorithm would be employed to alternate between the two steps of “E” and “M”. In the “E” step, the complete data, or the probability density function (PDF) of multivariate distribution, would be estimated based on Equation (3). Then, the obtained values would be used in the M step for determining parameters, maximizing the log likelihood, which is based on estimated parameters in E step.

In summary, the E step is based on Equation (4) and M is based on Equation (5). The density of multivariate normal distribution needs to be estimated based on Equation (3) for  $f_k(y_i|\hat{\theta}_k)$  to be used in Equation (4). That requires  $\Sigma_k$ , which would be estimated by Equation (1).

After estimating  $\hat{z}_{ik}$  for each latent class, Equation (4), the value would be used for M step. For  $\pi_k$ , a similar approach by employing the proportion of each class frequency could be employed. Now values of  $z_{ik}$  and  $\pi_k$  would be obtained easily. An increase in the maximum likelihood would be ensured in the covariance matrix estimation.

### 2.2. Hamiltonian Monte Carlo

One of the shortcomings of application of random walk in algorithms, such as Metropolis Hastings, is the samplers explore the target distribution extremely slowly [18]. In other words, the Metropolis Hastings (MH) takes a long time with large number of iterations to get to the true posterior density (e.g., often the acceptance rate is less than a quarter of all proposals) [19].

Hamiltonian Monte Carlo (HMC), on the other hand, could be employed by utilizing methods such as leapfrog to generate transition and eliminating the random walk behavior. The HMC algorithm has been identified as a powerful tool, which could efficiently compute model parameters. In the HMC method, we assume that the system has an energy level,  $E$ , being dependent on  $\theta$  and  $M$ , where  $\theta$  is the model’s parameter to be estimated, and  $M$  is the momentum. Energy itself depends on the position of a particle and the particle’s kinetic energy.

We would solve for the particle with potential energy of  $u(\theta)$ , and kinetic energy of  $k(M)$ , where  $u(\theta)$  is  $-\log f(\theta)$ , and the kinetic energy is  $0.5 p^T M^{-1} p$ . Here  $M$ , the auxiliary momentum variables, is added, which is known.

Therefore, we would have:

$$H(\theta, m) = u + k = -\log f(\theta) + 0.5m^T M^{-1}m \tag{6}$$

where  $m \sim N(0, M)$ . As we cannot exactly solve for part of a particle, we end up discretizing the path, so the algorithm would be small enough to make a precise enough approximation. Leapfrog would be used to approximate solutions for Hamiltonian equations by using a step size of  $\epsilon$ . The output of leapfrog is the proposed values of  $\theta^*$  and  $m^*$ . The gradient of the above, with respect to  $m$  and  $\theta$ , would be the Hamiltonian equations, which would be solved by the leapfrog algorithm as:

$$m(t + \epsilon/2) = m(t) + \epsilon/2 \frac{dm}{dt} = m(t) + \epsilon/2 \nabla_{\theta} \log f(\theta) \tag{7}$$

$$\theta(t + \epsilon) = \theta(t) + \epsilon \frac{d\theta}{dt} = \theta(t) + \epsilon M^{-1} m(t + \epsilon) \tag{8}$$

where  $\nabla_{\theta} \log f(\theta)$  is the gradient of log posterior density. Here, few steps would be used to move to a next proposal. In summary, leapfrog, which uses the above equations, would accelerate the path to the posterior  $\theta$ . After coming up with proposal  $\theta$ , we need to make a decision to reject or accept it. Therefore,  $\alpha$  would be created and compared to decide regarding accepting or rejecting the proposal. Here, instead of random walk for doing the

proposal, we are solving for part of a particle, using the deterministic equation of motion.  $\alpha$  would be written as:

$$\alpha = \frac{\log(f(\theta_2) - 0.5m_2^T M^{-1}m_2)}{\log(f(\theta_1) - 0.5m_1^T M^{-1}m_1)} \tag{9}$$

where subscript 2 is an updated version of parameters and 1 is the previous values. The above value in Equation (9) would be compared with  $u^*$ , where  $u^*$  is a random uniform number between 0 and 1. The  $\alpha$  equation is intuitively highlighting that when  $f(\theta_t) > f(\theta_{t-1})$ , the proposal has a higher probability, and we accept the proposal. Therefore, if  $\alpha$  is greater than  $u^*$ , we would accept the proposed value of  $\theta$  and  $m$  to be our next sample from our distribution, otherwise the value would remain as the previous values. In other words, if  $\alpha > u$  holds in Equation (9), then:

$$m^* \leftarrow -m_2 \ \& \ \theta^* \leftarrow \theta_2 \ \text{Otherwise} \ m^* \leftarrow m_1 \ \& \ \theta^* \leftarrow \theta_1. \tag{10}$$

where  $m^*$  and  $\theta^*$  are proposal values, which would be used in the next iteration. We flip the sign of momentum in Equation 10 as if we start at a particular value of  $\theta$  and  $m$ , then the part of deterministically leads to  $\theta^*$  and  $m^*$ , or  $P(\theta^*, m^* | \theta, m) = 1$ , but the reverse probability is 0. Therefore, what we do is switch the proposed  $m$  with its negative.

Every time we propose a new value of  $m$ , it changes our energy level, jumping up or down. Therefore, we need to propose a reasonable value of energy that could be represented by a joint distribution. For momentum, we assume  $m \sim N(0, M)$ , and  $M$  is defined as a unit diagonal matrix.

### 2.2.1. Model Formulation

For estimation of  $\alpha$  in Equation (9), the log likelihood or  $\log(f(\theta))$  is needed. The probability of the binary mixed effect model could be written as:

$$Pr(y = 1|X, Z) = \frac{\exp(X\beta + Zu)}{1 + \exp(X\beta + Zu)}, \text{ where } u \sim N(0, G) \tag{11}$$

where  $u$  follows multivariate normal distribution with mean 0 and variance covariance of  $G$ ,  $u \sim N(0, G)$ ,  $Z$  is a matrix where a column  $j$  in  $z_j$  would be filled by values of 1 if the value of random effect exists in  $j$  column, and 0 otherwise.  $u$  could be written as

$$u = G^{0.5}\tau \tag{12}$$

where  $\tau$  is the initial values of random effects, to the number of categories.

As the square root of  $G$  is required for Equation (12), LDL matrix decomposition would be used. Therefore, we have:

$$G = LDL^T = LD^{1/2}D^{1/2}L^T \tag{13}$$

Thus,

$$\sqrt{G} = LD^{1/2} \tag{14}$$

where  $L$  is an identity matrix of  $I_n$ ,  $D^{1/2}$  is a matrix with diagonal values of  $\lambda$ , where  $\lambda$  follows half-t student distribution [20]. Therefore, matrix of  $\sqrt{G}$  would be built based on Equation (14). Now, to finalize the  $u$  matrix, we need  $G^{1/2} \times \tau$ , where  $\tau$  is the parameters to be estimated for random effect, and already the initial values being updated through the leapfrog algorithm.

As the equation of half-t has a form of the power of  $-(\nu + 1)/2$ , transformation would be used to make it in a linear form by the help of Jacobian transformation. Due to transformation variables, and from Jacobian transformation, the processes,  $\log P(\beta|\sigma_\beta^2)$ ,  $\log P(\tau)$  and  $\log P(\xi)$ , would be added to the log posterior. The likelihood of  $\log \pi(\beta|\sigma_\beta^2)$

follows the multinomial probability density function with variance of  $\sigma_\beta^2$  so it would be written as:

$$\log P(\beta|\sigma_\beta^2) \rightarrow -\frac{\beta^T \beta}{2\sigma_\beta^2} \tag{15}$$

$\tau$  follows normal distribution probability density function (PDF) with  $N(0, 1)$ , so it would be written as

$$\log P(\tau) \rightarrow -0.5\tau^T \tau \tag{16}$$

After transforming the variance covariance of  $\lambda$  to make it in a form of linear model through Jacobian transformation, the probability of transformed  $\lambda$ ,  $\zeta = \log(\lambda)$  and based on PDF of half-t distribution would be written as:

$$\log P(\zeta) \rightarrow -0.5(v_\zeta + 1)\log\left(1 + \frac{1}{v_\zeta} \left(\frac{e^{2*\zeta}}{A_\zeta^2}\right)\right) + \zeta \tag{17}$$

where  $v_\zeta$  is parameter of half-t distribution, being the degree of freedom, and  $A_\zeta$  is the value of variance.

On the other hand, the log posterior,  $\log(f(\theta))$ , for logistic mixed effect would be written as:

$$\log f(y|X, Z, \beta, u) = (y - 1)^T(X\beta + Zu) - 1 \log(1 + \exp(-(X\beta + Zu))) \tag{18}$$

where all items were defined before.

Now, the full likelihood after considering the transformation should include additive terms being discussed above.

The gradients would be used in the leapfrog algorithm to direct the transitions of Markov Chain to the region of higher probability. The gradients are obtained by getting the derivatives of full log likelihood, which include log priors for transformed variables.

### 2.2.2. Model Parameters' Estimations

In this section, we present the simplified model implementation to estimate parameters, being based on the work in the literature review [21]. The process of parameters' estimation could be divided into 5 main sections and related subsections:

- General data preparation

Preparation of the matrix of input by QR decomposition, orthogonal matrix of Q and upper triangular matrix of R, to improve efficiency of HMC sampling. Creating value of  $\epsilon$  for all parameters to update the model parameters estimates (e.g., Equation (8)). Auxiliary log density function of momentum would be created by assigning the mean of zeroes and variance-covariance diagonal identity matrix of  $p \times p$ , where  $p$  is the number of parameters to be estimated. Non-informative initial values of 0 would be given to all parameters.

- Leapfrog algorithm

#### 1. Use of gradients

Derivative of full posterior distribution relative to  $\beta$ ,  $\tau$ , and  $\zeta$  would be obtained, and saved in a matrix for updating various parameters to be estimated. Based on Equation (7),  $m$  or momentum would be updated, and  $\theta$  would be updated based on Equation (8). It should be noted that the  $\epsilon$  would be set as small values for all parameters. The leapfrog would be conducted for 10 iterations nested inside the iteration of HMC sampling. The gradient in leapfrog would direct the sampling to the optimum values.

#### 2. Updating $\theta$ and $m$

In each iteration of leapfrog, the previous value of  $\theta$  and  $m$  would be used and updated.

- Estimating log posterior

We are still in the main loop of the first HMC sampling iteration. For this section, values would be estimated for Equation (9), related to its numerator and denominator. The numerator includes the new value of  $\theta$  that is just estimated, while the equation's denominator would use the old value of  $\theta$ , which for the first iteration is preassigned values of 0.

Here, full posterior, as discussed before, is dependent on Equation (18), and other priors' value. For Equation (18), parameters related to random effect,  $u$ , need to be prepared. That is based on Equation (12), which itself would be written based on Equation (13) or (14). Now,  $\log P(\beta|\sigma_\beta^2)$ ,  $\log P(\tau)$ ,  $\log P(\xi)$ , related to Equations (15)–(17), respectively, would be calculated and added to Equation (18). Recall that the process for numerator and denominator would be estimated separately for  $\theta$  and  $\theta_{t-1}$ .

- Estimation of  $\alpha$ : Whether to reject or accept the new proposal

As discussed, one of the advantages of the leapfrog algorithm is to increase the rate of acceptance, compared with random walk. Now, based on Equation (9), in case of  $\alpha > 1$ , we accept the new proposed  $\theta$  and  $-m$ , otherwise we reject it and keep  $\theta_{t-1}$  and  $m_{t-1}$ . The values would be saved and used for the next iteration. The process would be iterated for 6000 HMC sampling iterations, each with 10 leapfrog iterations. The final  $\theta$  would bind all  $\theta$ s for further analysis.

It should be noted that, for the HMC, we flip the distribution by getting the log of probability density. The idea is that we follow the ledge of the distribution, which tends to go to the bottom of the distribution based on the gravity. Therefore, by generating more samples from the bottom of the distribution, in reality, we draw from the pick of the distribution.

- Results preparation

The main output of the model is the  $\theta$ s value during iterations, Equation (10). Various quantiles of sampled  $\theta$  would be used after discarding the first N number of burning samples, highlighting certainty or uncertainty associated with parameters' estimates.

### 2.3. Goodness-of-Fit

As the widely applicable information criterion (WAIC) was used as a means of goodness-of-fit, this section would outline this parameter's estimation. The goodness-of-fit is based on the use of log-likelihood by the posterior simulations of the included parameters. First, the log pointwise predictive density (*lpd*) would be computed as [22]:

$$lpd = \sum_{i=1}^n \log\left(\frac{1}{S} \sum_{s=1}^S p(y_i|\theta^s)\right) \tag{19}$$

where  $\sum_{s=1}^S \log(p(y_i|\theta^s))$  in Equation (19) is the logarithm of the sum of the simulated draws  $\theta^s$  for  $S = 6000$ , where  $i$  is number of data points. Now, constant value of  $-\log(6000)$  would be deducted from  $\sigma$ . It should be noted, in Equation (19), the value of  $\frac{1}{S}$  is considered to account for the overfitting issue. Now, for estimating the WAIC, the expected log pointwise predictive density (*elpd*) is required, which would be written as:

$$elpd_{waic} = lpd - \hat{p}_{waic} \tag{20}$$

where  $\hat{p}_{waic}$  itself, the estimated effective number of parameters, could be written based on variance of posterior,  $var_{pos}$  [22]:

$$p_{waic} = \sum_{i=1}^n var(\log(y_i|\theta)) \tag{21}$$

Now we have everything needed for estimating  $elpd_{waic}$  in Equation (20). The WAIC itself could be written as  $-2 \times p_{waic}$ . In summary, the method uses out-of-sample expecta-

tions by using first log pointwise posterior predictive density and then using the correction of the effective number of parameters for overfitting adjustment [23]. Additionally, it should be noted that, based on above equations (e.g., Equation (20) or (21)), we have WAIC for each attribute and the final WAIC is the sum of all attributes. It is intuitive that the higher the variance of the  $p_{waic}$ , the lower the  $el\hat{p}d_{waic}$  and, consequently, the higher the values for WAIC. Therefore, for model comparison, similar to AIC, we favor a lower value of WAIC.

### 3. Data

The dataset used in this study was directly obtained from the Wyoming Department of Transportation (WYDOT). Here, various datasets including bicyclists, vehicles, drivers and a bulk dataset were aggregated, based on the common column of crash ID.

The dataset incorporated the information regarding crashes during the period 2011–2019. It is expected, due to the implementation of the finite mixture model and its application as a random effect, most of the similar and dissimilar observations were grouped together, minimizing the heterogeneity in the dataset. Here, the variable of cluster highlights the number of groupings being clarified based on variance-covariance matrices. The number of clusters is based on goodness-of-fit, Bayesian information criterion (BIC). The choice of 2 for numbers of clusters might be mainly due to a low number of observations,  $n = 749$ . The descriptive summary is presented in Table 1.

**Table 1.** Attributes and response of significant attributes in the statistical analysis.

Attributes	Mean	Variance	Min	Max
Alcohol involvement, alcohol was involved vs. others *	0.04	0.036	0 *	1
Bikers age, <20 as 0 *, others 1	0.54	0.249	0	1
Hit and run vs. others *	0.150	0.128	0	1
Bikers cross marked line at an intersection vs. others *	0.231	0.179	0	1
Vehicle maneuver, turning vs. others *	0.42	0.244	0	1
Number of lanes, ≤3 vs. others *	0.29	0.204	0	1
Light conditions, dark as 1 vs. others *	0.124	0.109	0	1
Posted speed limit, ≤20 vs. others *	0.92	0.073	0	1
Driver emotional condition, under emotional condition vs. others *	0.179	0.147	0	1
Bikers crossing entering the road vs. others *	0.606	0.239	0	1
Bikers travel with the direction of traffic vs. others *	0.208	0.165	0	1
Cluster ‘, discrete response	1.45	0.248	1	2
Bikers crash severity, binary	0.11	0.1	0	1

\* Reference category as the lowest category value; ‘objective cluster was used as a random effect.

### 4. Results

The HMC analysis was conducted using 6000 draws with three chains, where the first 1000 draws were discarded. It should be noted that  $\lambda$  in Table 2 is the diagonal value of  $D_{1/2}$ , where it was transformed by using the  $\exp(\xi)$ . The results sections were divided into three main subsections and its related variables. Those subsections include drivers’ actions, bikers’ actions and environmental and roadway characteristics.

**Table 2.** Modeling results of bikers crash severity, HMC technique.

Attributes	Effect Estimates		CI	
	Mean	SD	2.50%	97.50%
Intercept	−4.26	0.74	−5.66	−2.90
Alcohol involvement	1.27	0.49	0.32	2.20
Bikers' age	0.52	0.27	0.0008	1.08
Hit and run	−1.79	0.52	−2.82	−0.83
Bikers crossing marked line at an intersection	−1.74	0.48	−2.70	−0.87
Vehicle maneuver, turning	−0.96	0.30	−1.57	−0.41
Number of lanes	0.70	0.270	0.19	1.24
Light conditions	0.60	0.33	−6.89	0.22
Posted speed limit	1.16	0.64	0.031	2.45
Driver emotional condition	0.99	0.38	0.2	1.73
Bikers crossing or entering the road	1.03	0.41	0.27	1.85
Bikers travel with the direction of traffic	1.14	0.43	0.31	2.02
Random Effects				
$\tau_1$	−0.001	1.01	−2.061	2.030
$\tau_2$	−0.06	1.0	−1.93	1.98
$\lambda$	−18.01			
$WAIC_{random\ effect} = 173$ vs. $WAIC_{standard} = 463$				

#### 4.1. Drivers' Actions

This section outlines the actions taken by drivers whilst being involved in bicyclist crashes.

##### 4.1.1. Hit-and-Run

Hit-and-run crashes are related to when at-fault drivers leave the crash scene without helping the victims or reporting the incident to the authorities. The results highlighted that when the crash involved hit-and-run, those crashes are more likely to be associated with lower severity of bicyclists' crashes. The results seem counterintuitive and against the work in the literature review. The main reason in the literature for higher hit-and-run crash severity has been linked to the fact that those drivers would be involved in hit-and-run crashes when they perceived economic costs of crashes to be serious [24].

The counterintuitive obtained results of this study could be linked to a few factors: First, drivers experiencing more remorse and feeling guilty, when observing the severity is higher. Second, as these crashes are likelier to be caused by older drivers, those drivers probably did not notice they hit bicyclists due to those crashes being minor. For instance, for our case study, almost 80% of hit-and-run crashes involved those drivers aged over 45 years old. In addition, almost 70% of those drivers were aged at least 85 years old. Thus, it is likely that the results would be due to the older age. Here we did not consider the interaction terms due to a very low number of hit-and-run crashes, while considering older drivers.

##### 4.1.2. Driver Emotional Condition

This attribute highlights the driver's conditions at the time of crashes, compared to being "apparently normal". Those include feelings such as being fatigued, ill or being under emotional conditions such as being angry, sad or depressed. The result is in line with the work in the literature that stress, and emotional states elevate crash risk [25]. In another study, by means of a driving simulator, the impacts of feelings of anger on driving choices and abilities were examined [26]. The results highlighted that angry drivers show a higher likelihood of crossing through an intersection at a yellow traffic light and tend to drive faster.

#### 4.1.3. Driver Alcohol Involvement

Drivers who were impaired by alcohol are more likely to result in more severe crashes involving bicycles. The results are expected and shown in the literature review. The results could be linked to lack of judgment, aggression and increased risk-taking behavior [27].

#### 4.1.4. Vehicle Maneuver, Turning

The results highlighted that those drivers involved in bicyclists' crashes, while making a turn, are less likely to be involved in a higher severity crash. The results might be attributed to the fact that turning maneuvers are usually associated with lower speed and, consequently, lower crash severity. That is true especially due to the rural nature of Wyoming. Turning, in this study, includes turning right or left. The results somehow are in disagreement with the previous study that motorists turning left is associated with an increased probability for cyclist injuries [8].

### 4.2. *Bicyclists' Actions and Characteristics*

This subsection would outline some of the actions the bikers took while being involved in crashes. In addition, a single attribute related to the biker age is discussed at the end.

#### 4.2.1. Bikers Crossing/Entering the Roads

Sometimes bicyclists try to get into the flow of the traffic perpendicular to the dominant direction of travel. The result highlights the impact of the crossing behavior of bikers on the severity of crashes. For instance, previous studies highlighted that running the red light at the crosswalk is one of the common causes of bike crashes [28]. The impact in this study is intuitive and might be due to more interaction with motor vehicles. Additionally, it was found that when sidewalk cyclists enter the intersection, there is an elevated risk [29]. Therefore, for this study, we assigned a higher severity level due to a higher interaction between bikers and vehicles.

#### 4.2.2. Bicyclists Travel with the Direction of Traffic, Same Direction

For this type of driver actions, bicyclists move in the direction of the traffic flow. The impact is lower compared with crossing and entering the traffic, which might be related to less interaction with vehicles, while getting into a crash. Although this type of interaction, compared with types of collision when entering the roadway, has an increasing impact on the severity of crashes, its impact is less than crossing the roadway. It is intuitive to think that travelling in the same direction is likely to result in the rear-ending of bicycles.

#### 4.2.3. Bikers Cross Marked Line at an Intersection

Providing the marked line facilitates crossing safely for pedestrian and bikers, they could be used for an uncontrolled locations or an intersection equipped with a traffic light. The results highlighted that risk of injury severity for bicycle crossing would be reduced when there is assigned marked line, which is in line with a previous study [30].

#### 4.2.4. Biker's Age

Here the cutting point of 20 years old was chosen as it divides this age group into two almost equal categories. The results highlighted that when bikers aged 20 years old or older got hit by a motor vehicle, the bikers' crash severity would be increased. It is expected that there are many confounding factors to this impact. This especially holds true as the cutting point of the age category was very low. The low value of the cutting point is due to the use of bikes by the younger generation.

### 4.3. Environmental and Roadway Characteristics

#### 4.3.1. Light Condition

The results highlighted that biking at night, compared with the daytime, increases the crash severity of bikers,  $\hat{\beta} = 0.60$ . However, the uncertainty associated with this parameter should be taken into consideration since the confidence interval (CI) is wide, including 0.

#### 4.3.2. Posted Speed Limit

The posted speed limit of 20 mph or lower has been selected in Wyoming due to the posted speed limit in local areas serving residential or school areas. The result is expected as a higher speed limit is associated with a higher severity of crashes. The previous study highlights that a speed limit of 35 mph or higher is associated with a higher severity level for pedestrians [31].

#### 4.3.3. Number of Lanes

Here, the number of lanes for considered observations was between 1 and a maximum value of 4. The threshold here was set as a boundary of greater than 2. The results highlighted that, when the number of lanes increase, the severity of bicyclist injury increases as well. The result is expected, because as the number of lanes increase, the roadway provides a more unsafe environment for the bikers.

### 4.4. Model Performance

We compared the performance of the standard Bayesian technique with the technique considering the random effect. The result highlighted a significant improvement in the model fit by considering the objective hierarchy as our assigned random effect, compared with a model with no hierarchy, WAIC = 173 vs. WAIC = 463, for random effect and standard models, respectively. It should also be noted that, based on the discussion in the literature, WAIC is more valid than the deviance information criteria (DIC), as the averaging could be conducted over the posterior distribution rather than being conditioned on a point estimate. Which make it more valid, especially from the prediction point of view in the Bayesian context [23].

## 5. Discussion

Due to the importance of bikers' safety concerns, their safety has become an essential goal of transportation and safety engineers' policies. Biking makes the biking-road users especially vulnerable as they need to share the roads with other motor vehicle users. Thus, this study was conducted to examine factors associated with injury severity of bicyclists in bicycle–motor vehicle crashes. The first step in the reduction of the severity of bicyclists' crashes might be through identification of contributory factors to those crashes. However, for doing any analysis, it is important to account for the heterogeneity of the dataset, as not accounting for the heterogeneity might result in biased and even erroneous models' points' estimates. In traffic safety, most past studies assign a subjective hierarchy, such as highway system or seasonality. However, those subjective assignments are unlikely to account for an optimum amount of heterogeneity.

Therefore, to account for an optimal amount of heterogeneity in the dataset, we implemented the finite mixture model, being mainly based on the covariance matrices. This is one of the earliest studies employing the finite mixture model for assignments of an objective hierarchy. After allocation of the observations to their related clusters, the clusters were used in a main model for doing the final analysis. An extensive discussion has been made regarding the mathematical formulation and methodological approach of the implemented methods. The results of the implemented technique highlighted a significant improvement, compared with a standard Bayesian model.

The past studies highlighted the importance of accounting for differences across case studies in various geographic areas. The difference is acknowledged in this study. For instance, it was found that, in Wyoming, when hit-and-run drivers were involved in the

crash, the severity of bicyclist crashes seemed to be minor. That is against the findings in the literature review that hit-and-run crashes would result in increased crash severity. We discussed in the manuscript that the results could be mainly due to the fact that those bicyclists were hit by older drivers, and the result might be due to lack of attention by those drivers after the crash occurred.

The results of the identified model highlighted that various bikers' and drivers' actions, as well as environmental and roadway characteristics, could contribute to the severity of bicyclists' crash severity. These include factors increasing the biker's severity, such as alcohol involvement, older bikers, when bikers enter the roadway and travelling in the direction of traffic flow. The results also highlighted that there are some discrepancies across the identified results and the previous findings in the literature, which was discussed in the manuscript. Particular attention was paid regarding the model's parameters estimations.

### Recommendations

Bicyclists are one of the most vulnerable road users due to lack of protection, compared with motor vehicles. As a result, understanding the contributory factors to those crashes, by studying the patterns, could guide policies targeting the improvement of traffic safety. Based on the obtained results, the following recommendations could be given to the policy makers in the state:

- Various actions of bikers, such as crossing the roadway or the impact of the posted speed limit, on biker's crash severity could be due to a conflict between bikers and motor vehicles. Certain bicycle treatments, such as bike lanes and removal of on-street parking, could reduce the interactions between cyclists and motor vehicles, and consequently could reduce crashes, along with their severities.
- Based on the identified results of drivers' and cyclists' actions, it is expected that one of those parties did not follow road signs and signals. Educating the drivers and bikers to respect one another's space would be recommended.
- More regulations and law enforcement are recommended for both parties to reduce the crash frequency and severities. It is especially important to provide more marked lines for bikers, as it was found that conflicts between bikers and motor vehicles (e.g., crossing) increases the severity of bikers' crashes. Additionally, countermeasures would reduce the severity of bicyclists' crashes significantly. For instance, the areas assigned to bikers and pedestrians should be enforced to make sure vehicles do not violate the designated areas.
- More studies and investigations are needed to study the discrepancies between this study and previous studies. Especially more investigations are needed to study the reason behind the lower hit-and-run crash severity in the state.

**Author Contributions:** K.K. supervision, M.R. methodology. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors would like to acknowledge that this work is part of a project funded by the Wyoming Department of Transportation (WYDOT). The subject matter, all figures, tables and equations not previously copyrighted by outside sources are copyrighted by WYDOT, State of Wyoming and the University of Wyoming. All rights reserved copyrighted in 2016.

**Conflicts of Interest:** The authors declare no conflict of interest.

### References

1. Deenihan, G.; Caulfield, B. Estimating the health economic benefits of cycling. *J. Transp. Health* **2014**, *1*, 141–149. [[CrossRef](#)]
2. Oja, P.; Titze, S.; Bauman, A.; De Geus, B.; Krenn, P.; Reger-Nash, B.; Kohlberger, T. Health benefits of cycling: A systematic review. *Scand. J. Med. Sci. Sports* **2011**, *21*, 496–509. [[CrossRef](#)] [[PubMed](#)]
3. World Health Organization. *Global Status Report on Road Safety 2015*; World Health Organization: Geneva, Switzerland, 2015.
4. Pedestrian and Bicycle Information Center. National Bicycling and Walking Study: 15 Year Status Report. Available online: [https://www.pedbikeinfo.org/cms/downloads/15-year\\_report.pdf](https://www.pedbikeinfo.org/cms/downloads/15-year_report.pdf) (accessed on 1 March 2021).

5. National Highway Traffic Safety Administratio. Traffic Safety Facts 2017 Data: Pedestrians. Available online: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812806> (accessed on 1 March 2021).
6. DiGioia, J.; Watkins, K.E.; Xu, Y.; Rodgers, M.; Guensler, R. Safety impacts of bicycle infrastructure: A critical review. *J. Saf. Res.* **2017**, *61*, 105–119. [[CrossRef](#)]
7. Pucher, J.; Dijkstra, L. Promoting safe walking and cycling to improve public health: Lessons from the Netherlands and Germany. *Am. J. Public Health* **2003**, *93*, 1509–1516. [[CrossRef](#)]
8. Chen, P.; Shen, Q. Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accid. Anal. Prev.* **2016**, *86*, 239–246. [[CrossRef](#)]
9. Moore, D.N.; Schneider, W.H., IV; Savolainen, P.T.; Farzaneh, M. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accid. Anal. Prev.* **2011**, *43*, 621–630. [[CrossRef](#)]
10. Thomas, L.; Nordback, K.; Sanders, R. Bicyclist crash types on national, state, and local levels: A new look. *Transp. Res. Rec.* **2019**, *2673*, 664–676. [[CrossRef](#)]
11. Yuan, Q.; Yang, H.; Huang, J.; Kou, S.; Li, Y.; Theofilatos, A. What factors impact injury severity of vehicle to electric bike crashes in China? *Adv. Mech. Eng.* **2017**, *9*, 1–10. [[CrossRef](#)]
12. Reynolds, C.C.; Harris, M.A.; Teschke, K.; Crompton, P.A.; Winters, M. The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. *Environ. Health* **2009**, *8*, 47. [[CrossRef](#)]
13. Helak, K.; Jehle, D.; McNabb, D.; Battisti, A.; Sanford, S.; Lark, M.C. Factors Influencing Injury Severity of Bicyclists Involved in Crashes with Motor Vehicles: Bike Lanes, Alcohol, Lighting, Speed, and Helmet Use. *South. Med. J.* **2017**, *110*, 441–444. [[CrossRef](#)] [[PubMed](#)]
14. Lin, Z.; Fan, W. Modeling bicyclist injury severity in bicycle–motor vehicle crashes that occurred in urban and rural areas: A mixed logit analysis. *Can. J. Civ. Eng.* **2019**, *46*, 924–933. [[CrossRef](#)]
15. Carvajal, G.A.; Sarmiento, O.L.; Medaglia, A.L.; Cabrales, S.; Rodríguez, D.A.; Quistberg, D.A.; López, S. Bicycle safety in Bogotá: A seven-year analysis of bicyclists’ collisions and fatalities. *Accid. Anal. Prev.* **2020**, *144*, 105596. [[CrossRef](#)]
16. Bensmail, H.; Celeux, G. Regularized Gaussian discriminant analysis through eigenvalue decomposition. *J. Am. Stat. Assoc.* **1996**, *91*, 1743–1748. [[CrossRef](#)]
17. Fraley, C.; Raftery, A.E. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [[CrossRef](#)]
18. Betancourt, M.; Girolami, M. Hamiltonian Monte Carlo for hierarchical models. *Curr. Trends Bayesian Methodol. Appl.* **2015**, *79*, 2–4.
19. Roberts, G.O.; Gelman, A.; Gilks, W.R. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **1997**, *7*, 110–120. [[CrossRef](#)]
20. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534. [[CrossRef](#)]
21. Thomas, S.; Tu, W. Learning Hamiltonian Monte Carlo in R. *Am. Stat.* **2020**, 1–28. [[CrossRef](#)]
22. Vehtari, A.; Gelman, A.; Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **2017**, *27*, 1413–1432. [[CrossRef](#)]
23. Gelman, A.; Hwang, J.; Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **2014**, *24*, 997–1016. [[CrossRef](#)]
24. Zhu, S. Investigation of vehicle-bicycle hit-and-run crashes. *Traffic Inj. Prev.* **2020**, *21*, 506–511. [[CrossRef](#)] [[PubMed](#)]
25. Legree, P.J.; Heffner, T.S.; Psotka, J.; Martin, D.E.; Medsker, G.J. Traffic crash involvement: Experiential driving knowledge and stressful contextual antecedents. *J. Appl. Psychol.* **2003**, *88*, 15. [[CrossRef](#)] [[PubMed](#)]
26. Abdu, R.; Shinar, D.; Meiran, N. Situational (state) anger and driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2012**, *15*, 575–580. [[CrossRef](#)]
27. Ogden, E.J.; Moskowitz, H. Effects of alcohol and other drugs on driver performance. *Traffic Inj. Prev.* **2004**, *5*, 185–198. [[CrossRef](#)]
28. Guo, Y.; Osama, A.; Sayed, T. A cross-comparison of different techniques for modeling macro-level cyclist crashes. *Accid. Anal. Prev.* **2018**, *113*, 38–46. [[CrossRef](#)]
29. Wachtel, A.; Lewiston, D. Risk factors for bicycle-motor vehicle collisions at intersections. *ITE J.* **1994**, *64*, 30–35.
30. Jensen, S.U. Safety effects of blue cycle crossings: A before-after study. *Accid. Anal. Prev.* **2008**, *40*, 742–750. [[CrossRef](#)] [[PubMed](#)]
31. Zegeer, C.V.; Richard Stewart, J.; Huang, H.; Lagerwey, P. Safety effects of marked versus unmarked crosswalks at uncontrolled locations: Analysis of pedestrian crashes in 30 cities. *Transp. Res. Rec.* **2001**, *1773*, 56–68. [[CrossRef](#)]