Accident Analysis and Prevention 64 (2014) 41–51

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Bayesian methodology to estimate and update safety performance functions under limited data conditions: A sensitivity analysis☆

Shahram Heydari[a,*], Luis F. Miranda-Moreno[b,1], Dominique Lord[c,2], Liping Fu[a,3]

[a] Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue W., Waterloo, ON N2L 3G1, Canada
[b] Department of Civil Engineering and Applied Mechanics, McGill University, 817 Sherbrooke St. W., Montreal, QC H3A 2K6, Canada
[c] Zachary Department of Civil Engineering, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Article history:
Received 21 May 2013
Received in revised form 27 October 2013
Accepted 1 November 2013

Keywords:
Full Bayes road safety analysis
Prior distribution
SPF parameter estimation and update
Index of treatment effectiveness

ABSTRACT

In road safety studies, decision makers must often cope with limited data conditions. In such circumstances, the maximum likelihood estimation (MLE), which relies on asymptotic theory, is unreliable and prone to bias. Moreover, it has been reported in the literature that (*a*) Bayesian estimates might be significantly biased when using non-informative prior distributions under limited data conditions, and that (*b*) the calibration of limited data is plausible when existing evidence in the form of proper priors is introduced into analyses. Although the Highway Safety Manual (2010) (HSM) and other research studies provide calibration and updating procedures, the data requirements can be very taxing. This paper presents a practical and sound Bayesian method to estimate and/or update safety performance function (SPF) parameters combining the information available from limited data with the SPF parameters reported in the HSM. The proposed Bayesian updating approach has the advantage of requiring fewer observations to get reliable estimates. This paper documents this procedure. The adopted technique is validated by conducting a sensitivity analysis through an extensive simulation study with 15 different models, which include various prior combinations. This sensitivity analysis contributes to our understanding of the comparative aspects of a large number of prior distributions. Furthermore, the proposed method contributes to unification of the Bayesian updating process for SPFs. The results demonstrate the accuracy of the developed methodology. Therefore, the suggested approach offers considerable promise as a methodological tool to estimate and/or update baseline SPFs and to evaluate the efficacy of road safety countermeasures under limited data conditions.

1. Introduction

Safety performance functions (SPFs) often referred to as crash frequency models are an essential component of road safety studies. In practice, roadway or transportation agencies often need to estimate crash frequency models for limited data, that is, data with only a small number of observations and limited number of contributing factors (independent variables). In fact, limited data conditions frequently occur in road safety analyses, mainly due to the lack of funds required to involve a large sample of sites in developing SPFs and/or conducting before–after observational studies (Lord and Bonneson, 2005). Hence, practitioners often need to calibrate statistical models under these restrictions to obtain baseline SPFs. The MLE that relies on asymptotic theory has been shown to be unreliable for limited data conditions (Lord, 2006; Daziano et al., 2013). However, the full Bayes (FB) paradigm can be employed as a viable alternative to the MLE.

Some advantages of the FB context compared to its Frequentist counterpart are, first, that the available information (based on expert criteria, previous studies, etc.) related to the parameters of interest can be incorporated into the analysis by assigning prior distributions to these parameters. This is a vital advantage resulting in unbiased estimates for limited data (Lord and Miranda-Moreno, 2008; Miranda-Moreno et al., 2013; Heydari et al., 2013). By using suitable priors, thus, the sample size required to conduct a reliable road safety analysis may decrease. Second, Bayesian statistics have a natural characteristic of accommodating hierarchical models. Note that hierarchical models are capable of dealing with complex data structures and their use in the Bayesian methods is common and straightforward (Gelman et al., 2003). Third, solving complex

☆ A preliminary version of this paper has been presented at the International Road Safety and Simulation Conference in Rome, Italy (2013).

* Corresponding author. Tel.: +1 519 888 4567.
E-mail addresses: shahram.heydari@uwaterloo.ca (S. Heydari), luis.miranda-moreno@mcgill.ca (L.F. Miranda-Moreno), d-lord@tamu.edu (D. Lord), lfu@uwaterloo.ca (L. Fu).
[1] Tel.: +1 514 398 6589.
[2] Tel.: +1 979 458 3949.
[3] Tel.: +1 519 888 4567.

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.11.001

statistical models in the Frequentist framework requires further computational efforts and algorithms such as in a situation in which the MLE cannot solve a problem, and a simulation based solution is necessary. However, the Bayesian perspective, regardless of the model complexity, always applies the Bayes theorem to derive the posterior inference. Thus, the modeler can usually adopt a complicated statistical model, then run Markov chain Monte Carlo (MCMC) simulations to obtain the posteriors without any additional effort to invent or develop a method for this goal. Fourth, the uncertainty in the Frequentist statistics is mainly addressed via confidence intervals that do not imply, as believed by many practitioners, that an estimate occurs in this interval with a certain probability for a given dataset. Nevertheless, this implication can be made under the Bayesian approach. In other words, the Frequentist context cannot conclude, for an observed dataset, the probability for an estimate being in a certain interval. It can only state that a confidence interval contains the estimated value given that a considerable number of trials are repeated. Conversely, Bayesian statistics, for a given dataset, provide directly the probability that an estimate occurs in an interval. Based on this probability, a direct and explicit statement can be made, which is a natural interpretation of results. On the basis of the above-mentioned advantages, this paper adopts a Bayesian updating method.

When using FB methods in road safety studies, the majority of research papers employ non-informative priors to analyze accident data. As mentioned above, specifying informative priors can result in more-robust estimates under certain circumstances (Lord and Miranda-Moreno, 2008; Miranda-Moreno et al., 2013; Heydari et al., 2013). To specify alternative priors, for instance, Miranda-Moreno et al. (2013) built informative priors for the inverse dispersion parameter from the reported values in previous studies. However, doing so, the estimated priors vary from one practitioner to another since different studies are usually taken into account yielding disparate statistical inferences. Heydari et al. (2013), using a large number of quasi-simulated data to draw reliable statistical inferences, investigated the effect of prior specification (based on past evidence) on regression coefficient (e.g., traffic flow) estimates, inverse dispersion parameter estimates, hotspot identification, and goodness-of-fit. The authors concluded that the inverse dispersion parameter is the SPF parameter most sensitive to prior choice, while hotspot identification and deviance information criteria (DIC) are slightly affected by prior specification.

Following the aforementioned studies, this paper takes advantage of informative and semi-informative priors to calibrate and update SPFs. Note that great strides have been made in improving updating and re-calibration procedures (Persaud et al., 2002; Hadayeghi et al., 2006; Yu and Abdel-Aty, 2013; Connors et al., 2013; Wood et al., 2013). However, the data requirements for updating SPFs can be very taxing (for example, Wood et al. (2013) have expressed some concerns in this regard). We propose a practical and efficient Bayesian updating approach that has the advantage of requiring fewer observations to get reliable estimates. This approach combines the information available from limited data with specific SPF parameter values reported in the HSM to estimate and/or update SPFs. Related to the updating problem, Yu and Abdel-Aty (2013) compared different prior selection methods for Bayesian updating of SPFs. The authors grouped their case study into training and test datasets. A two-stage Bayesian updating procedure was then developed and stated to be able to provide the most accurate estimates. In practice, however, finding a training database with characteristics similar to the limited data under investigation may not be always feasible. Connors et al. (2013), based on a case study of rural single carriageway roads, compared the performance of different models (e.g., Poisson-gamma and Poisson-Weibull models) using both the MCMC and the MLE estimation techniques. The authors assumed three informative priors (gamma, log normal, and Weibull) for the inverse dispersion parameter, which have a mean of 1 and c_v (coefficient of variation) value of 0.8. Other issues such as goodness-of-fit measures and independency among observations in different years were also investigated in the latter paper. For a detailed discussion on the updating issues, including temporal transferability of SPFs, see Wood et al. (2013).

The scope of this research study was to offer a practical and efficient Bayesian method to estimate and update SPF parameters (which, in turn, allows practitioners to evaluate safety treatment effectiveness), especially under limited data conditions. This paper's methodology was tested through an extensive simulation exercise—instead of using a single case study—to draw reliable statistical inferences. For this purpose, we examined the estimation of the SPF parameters and indices of treatment effectiveness in before–after studies, using a large variety of prior assumptions. It was thus possible to compare a large number of prior distributions for model parameters (informative, semi-informative, and non-informative). Despite the complexity of the adopted simulation study, which was inevitable to validate the proposed method, our technique can be easily applied by practitioners. Note that by using this technique, transportation authorities will be able to conduct road safety studies even with a limited number of observations, thereby reducing the cost and time for the decision-making process.

2. Methodology

When dealing with limited data, one may use the HSM parameters (or any other similar guideline) to obtain a baseline model. Doing so, however, one may introduce bias in SPF parameter estimates since every dataset has its own traits. For example, a dataset from a particular jurisdiction in one part of a country may be characterized differently (e.g., different reporting criteria, etc.) than in another jurisdiction. As a result, the regression coefficients will never be exactly equal to those reported in the HSM. Therefore, a better practice would be to calibrate an SPF for each dataset separately. Now, if the number of observations is limited (i.e., a critical condition in which data is characterized by low mean values and/or small sample size), we suggest the use of HSM parameters as prior information (to estimate proper prior distributions) to calibrate an SPF for each specific database. To this end, a methodology is provided to successfully estimate and update SPF parameters from the HSM under limited data conditions. To include the low mean value problem in the analysis, fatal-injury accidents were used. Nevertheless, the proposed methodology can also be utilized for property damage only or total accidents.

In this paper, the authors examined the proposed methodology using the most critical condition; i.e., data including a small number of observations having a low mean value. From the HSM, we utilized the reported parameters associated with the contributing factors (e.g., traffic flow) and the function provided for the estimation of the dispersion parameter. Upon doing so, the expected magnitude of a parameter of interest can be introduced into the analysis. Since HSM parameters (see Table 1) are estimated through a comprehensive scientific study, these parameters can be used as a valid reference (prior belief) to build informative priors for unknown SPF parameters. The assumption, in this paper, was that the base conditions indicated in the HSM were satisfied (i.e., all crash modification

Table 1
Parameter values reported in the HSM for undivided 4-lane roadway segments.

	Constant (a_0)	Traffic flow (a_1)	C
Fatal-injury accidents	−9.410	1.094	1.796