ELSEVIER

Accident Analysis and Prevention 40 (2008) 1013–1017

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Methodology for estimating the variance and confidence intervals for the estimate of the product of baseline models and AMFs

Dominique Lord *

Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, USA

Received 9 August 2007; received in revised form 8 November 2007; accepted 21 November 2007

Abstract

This manuscript describes a methodology for estimating the variance and 95% confidence intervals (CI) for the estimate of the product between baseline models and Accident Modification Factors (AMFs). This methodology is provided for the upcoming Highway Safety Manual (HSM) currently in development in the United States (U.S.). The methodology is separated into two parts. The first part covers the proposed approach for estimating the variance of the estimate of the product between baseline models and AMFs. The second part presents the method for estimating the variance of baseline models. Several examples are presented to illustrate the application of the methodology.

Keywords: Crash prediction models; Baseline models; Variance estimation; Confidence intervals; Highway Safety Manual

1. Introduction

This manuscript describes a methodology for estimating the variance and 95% confidence intervals (CI) for the estimate of the product between baseline models and Accident Modification Factors (AMFs). This methodology is provided for the upcoming Highway Safety Manual (HSM) (see Hughes et al., 2005, for additional information) currently in development in the United States (U.S.). The HSM, which is near completion, is a document that will serve as a tool to help practitioners make planning, design, and operations decisions based on safety. The document will serve the same role for safety analysis that the Highway Capacity Manual (HCM) (TRB, 2000) serves for traffic-operations analyses. The purpose of the HSM is to provide the best factual information and tools in a useful and widely accepted form to facilitate explicit consideration of safety in the decision making process.

The technique described in the HSM consists of first developing baseline models using data that meet specific nominal conditions, such as 12-ft. lane and 8-ft. shoulder widths for segments or no turning lanes for intersections. These conditions usually reflect design or operational variables most commonly used by state transportation agencies (defined as state DOTs). Consequently, baseline models typically only include traffic flow as covariates (e.g., $\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2}$). The second component of the technique consists of multiplying the output of such models by AMFs to capture changes in geometric design and operational characteristics (Hughes et al., 2005). An important assumption about using this technique is that the AMFs are considered independent, which may not always be true in practice. The formulation of the technique is given by the following:

$$\mu_{\text{final}} = \mu_{\text{baseline}} \times \text{AMF}_1 \times \cdots \times \text{AMF}_n \quad (1)$$

where μ_{final} is the final predicted number of crashes per unit of time; μ_{baseline} is the baseline predicted number of crashes per unit of time (via a regression model); and $\text{AMF}_1 \times \cdots \times \text{AMF}_n$ are the accident modification factors assumed to be independent.

Recent discussions at various meetings related to the production of the Manual have shown that estimating the uncertainty associated with baseline models, AMFs, and the estimate of the product between the two have become very important in the eyes of the Task Force members, the committee responsible for the implementation of the Manual, as well as for potential HSM users. So far, the work in this area has only focused on estimating the uncertainty associated with AMFs (Bahar et al., 2007) and, to a lesser degree, with regression models

* Tel.: +1 979 458 3949; fax: +1 979 845 6481.
E-mail address: d-lord@tamu.edu.

0001-4575/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2007.11.008

(Wood, 2005; Agrawal and Lord, 2006) (the latter not in the context of the HSM however). There is therefore a need to fill this gap by providing a methodology for estimating the variance of the estimate of the product of baseline models and AMFs.

This document is divided into three sections. The second section describes the first part of the methodology, which consists of estimating the variance of the product of baseline models and AMFs. This section covers the background material on the product of random materials and presents two examples describing different combinations of baseline models, AMFs and their associated uncertainties. The third section describes the second part of the methodology and explains how to compute the variance of baseline models. An example illustrates the computation of the variance of baseline models as well as the 95% CI for the final product.

2. Estimating the variance of baseline models and AMFs

This section describes the first part of the methodology and is divided into two sub-sections. The first sub-section provides details about the theory behind the multiplication of independent random variables. The second sub-section presents the application of the proposed method for estimating the variance of the estimate of the product of baseline models and AMFs. Two examples are provided.

2.1. Computing the product of random variables

The estimation of the variance can be accomplished using the theory behind the multiplication of independent random variables (Ang and Tang, 1975; Browne, 2000). This theory states that the equations presented below will be exact independent of the type of distribution to which each random variable belongs. For the purpose of this description, we will define z as the product of independent random variables:

$$z = x_1 x_2 x_3 \cdots \tag{2}$$

where z is the product of independent random variables; and, x's is the random variable taken from any kind of distribution.

It should be pointed out that the mean and variance estimates are defined as $E[x] = \lambda$ and $E[(x-\lambda)^2] = \nu$ (second central moment), respectively.

2.1.1. Mean of a product

The mean of the product is the direct application of Eq. (2):

$$\begin{aligned} z &= x_1 x_2 x_3 \cdots \\ E[z] &= E[x_1]E[x_2]E[x_3] \cdots \\ \lambda_z &= \lambda_{x1}\lambda_{x2}\lambda_{x3} \cdots \end{aligned} \tag{3}$$

The mean or average of the product is simply the product of the mean value of the random variables.

2.1.2. Variance of a product

The variance of a product is obtained by taking the expectation square of z:

$$\begin{aligned} z^2 &= x_1^2 x_2^2 x_3^2 \cdots \\ E[z^2] &= E[x_1^2]E[x_2^2]E[x_3^2] \cdots \\ (\lambda_z^2 + \nu_z) &= (\lambda_{x1}^2 + \nu_{x1})(\lambda_{x2}^2 + \nu_{x2})(\lambda_{x3}^2 + \nu_{x3}) \cdots \end{aligned} \tag{4}$$

Note: $E[x^n] = E[((x-\lambda)+\lambda)^n]$; for $n = 2$, $E[x^2] = E[((x-\lambda)+\lambda)^2] = E[(x-\lambda)^2] + E[2\lambda(x-\lambda)] + E[\lambda^2] = \nu + \lambda^2$.

The variance ν_z is computed by first calculating the product on the right hand side and then subtracting the square of the mean λ_z^2 computed in Eq. (4):

$$\nu_z = (\lambda_{x1}^2 + \nu_{x1})(\lambda_{x2}^2 + \nu_{x2})(\lambda_{x3}^2 + \nu_{x3}) \cdots - \lambda_z^2 \tag{5}$$

Note that if all ν_x's equal zero, the variance ν_z will also equal zero:

$$\nu_z = (\lambda_{x1}^2)(\lambda_{x2}^2)(\lambda_{x3}^2) \cdots - \lambda_z^2 = \lambda_z^2 - \lambda_z^2 = 0 \tag{6}$$

2.2. Application of the theory

This section describes the application of the theory behind the multiplication of independent random variables. Two examples describing different values of predicted values, AMFs, and associated uncertainties are presented. The uncertainty associated with the baseline models can be estimated using the method described in the next section. For estimating the uncertainty related to AMFs, the reader is referred to the work of Bahar et al. (2007), which will be incorporated into the forthcoming HSM.

2.2.1. Example 1—one AMF

This example shows the application of a single AMF. Let x_1 represent the predicted value of a baseline model and x_2 an AMF:

x_1 = 5.0 crashes/year (standard deviation or S.D. = 2.0 crashes/year)
x_2 = 0.80 (S.D. = 0.10)

The mean is given by:

$$\lambda_z = 5.0 \times 0.80 = 4.0$$

The variance is given by:

$$\begin{aligned} \nu_z &= (5.0^2 + 4.0)(0.80^2 + 0.01) - 4.0^2 = (29)(0.65) - 16.0 \\ &= 18.85 - 16.0 = 2.85 \end{aligned}$$

The final value is estimated to be:

4.0 crashes/year (S.D. = 1.69 crashes/year)

2.2.2. Example 2—two AMFs

In this example, two AMFs are used. Let x_1 represent the predicted value of a baseline model and x_2 and x_3 independent