MALAYA JOURNAL OF MATEMATIK
Malaya J. Mat. **10(04)**(2022), 322–335.
http://doi.org/10.26637/mjm1004/003

Theoretical and empirical comparison of two discrete statistical models of crash frequencies

ISSA CHERIF GERALDO[*1] AND TCHILABALO ABOZOU KPANZOU[2]

[1] *Laboratoire d'Analyse, de Modélisations Mathématiques et Applications (LAMMA), Département de Mathématiques, Faculté des Sciences, Université de Lomé, 1 B.P. 1515 Lomé 1, Togo.*
[2] *Département de Mathématiques, Faculté des Sciences et Techniques, Université de Kara, Kara, Togo.*

***R*eceived 24 May 2022; *A*ccepted 27 September 2022**

Abstract. In this paper, we compare two discrete statistical models for the evaluation of a road safety measure. We give a much simpler proof of the expression of the maximum likelihood estimator for the more complex model and we demonstrate theoretical results on the measure of divergence between the two models. The results obtained on real data suggest that both models are very competitive.

AMS Subject Classifications: 62F10, 62F30, 62H10, 62P99.

Keywords: Statistical model, maximum likelihood, parameter estimation, road safety, Kullback-Leibler divergence.

Contents

[*] **Corresponding author.** Email address: **cherifgera@gmail.com** (I.C. GERALDO), **kpanzout@gmail.com** (T.A. KPANZOU)

1. Introduction

Road accidents are very unfortunate events that often cause damage of all kinds, the heaviest of which is the loss of lives. According to a report published in 2018 by the World Health Organization (WHO), it can be estimated that more than 1.35 million people are killed annually by road accidents (about 2.5% of deaths worldwide) and more than 50 million people are injured annually because of road accidents [17].

As the science of data collection, analysis and interpretation, statistics play an important role in the study of road accidents. Over the years, a plethora of statistical models have been proposed for accident data and several states of the art of these models have been made. Among them are the papers [7] and [8] in which the authors present the different existing methodological approaches as well as the strengths and weaknesses of each approach. It should be noted that statistical models very often depend on the context (available data and objective of the study).

One of the important contributions of statistics in the field of road safety concerns the statistical evaluation of the effect of a road safety measure (reduction or increase of the maximum speed allowed on a road, installation of speed bumps, installation of signs, transformation of intersections into roundabouts, modification of road markings, etc.). After a certain period of application of a certain measure, it is necessary and logical to seek to assess its effect in order to ascertain whether or not that measure has made it possible to reduce the number of accidents.

In this paper, we consider the case where accidents are classified by level of severity. Among the most widely used statistical models in this context, before-after models with control site are in a good position since they allow cause and effect interpretation [5]. These models simultaneously consider accidents by level of severity before and after the road safety measure not only at the site that underwent the measure (often referred to as the experimental site or treated site or treatment site) but also at another site (often called the control site or comparison site) with the same characteristics as the treatment site but where the measure was not applied. This avoids attributing erroneously to the measure any underlying trend due to factors other than the measure [15].

Let n be the total number of accidents at the experimental site, $r \in \mathbb{N}^*$ be the number of accident severity levels and $\boldsymbol{X} = (X_{11}, \ldots, X_{1r}, X_{21}, \ldots, X_{2r}) \in \mathbb{R}^{2r}$ be a discrete random vector where, for all $j = 1, \ldots, r$, X_{ij} is the discrete random variable representing the number of accidents of severity level j on the experimental site during the period i ($i = 1$ represents the period before the measure and $i = 2$ represents the period after). Modelling also requires the non-random vector $\boldsymbol{z} = (z_1, \ldots, z_r)$, where for all $j = 1, \ldots, r$,

$$z_j = \frac{\text{number of crashes of severity level } j \text{ on the control site in the "after" period}}{\text{number of crashes of severity level } j \text{ on the control site in the "before" period}}. \tag{1.1}$$

The objective of this paper is to theoretically and empirically compare two very relevant statistical models proposed respectively in [10] (referred to as Model 1 in the following) and [11] (referred to as Model 2 in the following). These two models represent the probability distribution of the vector $\boldsymbol{X}$ by a multinomial distribution

$$\boldsymbol{X} \rightsquigarrow \mathcal{M}(n, \boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{z})), \tag{1.2}$$

where

$$\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{z}) = (\pi_{11}(\boldsymbol{\theta}|\boldsymbol{z}), \ldots, \pi_{1r}(\boldsymbol{\theta}|\boldsymbol{z}), \pi_{21}(\boldsymbol{\theta}|\boldsymbol{z}), \ldots, \pi_{2r}(\boldsymbol{\theta}|\boldsymbol{z})) \tag{1.3}$$

is a vector function of an unknown parameter vector (to be estimated) $\boldsymbol{\theta} \in \mathbb{R}^{r+1}$ and $\boldsymbol{z}$ is such that, for all $i = 1, 2$ and $j = 1, \ldots, r$, $0 < \pi_{ij}(\boldsymbol{\theta}|\boldsymbol{z}) < 1$ and $\sum_{i=1}^{2} \sum_{j=1}^{r} \pi_{ij}(\boldsymbol{\theta}|\boldsymbol{z}) = 1$. The notation (1.2) means that for any realization $\boldsymbol{x} = (x_{11}, \ldots, x_{1r}, x_{21}, \ldots, x_{2r})$ of $\boldsymbol{X}$ such that $\sum_{i=1}^{2} \sum_{j=1}^{r} x_{ij} = n$, the probability function evaluated to the vector $\boldsymbol{x}$ is

$$P(\boldsymbol{x}) = \frac{n!}{\prod_{i=1}^{2} \prod_{j=1}^{r} x_{ij}!} \prod_{i=1}^{2} \prod_{j=1}^{r} (\pi_{ij}(\boldsymbol{\theta}|\boldsymbol{z}))^{x_{ij}}. \tag{1.4}$$

The difference between the two models considered in this paper lies in the definition of the function $\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{z})$ and the theoretical and empirical comparison that we propose is an innovation because until now, in the literature, the two models have been treated separately.

In addition to the introductory section, this paper has other sections organized as follows. Section 2 presents the two models and the estimation of the parameter vector $\boldsymbol{\theta}$ for each of them. Section 3 presents the main theoretical results of this paper. In this section, we give a much simpler proof of the expression of the maximum likelihood estimator for Model 1 and we demonstrate theoretical results on the measure of divergence between the two models. Section 4 presents the results of the comparison on real data. Some concluding remarks are given in Section 5.

2. Models and estimation of their parameters

2.1. Presentation of the models

The details of the construction of both models can be found in [10, 11] and are not presented here. The main question behind these models is: how to calculate the average effect of the measure on the number of accidents? This average effect, denoted α in the rest of this paper, is a strictly positive real number defined as the ratio of the number of accidents observed at the experimental site in the "after" period to the number of accidents that one would have expected to observe if the measure had no effect [11]. Consideration of the different types of accident severity introduces r positive secondary parameters $\beta_1, \ldots, \beta_r$ such as $\sum_{i=1}^{r} \beta_i = 1$. In each of the models, the vector parameter $\boldsymbol{\theta}$ takes the form $\boldsymbol{\theta} = (\alpha, \beta_1, \ldots, \beta_r)$.

In Model 1 [10], the components of the vector function $\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{z})$ are defined by:

$$\pi_{1j}^{(1)}(\boldsymbol{\theta}|\boldsymbol{z}) = \frac{\beta_j}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad \pi_{2j}^{(1)}(\boldsymbol{\theta}|\boldsymbol{z}) = \frac{\alpha\beta_j z_j}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad j = 1, \ldots, r, \tag{2.1}$$

whereas in Model 2 [11], they are defined by:

$$\pi_{1j}^{(2)}(\boldsymbol{\theta}|\boldsymbol{z}) = \frac{\beta_j}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad \pi_{2j}^{(2)}(\boldsymbol{\theta}|\boldsymbol{z}) = \frac{\alpha\beta_j \sum_{k=1}^{r} z_k\beta_k}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad j = 1, \ldots, r. \tag{2.2}$$

The value of the mean effect α can be interpreted by comparing it to 1 (for example, if $\alpha = 0.8 < 1$, then $1 - \alpha = 0.2 = 20\%$ and we can estimate the reduction in the number of accidents due to the measure at 20%). The parameters $\beta_1, \ldots, \beta_r$ are the respective probabilities associated with the severity levels.

2.2. Maximum likelihood estimation of parameters

Let $\boldsymbol{x} = (x_{11}, \ldots, x_{1r}, x_{21}, \ldots, x_{2r})$ be an observation of $\boldsymbol{X}$ such that $\sum_{i=1}^{2}\sum_{j=1}^{r} x_{ij} = n$. Applying the logarithm to Formula (1.4) and taking into account the definition of $\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{z})$ for each of the two models, one can verify that the log-likelihoods associated with the two models are respectively defined by

$$L_1(\boldsymbol{\theta}) = C + \sum_{j=1}^{r}\left\{x_{\bullet j}\log(\beta_j) + x_{2j}\log(\alpha) - x_{\bullet j}\log\left(1+\alpha\sum_{k=1}^{r} z_k\beta_k\right) + x_{2j}\log z_j\right\} \tag{2.3}$$

and

$$L_2(\boldsymbol{\theta}) = C + \sum_{j=1}^{r}\left\{x_{\bullet j}\log(\beta_j) + x_{2j}\log(\alpha) - x_{\bullet j}\log\left(1+\alpha\sum_{k=1}^{r} z_k\beta_k\right) + x_{2j}\log\left(\sum_{k=1}^{r} z_k\beta_k\right)\right\}, \tag{2.4}$$

where

$$C = \log\left(\frac{n!}{\prod_{i=1}^{2}\prod_{j=1}^{r} x_{ij}!}\right).$$

Remark 2.1. *Let* $\mathbb{S}_{r-1} = \{(\beta_1, \ldots, \beta_r) \in (\mathbb{R}_+^*)^r, \quad \sum_{i=1}^r \beta_i = 1\}$. *For all* $\boldsymbol{\theta} \in \mathbb{R}_+^* \times \mathbb{S}_{r-1}$, *the difference between the two log-likelihoods*

$$L_2(\boldsymbol{\theta}) - L_1(\boldsymbol{\theta}) = \sum_{j=1}^{r} x_{2j} \left\{ \log \left(\sum_{k=1}^{r} z_k \beta_k \right) - \log z_j \right\}$$

does not have a constant sign (see Figure 1). In the special case $z_1 = \cdots = z_r$, *we have* $L_2(\boldsymbol{\theta}) - L_1(\boldsymbol{\theta}) = 0$.

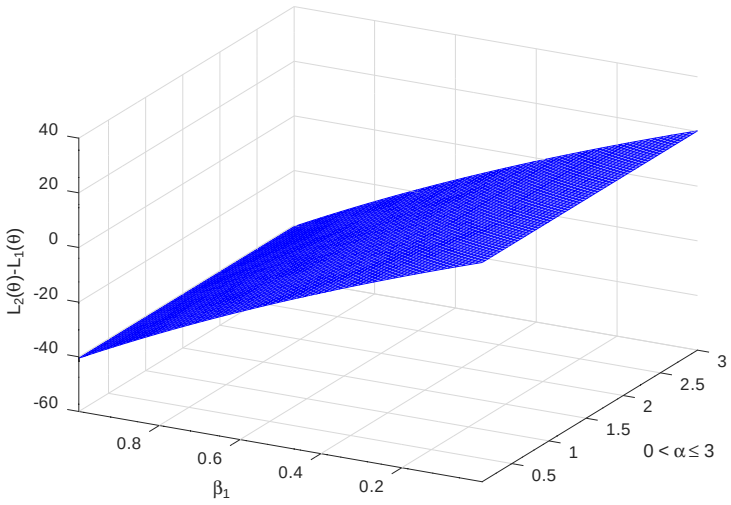

Figure 1: Example of representation of $L_2(\boldsymbol{\theta}) - L_1(\boldsymbol{\theta})$ for r =2.

The respective Maximum Likelihood Estimators (MLE) $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_r)$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\alpha}, \tilde{\beta}_1, \ldots, \tilde{\beta}_r)$ of the unknown parameter vector $\boldsymbol{\theta}$ are obtained by maximizing $L_1(\boldsymbol{\theta})$ and $L_2(\boldsymbol{\theta})$.

For Model 1, it was proved that an explicit form could not be obtained for $\boldsymbol{\theta}$ [9]. An estimation algorithm called cyclic algorithm (CA) was therefore proposed in [12]. Let $x_{1\bullet} = \sum_{j=1}^r x_{1j}$, $x_{2\bullet} = \sum_{j=1}^r x_{2j}$ and, for all $j = 1, \ldots, r$, $x_{\bullet j} = x_{1j} + x_{2j}$. The CA is written in iterative form

$$\begin{aligned} \alpha^{(k+1)} &= \frac{x_{2\bullet}}{x_{1\bullet} \left(\sum_{j=1}^r z_j \beta_j^{(k)} \right)} \\ \beta_j^{(k+1)} &= \frac{1}{1 - \frac{1}{n} \sum_{m=1}^{r} \frac{x_{\bullet m} \alpha^{(k+1)} z_m}{1 + \alpha^{(k+1)} z_m}} \times \frac{x_{\bullet j}}{n(1 + \alpha^{(k+1)} z_j)}, \quad j = 1, \ldots, r, \end{aligned} \tag{2.5}$$

where $\alpha^{(k+1)}$ and $\beta_j^{(k+1)}$ $(j = 1, \ldots, r)$ denote the respective estimates of α and β_j after $k+1$ iterations. This algorithm begins with an initial vector $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)}, \beta_1^{(0)}, \ldots, \beta_r^{(0)})$ and stops when two successive values $\boldsymbol{\theta}^{(k)} = (\alpha^{(k)}, \beta_1^{(k)}, \ldots, \beta_r^{(k)})$ and $\boldsymbol{\theta}^{(k+1)} = (\alpha^{(k+1)}, \beta_1^{(k+1)}, \ldots, \beta_r^{(k+1)})$ are such that $|L_1(\boldsymbol{\theta}^{(k+1)}) - L_1(\boldsymbol{\theta}^{(k)})| < \epsilon$ for a sufficiently small precision $\epsilon > 0$. The global convergence of the (2.5) algorithm to the MLE in the algorithmic sense (i.e. the convergence of the sequence $(\boldsymbol{\theta}^{(k)})$ to the MLE $\hat{\boldsymbol{\theta}}$ regardless of the initial vector $\boldsymbol{\theta}^{(0)}$) was demonstrated by [4] and the almost sure convergence (i.e. the strong convergence in the sense of random variables) of the EMV $\hat{\boldsymbol{\theta}}$ to the true unknown value $\boldsymbol{\theta}$ of the parameter was obtained by [3].

For Model 2, the exact analytical expression of MLE has been obtained and its almost sure convergence has been demonstrated [2] (see Lemma 2.2 below).

Lemma 2.2 (MLE in Model 2 [2]). *Let $\boldsymbol{X} = (X_{11}, \ldots, X_{1r}, X_{21}, \ldots, X_{2r})$ be a random vector with multinomial distribution defined by* (2.2) *and $\boldsymbol{\theta} = (\alpha, \beta_1, \ldots, \beta_r)$. The MLE $\tilde{\boldsymbol{\theta}} = (\tilde{\alpha}, \tilde{\beta}_1, \ldots, \tilde{\beta}_r)$ of $\boldsymbol{\theta}$ is defined by*

$$\begin{aligned} \tilde{\alpha} &= \frac{n\sum_{k=1}^{r} X_{2k}}{\Big(\sum_{k=1}^{r} X_{1k}\Big)\Big(\sum_{k=1}^{r} z_k(X_{1k} + X_{2k})\Big)} \\ \tilde{\beta}_j &= \frac{X_{1j} + X_{2j}}{n}, \quad j = 1, \ldots, r. \end{aligned} \tag{2.6}$$

2.3. Computation of the standard errors

Standard errors are very important in Statistics. Indeed, the MLE of α and β_j $(j = 1, \ldots, r)$ are random variables and therefore each has a standard deviation called standard error. These standard errors are given for Model 1 and Model 2 respectively by the following lemmas.

Lemma 2.3 ([14]). *Let $\overline{z} = \sum_{i=1}^{r} z_i\beta_i$, $\gamma_n = n/(1 + \alpha\overline{z})$, $\tau = \gamma_n^2\overline{z}/(n\alpha)$ and*

$$\boldsymbol{\Gamma} = \begin{bmatrix} \tau & \frac{\gamma_n^2 z_1}{n} & \cdots & \cdots & \frac{\gamma_n^2 z_r}{n} & 0 \\ \frac{\gamma_n^2 z_1}{n} & \frac{n\gamma_n\omega_1 - \gamma_n^2\alpha^2 z_1^2}{n} & -\frac{\gamma_n^2\alpha^2 z_1 z_2}{n} & \cdots & -\frac{\gamma_n^2\alpha^2 z_1 z_r}{n} & 1 \\ \frac{\gamma_n^2 z_2}{n} & -\frac{\gamma_n^2\alpha^2 z_2 z_1}{n} & \frac{n\gamma_n\omega_2 - \gamma_n^2\alpha^2 z_2^2}{n} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -\frac{\gamma_n^2\alpha^2 z_{r-1} z_r}{n} & \vdots \\ \frac{\gamma_n^2 z_r}{n} & -\frac{\gamma_n^2\alpha^2 z_r z_1}{n} & \cdots & -\frac{\gamma_n^2\alpha^2 z_r z_{r-1}}{n} & \frac{n\gamma_n\omega_r - \gamma_n^2\alpha^2 z_r^2}{n} & 1 \\ 0 & 1 & \cdots & \cdots & 1 & 0 \end{bmatrix} \tag{2.7}$$

be a non singular matrix of order $(r + 2) \times (r + 2)$ where for all $j = 1, \ldots, r$, $\omega_j = (1 + \alpha z_j)/\beta_j$. Let $\boldsymbol{W} = (W_{i,j})_{1 \leqslant i,j \leqslant r+1}$ be the matrix of order $(r + 1) \times (r + 1)$ composed of the first $r + 1$ rows and $r + 1$ columns of $\boldsymbol{\Gamma}^{-1}$. The approximate standard errors of $\hat{\alpha}$ and $\hat{\beta}_j$ $(j = 1, \ldots, r)$ in Model 1 are the square roots of the diagonal elements of $\boldsymbol{W}$:

$$\sigma_1(\hat{\alpha}) = \sqrt{W_{1,1}} \tag{2.8}$$

$$\sigma_1(\hat{\beta}_j) = \sqrt{W_{j+1,j+1}}, \quad j = 1, \ldots, r. \tag{2.9}$$

Lemma 2.4 ([2]). *Let $\overline{z^2} = \sum_{i=1}^{r} z_i^2\beta_i$ and $\gamma = 1/(1 + \alpha\overline{z})$. The approximate standard error of $\hat{\alpha}$ and the exact standard error of $\hat{\beta}_j$ $(j = 1, \ldots, r)$ in Model 2 are*

$$\sigma_2(\tilde{\alpha}) = \sqrt{\frac{\alpha}{n\gamma^2\overline{z}} + \frac{\alpha^2\overline{z^2}}{n\overline{z}^2} - \frac{\alpha^2}{n}} \tag{2.10}$$

and

$$\sigma_2(\tilde{\beta}_j) = \sqrt{\frac{\beta_j(1 - \beta_j)}{n}}, \tag{2.11}$$

respectively.

Remark 2.5. *In practice, the true values α and β_j $(j = 1, \ldots, r)$ are unknown and replaced by their respective estimates.*

3. Main theoretical results

3.1. A much simpler proof of the expression of the MLE in Model 1

It is proven (see [10] or [13]) that the MLE of the parameter vector of Model 1 satisfies the following system of non-linear equations:

$$\begin{cases} \displaystyle\sum_{j=1}^{r}\left(x_{2j}-\frac{x_{\bullet j}\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}{1+\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}\right)=0 \\ \displaystyle x_{\bullet j}-\frac{n\hat{\beta}_j(1+\hat{\alpha}z_j)}{1+\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}=0, \quad j=1,\ldots,r. \end{cases} \tag{3.1}$$

N'Guessan and Truffier [13] have transformed the second row of the system (3.1) into a system of r linear equations of unknowns $\beta_1,\ldots,\beta_r$ whose matrix depends on the parameter α. They then proved that this matrix is invertible and then inverted it analytically using the Schur complement. This enabled them to obtain the following theorem:

Theorem 3.1 ([13]). *The MLE $\hat{\boldsymbol{\theta}}=(\hat{\alpha},\hat{\beta}_1,\ldots,\hat{\beta}_r)$ of $\boldsymbol{\theta}$ in Model 1 is given by:*

$$\begin{cases} \displaystyle\hat{\alpha}=\frac{x_{2\bullet}}{x_{1\bullet}\left(\sum_{j=1}^{r}z_j\hat{\beta}_j\right)} \\ \displaystyle\hat{\beta}_j=\frac{1}{1-\frac{1}{n}\sum_{m=1}^{r}\frac{x_{\bullet m}\hat{\alpha}z_m}{1+\hat{\alpha}z_m}}\times\frac{x_{\bullet j}}{n(1+\hat{\alpha}z_j)}, \quad j=1,\ldots,r. \end{cases} \tag{3.2}$$

In this paper, we give a simpler proof of the expression of the MLE in Model 1 that does not require the use of Schur complement or any other technique for analytical inversion of block-defined matrices.

Proof. Since $\sum_{j=1}^{r}x_{\bullet j}=n$, the first line of the non-linear system (3.1) is equivalent to

$$x_{2\bullet}-\frac{n\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}{1+\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}=0$$

that can be rewritten as

$$x_{2\bullet}-n+\frac{n}{1+\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}=0$$

and then

$$\frac{n}{1+\hat{\alpha}\sum_{m=1}^{r}z_m\hat{\beta}_m}=x_{1\bullet} \tag{3.3}$$

The expression of $\hat{\alpha}$ as a function of $\hat{\beta}_j$ is a simple consequence of Equation (3.3). Replacing Equation (3.3) in the second row of the system (3.1), we get

$$x_{\bullet j}-\hat{\beta}_j(1+\hat{\alpha}z_j)x_{1\bullet}=0, \quad j=1,\ldots,r,$$

hence

$$\hat{\beta}_j=\frac{x_{\bullet j}}{(1+\hat{\alpha}z_j)x_{1\bullet}}, \quad j=1,\ldots,r.$$

The condition $\sum_{j=1}^{r}\hat{\beta}_j=1$ means that $\sum_{j=1}^{r}\frac{x_{\bullet j}}{(1+\hat{\alpha}z_j)}=x_{1\bullet}$, from where we have:

$$\hat{\beta}_j=\frac{1}{\displaystyle\sum_{m=1}^{r}\frac{x_{\bullet m}}{1+\hat{\alpha}z_m}}\times\frac{x_{\bullet j}}{1+\hat{\alpha}z_j}. \tag{3.4}$$

The equivalence between Equation (3.4) and the second line of (3.2) is obtained by noting that

$$\begin{aligned}
1-\frac{1}{n}\sum_{m=1}^{r}\frac{x_{\bullet m}\hat{\alpha}z_m}{1+\hat{\alpha}z_m} &= 1-\frac{1}{n}\sum_{m=1}^{r}\left(x_{\bullet m}-\frac{x_{\bullet m}}{1+\hat{\alpha}z_m}\right)\\
&= 1-\frac{1}{n}\sum_{m=1}^{r}x_{\bullet m}+\frac{1}{n}\sum_{m=1}^{r}\frac{x_{\bullet m}}{1+\hat{\alpha}z_m}\\
&= 1-\frac{n}{n}+\frac{1}{n}\sum_{m=1}^{r}\frac{x_{\bullet m}}{1+\hat{\alpha}z_m}\\
&= \frac{1}{n}\sum_{m=1}^{r}\frac{x_{\bullet m}}{1+\hat{\alpha}z_m}.
\end{aligned}$$

■

3.2. Measure of divergence between the two models

The notion of divergence makes it possible to quantify the "distance" between two probability distributions or to quantify the difficulty of discriminating between them. Among the most widely used divergences is the Kullback-Leibler (KL) divergence [6].

Let $m \in \mathbb{N}^*$, $\boldsymbol{p} = (p_1, \ldots, p_m) \in \mathbb{S}_{m-1}$, $\boldsymbol{q} = (q_1, \ldots, q_m) \in \mathbb{S}_{m-1}$ and

$$\mathbb{E}_n = \left\{\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{N}^m, \quad y_1 + \cdots + y_m = n\right\}.$$

Let P_n and Q_n be the probability functions associated with the multinomial distributions $\mathcal{M}(n, \boldsymbol{p})$ and $\mathcal{M}(n, \boldsymbol{q})$, respectively. That is, for any vector $\boldsymbol{y} = (y_1 \ldots, y_m) \in \mathbb{E}_n$, we have:

$$P_n(\boldsymbol{y}) = \frac{n!}{\prod_{i=1}^m y_i!}\prod_{i=1}^m p_i^{y_i} \quad \text{and} \quad Q_n(\boldsymbol{y}) = \frac{n!}{\prod_{i=1}^m y_i!}\prod_{i=1}^m q_i^{y_i}. \tag{3.5}$$

The Kullback-Leibler (KL) divergence between $\mathcal{M}(n, \boldsymbol{p})$ and $\mathcal{M}(n, \boldsymbol{q})$ is defined by

$$D_{\mathrm{KL}}\left(\mathcal{M}(n, \boldsymbol{p}) \| \mathcal{M}(n, \boldsymbol{q})\right) = \sum_{\boldsymbol{y} \in \mathbb{E}_n} P_n(\boldsymbol{y}) \log\left(\frac{P_n(\boldsymbol{y})}{Q_n(\boldsymbol{y})}\right) \tag{3.6}$$

and represents the mean information for discriminating for the distribution $\mathcal{M}(n, \boldsymbol{p})$ against $\mathcal{M}(n, \boldsymbol{q})$ when the true distribution is supposed to be $\mathcal{M}(n, \boldsymbol{p})$. The KL divergence is non-negative (greater than or equal to 0) and equals zero if and only if the two distributions are the same ($\boldsymbol{p} = \boldsymbol{q}$) but it is not symmetric [6].

We have the following result.

Theorem 3.2. *Let $m \in \mathbb{N}^*$, $\boldsymbol{p} = (p_1, \ldots, p_m) \in \mathbb{S}_{m-1}$ and $\boldsymbol{q} = (q_1, \ldots, q_m) \in \mathbb{S}_{m-1}$. The Kullback-Leibler (KL) divergence between $\mathcal{M}(n, \boldsymbol{p})$ and $\mathcal{M}(n, \boldsymbol{q})$ is*

$$D_{\mathrm{KL}}\left(\mathcal{M}(n, \boldsymbol{p}) \| \mathcal{M}(n, \boldsymbol{q})\right) = n\sum_{i=1}^m p_i \log\left(\frac{p_i}{q_i}\right). \tag{3.7}$$

Proof. We make a proof by induction. To save space, $D_{\mathrm{KL}}\left(\mathcal{M}(n, \boldsymbol{p}) \| \mathcal{M}(n, \boldsymbol{q})\right)$ will simply be denoted $d_n(\boldsymbol{p} \| \boldsymbol{q})$.

- Let $n = 1$. The elements of the set

$$\mathbb{E}_1 = \left\{\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{N}^m, \quad y_1 + \cdots + y_m = 1\right\}$$

are the m vectors $\boldsymbol{e}_{1,1} = (1, 0, 0, \ldots, 0)$, $\boldsymbol{e}_{1,2} = (0, 1, 0, \ldots, 0)$, …, $\boldsymbol{e}_{1,m} = (0, 0, \ldots, 0, 1)$ such that, for all $i = 1, \ldots, m$, $P_1(\boldsymbol{e}_{1,i}) = p_i$ and $Q_1(\boldsymbol{e}_{1,i}) = q_i$. Thus, according to the formula (3.6),

$$d_1(\boldsymbol{p}\|\boldsymbol{q}) = \sum_{\boldsymbol{e}_{1,i}\in\mathbb{E}_1} P_1(\boldsymbol{e}_{1,i}) \log\left(\frac{P_1(\boldsymbol{e}_{1,i})}{Q_1(\boldsymbol{e}_{1,i})}\right) = \sum_{i=1}^{m} p_i \log\left(\frac{p_i}{q_i}\right) \tag{3.8}$$

and so, the formula (3.7) is true for $n = 1$.

- Let $n \in \mathbb{N}^*$. Suppose Equation (3.7) is true for n and let us show that it is true for $n + 1$. From the classical results of discrete mathematics, we know that the respective cardinals of $\mathbb{E}_n$ and $\mathbb{E}_{n+1}$ are $c_n = C^n_{m+n-1} = \frac{(m+n-1)!}{n!\,(m-1)!}$ and $c_{n+1} = C^{n+1}_{m+n} = \frac{(m+n)!}{(n+1)!\,(m-1)!}$. Note by $e_{n,1}$, …, e_{n,c_n} and $e_{n+1,1}$, …, $e_{n+1,c_{n+1}}$ the elements of $\mathbb{E}_n$ and $\mathbb{E}_{n+1}$, respectively. It is known that the distribution $\mathcal{M}(n + 1, \boldsymbol{p})$ is obtained as the probability distribution of the sum of two independent random variables of respective distributions $\mathcal{M}(n, \boldsymbol{p})$ and $\mathcal{M}(1, \boldsymbol{p})$. On the other hand, KL divergence is additive for independent random variables [6]. So, noting $P_{n,1}$ the joint probability function associated with the distributions $\mathcal{M}(n, \boldsymbol{p})$ and $\mathcal{M}(1, \boldsymbol{p})$ and $Q_{n,1}$ the joint probability function associated with the distributions $\mathcal{M}(n, \boldsymbol{q})$ and $\mathcal{M}(1, \boldsymbol{q})$, we have:

$$d_{n+1}(\boldsymbol{p}\|\boldsymbol{q}) = \sum_{i=1}^{c_n}\sum_{j=1}^{m} P_{n,1}(\boldsymbol{e}_{n,i}, \boldsymbol{e}_{1,j}) \log\left(\frac{P_{n,1}(\boldsymbol{e}_{n,i}, \boldsymbol{e}_{1,j})}{Q_{n,1}(\boldsymbol{e}_{n,i}, \boldsymbol{e}_{1,j})}\right)$$

and then, by independence,

$$\begin{aligned} d_{n+1}(\boldsymbol{p}\|\boldsymbol{q}) &= \sum_{i=1}^{c_n}\sum_{j=1}^{m} P_n(\boldsymbol{e}_{n,i})P_1(\boldsymbol{e}_{1,j}) \log\left(\frac{P_n(\boldsymbol{e}_{n,i})P_1(\boldsymbol{e}_{1,j})}{Q_n(\boldsymbol{e}_{n,i})Q_1(\boldsymbol{e}_{1,j})}\right) \\ &= \sum_{i=1}^{c_n}\sum_{j=1}^{m} P_n(\boldsymbol{e}_{n,i})\, p_j \log\left(\frac{P_n(\boldsymbol{e}_{n,i})\, p_j}{Q_n(\boldsymbol{e}_{n,i})\, q_j}\right). \end{aligned}$$

So, we can write

$$\begin{aligned} d_{n+1}(\boldsymbol{p}\|\boldsymbol{q}) &= \sum_{i=1}^{c_n}\sum_{j=1}^{m} P_n(\boldsymbol{e}_{n,i})\, p_j \log\left(\frac{P_n(\boldsymbol{e}_{n,i})}{Q_n(\boldsymbol{e}_{n,i})}\right) + \sum_{i=1}^{c_n}\sum_{j=1}^{m} P_n(\boldsymbol{e}_{n,i})\, p_j \log\left(\frac{p_j}{q_j}\right) \\ &= \sum_{j=1}^{m} p_j \left\{\sum_{i=1}^{c_n} P_n(\boldsymbol{e}_{n,i}) \log\left(\frac{P_n(\boldsymbol{e}_{n,i})}{Q_n(\boldsymbol{e}_{n,i})}\right)\right\} + \sum_{i=1}^{c_n} P_n(\boldsymbol{e}_{n,i}) \left\{\sum_{j=1}^{m} p_j \log\left(\frac{p_j}{q_j}\right)\right\} \\ &= \sum_{j=1}^{m} p_j\, d_n(\boldsymbol{p}\|\boldsymbol{q}) + \sum_{i=1}^{c_n} P_n(\boldsymbol{e}_{n,i})\, d_1(\boldsymbol{p}\|\boldsymbol{q}) \\ &= d_n(\boldsymbol{p}\|\boldsymbol{q}) + d_1(\boldsymbol{p}\|\boldsymbol{q}) \end{aligned}$$

because $\sum_{j=1}^{m} p_j = 1$ and $\sum_{i=1}^{c_n} P_n(\boldsymbol{e}_{n,i}) = 1$ (the total sum of a probability function is equal to 1). Combining Equation (3.8) and the assumption that Equation (3.7) is true for n, we obtain:

$$d_{n+1}(\boldsymbol{p}\|\boldsymbol{q}) = (n + 1)\sum_{i=1}^{m} p_i \log\left(\frac{p_i}{q_i}\right),$$

which completes the proof.

■

We have the following three corollaries.

Corollary 3.3. *Let $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_r)$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\alpha}, \tilde{\beta}_1, \ldots, \tilde{\beta}_r)$ be the MLE of $\boldsymbol{\theta}$ in Model 1 and Model 2 respectively. The KL divergences between Model 1 and Model 2 are respectively given by:*

$$D_{\mathrm{KL}}(1\|2) = n\sum_{j=1}^{r}\left\{\pi_{1j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})\log\left(\frac{\pi_{1j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})}{\pi_{1j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})}\right) + \pi_{2j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})\log\left(\frac{\pi_{2j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})}{\pi_{2j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})}\right)\right\}$$
$$D_{\mathrm{KL}}(2\|1) = n\sum_{j=1}^{r}\left\{\pi_{1j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})\log\left(\frac{\pi_{1j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})}{\pi_{1j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})}\right) + \pi_{2j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})\log\left(\frac{\pi_{2j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})}{\pi_{2j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})}\right)\right\}.$$

Proof. This is a direct application of the formula (3.7). For example, to calculate $D_{\mathrm{KL}}(1\|2)$, we apply the formula (3.7) to $\boldsymbol{p} = (\pi_{11}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}), \ldots, \pi_{1r}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}), \pi_{21}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}), \ldots, \pi_{2r}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}))$ and $\boldsymbol{q} = (\pi_{11}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}), \ldots, \pi_{1r}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}), \pi_{21}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}), \ldots, \pi_{2r}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}))$ where the $\pi_{ij}^{(1)}$'s and $\pi_{ij}^{(2)}$'s, $i = 1, 2$, $j = 1, \ldots, r$, are defined by Equations (2.1) and (2.2). ■

Corollary 3.4. *$D_{\mathrm{KL}}(1\|2) \geqslant 0$, $D_{\mathrm{KL}}(2\|1) \geqslant 0$ with equality if and only if the z_j's are all equal i.e. $z_1 = \cdots = z_r$.*

Proof. The non-negativity of D_{KL} is given by Theorem 3.1 of the second chapter of [6]. According to the same theorem, the KL divergence is zero if and only if the models are the same. In our context, this means that $D_{\mathrm{KL}}(1\|2) = 0$ and $D_{\mathrm{KL}}(2\|1) = 0$ if and only if

$$\pi_{1j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}) = \pi_{1j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}) \quad \text{and} \quad \pi_{2j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z}) = \pi_{2j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z}), \quad j = 1, \ldots, r.$$

We have:

$$\begin{cases} \dfrac{\hat{\beta}_j}{1 + \hat{\alpha}\sum_{k=1}^{r} z_k\hat{\beta}_k} = \dfrac{\tilde{\beta}_j}{1 + \tilde{\alpha}\sum_{k=1}^{r} z_k\tilde{\beta}_k}, & j = 1, \ldots, r \\ \dfrac{\hat{\alpha}\hat{\beta}_j z_j}{1 + \hat{\alpha}\sum_{k=1}^{r} z_k\hat{\beta}_k} = \dfrac{\tilde{\alpha}\tilde{\beta}_j\sum_{k=1}^{r} z_k\tilde{\beta}_k}{1 + \tilde{\alpha}\sum_{k=1}^{r} z_k\tilde{\beta}_k}, & j = 1, \ldots, r. \end{cases} \tag{3.9}$$

Dividing the second line by the first, we have:

$$\hat{\alpha} z_j = \tilde{\alpha}\sum_{k=1}^{r} z_k\tilde{\beta}_k, \quad j = 1, \ldots, r.$$

In this last equality, the second member does not depend on the index j so we deduce that $z_1 = \cdots = z_r$. ■

Although interesting, the divergences $D_{\mathrm{KL}}(1\|2)$ and $D_{\mathrm{KL}}(2\|1)$ do not really allow to choose the model that fits better to the observed data. It is known that, for a vector $\boldsymbol{x} = (x_{11}, \ldots, x_{1r}, x_{21}, \ldots, x_{2r})$ such that $\sum_{i=1}^{2}\sum_{j=1}^{r} x_{ij} = n$, the observed distribution is

$$\boldsymbol{\pi}^* = \left(\frac{x_{11}}{n}, \ldots, \frac{x_{1r}}{n}, \frac{x_{21}}{n}, \ldots, \frac{x_{2r}}{n}\right) \tag{3.10}$$

so we found it more interesting to compare the observed distribution and the distributions estimated by the two models using the KL divergence. By applying Equation (3.7) to $\boldsymbol{p} = \boldsymbol{\pi}^*$ and $\boldsymbol{q} = \boldsymbol{\pi}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})$ and $\boldsymbol{q} = \boldsymbol{\pi}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})$ respectively, we have the following result.

Corollary 3.5. *The KL divergences from the observed distribution to the distributions estimated by the two models are respectively given by:*

$$D_{\mathrm{KL}}(*\|1) = n\sum_{j=1}^{r}\left\{\frac{x_{1j}}{n}\log\left(\frac{x_{1j}}{n\cdot\pi_{1j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})}\right)+\frac{x_{2j}}{n}\log\left(\frac{x_{2j}}{n\cdot\pi_{2j}^{(1)}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})}\right)\right\} \tag{3.11}$$

$$D_{\mathrm{KL}}(*\|2) = n\sum_{j=1}^{r}\left\{\frac{x_{1j}}{n}\log\left(\frac{x_{1j}}{n\cdot\pi_{1j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})}\right)+\frac{x_{2j}}{n}\log\left(\frac{x_{2j}}{n\cdot\pi_{2j}^{(2)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{z})}\right)\right\}. \tag{3.12}$$

The best model will then be the one with a smaller divergence.

4. Empirical comparison on real data

In this section, we compare the fit of the two models on real data. For each dataset, we estimate the parameters of both models using Algorithm (2.5) and Formula (2.6) respectively. Afterwards, we calculate the KL divergences $D_{\mathrm{KL}}(*\|1)$ and $D_{\mathrm{KL}}(*\|2)$ and other indicators among the most used for the comparison of two models which are the Akaike Information Criterion (AIC), the Corrected AIC (AICc) and the Bayesian Information Criterion (BIC) respectively defined by:

$$\mathrm{AIC} = 2k - 2\log L, \tag{4.1}$$

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1} \tag{4.2}$$

and

$$\mathrm{BIC} = -2\log L + k\log n, \tag{4.3}$$

where k is the number of parameters of the model i.e. $k = r+1$. The best model is the one with the smallest values for all indicators.

4.1. Evaluation of an unspecified road safety measure

These data (see Table 1) are taken from [1]. Unspecified concrete measures were implemented in 2004 at an experimental site in Accra (Ghana) to improve safety. There are three crash types: Fatal, Hospitalised and Injured.

Table 1: Dataset 1: Before and after crashes data from an experimental site in Ghana

	Before period (3 years)			After period (3 years)		
	Fatal	Hospitalised	Injured	Fatal	Hospitalised	Injured
Experimental site	8	23	23	3	6	16
Control site	33	58	69	27	36	62

Dividing for each type of accident, the number of accidents after by the number of accidents before in the control area, one can obtain the control coefficients in Table 2.

Table 2: Control coefficients related to Dataset 1

z_1 (Fatal)	z_2 (Hospitalised)	z_3 (Injured)
0.8182	0.6207	0.8986

The results of the comparison of the two models are given by Table 3.

According to the results in Table 3, Model 1 is the best because it has smaller values for AIC, AICc, BIC and D_{KL}. According to this model, we have $\hat{\alpha} = 0.5946$ and we can estimate at $(1-0.5946) = 0.4054 = 40.54\%$ the reduction in the number of accidents after application of the measure.

Table 3: Models comparison results for Dataset 1 (Standard errors are in parentheses).

	Model 1	Model 2
$\hat{\alpha}$	0.5946 (0.1443)	0.5895 (0.1430)
$\hat{\beta}_1$	0.1370 (0.0386)	0.1392 (0.0390)
$\hat{\beta}_2$	0.3923 (0.0588)	0.3671 (0.0542)
$\hat{\beta}_3$	0.4707 (0.0560)	0.4937 (0.0562)
AIC	261.2306	263.1768
AICc	261.7712	263.7173
BIC	270.7084	272.6546
D_{KL}	$D_{\mathrm{KL}}(*\|1) = 0.7050$	$D_{\mathrm{KL}}(*\|2) = 1.6781$

4.2. Evaluation of the modification of ground markings

The data (see Table 4) come from [13]. A road modification was carried out in 1999 on national road 17 (RN17) in France. It consisted of the modification of ground markings of this three-lane two-way road so that it is impossible to overtake simultaneously in both directions. Accidents are classified into three categories: Fatal, Serious and Minor.

Table 4: Dataset 2

	Before period (4 years)			After period (4 years)		
	Fatal	Serious	Minor	Fatal	Serious	Minor
Treated site	4	4	16	1	1	7
Control site	27	64	182	14	27	102

Table 5: Control coefficients related to Dataset 2.

z_1	z_2	z_3
0.5190	0.4220	0.5600

The results of the comparison of the two models are given by Table 6.

Table 6: Models comparison results for Dataset 2 (Standard errors are in parentheses).

	Model 1	Model 2
$\hat{\alpha}$	0.7054 (0.2760)	0.7037 (0.2753)
$\hat{\beta}_1$	0.1525 (0.0632)	0.1515 (0.0624)
$\hat{\beta}_2$	0.1605 (0.0664)	0.1515 (0.0624)
$\hat{\beta}_3$	0.6870 (0.0854)	0.6970 (0.0800)
AIC	100.8238	101.0209
AICc	102.2524	102.4495
BIC	106.8098	107.0069
D_{KL}	$D_{\mathrm{KL}}(*\|1) = 0.1003$	$D_{\mathrm{KL}}(*\|2) = 0.1988$

According to the results of Table 6, Model 1 is the best because it has smaller values for the AIC, AICc, BIC and D_{KL}. According to this model, we have $\hat{\alpha} = 0.7054$ and we can estimate at $(1-0.7054) = 0.2946 = 29.46\%$ the reduction in the number of accidents after application of the measure.

4.3. Evaluation of the presence of display panels on roadsides

The data (see Table 7) come from [14]. The experimental site is the Turcot Interchange (Canada). The

modification consisted of the installation of roadside billboards in July 1995. The control site consists of two interchanges. There are three categories of accidents: Fatal or severe, Minor and Property damages only (PDO).

Table 7: Dataset 3

	Before period (1 year)			Period after (1 year)		
	Fatal or Severe	Minor	PDO	Fatal or Severe	Minor	PDO
Experimental site	4	20	133	3	29	143
Control site	2	26	154	9	37	239

The control coefficients are given by Table 8.

Table 8: Control coefficients related to Dataset 3.

z_1	z_2	z_3
4.5	1.423	1.552

The results of the comparison of the two models are given by Table 9.

Table 9: Models comparison results for Dataset 3 (Standard errors sont entre parenthèses).

	Model 1	Model 2
$\hat{\alpha}$	0.7130 (0.0786)	0.6988 (0.0775)
$\hat{\beta}_1$	0.0106 (0.0040)	0.0211 (0.0079)
$\hat{\beta}_2$	0.1549 (0.0207)	0.1476 (0.0195)
$\hat{\beta}_3$	0.8345 (0.0283)	0.8313 (0.0206)
AIC	814.7443	810.7806
AICc	814.8666	810.9029
BIC	829.9649	826.0011
D_{KL}	$D_{\mathrm{KL}}(*\|1) = 2.5778$	$D_{\mathrm{KL}}(*\|2) = 0.5959$

According to the results of Table 9, Model 2 is the best because it has smaller values for AIC, AICc, BIC and D_{KL}. According to this model, we have $\hat{\alpha} = 0.6988$ and $(1 - 0.6988) = 0.3012 = 30.12\%$ so we can estimate that the measure led to a decrease of 30.12% in the number of accidents.

4.4. Evaluation of the increase in speed limit on Arizona's rural interstate

Data are extracted from [16]. There are three severity levels for accidents: Fatal, Injury and property-damage-only (PDO). The speed limit on Arizona's rural interstate was raised to 65 mph in 1987. The treatment site (rural interstate) represents the portions of the Arizona interstate system that had the speed limit raised to 65 mph and the control site (urban interstate) represents the portions that have the speed limit maintained at 55 mph. In [16], the period before covers about four years and the period after covers one year. In order to respect the basic principle of model construction which requires the periods before and after to have approximately the same duration, we have chosen just one year before the measure and one year after the measure. This gives the following table:

Table 10: Dataset 4 (Arizona)

	Before period (1 year)			After period (1 year)		
	PDO	Injury	Fatal	PDO	Injury	Fatal
Treatment site	1669	1047	97	1969	1322	117
Control site	2105	803	13	2217	737	15

Table 11: Control coefficients related to Dataset 4.

z_1 (PDO)	z_2 (Injury)	z_3 (Fatal)
1.0532	0.9178	1.1538

The control coefficients related to Dataset 4 are given by Table 11.
The results of the comparison of the two models are given by Table 12.

Table 12: Models comparison results for Dataset 4 (Standard errors are in parentheses).

	Model 1	Model 2
$\hat{\alpha}$	1.2087 (0.0308)	1.2054 (0.0307)
$\hat{\beta}_1$	0.5690 (0.0069)	0.5848 (0.0062)
$\hat{\beta}_2$	0.3993 (0.0068)	0.3808 (0.0062)
$\hat{\beta}_3$	0.0318 (0.0021)	0.0344 (0.0023)
AIC	18509.8110	18495.2896
AICc	18509.8174	18495.2960
BIC	18536.7537	18522.2323
D_{KL}	$D_{\mathrm{KL}}(*\|1) = 8.0800$	$D_{\mathrm{KL}}(*\|2) = 0.8193$

According to the results of Table 12, Model 2 is the best because it has smaller values for the AIC, AICc, BIC, and D_{KL}. According to this model, we have $\hat{\alpha} = 1.2054$ and $(1 - 1.2054) = -0.2054 = -20.54\%$ so we can estimate that the measure has led to an increase of 20.54% in the number of accidents.

5. Conclusion

In this paper, we compared two discrete statistical models for the evaluation of a road safety measure applied on an experimental site (treatment site) where the accidents are classified by severity in r categories. In order to take into account the effects external to the measure and which could influence the number of accidents, the treatment site is associated with a control site where the measure was not applied. The two models are multinomial models coming respectively from [10] (Model 1) and [11] (Model 2).

There certainly exist results on the maximum likelihood estimator (MLE) of the parameter vector for each of the models. But for Model 1 (considered in this work as the most complex given the form of the MLE), previous works have used the notion of Schur complement for the exact analytical inversion of a matrix involved in the resolution of likelihood equations. In our work, we have given a much simpler proof of the expression of the MLE for Model 1 without using neither Schur complement nor any other technique of exact analytical inversion of block matrices. We then obtained theoretical results on the measure of divergence between the two models. The results obtained on real data suggest that both models are competitive and that none of them is systematically better than the other.

References

[1] AFRICAN DEVELOPMENT BANK, *Road safety manuals for Africa – Existing Roads: Reactive Approaches*, (2014).

[2] I. C. GERALDO, On the maximum likelihood estimator for a discrete multivariate crash frequencies model, *Afr. Stat.*, **15(2)**(2020), 2335–2348.

[3] I. C. GERALDO, A. N'GUESSAN AND K. E. GNEYOU, A note on the strong consistency of a constrained maximum likelihood estimator used in crash data modeling, *C. R. Math. Acad. Sci. Paris*, **353(12)**(2015), 1147–1152.

[4] I. C. GERALDO, A. N'GUESSAN AND K. E. GNEYOU, Global convergence and ascent property of a cyclic algorithm used for statistical analysis of crash data, *Afr. Stat.*, **13(2)**(2018), 1631–1643.

[5] E. HAUER, Cause, effect and regression in road safety: A case study, *Accident Analysis and Prevention*, **42(4)**(2010), 1128–1135.

[6] S. KULLBACK, *Information Theory and Statistics*, Peter Smith, Gloucester (Massachusetts), (1978).

[7] D. LORD AND F. MANNERING, The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives, *Transportation Research Part A*, **44(5)**(2010), 291–305.

[8] F. L. MANNERING AND C. R. BHAT, Analytic methods in accident research: Methodological frontier and future directions, *Analytic Methods in Accident Research*, **1**(2014), 1–22.

[9] A. N'GUESSAN, Analytical existence of solutions to a system of nonlinear equations, with application, *J. Comput. Appl. Math.*, **234(1)**(2010), 297–304.

[10] A. N'GUESSAN, A. ESSAI AND C. LANGRAND, Estimation multidimensionnelle des contrôles et de l'effet moyen d'une mesure de sécurité routière, *Revue de statistique appliquée*, **49(2)**(2001), 85–102.

[11] A. N'GUESSAN, A. ESSAI AND M. N'ZI, An estimation method of the average effect and the different accident risks when modelling a road safety measure: A simulation study, *Comput. Statist. Data Anal.*, **51(2)**(2006), 1260–1277.

[12] A. N'GUESSAN AND I.C. GERALDO, A cyclic algorithm for maximum likelihood estimation using Schur complement, *Numer. Linear Algebra Appl.*, **22(6)**(2015), 1161–1179.

[13] A. N'GUESSAN AND M. TRUFFIER, Impact d'un aménagement de sécurité routière sur la gravité des accidents de la route, *J. SFdS*, **49**(2008), 23–41.

[14] A. N'GUESSAN, M. TRUFFIER, C. LANGRAND AND F. BELLAVANCE, Application d'une méthode statistique pour évaluer l'impact de certaines modifications des conditions de la route : étude de cas, *Pub. IRMA Lille*, **66(III)**(2006), 1–20.

[15] K.W. OGDEN, The effects of paved shoulders on accidents on rural highways, *Accident Analysis & Prevention*, **29(3)**(1997), 353–362.

[16] J. UPCHURCH, Arizona's experience with the 65-mph speed limit, *Transportation Research Record*, **1244**(1989), 1–6.

[17] WORLD HEALTH ORGANIZATION, *Global status report on road safety 2018*. World Health Organization, Geneva, (2018).