Afrika Statistika

Afrika Statistika
Vol. 15 (2), 2020, pages 2335 - 2348.
DOI: http://dx.doi.org/10.16929/as/2020.2325.161

ISSN 2316-090X

On the maximum likelihood estimator for a discrete multivariate crash frequencies model

Issa Cherif Geraldo

Department of Mathematics, Faculty of Sciences, Université de Lomé, 1 B.P. 1515 Lomé 1, Togo.

Received on June 18, 2020. Accepted on October 18, 2020

Abstract. In this paper, we study the maximum likelihood estimator (MLE) of the parameter vector of a discrete multivariate crash frequencies model used in the statistical analysis of the effectiveness of a road safety measure. We derive the closed-form expression of the MLE afterwards we prove its strong consistency and we obtain the exact variance of the components of the MLE except one component whose variance is approximated via the delta method.

Résumé. Dans cet article, nous étudions l'estimateur du maximum de vraisemblance (EMV) du vecteur de paramètres d'un modèle discret multivarié utilisé dans l'analyse statistique de l'efficacité d'une mesure de sécurité routière. Nous obtenons l'expression analytique exacte de l'EMV après quoi nous prouvons sa forte consistance et nous obtenons la variance exacte des composantes de l'EMV, sauf pour une composante dont la variance est approximée par la méthode delta.

Key words: Maximum likelihood; parameter estimation; strong consistency; almost sure convergence; variance estimation
AMS 2010 Mathematics Subject Classification Objects : 62F10; 62F12; 62H12.

Issa Cherif Geraldo: cherifgera@gmail.com

The author.

Issa Cherif Geraldo, Ph.D, is a Lecturer-Researcher at Université de Lomé, Lomé, Togo (Department of Mathematics).

1. Introduction and motivation

Let n and r be two positive integers and $\mathbf{X} = (X_{11}, \dots, X_{1r}, X_{21}, \dots, X_{2r}) \in \mathbb{R}^{2r}$ be a discrete random vector with multinomial distribution $\mathcal{M}(n, \boldsymbol{\pi})$ where $\boldsymbol{\pi} = (\pi_{11}, \dots, \pi_{1r}, \pi_{21}, \dots, \pi_{2r})^{\mathrm{T}}$ is a vector of cell probabilities such that the sum of its components equals one. Such a modelling is particularly used in the field of road safety to estimate how much crash frequencies have been affected by a road safety measure (design change or intervention) on a given treatment site (N'Guessan and Langrand, 1993; N'Guessan et al., 2001, 2006a,b; N'Guessan and Truffier, 2008). In that case, r represents the total number of accident severity levels (for example, if accidents are categorized by severity level as property damage only, minor injury, severe injury and fatal accident, then $r = 4$), n is the total number of crashes in both periods and for all $j = 1, \dots, r$, X_{ij} (resp. π_{ij}) represents the number (resp. the risk) of crashes of severity level j on the site in time period i ($i = 1$ for the period before the application of the road safety measure and $i = 2$ for the period after). One of the benefits of this kind of before-after studies is that they allow cause-effect interpretations (Hauer, 2010).

The estimation of the parameters π_{ij}'s is not of direct interest to researchers or decision-makers. One is more interested in answering the following question: how did the measure affect the number of accidents? To this purpose, it seems interesting to estimate the measure's efficiency index that is a positive real number denoted α and defined by N'Guessan et al. (2006a) as the ratio of the total number of accidents observed in the after period to the total number of accidents expected in the same period if the measure had no effect i.e. if the treatment site behaved like its control area. The interpretation of the mean effect α can be done by comparing α to 1 through a statistical test (see N'Guessan and Truffier (2008) for more details). For example, if $\alpha < 1$, then it could be concluded that the measure has enabled to reduce the number of accidents occurring on the treatment site. The main challenge is then to find a link function between the parameters π_{ij} and α.

In order to take into account any underlying trend in crash frequencies which may erroneously be attributed to the measure, the treatment site is paired with a control site of similar conditions (geometric characteristics, traffic flow, accident exposure, roadside conditions, etc.) and where the measure was not applied (Ogden, 1997). The accidents data of the control site over both periods of time are represented by a non-random vector $\mathbf{Z} = (z_1, \dots, z_r)^{\mathrm{T}}$ where z_j is the ratio of the number of crashes of severity level j in the "after" period to the number of crashes of the same severity level in the "before" period.

The simultaneous consideration of different severity levels and control site data introduces several secondary parameters. N'Guessan et al. (2001) proposed the link functions

$$\pi_{1j} = \frac{\beta_j}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad \pi_{2j} = \frac{\alpha\beta_j z_j}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad j=1,\dots,r. \tag{1}$$

where $\beta_1, \dots, \beta_r$ are positive additional secondary parameters such that $\sum_{i=1}^{r}\beta_i = 1$ and for all $j = 1,\dots,r$, β_j represents the probability that a crash occurring in an area similar to the treatment site has a severity level j. Later, N'Guessan et al. (2006a) proposed the link functions

$$\pi_{1j} = \frac{\beta_j}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad \pi_{2j} = \frac{\alpha\beta_j \sum_{k=1}^{r} z_k\beta_k}{1+\alpha\sum_{k=1}^{r} z_k\beta_k}, \quad j=1,\dots,r. \tag{2}$$

The main difference between link functions (1) and (2) is the definition of π_{2j}, motivated in model (2) by the fact that the mean value $\sum_{k=1}^{r} z_k\beta_k$ is considered as more stable and less sensitive to errors than the control coefficients $z_1, \dots, z_r$ taken individually. The models thus defined have a parameter vector $\boldsymbol{\theta} = (\alpha, \beta_1, \dots, \beta_r)$ such that $\alpha > 0$, $\beta_i > 0$ for all $i = 1,\dots,r$ and $h(\boldsymbol{\theta}) = 0$ where h is the function from $\mathbb{R}^{r+1}$ to $\mathbb{R}$ defined by $h(\boldsymbol{\theta}) = \sum_{i=1}^{r}\beta_i - 1$.

Model (1) has been the subject of several works. N'Guessan (2010) studied the analytical existence of the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_r)$ of $\boldsymbol{\theta} = (\alpha, \beta_1, \dots, \beta_r)$ and proved that, although an explicit closed-form expression of $\hat{\boldsymbol{\theta}}$ cannot be obtained, it is possible to write $\hat{\alpha}$ as a function of the $\hat{\beta}_j$'s and vice versa. N'Guessan and Langrand (2005) have obtained the explicit asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$. A cyclic algorithm has been developed by N'Guessan and Geraldo (2015) for the numerical estimation of $\hat{\boldsymbol{\theta}}$ and the convergence of this algorithm has been proved by Geraldo et al. (2018). Geraldo et al. (2015) have demonstrated the strong consistency of the MLE, that is, $\hat{\boldsymbol{\theta}}$ converges almost surely (a.s.) to the true value $\boldsymbol{\theta} = (\alpha, \beta_1, \dots, \beta_r)$ of the vector parameter when the sample size n tends to infinity.

Although the MLE for model (2) has been shown to perform well in the numerical simulation studies of N'Guessan et al. (2006a), the exact expression of the MLE and the theoretical justification of its strong consistency have not been established yet. So the aim of this paper is to fill this gap by making a comprehensive study of the maximum likelihood estimator for the model (2).

To achieve our goal, we prove in Section 2 that the likelihood equations (obtained by setting the partial derivatives of the log-likelihood to zero) have a unique closed-form solution. Afterwards, in Section 3, we prove the strong consistency of the MLE $\hat{\boldsymbol{\theta}}$. We also derive the approximated variance of $\hat{\alpha}$ using the delta method and the exact variance of the $\hat{\beta}_j$'s in Section 4. In the last section of the paper (Section 5), we discuss the possible extension of our work to the case where the road safety measure is applied, no longer on a single site but on s ($s > 1$) different sites.

To make it easier for the reader to identify the elements involved in the matrix operations carried out in the paper, the multidimensional objects (vectors and matrices) are denoted in bold. The almost sure (a.s.) convergence is denoted by the symbol $\xrightarrow{a.s.}$. The vector $\boldsymbol{\pi}$ and its components π_{ij} will be denoted sometimes $\boldsymbol{\pi}(\boldsymbol{\theta})$ and $\pi_{ij}(\boldsymbol{\theta})$ to emphasize their dependence on the parameter vector $\boldsymbol{\theta}$.

2. Closed-form expression of the maximum likelihood estimator

Let us start with the following lemma.

Lemma 1. *Let* $\mathbf{x} = (x_{11}, \dots, x_{1r}, x_{21}, \dots, x_{2r})$ *be a vector of observed data from model* (2) *such that* $\sum_{i=1}^{2}\sum_{j=1}^{r} x_{ij} = n$. *The MLE* $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_r)$, *if it exists, is solution to the system of non-linear equations*

$$\begin{cases} \sum_{j=1}^{r}\left(x_{2j} - \frac{\alpha\bar{z}(\boldsymbol{\beta})x_{\cdot j}}{1+\alpha\bar{z}(\boldsymbol{\beta})}\right) = 0 \\ \\ x_{\cdot j} - \frac{n\beta_j(1+\alpha z_j)}{1+\alpha\bar{z}(\boldsymbol{\beta})} - \frac{x_{2\cdot}\beta_j(\bar{z}(\boldsymbol{\beta}) - z_j)}{\bar{z}(\boldsymbol{\beta})} = 0, \quad j = 1, \dots, r \end{cases} \tag{3}$$

where $x_{\cdot j} = x_{1j} + x_{2j}$, $x_{2\cdot} = \sum_{k=1}^{r} x_{2k}$ *and* $\bar{z}(\boldsymbol{\beta}) = \sum_{k=1}^{r} z_k\beta_k$.

Proof. It is inspired from (N'Guessan et al., 2006a, Appendix B). One shows that the log-likelihood is given, up to an irrelevant additive constant, by

$$\ell(\boldsymbol{\theta}) = \sum_{j=1}^{r}\left\{x_{\cdot j}\log(\beta_j) + x_{2j}\log(\alpha) - x_{\cdot j}\log\left(1+\alpha\sum_{k=1}^{r} z_k\beta_k\right) + x_{2j}\log\left(\sum_{k=1}^{r} z_k\beta_k\right)\right\}.$$

The maximization of $\ell(\boldsymbol{\theta})$ under the constraint $h(\boldsymbol{\theta}) = 0$ (where $h(\boldsymbol{\theta}) = \sum_{j=1}^{r}\beta_j - 1$) is equivalent to the maximization of $\mathcal{L}(\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}) - \lambda h(\theta)$ where λ is a Lagrange multiplier. The first line of (3) is easily obtained from $\partial\mathcal{L}/\partial\alpha = 0$. For all $j = 1, \dots, r$, we also have,

$$\frac{\partial\mathcal{L}}{\partial\beta_j} = \frac{1}{\beta_j}\left(x_{\cdot j} - \frac{n\alpha\beta_j z_j}{1+\alpha\bar{z}(\boldsymbol{\beta})} + \frac{x_{2\cdot}\beta_j z_j}{\bar{z}(\boldsymbol{\beta})} - \lambda\beta_j\right) = 0. \tag{4}$$

After multiplication by β_j and summation on the index j, we get

$$\lambda = \frac{n}{1+\alpha\bar{z}(\boldsymbol{\beta})} + x_{2\cdot}$$

and the second line of (3) is then obtained by substitution of the expression of λ in (4).

Let us now give the first main result of the paper.

Theorem 1. *Let* $\mathbf{X} = (X_{11}, \ldots, X_{1r}, X_{21}, \ldots, X_{2r})$ *be a random vector with multinomial distribution* $\mathcal{M}(n, \boldsymbol{\pi}(\boldsymbol{\theta}))$ *where* $\boldsymbol{\pi}(\boldsymbol{\theta})$ *is defined by* (2) *and* $\boldsymbol{\theta} = (\alpha, \beta_1, \ldots, \beta_r)$*. The MLE* $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_r)$ *of* $\boldsymbol{\theta}$ *is given by*

$$\hat{\alpha} = \frac{n \sum_{k=1}^{r} X_{2k}}{\left(\sum_{k=1}^{r} X_{1k}\right)\left(\sum_{k=1}^{r} z_k (X_{1k} + X_{2k})\right)} \tag{5}$$

$$\hat{\beta}_j = \frac{X_{1j} + X_{2j}}{n}, \qquad j = 1, \ldots, r. \tag{6}$$

Proof. From the first line of (3), we have the following equivalences:

$$\begin{aligned}\sum_{j=1}^{r}\left(x_{2j} - \frac{\alpha \overline{z}(\boldsymbol{\beta}) x_{\cdot j}}{1 + \alpha \overline{z}(\boldsymbol{\beta})}\right) = 0 &\iff \sum_{j=1}^{r} x_{2j} - \frac{n \alpha \overline{z}(\boldsymbol{\beta})}{1 + \alpha \overline{z}(\boldsymbol{\beta})} = 0 \\ &\iff \sum_{j=1}^{r} x_{2j} - \left(n - \frac{n}{1 + \alpha \overline{z}(\boldsymbol{\beta})}\right) = 0 \\ &\iff \frac{n}{1 + \alpha \overline{z}(\boldsymbol{\beta})} = \sum_{j=1}^{r} x_{1j}\end{aligned}$$

because $\sum_{j=1}^{r}(x_{1j} + x_{2j}) = n$. Thus

$$\overline{z}(\boldsymbol{\beta}) = \frac{\sum_{j=1}^{r} x_{2j}}{\alpha \sum_{j=1}^{r} x_{1j}} = \frac{x_{2\cdot}}{\alpha(n - x_{2\cdot})}. \tag{7}$$

After substitution of (7) in the second line of (3), we get, for all $j = 1, \ldots, r$,

$$x_{\cdot j} - \frac{n \beta_j (1 + \alpha z_j)}{1 + \dfrac{x_{2\cdot}}{n - x_{2\cdot}}} - x_{2\cdot} \beta_j + \frac{x_{2\cdot} \beta_j z_j}{\dfrac{x_{2\cdot}}{\alpha(n - x_{2\cdot})}} = 0$$

which yields

$$x_{\cdot j} - \beta_j (1 + \alpha z_j)(n - x_{2\cdot}) - x_{2\cdot} \beta_j + \alpha \beta_j z_j (n - x_{2\cdot}) = 0.$$

After simplification, we get

$$x_{\cdot j} - n \beta_j = 0$$

hence the expression of $\hat{\beta}_j$. The expression of $\hat{\alpha}$ is then easily obtained after substitution of $\hat{\beta}_j$ in (7).

Remark 1. It is easy to check that the MLE $\hat{\boldsymbol{\theta}}$ satisfies the conditions $\hat{\alpha} > 0$, $0 < \hat{\beta}_j < 1$ for $j = 1, \ldots, r$ and $\sum_{i=1}^{r} \hat{\beta}_i = 1$.

3. Strong consistency of the MLE

The strong consistency of the estimator $\hat{\boldsymbol{\theta}}$ is a very desirable property. This property guarantees that, if $\mathbf{X} = (X_{11}, \dots, X_{1r}, X_{21}, \dots, X_{2r})$ originates from the model (2) with true unknown vector parameter $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}}$ converges almost surely (a.s.) to $\boldsymbol{\theta}$ when the sample size n tends to $+\infty$. Since the MLE $\hat{\boldsymbol{\theta}}$ is available in closed-form, the study of its strong convergence can be done directly by using its closed-form expression and the properties of the underlying multinomial distribution.

We first recall the continuous mapping theorem that will be very useful in the proof of the main theorem on the strong consistency of the MLE.

Lemma 2 (Van der Vaart (1998)). *Let $\mathbf{Y}_n = (Y_{n,1}, \dots, Y_{n,k})$ and $\mathbf{Y} = (Y_1, \dots, Y_k)$ be $k-$dimensional random vectors and g be a mapping from $\mathbb{R}^k$ to $\mathbb{R}^m$ continuous at every point of a set $\mathbb{A}$ such that $\mathrm{P}(\mathbf{Y} \in \mathbb{A}) = 1$. If $\mathbf{Y}_n \xrightarrow{a.s.} \mathbf{Y}$ then $g(\mathbf{Y}_n) \xrightarrow{a.s.} g(\mathbf{Y})$.*

Let us now give the consistency theorem.

Theorem 2. *Let $\mathbf{X} = (X_{11}, \dots, X_{1r}, X_{21}, \dots, X_{2r})$ be a random vector with the multinomial distribution $\mathcal{M}(n; \boldsymbol{\pi}(\boldsymbol{\theta}))$ where $\boldsymbol{\pi}(\boldsymbol{\theta})$ is defined by (2) and $\boldsymbol{\theta} = (\alpha, \beta_1, \dots, \beta_r)$. Then, the MLE $\hat{\boldsymbol{\theta}}$ defined by Theorem 1 converges a.s. to $\boldsymbol{\theta}$ as n tends to $+\infty$.*

Proof. We know that the almost sure (a.s.) convergence of a random vector is equivalent to the a.s. convergence of each of its components. To prove that $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_r)$ converges a.s. to $\boldsymbol{\theta} = (\alpha, \beta_1, \dots, \beta_r)$, it is sufficient to prove that: (1) $\hat{\alpha} \xrightarrow{a.s.} \alpha$ and, (2) for all $j = 1, \dots, r$, $\hat{\beta}_j \xrightarrow{a.s.} \beta_j$.

(1) We can write

$$\hat{\alpha} = \frac{\displaystyle\sum_{k=1}^{r} \frac{X_{2k}}{n}}{\left(\displaystyle\sum_{k=1}^{r} \frac{X_{1k}}{n}\right) \times \displaystyle\sum_{k=1}^{r} z_k \left(\frac{X_{1k}}{n} + \frac{X_{2k}}{n}\right)} = g\left(\frac{X_{11}}{n}, \dots, \frac{X_{1r}}{n}, \frac{X_{21}}{n}, \dots, \frac{X_{2r}}{n}\right)$$

where g is the continuous function defined from $\mathbb{R}^{2r}$ to $\mathbb{R}$ by

$$g(b_1, \dots, b_r, a_1, \dots, a_r) = \frac{\sum_{k=1}^{r} a_k}{\left(\sum_{k=1}^{r} b_k\right)\left(\sum_{k=1}^{r} z_k (a_k + b_k)\right)}. \tag{8}$$

As $\mathbf{X}/n \xrightarrow{a.s.} \boldsymbol{\pi}$ when $n \to \infty$, we apply Lemma 2 and get

$$\hat{\alpha} \quad \xrightarrow{a.s.} \quad g(\pi_{11}(\boldsymbol{\theta}), \dots \pi_{1r}(\boldsymbol{\theta}), \pi_{21}(\boldsymbol{\theta}), \dots, \pi_{2r}(\boldsymbol{\theta})) = \alpha.$$

(2) For all $j = 1, \dots, r$, we have

$$\hat{\beta}_j = g_j\left(\frac{X_{11}}{n}, \dots, \frac{X_{1r}}{n}, \frac{X_{21}}{n}, \dots, \frac{X_{2r}}{n}\right)$$

where g_j is the continuous mapping defined from $\mathbb{R}^{2r}$ to $\mathbb{R}$ by

$$g_j(b_1, \dots, b_r, a_1, \dots, a_r) = b_j + a_j.$$

As $\mathbf{X}/n \xrightarrow{a.s.} \boldsymbol{\pi}$ when $n \to \infty$, we apply Lemma 2 here too and get

$$\hat{\beta}_j \xrightarrow{a.s.} g_j(\pi_{11}(\boldsymbol{\theta}), \dots, \pi_{1r}(\boldsymbol{\theta}), \pi_{21}(\boldsymbol{\theta}), \dots, \pi_{2r}(\boldsymbol{\theta})) = \pi_{1j}(\boldsymbol{\theta}) + \pi_{2j}(\boldsymbol{\theta}) = \beta_j.$$

4. Computation of the variance of the MLE's components

The estimation of the parameters of the model cannot be complete without the estimation of the variance of the estimators $\hat{\alpha}$, $\hat{\beta}_1$, ..., $\hat{\beta}_r$ which allows to have an idea on their variability.

Theorem 3. *Let $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_r)$ be the MLE defined by Theorem 1. Then, for all $j = 1, \dots, r$, the exact variance of the MLE $\hat{\beta}_j$ is*

$$\mathrm{var}(\hat{\beta}_j) = \frac{\beta_j(1 - \beta_j)}{n}. \tag{9}$$

Proof. For any $j = 1, \dots, r$, the random variable $X_{1j} + X_{2j}$ follows the binomial distribution $\mathcal{B}(n, \pi_{1j} + \pi_{2j})$. This simple property of the multinomial distribution will not be demonstrated here and we rather refer the reader to (Wasserman, 2004, pages 53-54, 235-237). We then have

$$\begin{aligned} \mathrm{var}(\hat{\beta}_j) &= \mathrm{var}\left(\frac{X_{1j} + X_{2j}}{n}\right) = \frac{1}{n^2}\mathrm{var}(X_{1j} + X_{2j}) \\ &= \frac{1}{n^2}\Big(n(\pi_{1j} + \pi_{2j})(1 - \pi_{1j} - \pi_{2j})\Big) \end{aligned}$$

where $\pi_{1j} + \pi_{2j} = \beta_j$.

Direct calculation of the variance of the estimator $\hat{\alpha}$ seems impossible because $\hat{\alpha}$ is expressed as a quotient of random variables whose exact distribution cannot be determined accurately. However, an approximate variance can be obtained using the delta method (Lo et al., 2016, Proposition 35). This method enables to approximate the variance of a function of a random variable (or a random vector) whose exact variance is known. The delta method is recalled by the following lemma.

Lemma 3 ((Lo et al., 2016, Proposition 35)). *Let $\mathbf{Y} \in \mathbb{R}^d$ be a random vector with mathematical expectation $\boldsymbol{\mu} \in \mathbb{R}^d$ and variance-covariance matrix $\boldsymbol{\Sigma}$ of order $d \times d$. For any differentiable function $g : \mathbb{R}^d \to \mathbb{R}$,*

$$\mathrm{var}(g(\mathbf{Y})) \approx (\boldsymbol{\nabla}_g(\boldsymbol{\mu}))^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\nabla}_g(\boldsymbol{\mu})) \tag{10}$$

where $\boldsymbol{\nabla}_g(\boldsymbol{\mu}) \in \mathbb{R}^d$ is the gradient of g (the vector of first partial derivatives) evaluated at the point $\boldsymbol{\mu}$.

Before giving the approximate variance of $\hat{\alpha}$, we recall the first-order and second-order moments of the multinomial distribution through the following lemma.

Lemma 4. *Let* $\mathbf{X} = (X_{11}, \ldots, X_{1r}, X_{21}, \ldots, X_{2r})$ *be a random vector with multinomial distribution* $\mathcal{M}(n; \boldsymbol{\pi})$ *where* $\boldsymbol{\pi} = (\pi_{11}, \ldots, \pi_{1r}, \pi_{21}, \ldots, \pi_{2r})^{\mathrm{T}}$. *Let* $\boldsymbol{\pi}_1 = (\pi_{11}, \ldots, \pi_{1r})^{\mathrm{T}} \in \mathbb{R}^r$ *and* $\boldsymbol{\pi}_2 = (\pi_{21}, \ldots, \pi_{2r})^{\mathrm{T}} \in \mathbb{R}^r$. *The mathematical expectation of* $\mathbf{X}/n$ *is* $\mathrm{E}(\mathbf{X}/n) = \boldsymbol{\pi}$ *and its variance-covariance matrix is*

$$\boldsymbol{\Sigma} = \frac{1}{n}\left(\boldsymbol{\Delta} - \boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}}\right)$$

where

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_1 & \mathbf{0}_{r,r} \\ \mathbf{0}_{r,r} & \boldsymbol{\Delta}_2 \end{pmatrix},$$

$\boldsymbol{\Delta}_1$ *(resp.* $\boldsymbol{\Delta}_2$*) is the diagonal matrix of order* $r \times r$ *whose diagonal elements are the components of* $\boldsymbol{\pi}_1$ *(resp.* $\boldsymbol{\pi}_2$*) and* $\mathbf{0}_{r,r}$ *is the null matrix of order* $r \times r$.

Proof. From the classical results on the multinomial distribution (see for example (Wasserman, 2004, page 53)), we know that $\mathrm{E}(\mathbf{X}) = n\boldsymbol{\pi}$ and $\mathrm{var}(\mathbf{X}) = n\left(\boldsymbol{\Delta} - \boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}}\right)$. Lemma 4 is then easily deduced from the equalities $\mathrm{E}(\mathbf{X}/n) = \mathrm{E}(\mathbf{X})/n$ and $\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{X}/n) = \mathrm{var}(\mathbf{X})/n^2$.

The following lemma gives some intermediate results that will be needed to prove the theorem on the approximate variance of $\hat{\alpha}$. For simplicity, $\bar{z}(\boldsymbol{\beta})$ is denoted $\bar{z}$.

Lemma 5. *Let* g *be the continuous function defined from* $\mathbb{R}^{2r}$ *to* $\mathbb{R}$ *by Equation (8). For any vector* $\mathbf{y} \in \mathbb{R}^r$, *let* $\|\mathbf{y}\|^2_{\boldsymbol{\Delta}_1} = \mathbf{y}^{\mathrm{T}}\boldsymbol{\Delta}_1\mathbf{y}$ *and* $\|\mathbf{y}\|^2_{\boldsymbol{\Delta}_2} = \mathbf{y}^{\mathrm{T}}\boldsymbol{\Delta}_2\mathbf{y}$. *Let* $\gamma = \dfrac{1}{1+\alpha\bar{z}}$ *and* $\overline{z^2} = \sum_{j=1}^{r}\beta_j z_j^2$.

(1)

$$\boldsymbol{\nabla}_g(\boldsymbol{\pi}) = \frac{1}{\gamma\bar{z}}\Big(\underbrace{-\alpha\gamma\mathbf{Z}^{\mathrm{T}} - \alpha\bar{z}\mathbf{1}_r^{\mathrm{T}}}_{\in\mathbb{R}^r}, \underbrace{\mathbf{1}_r^{\mathrm{T}} - \alpha\gamma\mathbf{Z}^{\mathrm{T}}}_{\in\mathbb{R}^r}\Big)^{\mathrm{T}}$$

where $\mathbf{1}_r = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^r$,

$$\begin{aligned} -\alpha\gamma\mathbf{Z}^{\mathrm{T}} - \alpha\bar{z}\mathbf{1}_r^{\mathrm{T}} &= (-\alpha\gamma z_1 - \alpha\bar{z}, \ldots, -\alpha\gamma z_r - \alpha\bar{z})^{\mathrm{T}} \in \mathbb{R}^r \\ \mathbf{1}_r^{T} - \alpha\gamma\mathbf{Z}^{\mathrm{T}} &= (1 - \alpha\gamma z_1, \ldots, 1 - \alpha\gamma z_r)^{\mathrm{T}} \in \mathbb{R}^r. \end{aligned}$$

(2) $\|\alpha\gamma\mathbf{Z} + \alpha\bar{z}\mathbf{1}_r\|^2_{\boldsymbol{\Delta}_1} = \alpha^2\gamma^3\overline{z^2} + \alpha^2\gamma\bar{z}^2 + 2\alpha^2\gamma^2\bar{z}^2$.
(3) $\|\mathbf{1}_r - \alpha\gamma\mathbf{Z}\|^2_{\boldsymbol{\Delta}_2} = \alpha\gamma\bar{z} + \alpha^3\gamma^3\bar{z}\overline{z^2} - 2\alpha^2\gamma^2\bar{z}^2$.
(4) $\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi} = -\alpha$.

Proof.

(1) For all $i = 1, \ldots, r$, we have

$$\frac{\partial g}{\partial b_i} = \frac{-\left(\sum_{k=1}^r a_k\right)\left[\left(\sum_{k=1}^r z_k(a_k+b_k)\right) + z_i\sum_{k=1}^r b_k\right]}{\left(\sum_{k=1}^r b_k\right)^2\left(\sum_{k=1}^r z_k(a_k+b_k)\right)^2}$$

$$\frac{\partial g}{\partial a_i} = \frac{\left(\sum_{k=1}^r b_k\right)\left(\sum_{k=1}^r z_k(a_k+b_k)\right) - z_i\left(\sum_{k=1}^r b_k\right)\left(\sum_{k=1}^r a_k\right)}{\left(\sum_{k=1}^r b_k\right)^2\left(\sum_{k=1}^r z_k(a_k+b_k)\right)^2}$$

and, therefore,

$$\frac{\partial g}{\partial b_i}(\boldsymbol{\pi}) = \frac{-\left(\sum_{k=1}^r \pi_{2k}\right)\left[\left(\sum_{k=1}^r z_k(\pi_{2k}+\pi_{1k})\right) + z_i\sum_{k=1}^r \pi_{1k}\right]}{\left(\sum_{k=1}^r \pi_{1k}\right)^2\left(\sum_{k=1}^r z_k(\pi_{2k}+\pi_{1k})\right)^2}$$

$$\frac{\partial g}{\partial a_i}(\boldsymbol{\pi}) = \frac{\left(\sum_{k=1}^r \pi_{1k}\right)\left(\sum_{k=1}^r z_k(\pi_{2k}+\pi_{1k})\right) - z_i\left(\sum_{k=1}^r \pi_{1k}\right)\left(\sum_{k=1}^r \pi_{2k}\right)}{\left(\sum_{k=1}^r \pi_{1k}\right)^2\left(\sum_{k=1}^r z_k(\pi_{2k}+\pi_{1k})\right)^2}.$$

Since $\pi_{2k} + \pi_{1k} = \beta_k$ and

$$\sum_{k=1}^r z_k(\pi_{2k}+\pi_{1k}) = \sum_{k=1}^r z_k\beta_k = \overline{z}, \quad \sum_{k=1}^r \pi_{1k} = \gamma, \quad \sum_{k=1}^r \pi_{2k} = \alpha\gamma\overline{z},$$

we can write

$$\frac{\partial g}{\partial b_i}(\boldsymbol{\pi}) = \frac{-\alpha\gamma\overline{z}\left(\overline{z}+z_i\gamma\right)}{\gamma^2\overline{z}^2} = \frac{-\alpha\left(\overline{z}+\gamma z_i\right)}{\gamma\overline{z}}, \quad \frac{\partial g}{\partial a_i}(\boldsymbol{\pi}) = \frac{\gamma\overline{z} - z_i\gamma\alpha\gamma\overline{z}}{\gamma^2\overline{z}^2} = \frac{1-\alpha\gamma z_i}{\gamma\overline{z}}.$$

(2)

$$\begin{aligned}
\|\alpha\gamma\mathbf{Z} + \alpha\overline{z}\mathbf{1}_r\|^2_{\boldsymbol{\Delta}_1} &= \alpha^2\,\|\gamma\mathbf{Z} + \overline{z}\mathbf{1}_r\|^2_{\boldsymbol{\Delta}_1} = \alpha^2\left(\gamma\mathbf{Z}^{\mathrm{T}} + \overline{z}\mathbf{1}_r^{\mathrm{T}}\right)\boldsymbol{\Delta}_1\left(\gamma\mathbf{Z} + \overline{z}\mathbf{1}_r\right) \\
&= \alpha^2\sum_{j=1}^r \pi_{1j}(\gamma z_j + \overline{z})^2 = \alpha^2\gamma\sum_{j=1}^r \beta_j(\gamma z_j + \overline{z})^2 \\
&= \alpha^2\gamma\left\{\gamma^2\sum_{j=1}^r \beta_j z_j^2 + \overline{z}^2\sum_{j=1}^r \beta_j + 2\gamma\overline{z}\sum_{j=1}^r \beta_j z_j\right\} \\
&= \alpha^2\gamma\left(\gamma^2\overline{z^2} + \overline{z}^2 + 2\gamma\overline{z}^2\right)
\end{aligned}$$

(3)

$$\begin{aligned}
\|\mathbf{1}_r - \alpha\gamma\mathbf{Z}\|^2_{\mathbf{\Delta}_2} &= \left(\mathbf{1}_r^{\mathrm{T}} - \alpha\gamma\mathbf{Z}^{\mathrm{T}}\right)\mathbf{\Delta}_2\left(\mathbf{1}_r - \alpha\gamma\mathbf{Z}\right) \\
&= \sum_{j=1}^{r} \pi_{2j}(1-\alpha\gamma z_j)^2 \\
&= \sum_{j=1}^{r} \pi_{2j} + \alpha^2\gamma^2 \sum_{j=1}^{r} \pi_{2j} z_j^2 - 2\alpha\gamma \sum_{j=1}^{r} \pi_{2j} z_j \\
&= \sum_{j=1}^{r} \alpha\gamma\overline{z}\beta_j + \alpha^2\gamma^2 \sum_{j=1}^{r} \alpha\gamma\overline{z}\beta_j z_j^2 - 2\alpha\gamma \sum_{j=1}^{r} \alpha\gamma\overline{z}\beta_j z_j \\
&= \alpha\gamma\overline{z} + \alpha^3\gamma^3\overline{z}\overline{z^2} - 2\alpha^2\gamma^2\overline{z}^2.
\end{aligned}$$

(4)

$$\begin{aligned}
\mathbf{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi} &= \frac{1}{\gamma\overline{z}}\Big(-\alpha\gamma\mathbf{Z}^{\mathrm{T}} - \alpha\overline{z}\mathbf{1}_r^{\mathrm{T}}, \mathbf{1}_r^{\mathrm{T}} - \alpha\gamma\mathbf{Z}^{\mathrm{T}}\Big)\begin{pmatrix}\boldsymbol{\pi}_1\\ \boldsymbol{\pi}_2\end{pmatrix} \\
&= \frac{1}{\gamma\overline{z}}\left\{-\alpha\gamma\mathbf{Z}^{\mathrm{T}}\boldsymbol{\pi}_1 - \alpha\overline{z}\mathbf{1}_r^{\mathrm{T}}\boldsymbol{\pi}_1 + \mathbf{1}_r^{\mathrm{T}}\boldsymbol{\pi}_2 - \alpha\gamma\mathbf{Z}^{\mathrm{T}}\boldsymbol{\pi}_2\right\} \\
&= \frac{1}{\gamma\overline{z}}\left\{-\alpha\gamma\sum_{k=1}^{r} z_k\pi_{1k} - \alpha\overline{z}\sum_{k=1}^{r}\pi_{1k} + \sum_{k=1}^{r}\pi_{2k} - \alpha\gamma\sum_{k=1}^{r} z_k\pi_{2k}\right\} \\
&= \frac{1}{\gamma\overline{z}}\left\{-\alpha\gamma\sum_{k=1}^{r} z_k(\pi_{1k}+\pi_{2k}) - \alpha\overline{z}\sum_{k=1}^{r}\pi_{1k} + \sum_{k=1}^{r}\pi_{2k}\right\} \\
&= \frac{1}{\gamma\overline{z}}\left\{-\alpha\gamma\sum_{k=1}^{r} z_k\beta_k - \alpha\overline{z}\sum_{k=1}^{r}\pi_{1k} + \sum_{k=1}^{r}\pi_{2k}\right\} \\
&= \frac{1}{\gamma\overline{z}}\left\{-\alpha\gamma\overline{z} - \alpha\gamma\overline{z} + \alpha\gamma\overline{z}\right\} \\
&= -\alpha
\end{aligned}$$

Theorem 4. *Let $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_r)$ be the MLE defined by Theorem 1. The approximate variance of $\hat{\alpha}$ is*

$$\mathrm{var}(\hat{\alpha}) \approx \frac{\alpha}{n\gamma^2\overline{z}} + \frac{\alpha^2\overline{z^2}}{n\overline{z}^2} - \frac{\alpha^2}{n} \tag{11}$$

where $\gamma = \dfrac{1}{1+\alpha\overline{z}}$ and $\overline{z^2} = \sum_{i=1}^{r}\beta_i z_i^2$.

Proof. Recall that

$$\hat{\alpha} = g\left(\frac{\mathbf{X}}{n}\right) = g\left(\frac{X_{11}}{n}, \ldots, \frac{X_{1r}}{n}, \frac{X_{21}}{n}, \ldots, \frac{X_{2r}}{n}\right)$$

where g is defined by (8). From Lemmas 3 and 4,

$$\begin{aligned}
\mathrm{var}(\hat{\alpha}) &\approx \frac{1}{n}\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\left(\boldsymbol{\Delta}-\boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}}\right)\boldsymbol{\nabla}_g(\boldsymbol{\pi})\\
&= \frac{1}{n}\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\nabla}_g(\boldsymbol{\pi}) - \frac{1}{n}\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{\nabla}_g(\boldsymbol{\pi})\\
&= \frac{1}{n}\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\nabla}_g(\boldsymbol{\pi}) - \frac{1}{n}\boldsymbol{\nabla}g(\pi)^{T}\boldsymbol{\pi}\left(\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi}\right)^{\mathrm{T}}\\
&= \frac{1}{n}\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\nabla}_g(\boldsymbol{\pi}) - \frac{1}{n}\left(\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi}\right)^2
\end{aligned}$$

because $\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi} \in \mathbb{R}$. On the one hand, we have

$$\begin{aligned}
\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\nabla}_g(\boldsymbol{\pi}) &= \frac{1}{\gamma^2\overline{z}^2}\Big(-\alpha\gamma\mathbf{Z}^{\mathrm{T}}-\alpha\overline{z}\mathbf{1}_r^{\mathrm{T}},\mathbf{1}_r^{\mathrm{T}}-\alpha\gamma\mathbf{Z}^{\mathrm{T}}\Big)\begin{pmatrix}\boldsymbol{\Delta}_1 & \mathbf{0}_{r,r}\\ \mathbf{0}_{r,r} & \boldsymbol{\Delta}_2\end{pmatrix}\begin{pmatrix}-\alpha\gamma\mathbf{Z}-\alpha\overline{z}\mathbf{1}_r\\ \mathbf{1}_r-\alpha\gamma\mathbf{Z}\end{pmatrix}\\
&= \frac{1}{\gamma^2\overline{z}^2}\left\{\|\alpha\gamma\mathbf{Z}+\alpha\overline{z}\mathbf{1}_r\|^2_{\boldsymbol{\Delta}_1}+\|\mathbf{1}_r-\alpha\gamma\mathbf{Z}\|^2_{\boldsymbol{\Delta}_2}\right\}
\end{aligned}$$

and, from Lemma 5,

$$\begin{aligned}
\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\nabla}_g(\boldsymbol{\pi}) &= \frac{1}{\gamma^2\overline{z}^2}\left\{\alpha^2\gamma^3\overline{z^2}+\alpha^2\gamma\overline{z}^2+2\alpha^2\gamma^2\overline{z}^2+\alpha\gamma\overline{z}+\alpha^3\gamma^3\overline{z}\overline{z^2}-2\alpha^2\gamma^2\overline{z}^2\right\}\\
&= \frac{1}{\gamma^2\overline{z}^2}\left\{\alpha^2\gamma^3\overline{z^2}+\alpha^2\gamma\overline{z}^2+\alpha\gamma\overline{z}+\alpha^3\gamma^3\overline{z}\overline{z^2}\right\}\\
&= \frac{\alpha}{\gamma\overline{z}^2}\left\{\alpha\gamma^2\overline{z^2}+\alpha\overline{z}^2+\overline{z}+\alpha^2\gamma^2\overline{z}\overline{z^2}\right\}\\
&= \frac{\alpha}{\gamma\overline{z}^2}\left\{(\alpha\gamma^2\overline{z^2}+\overline{z})(1+\alpha\overline{z})\right\}\\
&= \frac{\alpha}{\gamma\overline{z}^2}\left\{(\alpha\gamma^2\overline{z^2}+\overline{z})\frac{1}{\gamma}\right\}\\
&= \frac{\alpha^2\overline{z^2}}{\overline{z}^2}+\frac{\alpha}{\gamma^2\overline{z}}.
\end{aligned}$$

On the other hand, by Lemma 5, $\boldsymbol{\nabla}_g(\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{\pi} = -\alpha$. Finally, we have

$$\mathrm{var}(\hat{\alpha}) \approx \frac{1}{n}\left\{\frac{\alpha^2\overline{z^2}}{\overline{z}^2}+\frac{\alpha}{\gamma^2\overline{z}}\right\}-\frac{1}{n}(-\alpha)^2$$

and the proof is completed.

Remark 2. As the variance of a random variable is always positive, some authors (see for example (Casella and Berger, 2002, Theorem 5.5.28)) add the condition $\boldsymbol{\nabla}_g(\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\nabla}_g(\boldsymbol{\mu}) > 0$ in Lemma 3. This should be straightforward if the matrix $\boldsymbol{\Sigma}$ was positive definite, but this is not the case here because the variance-covariance matrix $\boldsymbol{\Sigma}$ of the multinomial distribution is not even invertible (the sum of its elements by row or by column is zero). Fortunately, for the approximation of the variance of $\hat{\alpha}$, we have

$$\frac{\alpha}{n\gamma^2\overline{z}} + \frac{\alpha^2\overline{z^2}}{n\overline{z}^2} - \frac{\alpha^2}{n} = \frac{\alpha}{n\gamma^2\overline{z}} + \frac{\alpha^2}{n}\left(\frac{\overline{z^2}}{\overline{z}^2} - 1\right)$$

and

$$\begin{aligned} 0 \leqslant \sum_{i=1}^{r} \beta_i (z_i - \overline{z})^2 &= \sum_{i=1}^{r} \beta_i z_i^2 + \overline{z}^2 \sum_{i=1}^{r} \beta_i - 2\overline{z} \sum_{i=1}^{r} \beta_i z_i \\ &= \overline{z^2} + \overline{z}^2 - 2\overline{z}^2 \\ &= \overline{z^2} - \overline{z}^2 \end{aligned}$$

hence the quantity (11) is strictly positive.

5. Discussion

In this paper, we derived the closed-form expression of the constrained maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\boldsymbol{\beta}})$ of the parameter vector $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ of a multivariate discrete crash data model where $\alpha > 0$ is the parameter of interest and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)$ is a vector of probabilities such that $\sum_{j=1}^{r} \beta_j = 1$. We proved that the MLE is strongly consistent and we derived the approximated variance of $\hat{\alpha}$ using the delta method and the exact variance of the components of $\hat{\boldsymbol{\beta}}$. The model considered in this paper is motivated by the statistical analysis of the effectiveness of a road safety measure applied to a given target site paired with a control site where the measure was not applied.

It should be interesting to consider the general case (N'Guessan et al., 2006a) where the road safety measure is applied, no longer on a single site but on s ($s > 1$) different treatment sites, each being paired with a control site. In this case, the parameter vector still has the form $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ but $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s)$ where, for all $k = 1, \ldots, s$, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kr})$ and $\sum_{j=1}^{r} \beta_{kj} = 1$. This simultaneous consideration of several treatment sites considerably increases the complexity of the model and, a priori, does not allow to envisage obtaining a closed-form expression of the MLE. However, one could consider studying the strong consistency directly from the likelihood equations as in Geraldo et al. (2015) and obtain the approximate variance of the components of the MLE using other methods like the analytical inversion of the Fisher information matrix (Aitchison and Silvey, 1958; Neuenschwander and Flury, 1997; N'Guessan and Langrand, 2005).

Acknowledgments

The author thanks the editorial team and, most especially, the editor for the consideration given to this paper. The author also thanks the anonymous reviewer for his careful reading of the manuscript and his insightful comments and recommendations that helped to improve the quality of the paper.

References

Aitchison, J. and Silvey, S. D. (1958). Maximum-Likelihood Estimation of Parameters Subject to Restraints. *The Annals of Mathematical Statistics*, 29(3):813–828.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Thomson Learning, 2nd edition.

Geraldo, I. C., N'Guessan, A., and Gneyou, K. E. (2015). A note on the strong consistency of a constrained maximum likelihood estimator used in crash data modelling. *Comptes Rendus Mathematique*, 353(12):1147–1152.

Geraldo, I. C., N'Guessan, A., and Gneyou, K. E. (2018). Global convergence and ascent property of a cyclic algorithm used for statistical analysis of crash data. *Afrika Statistika*, 13(2):1631–1643.

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

Lo, G. S., Ngom, M., and Kpanzou, T. A. (2016). *Weak Convergence (IA). Sequences of random vectors*. Statistics and Probability African Society (SPAS) Books Series, Calgary, Alberta. DOI: 10.16929/sbs/2016.0001, ISBN: 978-2-9559183-1-9.

Neuenschwander, B. E. and Flury, B. D. (1997). A note on Silvey's (1959) Theorem. *Statistics & Probability Letters*, 36:307–317.

N'Guessan, A. (2010). Analytical Existence of solutions to a system of non-linear equations with application. *Journal of Computational and Applied Mathematics*, 234:297–304.

N'Guessan, A., Essai, A., and Langrand, C. (2001). Estimation multidimensionnelle des contrôles et de l'effet moyen d'une mesure de sécurité routière. *Revue de statistique appliquée*, 49(2):85–102.

N'Guessan, A., Essai, A., and N'zi, M. (2006a). An estimation method of the average effect and the different accident risks when modelling a road safety measure: A simulation study. *Computational Statistics & Data Analysis*, 51(2):1260–1277.

N'Guessan, A. and Geraldo, I. C. (2015). A cyclic algorithm for maximum likelihood estimation using Schur complement. *Numerical Linear Algebra with Applications*, 22(6):1161–1179.

N'Guessan, A. and Langrand, C. (1993). Sur la distribution asymptotique de certaines statistiques utilisées dand le domaine de la sécurité routière. *C. R. Acad. Sci. Paris, Ser. I*, 317:401–404.

N'Guessan, A. and Langrand, C. (2005). A covariance components estimation procedure when modelling a road safety measure in terms of linear constraints. *Statistics*, 39(4):303–314.

N'Guessan, A. and Truffier, M. (2008). Impact d'un aménagement de sécurité routière sur la gravité des accidents de la route. *Journal de la Société Française de Statistique*, 149(3):23–41.

N'Guessan, A., Truffier, M., Langrand, C., and Bellavance, F. (2006b). 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):1–20.

Ogden, K. W. (1997). The effects of paved shoulders on accidents on rural highways. *Accident Analysis and Prevention*, 29(3):353–362.

Proschan, M. A. and Shaw, P. A. (2016). *Essentials of Probability Theory for Statisticians*. Chapman and Hall/CRC Texts in Statistical Science. CRC Press, New York.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge (United Kingdom).

Wasserman, L. (2004). *All of Statistics: a concise course in statistical inference*. Springer, New York.