C. R. Math. Rep. Acad. Sci. Canada Vol. **27**, (3), 2005 pp. 84–91

# A CONFIDENCE INTERVAL ESTIMATION PROBLEM USING THE SCHUR COMPLEMENT APPROACH, WITH APPLICATION

ASSI N'GUESSAN AND FRANÇOIS BELLAVANCE

Presented by Vlastimil Dlab, FRSC

ABSTRACT. We present a method based on the Schur complement approach to build asymptotic confidence intervals linked to the maximum likelihood estimator of a vector of parameters under constraints. This approach makes it possible to obtain the formal expression of the standard error of each component of the vector without direct inversion of the Fisher information matrix. We then give an application of this method to the modelling and the confidence interval estimation of the average effect of a road safety measure and the accident risks of different types.

RÉSUMÉ. Nous proposons une méthode basée sur la technique du complément de Schur pour construire des intervalles de confiance asymptotiques relatifs à l'estimateur du maximum de vraisemblance d'un vecteur paramètre soumis à des contraintes. Cette méthode permet d'obtenir l'expression formelle de l'écart-type de chaque composante du vecteur sans inverser directement la matrice d'information de Fisher. Nous indiquons ensuite une application de cette méthodologie à la modélisation et à l'estimation par intervalle de confiance de l'effet moyen d'une mesure de sécurité routière et de différents risques d'accident.

1. **Introduction and assumptions.** Let's consider $\{(X_{1k}, X_{2k}), 1 \leq k \leq s\}$, a collection of $s$ pairs of discreet independent random vectors where $X_{1k} = (X_{11k}, X_{12k}, \dots, X_{1rk})^T$ and $X_{2k} = (X_{21k}, X_{22k}, \dots, X_{2rk})^T$ are $r \times 1$ random vectors with $r > 1$. We assume that, for fixed $k$, each pair $(X_{1k}, X_{2k})$ has a multinomial distribution denoted by $\mathcal{M}(n_k\, ; \Pi_{1k}(\theta), \Pi_{2k}(\theta))$, where $n_k$ ($n_k > 0$) is a given integer, $\Pi_{tk}(\theta) = (\pi_{t1k}(\theta), \pi_{t2k}(\theta), \dots, \pi_{trk}(\theta))$ $t = 1, 2$, is a $r \times 1$ vector of cell probabilities, $\theta = (\gamma, \beta^T)^T$ an unknown vector of parameters whose dimension is $(1 + sr) \times 1$, $\gamma$ being a real number, $\beta = (\beta_1^T, \beta_2^T, \dots, \beta_s^T)^T$ being a vector whose dimension is $(sr) \times 1$, and $\beta_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{rk})^T \in \mathbb{R}^r$ with $0 < \beta_{jk} < 1$. We propose an estimation method for the asymptotic confidence interval of each element of $\theta$. This estimation method is based on the formal inversion of the Fisher information matrix using Schur complement (see Ouellette [11], Zhang [16]). There are certainly other methods (see for instance Don [3], Rao and Yanai [12], Tanabe and Sagae [14]) to invert the Fisher information matrix but the one we suggest is based on the Schur complement approach and

Received by the editors on March 14, 2005.

AMS subject classification: 15A09, 62F12, 62F25, 62F30, 62H10, 62H12, 62P30.

Keywords: logistic multinomial model, constrained maximum likelihood, confidence interval, Fisher information matrix, formal asymptotic variance, Schur complement, road safety measure, accident risk, accident data.

generalizes N'Guessan's results [7].

Let $\mathcal{L}(\theta) = \sum_{k=1}^{s} \mathcal{L}_k(\theta)$ the log-likelihood linked to the $s$ pairs with $\mathcal{L}_k(\theta)$, the one associated to pair $k$, and we assume for the rest of the work that the following conditions are verified:

(A1) $\forall k$, $\mathcal{L}_k(\theta)$ is continuously differentiable such that

$$\frac{\partial \mathcal{L}_k(\theta)}{\partial \beta_m} \equiv \mathbf{0}_r, \quad k \neq m, \quad k, m = 1, 2, \ldots, s$$

where $\mathbf{0}_r$ is a $r \times 1$ vector of zeros;

(A2) Fisher information matrix $J_\theta = \mathbb{E}(-\frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta\partial\theta^T})$ is thus partitioned:

$$J_\theta = \begin{bmatrix} \tau_\theta & U_\theta^T \\ U_\theta & B_\theta \end{bmatrix}$$

where $\tau_\theta$ is a real number, $U_\theta$ is a $(sr) \times 1$ vector and $B_\theta$ a $(sr) \times (sr)$ matrix, respectively defined by

$$\tau_\theta = \mathbb{E}\Big(-\frac{\partial^2 \mathcal{L}(\theta)}{\partial\gamma\partial\gamma}\Big), \quad U_\theta = (U_{\theta,1}^T, U_{\theta,2}^T, \ldots, U_{\theta,s}^T)^T,$$
$$B_\theta = \text{block-diag}(B_{\theta,1}, B_{\theta,2}, \ldots, B_{\theta,s})$$

with, for a fixed value of $k$, $U_{\theta,k}$ a $r \times 1$ vector of components $\mathbb{E}(-\frac{\partial^2 \mathcal{L}(\theta)}{\partial\gamma\partial\beta_{jk}})$ and $B_{\theta,k}$, an $r \times r$ matrix with elements $\mathbb{E}(-\frac{\partial^2 \mathcal{L}(\theta)}{\partial\beta_{jk}\partial\beta_{mk}})$, $(j = 1, 2, \ldots, r;$ $m = 1, 2, \ldots, r)$;

(A3) $\forall k$, there exists $\Omega_{\theta,k}$, a nonsingular $r \times r$ diagonal matrix, $V_{\theta,k}$, a $r \times 1$ vector, $a_k$ and $b_k$ two strictly positive real numbers such that

(i)
$$B_{\theta,k} = a_k(\Omega_{\theta,k} - V_{\theta,k}V_{\theta,k}^T),$$

(ii)
$$U_{\theta,k} = b_k V_{\theta,k};$$

(A4) $\forall k$ $(k = 1, 2, \ldots, s)$, we set

$$\Omega_{\theta,k}^{(0)} = \begin{bmatrix} \Omega_{\theta,k} & -V_{\theta,k} \\ V_{\theta,k}^T & 0 \end{bmatrix}, \quad B_{\theta,k}^{(0)} = \begin{bmatrix} B_{\theta,k} & -1_r \\ 1_r^T & 0 \end{bmatrix}$$

two $(r+1) \times (r+1)$ matrices and we assume that

(i)
$$0 < (\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k}) < 1,$$

(ii)

$$0 < \sum_{k=1}^{s} \frac{b_k^2}{a_k} \frac{(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})}{1-(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})} < \tau_\theta,$$

(iii)

$$0 < (B_{\theta,k}^{(0)}/B_{\theta,k}),$$

where $(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})$ (resp. $(B_{\theta,k}^{(0)}/B_{\theta,k})$) is the Schur complement of $\Omega_{\theta,k}$ in $\Omega_{\theta,k}^{(0)}$ (resp. the Schur complement of $B_{\theta,k}$ in $B_{\theta,k}^{(0)}$);

(A5) $\forall k$, there exists $h_k\colon \mathbb{R}^{1+sr} \mapsto \mathbb{R}$, $\theta \mapsto h_k(\theta)$ a continuously differentiable function such that

(i)

$$h_k(\theta) = 0,$$

(ii)

$$\frac{\partial h_k}{\partial \gamma} \equiv \mathbf{0}_s,$$

(iii)

$$\frac{\partial h_k}{\partial \beta_m} \equiv \delta_{km}\mathbf{1}_r,$$

$k, m = 1, 2, \ldots, s$ with $\delta_{km}$ being the Kronecker symbol, $\mathbf{0}_s = (0, \ldots, 0)^T \in \mathbb{R}^s$, and $\mathbf{1}_r = (1, \ldots, 1)^T \in \mathbb{R}^r$.

## 2. Preliminary results using Schur complement.

THEOREM 2.1. *Under* (A1) *to* (A4), *we show that*

(i) $(J_\theta/B_\theta) > 0$,
(ii) $(J_\theta/\tau_\theta)^{-1} = B_\theta^{-1} + (J_\theta/B_\theta)^{-1} B_\theta^{-1} U_\theta U_\theta^T B_\theta^{-1}$,

*where* $(J_\theta/B_\theta)$ *(resp.* $(J_\theta/\tau_\theta)$*) is the Schur complement of* $B_\theta$ *(resp. of* $\tau_\theta$*) in* $J_\theta$ *and is defined by* $(J_\theta/B_\theta) = \tau_\theta - U_\theta^T B_\theta^{-1} U_\theta$ *(resp.* $(J_\theta/\tau_\theta) = B_\theta - U_\theta \tau_\theta^{-1} U_\theta^T$*).*

THEOREM 2.2. *Let* $H_\theta = (H_\gamma, H_\beta)$ *the* $s \times (1+sr)$ *matrix with* $H_\gamma = (\frac{\partial h_1}{\partial \gamma}, \frac{\partial h_2}{\partial \gamma}, \ldots, \frac{\partial h_s}{\partial \gamma})^T$, $H_\beta = (\frac{\partial h_1}{\partial \beta}, \frac{\partial h_2}{\partial \beta}, \ldots, \frac{\partial h_s}{\partial \beta})^T$, *being respectively* $s \times 1$ *and* $s \times (sr)$ *matrices,* $\xi_\theta = H_\beta B_\theta^{-1} U_\theta$ *being an* $s \times 1$ *vector,*

$$\Lambda_\theta = \operatorname{diag}\big((B_{\theta,1}^{(0)}/B_{\theta,1}), \ldots, (B_{\theta,s}^{(0)}/B_{\theta,s})\big)$$

*an* $s \times s$ *matrix,*

$$\Lambda_\theta^{(1)} = \begin{bmatrix} \Lambda_\theta & -\xi_\theta \\ \xi_\theta^T & (J_\theta/B_\theta) \end{bmatrix}, \quad \Gamma_\theta = \begin{bmatrix} J_\theta & H_\theta^T \\ H_\theta & \bigcirc_{s,s} \end{bmatrix}$$

*being respectively* $(s+1)\times(s+1)$ *and* $(1+sr+s)\times(1+sr+s)$ *matrices with* $\bigcirc_{s,s}$ *the* $s\times s$ *matrix with all entries equal to zero. Then, under conditions* (A1) *to* (A5), *we show that*

(i) $R_\theta = -(\Gamma_\theta/J_\theta)$, *the negative of the Schur complement of* $J_\theta$ *in* $\Gamma_\theta$, *is a nonsingular* $s\times s$ *matrix and* $R_\theta^{-1} = \Lambda_\theta^{-1} - \Lambda_\theta^{-1}\xi_\theta(\Lambda_\theta^{(1)}/\Lambda_\theta)^{-1}\xi_\theta^T\Lambda_\theta^{-1}$.

(ii) *Matrix* $\Gamma_\theta$ *is nonsingular and*

$$(1)\qquad \Gamma_\theta^{-1} = \begin{bmatrix} W_\theta & J_\theta^{-1}H_\theta^T R_\theta^{-1} \\ R_\theta^{-1}H_\theta J_\theta^{-1} & -R_\theta^{-1}\end{bmatrix}, \quad W_\theta = \begin{bmatrix} W_\theta(1,1) & W_\theta^T(2,1) \\ & W_\theta(2,2)\end{bmatrix}$$

*where the diagonal components* $W_\theta(1,1)$ *and* $W_\theta(2,2)$ *of matrix* $W_\theta$ *are respectively a scalar and an* $(sr)\times(sr)$ *matrix defined as follows:*

$$(2)\qquad \begin{aligned} W_\theta(1,1) &= (\Lambda_\theta^{(1)}/\Lambda_\theta)^{-1} \\ W_\theta(2,2) &= (J_\theta/\tau_\theta)^{-1} - (J_\theta/\tau_\theta)^{-1}H_\beta^T R_\theta^{-1}H_\beta(J_\theta/\tau_\theta)^{-1} \\ W_\theta(2,1) &= -(J_\theta/B_\theta)^{-1}[B_\theta^{-1}U_\theta - (J_\theta/\tau_\theta)^{-1}H_\beta^T R_\theta^{-1}\xi_\theta]. \end{aligned}$$

Remark 1. We give an outline of the proofs of Theorems 2.1 and 2.2.

From matrix manipulations and (A3) (i) and (A4) (i) we obtain

$$(3)\qquad (J_\theta/B_\theta) = \tau_\theta - \sum_{k=1}^{s}\frac{b_k^2}{a_k}\frac{(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})}{1-(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})}.$$

Now using (A4) (i) and (A3) (ii), we get part (i) of Theorem 2.1. The rest of the proof uses theorems in N'Guessan [7, Section 2] and N'Guessan and Langrand [10, Section 3] and general applications of well-known results of Schur complement about the inverse of a partitioned matrix.

## 3. Asymptotic confidence interval.

Theorem 3.1. *Let* $M_\theta = J_\theta^{-1}\big(J_\theta + (\frac{\partial^2\mathcal{L}(\theta)}{\partial\theta\partial\theta^T})\big)$ *a* $(1+sr)\times(1+sr)$ *matrix,* $\Phi_\theta = J_\theta^{-1}H_\theta^T R_\theta^{-1}$ *a* $(1+sr)\times s$ *matrix,* $\lambda$ *the* $s\times 1$ *vector of the Lagrange multipliers,* $\theta^0$ *the true value of the vector of parameters* $\theta$, *and we suppose that, in addition to conditions* (A1) *to* (A5),

(A6) $(\hat\theta,\hat\lambda)$, *the maximum likelihood estimators of* $\theta$ *and* $\lambda$, *exist and converge.*

(A7) $M_{\hat\theta} \xrightarrow{\text{Proba}} 0$, $\Phi_{\hat\theta}-\Phi_{\theta^0} \xrightarrow{\text{Proba}} 0$, $|\Phi_{\theta^0}|<\infty$ *when* $n_k$ *and* $N=\sum_{k=1}^{s} n_k \longrightarrow +\infty$.

(A8) $(\frac{\partial L}{\partial\theta})_{\theta^0} \overset{\mathcal{D}\rangle\int\sqcup\nabla.}{\sim} \mathcal{N}(0,J_{\theta^0})$.

*Then the asymptotic confidence intervals, at the* $(1-\alpha)$ *confidence level, for the* $\theta$ *parameter components are given by the following formal expressions:*

$$
\begin{aligned}
IC_{1-\alpha}(\gamma) &= \left[\hat{\gamma} - \frac{z_{\alpha/2}}{(\Lambda_\theta^{(1)}/\Lambda_\theta)^{1/2}}\ ;\ \hat{\gamma} + \frac{z_{\alpha/2}}{(\Lambda_\theta^{(1)}/\Lambda_\theta)^{1/2}}\right] \\
IC_{1-\alpha}(\beta_{jk}) &= [\max(0;\hat{\beta}_{jk} - z_{\alpha/2}\hat{\sigma}_{jk})\ ;\ \min(1;\hat{\beta}_{jk} + z_{\alpha/2}\hat{\sigma}_{jk})] \\
&\qquad (k = 1, 2, \ldots, s; j = 1, 2, \ldots, r)
\end{aligned}
\tag{4}
$$

*where* $z_{\alpha/2}$ *and* $\hat{\sigma}^2_{jk}$ *are respectively obtained through the standard normal law table and the elements of the main diagonal of the* $(sr)\times(sr)$ *matrix* $(J_\theta/\tau_\theta)^{-1} - (J_\theta/\tau_\theta)^{-1}H_\beta^T R_\theta^{-1} H_\beta (J_\theta/\tau_\theta)^{-1}$.

REMARK 2. The main point of the proof consists in applying the mean value theorem, in the neighbourhood of $\theta^0$, to the constrained maximum likelihood equations and uses a version of a theorem in Crowder [2, Section 3] to show that asymptotically

$$
\begin{pmatrix} \hat{\theta} - \theta^0 \\ \eta_{\hat{\theta}} \end{pmatrix} \overset{\mathcal{D}\rangle\int\sqcup\nabla.}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} W_{\theta^0} & 0 \\ 0 & R_{\theta^0} \end{pmatrix}\right)
\tag{5}
$$

with $\eta_{\hat{\theta}} = R_\theta \hat{\lambda}$ being a $s \times 1$ vector and $W_\theta = J_\theta^{-1} - J_\theta^{-1} H_\theta^T R_\theta^{-1} H_\theta J_\theta^{-1}$ a $(1+sr)\times(1+sr)$ matrix obtained by inverting matrix $\Gamma_\theta$ with the Schur complement method. We then use conditions (A1) to (A5) along with Theorems 2.1 and 2.2 to deduce the asymptotic variance structure of the components of $\hat{\theta}$. Part of the method proposed here is clearly related to the constrained maximum likelihood estimation (RMLE) method which is an approach to estimation that maximizes the logarithm of a likelihood over a restricted space. This approach of estimation is the well-known applied mathematical method that maximizes a function subjected to restraints. The general statistical framework is well known and is not discussed here (see for instance Aitchison and Sylvey [1], Sylvey [13], Don [3], Magnus [5, p. 172]). However one recalls that the size of matrix $\Gamma_\theta$ increases very quickly with $r$ and $s$. The main strength of the Schur complement approach is its applicability to any values of $s$ and $r$, *i.e.*, to any dimension of matrix $\Gamma_\theta$, without being obliged to invert it.

## 4. Application: Evaluation of a road safety measure.

*4.1. A logistic multinomial model.* We apply the above results to a statistical road accident data modelling problem when a road safety measure is applied on different sites. We consider a multidimensional combination of road accident frequencies before and after the introduction of a road safety measure (crossroad lay-out, surface of a motorway section, *etc.*) at $s$ $(s > 0)$ experimental sites. Each site counts $r$ $(r > 1)$ accident types (fatal accidents, seriously injured people, slightly injured people, material damage only) over a fixed period of time. To model the average effect of the safety measure and the accident risks of different types, let's consider the period before (resp. after) the introduction of the safety measure (or change) and note $X_{1k} = (X_{11k}, X_{12k}, \dots, X_{1rk})^T$ (resp. $X_{2k} = (X_{21k}, X_{22k}, \dots, X_{2rk})^T$) the random vector giving the $r$ accident types number on experimental site $k$. In order to take some external factors into account (such as traffic flow, speed limit variation, weather conditions, ...), each experimental site is matched to a control site where the safety measure was not applied. We further assume that the vector of accident data for control site $k$ ($k = 1, 2, \dots, s$), over the same time period, is fixed and known and is given by $y_k = (y_{1k}, y_{2k}, \dots, y_{rk})^T$ where $y_{jk}$ denotes the ratio of the number of accidents of type $j$ for the period after to the period before in control site $k$ and uses a version of a control ratio defined by Tanner [15].

We then assume that $\{(X_{1k}, X_{2k}), 1 \le k \le s\}$ is a collection of $s$ independent $2r$-dimensional random vector and that, for fixed $k$, each couple $(X_{1k}, X_{2k})$ has a logistic multinomial distribution (see Hosmer and Lemeshow [4], Mehta and Patel [6]) denoted by $\mathcal{LM}(n_k\,; \Pi_{1k}(\theta), \Pi_{2k}(\theta))$, where $n_k$ ($n_k > 0$) is the observed total accident number in experimental site $k$,

$$\Pi_{tk}(\theta) = (\pi_{t1k}(\theta), \pi_{t2k}(\theta), \dots, \pi_{trk}(\theta))$$

is the $r \times 1$ vector of cell probabilities with component $\pi_{tjk} \colon \mathbb{R}^{1+sr} \mapsto ]0;1[$ given by a logistic link as follows:

$$\begin{aligned} \pi_{1jk}(\theta) &= \frac{\beta_{jk}}{1 + e^{\gamma} \sum_{m=1}^{r} y_{mk}\beta_{mk}}; \\ \pi_{2jk}(\theta) &= \frac{y_{jk}\beta_{jk}e^{\gamma}}{1 + e^{\gamma} \sum_{m=1}^{r} y_{mk}\beta_{mk}}, \quad (j = 1, 2, \dots, r), \end{aligned} \tag{6}$$

and where $\theta = (\gamma, \beta^T)^T$ is the $(1 + sr) \times 1$ unknown vector of parameters with $\gamma$ the logarithm of the average effect, $\beta = (\beta_1^T, \beta_2^T, \dots, \beta_s^T)^T$ with $\beta_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{rk})^T$ a $r \times 1$ vector of accident risks connected to the control site $k$ such that

$$\beta_k \in \mathcal{S}_k^{(r-1)} = \Big\{(\beta_{1k}, \beta_{2k}, \dots, \beta_{rk})^T \in \mathbb{R}^r, \beta_{jk} > 0, \sum_{j=1}^{r} \beta_{jk} = 1\Big\}. \tag{7}$$

The logistic link between $\pi_{tjk}(\theta)$ and vector $\theta$ comes from the combination of the multinomial model built by N'Guessan *et al.* [9] and used by N'Guessan [8] and the functional $g \colon \mathbb{R}^{1+sr} \mapsto \mathbb{R}^{1+sr}$, $\theta \mapsto g(\theta) = (e^{\gamma}, \beta^T)^T$ which leaves the vector $\beta$ invariant and transforms $\gamma$ in $e^{\gamma}$. Likewise, the fact that the sub-vector $\beta_k$ belongs to the simplex $\mathcal{S}_k^{(r-1)}$ means that the vector of parameters $\theta$ has the following restraints ($0 < \beta_{jk} < 1$) and the linear constraints $h_k(\theta) = 0$ with

$$h_k(\theta) = \langle \mathbf{1}_r, \beta_k \rangle - 1, \quad (k = 1, 2, \dots, s) \tag{8}$$

where $\langle\ ,\ \rangle$ is the classic inner product. The latter functions show that assumptions (A5) is satisfied and

$$H_\gamma = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^s; \quad H_\beta = \begin{pmatrix} \mathbf{1}_r^T & \mathbf{0}_r^T & \dots & \mathbf{0}_r^T \\ \mathbf{0}_r^T & \mathbf{1}_r^T & \ddots & \mathbf{0}_r^T \\ \vdots & \ddots & \ddots & \mathbf{0}_r^T \\ \mathbf{0}_r^T & \dots & \mathbf{0}_r^T & \mathbf{1}_r^T \end{pmatrix}_{s\times(sr)}.$$

Using model (6), we show that assumptions (A1) to (A4) are also satisfied with (9)

$$\mathcal{L}(\theta) = \text{constant} + \sum_{k=1}^{s}\sum_{j=1}^{r}\Big\{x_{.jk}\log_e(\beta_{jk}) + x_{2jk}\gamma - x_{.jk}\log_e(1+e^\gamma\sum_{m=1}^{r} y_{mk}\beta_{mk})\Big\}$$

where $x_{.jk} = x_{1jk} + x_{2jk}$ and $x_{1jk}$ (resp. $x_{2jk}$) stands for the number of accidents of type $j$ on site $k$ before (resp. after) the introduction of the safety measure.

*4.2. Asymptotic confidence interval for the average effect.* Taking the second partial derivatives of the negative of $\mathcal{L}(\theta)$ and using the usual statistical expectation operator, we obtain

$$\begin{aligned} a_k &= \frac{n_k}{1+e^\gamma\langle y_k, \beta_k\rangle}, \quad b_k = \frac{a_k}{(1+e^\gamma\langle y_k,\beta_k\rangle)^{1/2}}, \\ \Omega_{\theta,k} &= \mathrm{Diag}\Big(\frac{1+y_{1k}e^\gamma}{\beta_{1k}}, \dots, \frac{1+y_{rk}e^\gamma}{\beta_{rk}}\Big), \\ V_{\theta,k} &= \frac{e^\gamma}{(1+\langle y_k,\beta_k\rangle e^\gamma)^{1/2}}(y_{1k},\dots,y_{rk})^T. \end{aligned} \tag{10}$$

Using the latter expressions, we obtain the explicit expression of the asymptotic variance of $\hat{\gamma}$ as follows

$$\sigma^2(\hat\gamma) = \frac{1}{(\Lambda_\theta^{(1)}/\Lambda_\theta)}, \quad (\Lambda_\theta^{(1)}/\Lambda_\theta) = (J_\theta/B_\theta) + (\Lambda_\theta^{(0)}/\Lambda_\theta), \tag{11}$$

with

$$\begin{aligned} (J_\theta/B_\theta) &= \sum_{k=1}^{s} a_k\Big\{\frac{e^\gamma\langle y_k,\beta_k\rangle}{1+e^\gamma\langle y_k,\beta_k\rangle} - \frac{(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})}{1-(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})}\Big\}, \\ (\Lambda_\theta^{(0)}/\Lambda_\theta) &= \sum_{k=1}^{s}\frac{a_k}{n_k}\Big\{\frac{1}{(B_{\theta,k}^{(0)}/B_{\theta,k})}\frac{\Delta_{\theta,k}^2}{[1-(\Omega_{\theta,k}^{(0)}/\Omega_{\theta,k})]^2}\Big\} \end{aligned}$$

and

$$\Lambda_\theta^{(0)} = \begin{bmatrix} \Lambda_\theta & -\xi_\theta \\ \xi_\theta^T & 0 \end{bmatrix}$$

an $(s+1)\times(s+1)$ matrix, $\Delta_{\theta,k} = \mathbf{1}_r^T \Omega_{\theta,k}^{-1} V_{\theta,k}$ a scalar. Thereafter, we deduce the asymptotic confidence interval. We can also get the confidence interval for the other components of the vector of parameters $\theta$. We only need matrix manipulations (see N'Guessan [8], N'Guessan and Langrand [10] for useful technical details).

Acknowledgement. Part of this work was set up while the first author was a guest researcher at CRT (Centre for Research on Transportation), University of Montreal, whom we would like to thank for their technical and material help.

## References


**1**. J. Aitchison and S. D. Silvey, *Maximum likelihood estimation of parameters subject to restraints.* Ann. Math. Statist. **29** (1958), 813–829.
**2**. M. Crowder, *On the constrained maximum likelihood estimation with non i.i.d. observations.* Ann. Inst. Statist. Math. **36** (1984), 239–249.
**3**. F. J. H. Don, *The use of generalized inverses in restricted maximum likelihood.* Linear Algebra Appl. **70** (1985), 225–240.
**4**. D. W. Hosmer and S. Lemeshow, *Applied logistic regression.* 2nd edition, John Wiley & Sons, 2000.
**5**. J. R. Magnus, *Linear Structures.* Charles Griffin & Co., London, 1988.
**6**. C. R. Mehta and N. R. Patel, *Exact logistic regression: theory and examples.* Statistics in Medicine **14** (1995), 2143–2160.
**7**. A. N'Guessan, *Constrained covariance matrix estimation in road accident modelling with Schur complements.* C. R. Acad. Sci. Paris Ser. I **337** (2003), 219–222.
**8**. ______, *Constrained estimation of a road safety covariance matrix using Schur complements.* Technical Report, Centre for Research on Transportation, Université de Montréal, CRT-2003-12.
**9**. A. N'Guessan, A. Essai and C. Langrand, *Estimation multidimensionnelle des contrôles et de l'effet moyen d'une mesure de sécurité routière.* Rev. Statist. Appl. (2) **XLIX** (2001), 83–100.
**10**. A. N'Guessan and C. Langrand, *A Schur complement approach for computing subcovariance matrices arising in a road safety measure modelling.* J. Comput. Appl. Math. **177** (2005), 331–345.
**11**. D. V. Ouellette, *Schur complements and statistics.* Linear Algebra Appl. **36** (1981), 187–295.
**12**. C. R. Rao and H. Yanai, *Generalized inverses of partitioned matrices useful in statistical aplications.* Linear Algebra Appl. **70** (1985), 105–115.
**13**. S. D. Silvey, *The Lagrangian multiplier test.* Ann. Math. Statist. **30** (1959), 389–407.
**14**. K. Tanabe and M. Sagae, *An Exact Cholesky Decomposition and the Generalized Inverse of the Variance-Covaraince Matrix of the Multinomial Distribution, with Applications.* J. Roy. Statist. Soc. Ser. B (1) **54** (1992), 211–219.
**15**. J. C. Tanner, *A problem in the combination of accident frequencies.* Biometrika **45** (1958), 331–342.
**16**. F. Zhang (ed.), *Schur complement and its applications.* Springer-Verlag, 2005.



*Polytech'Lille and Laboratoire*
*Paul Painlevé CNRS, UMR 8542*
*Université de Lille 1*
*Bât. Polytech'Lille*
*59655 Villeneuve d'Ascq Cedex*
*France*
*email: assi.nguessan@polytech-lille.fr*

*HEC Montréal and Laboratory on*
*Transportation Safety*
*Centre for Research on Transportation*
*Université de Montréal,*
*P.O. Box 6128*
*Succ. Centre-ville*
*Montréal, Québec*
*H3C 3J7*
*email: francois.bellavance@hec.ca*