HAL
open science

A MM algorithm for constrained estimation in a road safety measure modelling

Assi N'Guessan, Abdallah Mkhadri, Bezza Hafidi

To cite this version:

Assi N'Guessan, Abdallah Mkhadri, Bezza Hafidi. A MM algorithm for constrained estimation in a road safety measure modelling. Communications in Statistics - Simulation and Computation, 2010, 39 (05), pp.1057-1071. 10.1080/03610911003778119 . hal-00591696

HAL Id: hal-00591696

https://hal.science/hal-00591696v1

Submitted on 10 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MM algorithm for constrained estimation in a road safety measure modelling

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2009-0179.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	12-Mar-2010
Complete List of Authors:	N'GUESSAN, Assi; Université de Lille 1, EPU, dept. GIS Mkhadri, Abdallah; Université Cadi-Ayyad, Dept. Maths Hafidi, Bezza; Université Inb Zohr
Keywords:	Road Safety Measure, Accident data, Multinomial model, Constrained Maximum Likelihood, MM algorithm, GP-EM algorithm, Monte Carlo Simulation
Abstract:	We model the effect of a road safety measure on a set of target sites with a control area for each site, and we suppose that the accident data recorded at each site are classified in different mutually exclusive types. In this paper, we propose an MM algorithm for obtaining the constrained maximum likelihood estimates of the parameter vector. We compare it with a GP-EM algorithm, based on gradient projections. The performance of the algorithms is examined through a simulation study of road safety data.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

revisedVersion-IDLSSP-2009-0179.tex

scholarONE Manuscript Central

For Peer Review Only

A MM algorithm for constrained estimation in a road safety measure modelling

Abdallah Mkhadri[1] * Assi N'Guessan[2]† Bezza Hafidi[3]

[1] Département de mathématiques, Université Cadi-Ayyad, Marrakech, Morocco - mkhadri@ucam.ac.ma
[2] Laboratoire Paul Painlevé UMR CNRS 8524 and Ecole Polytechnique, Université de Lille 1, Lille, France - Assi.Nguessan@polytech-lille.fr
[3] Université Inb Zohr, Agadir, Morocco - b.hafidi@ucam.ac.ma

Abstract

We model the effect of a road safety measure on a set of target sites with a control area for each site, and we suppose that the accident data recorded at each site are classified in different mutually exclusive types. We adopt the before-after technique and we assume that at any one target site the total number of accidents recorded is multinomially distibuted between the periods and types of accidents. In this paper, we propose an MM algorithm for obtaining the constrained maximum likelihood estimates of the parameter vector. We compare it with a GP-EM algorithm, based on gradient projections. The performance of the algorithms is examined through a simulation study of road safety data.

Keys Words Road Safety Measure, Multinomial model, Constrained Maximum Likelihood, MM algorithm, GP-EM algorithm, Standard errors, Monte Carlo Simulation.

1 Introduction

Most statistical studies involving data collecting (random phenomenon modelling, experiment planning, opinion poll, etc) not only bring out the problems of parameter estimation (looking for optimal solutions) but also those related to the evaluation of the accuracy of those estimations. In many of the statistical evaluation problems, and particularly in multivariate statistics, interest parameters are not functionally independent, which means there are relations - constraints - between them (Aitchison and Silvey 1958, Crowder 1984). These constraints add further difficulties to bring out the solutions (estimations) and their accuracy (standard errors). But, sometimes incorporation of constraints leads to better or more acceptable estimates (cf. Jamshidian 2004 for details).

Most problems in statistics involve the optimization of a function such as a likelihood or a sum of squares. EM (*Expectation Maximization*) algorithms (Dempster, Laird and Rubin 1977) are

*correspondance author
†correspondance author

For Peer Review Only

the most useful algorithms for maximum likelihood (ML) estimation since they consistently drive the likelihood uphill by maximizing a simple surrogate function for the loglikelihood. They are essentially based on the notion of missing data principle. An EM algorithm operates by identifying the theoritical complete data space. In its E step, the conditional expectation of the complete data loglikelihood is computed with respect to the observed data. The surrogate function created by the E step is a minorizing function. In the M step, this minorizing function is maximized with respect to the parameters of the underlying model. But, iterative optimization of a surrogate function as exemplified by an EM algorithm does not necessarily require missing data. In fact, every EM algorithm is a special case of the more general class of MM (Minorization Majorization) optimization algorithms (cf. Lange, Hunter and Yang 2000 and Hunter and Lange 2004), which typically exploit convexity rather than missing data principle in minorizing or majorizing an objective function.

On the other hand, Jamshidian (2004) proposed a globally convergent generalized gradient projection algorithm (GP) adapted to any maximum likelihood estimation problem that requires incorporation of linear equality and inequality constraints. Global convergent means that it converges to local maximizer of a non linear functional from almost any starting value (cf. Jamshidian 2004). As a special case, GP algorithm can be applied for solving restricted ML problems where the EM algorithm is applicable. As claimed by the author, GP algorithm is simple to implement and is more efficient than its competitors.

In this paper and for multidimensional statistical modelling of a road safety measure, we propose an MM algorithm which cycles through the components of the vector parameter and updates one component at a time which leads to closeform solutions of the parameters. It is simple to implement without any inversion matrix. Moreover, the inequality constraints are integrated easily. Standard errors are obtained by exploiting quantities readily available by running the algorithm. Its performance is compared with the GP algorithm, which is adapted to this setting, through a simulation study of road safety data.

The paper is organized as follows: Section 2 is devoted to the description of road safety data and problem formulation. A statistical model based on multinomial law depending on an unknown vector of parameters is described. In Section 3, we propose an MM algorithm for restricted maximum likelihood estimation of the vector of parameters. In the same section, a method of estimation of standard errors, based on the code of the MM algorithm, is described. Section 4 gives a brief description of GP algorithm and the method of estimation of its standard errors. Illustrative numerical experiments are presented in Section 5. We also analyse, the performance of the two algorithms via the Mean Square Error (MSE). Concluding remarks in Section 6 end the paper.

2 Modelling road accident data

2.1 Data and problem formulation

The data comes from a road safety measure (crossroad lay-out, surface of a motorway section, etc...) that has been applied to s sites. We suppose that the total accident numbers is known for each

For Peer Review Only

site before and after the measure for significant periods (often equal and generally longer than one year). The problem is: how can one estimate the measure's mean effect for the whole set of s sites, when each site counts $r(r > 1)$ accident types ?

We denote by y_{1jk} (resp. y_{2jk}) the number of type j accidents on site k before (resp. after) the setting up of the measure, n_k the total accident number on site k and $N = \sum_{k=1}^{s} n_k$. In order to take some external factors into account (such as traffic flow, speed limit variation, weather conditions,...), let's also suppose that, for the same periods, we know the data for areas where the measure is not applied but which are linked to the target sites thus enabling to obtain the control coefficient denoted by z_{jk} , $j = 1, \ldots, r; k = 1, \ldots, s$.

Here z_{jk} is the multiplying coefficient which has to be used in the control area linked to site k and for accident type j to convert the accident number in the before period into the accident number in the after period. One then has to simultaneously combine data crash of experimental sites and control areas in order to efficiently estimate the mean effect of the measure and the different global risk of accident relatively to all sites and accident types.

2.2 Statistical model

Different statistical models can be used to model the mean effect of a road safety measure and the global risks of different accident. The model used in this paper has been proposed for modelling such data set by N'Guessan and al. (2001). To formulate it, let's consider the period before (resp. after) the measure and note Y_{1jk} (resp. Y_{2jk}) the random variable (r.v.) giving the type j accident number on target site k. Then, we consider the joint random vector

$$\mathbf{Y}^{(k)} = (Y_{11k}, \cdots, Y_{1rk}, Y_{21k}, \cdots, Y_{2rk}) \qquad k = 1, \ldots, s$$

whose observed value is $\mathbf{y}^{(k)} = (y_{11k}, \cdots, y_{1rk}, y_{21k}, \cdots, y_{2rk})$. We then assume that for a fixed k, the random vector $\mathbf{Y}^{(k)}$ follows the multinomial law:

$$I\!P[\mathbf{Y}^{(k)} = \mathbf{y}^{(k)}] = \left(\frac{n_k!}{\prod_{t=1}^{2} \prod_{j=1}^{r} y_{tjk}!} \right) \prod_{t=1}^{2} \prod_{j=1}^{r} q_{tjk}^{y_{tjk}} \tag{1}$$

where

$$q_{1jk} = \frac{p_{jk}}{1 + \theta_0 z_k}, \quad q_{2jk} = \frac{\theta_0 z_{jk} p_{jk}}{1 + \theta_0 z_k}, \quad j = 1, 2, \cdots, r \tag{2}$$

with θ_0 ($\theta_0 > 0$), the parameter mean effect of the road safety measure for all sites and accident types, the p_{jk} ($0 < p_{jk} < 1$) values being auxiliary parameters linked to control areas and target sites such as

$$\sum_{j=1}^{r} p_{jk} = 1, \qquad z_k = \sum_{j=1}^{r} z_{jk} p_{jk}. \tag{3}$$

For Peer Review Only

Through construction of the pattern, two types of parameters can be spotted out among those linked to s target sites and r accident types: θ_0, the interest parameter, and the set of auxiliary parameters $\mathbf{p}_k = (p_{1k}, p_{2k}, \cdots, p_{rk})^t$, $k = 1, 2, \cdots, s$ where t stands for the transpose.
In this paper we focus on the estimation of the parameter vector $\theta = (\theta_0, \mathbf{p}_1^t, \mathbf{p}_2^t, \ldots, \mathbf{p}_s^t)^t \in \mathbb{R}^{1+sr}$ such that $\sum_{j=1}^r p_{jk} = 1$ for $k = 1, \ldots, s$. Then the log-likelihood function, to one additive constant, is given by

$$\mathcal{L}(\theta)) = \sum_{k=1}^{s}\sum_{j=1}^{r}\{y_{.jk}\ln(p_{jk}) + y_{2jk}\ln(\theta_0) - y_{.jk}\ln(1 + \theta_0 < z_k, \mathbf{p}_k >)\}. \tag{4}$$

where $\theta = (\theta_0, \mathbf{p}_1^t, \mathbf{p}_2^t, \ldots, \mathbf{p}_s^t)^t$, $y_{.jk} = \sum_{i=1}^2 y_{ijk}$ and $< .,. >$ is the classical inner product.
Now, using iterative procedures, the restricted maximum likelihood estimator (RMLE) $\hat{\theta}$ can be obtained (cf. N'Guessan et al. 2001). However, the whole parameter vector is updated at one with an inversion matrix in each iteration of the Newton-Raphson algorithm. In the following Section, we describe two algorithms which are simple to implement. The equality constraints are integrated easily and their standard errors are obtained by exploiting quantities readily available by running the two algorithms. The first one is a new MM algorithm which cycles through the components of the vector of parameters and updates one component at a time which leads to closeform solutions of the parameters. The second one is the generalized gradient projection algorithm (GP) (Jamshidian 2004) adapted to any maximum likelihood estimation problem that requires incorporation of linear equality and inequality constraints.

3 An MM algorithm for constrained estimation

Let $\underline{\alpha} = (\alpha_0, \alpha)^t$, with $\alpha = (\mathbf{p}_1^t, \mathbf{p}_2^t, \ldots, , \mathbf{p}_s^t)$ and $\alpha_0 = \theta_0$. In this section, we propose an MM algorithm for maximizing $\mathcal{L}(\underline{\alpha})$ (Eq. (4)) under the same constraints

$$\text{(A1)} \qquad h_k(\underline{\alpha}) = 0 \quad k = 1, 2, \ldots, s,$$

where functions $h_k : \mathbb{R}^{1+sr} \mapsto \mathbb{R}$ are given by:

$$\text{(A2)} \qquad h_k(\alpha) =< \mathbf{1}_r, \alpha_k > -1$$

with $\mathbf{1}_r = (1, \ldots, 1)^t \in \mathbb{R}^r$.

3.1 Main results

Let $\underline{\alpha}^{(m)}$ represent a fixed value of the vector parameter $\underline{\alpha}$, and let $g_1(\underline{\alpha}|\underline{\alpha}^{(m)})$ denote a real-valued function of $\underline{\alpha}$, whose form depends on $\underline{\alpha}^{(m)}$, defined by

$$g_1(\underline{\alpha}|\underline{\alpha}^{(m)}) = cte + \sum_{k=1}^{s}\sum_{j=1}^{r}\{y_{.jk}\ln(\alpha_{jk}) + y_{2jk}\ln(\alpha_0) - y_{.jk}\ln(1 + \alpha_0^{(m)} < z_k, \alpha_k^{(m)} >)\} \tag{5}$$

$$- \sum_{k=1}^{s} y_{..k}\frac{1 + \alpha_0 < z_k, \alpha_k >}{1 + \alpha_0^{(m)} < z_k, \alpha_k^{(m)} >} + y_{...}, \tag{6}$$

For Peer Review Only

where $y_{...} = \sum_{t=1}^{2}\sum_{j=1}^{r}\sum_{k=1}^{s} y_{tjk}$. The following lemma states that $g_1(\underline{\alpha}|\underline{\alpha}^{(m)})$ is a minorizing function of $\mathcal{L}(\underline{\alpha})$.

Lemma 3.1 *For any vector parameter $\underline{\alpha}$, we have*

$$g_1(\underline{\alpha}|\underline{\alpha}^{(m)}) \leq \mathcal{L}(\underline{\alpha}) \quad \textit{and} \quad g_1(\underline{\alpha}^{(m)}|\underline{\alpha}^{(m)}) = \mathcal{L}(\underline{\alpha}^{(m)}).$$

The proof of this lemma is based on the simple application of the convexity of the function $(-\ln)$ (cf. Mkhadri and N'Guessan, 2008). Now, we will introduce the constraints $\alpha_0 > 0, \alpha_{jk} \geq 0$ and $\sum_{j=1}^{r}\alpha_{jk} = 1$. But, we first begin by introducing the non negative constraints (i) $\alpha_0 > 0$ and (ii) $\alpha_{jk} \geq 0$ for $j = 1, \ldots, r$ and $k = 1, \ldots, s$. The maximization of $g_1(\underline{\alpha}|\underline{\alpha}^{(m)})$ under constraints (i) and (ii) is equivalent to the minimization of $-g_1(\underline{\alpha}|\underline{\alpha}^{(m)})$ subject to the same constraints. The lemma below gives us a new majorizing function of $-\mathcal{L}(\underline{\alpha})$ to be minimized under the constraints (i) and (ii).

Lemma 3.2 *Let w be a non negative real scalar $(w > 0)$ and $\alpha_{0k} = \alpha_0$, the new function*

$$g_2(\underline{\alpha}|\underline{\alpha}^{(m)}) = -g_1(\underline{\alpha}|\underline{\alpha}^{(m)}) + w\sum_{k=1}^{s}\sum_{j=0}^{r}\{\alpha_{jk}^{(m)} ln\frac{\alpha_{jk}^{(m)}}{\alpha_{jk}} + (\alpha_{jk} - \alpha_{jk}^{(m)})\} \tag{7}$$

satisfies : $g_2(\underline{\alpha}|\underline{\alpha}^{(m)}) \geq -g_1(\underline{\alpha}|\underline{\alpha}^{(m)})$ *and* $g_2(\underline{\alpha}^{(m)}|\underline{\alpha}^{(m)}) = -g_1(\underline{\alpha}^{(m)}|\underline{\alpha}^{(m)})$.

Proof. The proof is based on the arguments in Hunter and Lange (2004, p.21) for handling the constraints in MM algorithms, which we recall here. Consider the problem of minimizing a function $f(\theta)$ subject to the constraints $v_i(\theta) \geq 0$ for $1 \leq i \leq q$, where each $v_i(\theta)$ is a concave and differentiable function. Since $-v_i(\theta)$ is convex, then

$$-v_i(\theta) \geq -v_i(\theta^{(m)}) + \nabla_{v_i}(\theta^{(m)})^t(\theta^{(m)} - \theta),$$

where t stands for the transpose. The latter is based on the fact that any linear function tangent to the graph of a convex function is a minimizer at the point of tangency. Then, application of the similar inequality $-\ln(y) + \ln(x) \geq x^{-1}(x - y)$, for $x > 0$ and $y > 0$, implies that

$$v_i(\theta^{(m)})[-\ln v_i(\theta) + \ln v_i(\theta^{(m)})] \geq v_i(\theta^{(m)}) - v_i(\theta).$$

Adding the last two inequalities, we can see that

$$v_i(\theta^{(m)})[-\ln v_i(\theta) + \ln v_i(\theta^{(m)})] + \nabla_{v_i}(\theta^{(m)t}(\theta - \theta^{(m)}) \geq 0,$$

with equality when $\theta = \theta^{(m)}$. Summing over j and multiply by a positive tuning parameter w, the function

$$h(\theta|\theta^{(m)}) = f(\theta) + w\sum_{i=1}^{q}[v_i(\theta^{(m)})\frac{v_i(\theta^{(m)})}{v_i(\theta)} + (\theta - \theta^{(m)})^t\nabla_{v_i}(\theta^{(m)})]$$

majorizes $f(\theta)$ at $\theta^{(m)}$. Then, the majorizing function, at $\theta^{(m)}$, of $-g_1(\underline{\alpha}|\underline{\alpha}^{(m)})$ subject to the $(rs+1)$ inequality constraints $v_0(\underline{\alpha}) = \alpha_0 > 0$ and $v_i(\underline{\alpha}) = \alpha_{jk} \geq 0$, for $i = 1, \ldots, rs$ $(j = 1, \ldots, r; k = 1, \ldots, s)$ is defined by

$$g_2(\underline{\alpha}|\underline{\alpha}^{(m)}) = -g_1(\underline{\alpha}|\underline{\alpha}^{(m)}) + w\sum_{k=1}^{s}\sum_{j=0}^{r}\{\alpha_{jk}^{(m)}\ln\frac{\alpha_{jk}^{(m)}}{\alpha_{jk}} + (\alpha_{jk} - \alpha_{jk}^{(m)})\},$$

For Peer Review Only

with $\alpha_{0k}^{(m)} = \alpha_0^{(m)}$ by convention, which ends the proof. ■

We will now minimize $g_2(\underline{\alpha}|\underline{\alpha}^{(m)})$ while enforcing the constraint $\sum_{j=1}^{r} \alpha_{jk} = 1$, for $k = 1, \ldots, s$, by introducing a Lagrange multiplier and looking for a stationary point of the Lagrangian

$$g_2(\underline{\alpha}|\underline{\alpha}^{(m)}) + \sum_{k=1}^{s} \lambda_k (\sum_{j=1}^{r} \alpha_{jk} - 1). \tag{8}$$

The next theorem outlines the closeform solutions of the minimization problem of the latter function (8).

Theorem 3.3 *Let's assume that all the components of the vector parameter $\underline{\alpha}$, at the (m)th iteration, are remaining fixed and equal $\underline{\alpha}^{(m)}$, then*

i) the $(m+1)$th iterate of α_0 is given

$$\alpha_0^{(m+1)} = \frac{y_{2..} + ws\alpha_0^{(m)}}{ws + \sum_{k=1}^{s} a_k^{(m)} y_{..k} < z_k, \alpha_k^{(m)} >}. \tag{9}$$

ii) the $(m+1)$th iteration of the jk component of α ($j = 1, \ldots, r$ and $k = 1, \ldots, s$) is defined by

$$\alpha_{jk}^{(m+1)} = \frac{y_{.jk} + w\alpha_{jk}^{(m)}}{w + y_{..k} + a_k^{(m)} y_{..k} \alpha_0^{(m+1)} [z_{jk} - < z_k, \alpha_k^{(m)} >]}, \tag{10}$$

where

$$a_k^{(m)} = \frac{1}{1 + \alpha_0^{(m)} < z_k, \alpha_k^{(m)} >}.$$

Proof. Computation of the score functions of the function (8) with respect to the parameters α_0 and α_{jk}, for $j = 1, \ldots, r; k = 1, \ldots, s$, respectively, and equalizing to zero leads to the following system equations

$$\begin{aligned} \frac{-y_{2..}}{\alpha_0} + \sum_{k=1}^{s} a_k^{(m)} y_{..k} < z_k, \alpha_k^{(m)} > + w \sum_{k=1}^{s} \{1 - \frac{\alpha_{0k}^{(m)}}{\alpha_{0k}}\} &= 0 \\ \frac{-y_{.jk}}{\alpha_{jk}} + a_k^{(m)} y_{..k} \alpha_0 z_{jk} + w(1 - \frac{\alpha_{jk}^{(m)}}{\alpha_{jk}}) + \lambda_k &= 0. \end{aligned}$$

Now, putting by convention $\alpha_{0k} = \alpha_0$ and $\alpha_{0k}^{(m)} = \alpha_0^{(m)}$ in the first equation, then the result of equation (9) follows easily. The second equation can be written as

$$-y_{.jk} + \alpha_{jk}[a_k^{(m)} y_{..k} \alpha_0 z_{jk} + \lambda_k + w] - w\alpha_{jk}^{(m)} = 0.$$

Replacing α_0 and α_{jk} by $\alpha_0^{(m)}$ and $\alpha_{jk}^{(m)}$ respectively, and summing on j reveals that

$$\lambda_k = +y_{..k} - a_k^{(m)} y_{..k} \alpha_0^{(m+1)} < z_k, \alpha_k^{(m)} >,$$

and yields the update (10). ■

Remark 1 Since the $1 + sr$ components are computed one by one and in closeform, it makes the MM algorithm easy to implement. So our MM algorithm is a so-called Cyclic MM algorithm (cf. Hunter

For Peer Review Only

and Lange 2004) which cycles through the parameters updating one at a time instead of updating the whole vector at once as in N'Guessan et al. (2001). Moreover, a cyclic MM algorithm always drives the objective function in the right direction; indeed, every iteration of a cyclic MM algorithm is simply an MM iteration on a reduced parameter set (cf. Hunter and Lange 2004).

From equations (9) and (10) in the theorem, it is easy to see that all the components of the parameter $\underline{\alpha}$ are positives for any fixed positive value of tuning parameter w. The boundness of the components of α by 1 are guaranted if each component $y_{.jk}$ ($j = 1, \ldots, r$ and $k = 1, \ldots, s$) satisfies a supplementary inequality condition which seems to be often satisfied in practice. The following corollary summarizes these results.

Corollary 3.4 *i) At the (m) step of the MM algorithm and for all $w > 0$, we have $\alpha_0^{(m)} > 0$ and $\alpha_{jk}^{(m)} \geq 0$, for all $j = 1, \ldots, r; k = 1, \ldots, s$.*
ii) If we assume that $0 < \alpha_{jk}^{(m)} < 1$ and $y_{.jk} \leq y_{..k} + a_k^{(m)} y_{..k} \alpha_0^{(m+1)} [z_{jk} - < z_k, \alpha_k^{(m)} >]$ for all $j = 1, \ldots, r; k = 1, \ldots, s$, then $0 < \alpha_{jk}^{(m+1)} \leq 1$ for all $j = 1, \ldots, r; k = 1, \ldots, s$,.

Proof. *i)* The result is obvious from the equations (9) and (10), since all terms of these equations are positive.
ii) Let $0 < \alpha_{jk}^{(m)} < 1$ and $y_{.jk} \leq y_{..k} + a_k^{(m)} y_{..k} \alpha_0^{(m+1)} [z_{jk} - < z_k, \alpha_k^{(m)} >]$ for all $j = 1, \ldots, r; k = 1, \ldots, s$, then it's easy to show from equation (10) that $0 < \alpha_{jk}^{(m+1)} \leq 1$ for all $j = 1, \ldots, r; k = 1, \ldots, s$. ■

Remark 2 The latter inequality condition on $y_{.jk}$ ($j = 1, \ldots, r; k = 1, \ldots, s$) is difficult to establish analytically, since it depends on the values of the observations $(y_{.jk}, z_{jk})$ for $j = 1, \ldots, r$ and $k = 1, \ldots, s$. But, it seems to be often satisfied in practice as shown in our experimental study.

3.2 Standard error estimates

Besides computing point estimates, statistical inference requires measures of uncertainty, for example (asymptotic) variance-covariance matrix of the estimates which is equal to the inverse of the expected information matrix. In practice, the expected information matrix is well-approximated by the observed information matrix $-\nabla^2 \mathcal{L}(\underline{\alpha}) = H(\underline{\alpha})$ computed by differentiating the loglikelihood $\mathcal{L}(\underline{\alpha})$ twice. Thus, a standard error of the maximum likelihood estimator (MLE) $\hat{\underline{\alpha}}$ can be obtained by taking square roots of the diagonal terms of the inverse of $H(\hat{\underline{\alpha}})$. We can also use direct calculation of the information matrix via EM algorithm (see for instance Oakes, 1999). But,in some problems, however, direct computation of $\nabla^2 \mathcal{L}(\underline{\alpha})$ is difficult.
Estimating standard errors of $\underline{\alpha}$ when applying the EM algorithm appears in Jamshidian and Jennrich (2000). Two approaches, called NDS and NDM, are often considered in this context. The first numerically differentiates the Fisher score vector at MLE $\hat{\underline{\alpha}}$ to give the observed information matrix $H(\hat{\underline{\alpha}})$. The second numerically differentiates the EM operator $M(\underline{\alpha})$ and uses an identity that relates the Jacobian of $M(\underline{\alpha})$ at $\hat{\underline{\alpha}}$. A specific NDM algorithm called the SEM (supplemented EM) algorithm was introduced by Meng and Rubin (1991). To overcome some problems of SEM algorithm when the number of parameters is large, two alternative NDM algorithms called FDM

For Peer Review Only

and REM have been recently proposed by Jamshidian and Jennrich (2000) and are based on the forward difference and Richardson extrapolation methods of numerical differentiation. They also proposed a Richardson extrapolation implementation of the NDS approach and they called it RES. On the other hand, Hunter and Lange (2004) described a procedure similar to SEM algorithm for the computation of standard errors of the MM algorithm, we call it SuppMM. But, for many problems, SuppMM can be as numerically instable as SEM. So, we adapt here the three alternative NDM and NDS algorithms for obtaining standard errors estimates of $\hat{\underline{\alpha}}$. We call them hereunder MM-FDM, MM-REM and MM-RES, respectively. A detailed comparison of these four algorithms, based on a simulation study of road safety data, is considered in the next Section.

4 Gradient projection algorithm for linear equality constraints

An alternative to MM algorithm is the generalized gradient (GP) algorithm proposed by Jamshidian (2004) which is adapted to any restricted maximum likelihood problem. A brief description of the latter algorithm is described here. Consider the following restricetd ML problem

$$max_{\theta\in\Omega}\ell(\theta) \quad \text{where} \quad \Omega = \{\theta \in \mathbb{R} : A\theta = \mathbf{b}\}, \tag{11}$$

where A ia an m by p $(m < p)$ matrix of rank m, and $\ell(\theta)$ is the log-likelihood function which is supposed to be sufficiently smooth. Let W denote a positive definite matrix, $g(\theta)$ the gradient of $\ell(\theta)$. The generalized gradient of ℓ in the metric W is given by $\tilde{\mathbf{g}}(\theta_r) = W^{-1}g(\theta_r)$. Given a feasible initial value $\theta_r \in \Omega$, the GP algorithm cycles through the following steps until it converges to $\hat{\theta}$:
Step 1: Compute $\mathbf{d} = P_W\tilde{\mathbf{g}}(\theta_r)$, where $P_W = I - W^{-1}A^t(AW^{-1}A^t)^{-1}A$ and I is the identity matrix. If $\mathbf{d} = 0$, stop and declare convergence.
Step 2: Obtain a new point $\tilde{\theta}_r = \theta_r + \alpha\mathbf{d}$ by choosing $\alpha = argmax_\alpha\ell(\theta_r + \alpha\mathbf{d})$ (or alternatively a smallest integer $k \geq 0$ such that $\ell(\theta_r + (1/2)^k\mathbf{d}) > \ell(\theta_r)$.
Step 3: Replace θ_r by $\tilde{\theta}_r$ and go to Step 1.
It is shown that $\mathbf{d}$ is an ascent and feasible direction. Thus, it is guaranteed that a small enough step from θ_r in the direction of $\mathbf{d}$ results in a new feasible point θ_r such that $\ell(\tilde{\theta}_r) > \ell(\theta_r)$.
Jamshidian(2004) argued that, generally, the GP algorithm is superior to expectation-restricted-maximization (ERM), used to deal with constraints in the EM algorithm setting, in terms of simplicity of implementation and time to converge. Standard error estimates (called SE-GP) are obtained by taking square roots of the diagonal terms of the inverse of $H(\hat{\underline{\alpha}})$. In the following, numerical experiments are given to compare the finite-sample performance of MM and GP algorithms.

5 Numerical simulated studies

5.1 Data simulation Principle

In this section, we report on some numerical studies for analyzing the practical behavior of MM and GP algorithms for the estimation of the parameter vector $\theta = (\theta_0, \mathbf{p}_1^t, \mathbf{p}_2^t, \ldots, \mathbf{p}_s^t)^t$. Moreover,

For Peer Review Only

a comparison between three methods for the estimation of standard errors, following numerical differentiation methods of Section 3 which use the code of the two algorithms, is presented. Our numerical studies are based on Monte Carlo numerical experiments on simulated data sets of Target sites and control areas.

Given s (the number of sites) and r (the number of accidents types), we generate the coefficients z_{jk} $(j = 1, \ldots, r; k = 1, \ldots, s)$ from a uniform random variable $U_{]\frac{1}{2},\frac{5}{2}[}$. We pose the true valeur of θ_0 as the mean of a uniform random variable $U_{[0,1]}$ and suppose that the true value of vector $\mathbf{p}_k$, noted $\mathbf{p}_k^0 = (p_{1k}^0, \ldots, p_{rk}^0)^t$ such that $\sum_{j=1}^r p_{jk}^0 = 1; k = 1, \ldots, s$, comes from a uniform random variable $U_{[\alpha,1-\alpha]}$ (with $\alpha = 10^{-5}$). Using those values, we then define the true mean control coefficients

$$z_k^0 = \sum_{j=1}^r z_{jk} p_{jk}^0, \qquad (k = 1, \ldots, s),$$

and then the true probabilities

$$q_{1jk}^0 = \frac{p_{jk}^0}{1 + \theta^0 z_k^0}, \quad q_{2jk}^0 = \frac{z_{jk}\theta^0 p_{jk}^0}{1 + \theta^0 z_k^0}, \quad (k = 1, \ldots, r)$$

linked to the multinomial distribution of $r.v.$ $\mathbf{Y}^{(k)}$.

Finally, for a set of values of k ($k = 1, \ldots, s$), one generates the total number n_k of site k and then randomly shares n_k between the before and after periods using probabilities q_{1jk}^0 and q_{2jk}^0. The observed values of y_{tjk}, with $\sum_{t=1}^2 \sum_{j=1}^r y_{tjk} = n_k, k = 1, \ldots, s$, are then found. For our numerical experiments, several tables of differents sizes and with different numbers have been generated by setting the values of $s = 4$ and $r = 3$ and by choosing n_k in different accident brackets $B_1, \ldots, B_7$:

$$\{]0, 30[,]30, 50[,]300, 1000[,]3000, 5000[,]7000, 9000[,]9000, 11000[,]11000, 13000[\}.$$

These data are modelled using the multinomial model (1) with unknown vector parameters $\theta = (\theta_0, \mathbf{p}_1^t, \mathbf{p}_2^t, \ldots, \mathbf{p}_s^t)^t$. The model parameters are then estimated by the MM and GP algorithms defined via the equations (9), (10) and (11) respectively.

5.2 Numerical comparison of MM and GP algorithms

In order to compare the two algorithms for the estimatiom the true value of the parameters' vector, we have used the different tables of simulated accident data. The results presented here only concern the particular case when $s = 4$, $r = 3$ and n_k belonging to each of the seven accident brackets. The variation of this value of n_k enables to compare both algorithms at one time for small and large values of accident numbers.

==========

Tables 1 and 2 go here

==========

Table 1 (resp. 2) shows the parameter vector's estimation with the MM algorithm (resp. GP) for s=4 and r=3. For Table 1, for example, the values which are under column θ_0 correspond to the

For Peer Review Only

true value of the parameters whereas those under column B_1 to B_7 represent the estimation with the MM algorithm. Same thing for Table 2 which is relative to the GP algorithm.

For all obtained simulations, and especially in the example of Tables 1 and 2, we reach the fact that the estimation values given by the MM algorithm (column B_1 to B_7 of Table 1) are much closer to the true parameter value than those coming from the GP algorithm (column B_1 to B_7 of Table 2). This proximity to the true parameter value of the solutions of the MM algorithm is even more important when number n_k varies from bracket B_1 (small accident number) to bracket B_7 (high accident number). In particular, we note that the first component of vector $\hat{\theta}$ numerically converges towards the true value for the MM algorithm whereas it remains fairly far from it for the GP algorithm. The same remark is valuable for the other components of the parameters' vector in the case of an upward progression in the accident brackets.

==========

Tables 3 to 5 go here

==========

Tables 3 to 5 represent the standard errors associated to each component and obtained through the estimation methods described in section 3.2. In relation with the estimation of the standard errors associated to each of the components of the parameters' vector, we observe, through the carried out simulations, that the values obtained with the MM-RES method are on the whole lower with a more stable order of magnitude from one accident bracket to another.

Excepting bracket B_1 where the approximation methods give standard errors the order of magnitude of which varies between 10^{-1} and 10^{-2}, we can note, as early as bracket B_6, if not B_2, that the order of magnitude of the MM-RES becomes stable at 10^{-3} whereas the other approximation methods have an order of magnitude whose value goes on varying between 10^{-2} and 10^{-3}. If we only consider the higher numbers of accidents, i.e. accident numbers belonging to bracket B_4 and up, the simulations show that the estimated values of the standard errors with the SuppMM and MM-REM methods are similar (results not reported here). Nevertheless, these values remain, for the most part, higher than those obtained through the GP and MM-RES methods and smaller than those obtained through the MM-FDM (results not reported here).

==========

Table 6 goes here

==========

In parallel with the comparison of the estimated values $\hat{\theta}$, we have assessed the two algorithms in terms of iteration number and CPU times. Table 6 summarizes our comparison of iteration number and CPU times of the MM and GP algorithms on 50 simulations in each accident bracket. The mean number of iterations for both algorithms are about the same in all accident brackets. While for the mean CPU times, the MM algorithm is surprisingly on average faster than the GP algorithm.

For Peer Review Only

In addition to the numerical comparison, we also analyse the convergence of the solutions provided from the two algorithms. For the convergence, we use the mean square error (MSE) defined by:

$$MSE(\hat{\theta}, \theta^0) = (1 + sr)^{-1} \sum_{m=1}^{1+sr} (\hat{\theta}_m - \theta_m^0)^2.$$

To achieve this, we have simulated 1000 accident data sets per bracket using $s = 4$ and $r = 3$. We then get 1000 estimates for each of the two algorithms.

==========

Figure 1 goes here

==========

Figure 1 represents the cases of small accident's numbers. A similar figure (not reported here) is obtained with high accident's numbers. The Figure 1 shows that the two algorithms are of the same kind. They also provide solutions which converge to the true vector value when the total accident's number per experimental site becomes large. Whatever the algorithm may be, the order of magnitude of the $MSE(\hat{\theta}, \theta^0)$ goes from 10^{-2} to 10^{-5} when the total accident's number per experimental site passes from bracket B_1 to bracket B_7. The convergence's results of the two algorithms are similar to those obtained by the classic Newton-Raphson (N'Guessan et al 2001).

6 Conclusion

We presented an MM algorithm for the estimation of the parameter vector of the constrained maximum likelihood function of the multinomial before-after law modelling the effect of a road safety measure. Generally, iterative algorithms, like Newton-Raphson, are used for obtaining the constrained maximum likelihood estimates of the parameter vector (cf N'Guessan et al. 2001). However, the whole parameter vector is updated at one with an inversion matrix in each iteration of the Newton-Raphson algorithm. Instead, our MM algorithm cycles through the components of the parameter vector and updates one component at a time. It is simple to program without any inversion matrix and it leads to closeform solutions of the parameters. Moreover, the inequality constraints are integrated easily. Standard errors are obtained by exploiting quantities readily available by running the algorithm.

On the other hand, Jamshidian (2004) proposed a globally convergent generalized gradient projection algorithm (GP) adapted to any maximum likelihood estimation problem that requires incorporation of linear equality and inequality constraints. Global convergent means that it converges to a local maximizer of a nonlinear functional from almost any starting value (cf. Jamshidian 2004). As a special case, GP algorithm was applied for solving restricted ML problem on road safety measure. The performance of the two algorithm is examined through a simulation study in the cases of 4 sites and 3 types of accidents. Limited comparisons suggest that the MM algorithm is competitive in statistical accurancy and computational speed with the best currently available algorithm. It

For Peer Review Only

is shown that the MM estimates of the vector parameter are relatively close to those obtained by N'Guessan et al (2001), while the corresponding standard error estimates obtained by MM-RES algorithm are generally very small.

Acknowledgements

The authors thank the Editor and the referee for their helpful suggestions. This article was partially written while the first author was visiting Ecole Polytechnique Universitaire de Lille 1. The hospitality of the school is gratefully acknowledged.

References

Aitchison, J. and Silvey, S. D. (1958). Maximum Likelihood Estimation of parameters subject to restraints. *Annals of mathematical statistics*, 29,813-829.

Crowder M. (1984). On the constrained maximum likelihood estimation with non i.i.d observations *Ann. Inst. Statist. Math*, **36**, A, 239-249.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation for incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society.*, Series B **39**, 1-38.

Hunter, D. R. and Lange, K. (2004). A tutorial on MM algorithms. *The American Statisistician*, **58**, 30-37.

Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. *Computational Statistics and Data Analysis*, **45**, 137-157.

Jamshidian, M. and Jennrich, R. I. (2000). Standards Errors for EM Estmation. *Journal of the Royal Statistical Society,* Ser. B., 62, 257-270.

Lange, K., Hunter, D. R. and Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussion). *J. Comput. Graphical Statist.*, **9**, 1-20.

Meng,X-L. and Rubin, D. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *J. Amer. Statit. Ass.*, **86**, 899-909.

Mkhadri A., N'Guessan, A. (2008). On the Use of the MM algorithm for constrained estimation in a road safety measure modelling. *Technical Report, Publications de l'Institut de Recherche en Mathématiques Avancées, (IRMA)* **68** (III).

N'Guessan, A., Essai, A. and Langrand, C. (2001). Estimation multidimensionnelle des contrôles et de l'effet moyen d'une mesure de sécurité routière. *Rev. Statistique Appliquée,* XLIX(2), 83-100.

Oakes, D. (1999). Direct calculation of the information matrix via the EM Algorithm. *J. Roy. Statist. Soc. B*, **61**, 479-482.

For Peer Review Only

Table 1: Value of $\hat{\theta}$ for $s = 4$ and $r = 3$ obtained by MM algorithm.

θ^0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
5.000e-01	4.427e-01	4.427e-01	5.217e-01	4.933e-01	5.006e-01	4.987e-01	5.015e-01
3.413e-01	3.413e-01	3.205e-01	3.794e-01	3.373e-01	3.439e-01	3.368e-01	3.288e-01
2.469e-01	2.469e-01	3.051e-01	2.626e-01	2.400e-01	2.459e-01	2.557e-01	2.508e-01
4.118e-01	4.118e-01	3.745e-01	3.580e-01	4.227e-01	4.102e-01	4.075e-01	4.204e-01
1.699e-01	1.699e-01	2.674e-01	1.716e-01	1.693e-01	1.682e-01	1.738e-01	1.688e-01
3.760e-01	3.760e-01	2.280e-01	3.793e-01	3.717e-01	3.795e-01	3.737e-01	3.736e-01
4.540e-01	4.540e-01	5.045e-01	4.491e-01	4.590e-01	4.523e-01	4.525e-01	4.576e-01
4.636e-01	4.636e-01	4.937e-01	4.623e-01	4.614e-01	4.674e-01	4.589e-01	4.666e-01
8.520e-02	8.520e-02	2.520e-02	9.220e-02	8.330e-02	8.190e-02	8.380e-02	8.450e-02
4.512e-01	4.512e-01	4.811e-01	4.455e-01	4.553e-01	4.507e-01	4.573e-01	4.489e-01
4.954e-01	4.954e-01	5.924e-01	4.932e-01	4.967e-01	5.016e-01	4.943e-01	4.976e-01
2.553e-01	2.553e-01	1.065e-01	2.603e-01	2.578e-01	2.469e-01	2.618e-01	2.586e-01
2.493e-01	2.493e-01	3.011e-01	2.465e-01	2.455e-01	2.515e-01	2.439e-01	2.437e-01

Table 2: Value of $\hat{\theta}$ for $s = 4$ and $r = 3$ obtained by GP algorithm.

θ^0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
5.000e-01	5.534e-01	4.922e-01	6.281e-01	5.734e-01	5.645e-01	5.719e-01	5.791e-01
3.413e-01	3.325e-01	3.427e-01	3.242e-01	3.363e-01	3.322e-01	3.331e-01	3.331e-01
2.469e-01	2.436e-01	2.481e-01	2.479e-01	2.542e-01	2.433e-01	2.462e-01	2.478e-01
4.118e-01	4.022e-01	4.136e-01	3.977e-01	4.059e-01	4.000e-01	4.016e-01	4.012e-01
1.699e-01	1.660e-01	1.706e-01	1.753e-01	1.638e-01	1.642e-01	1.645e-01	1.634e-01
3.760e-01	3.654e-01	3.757e-01	3.908e-01	3.688e-01	3.690e-01	3.703e-01	3.684e-01
4.540e-01	4.434e-01	4.549e-01	4.600e-01	4.420e-01	4.433e-01	4.438e-01	4.413e-01
4.636e-01	4.539e-01	4.646e-01	4.436e-01	4.549e-01	4.577e-01	4.562e-01	4.555e-01
8.520e-02	1.662e-01	7.500e-02	1.871e-01	1.561e-01	1.682e-01	1.644e-01	1.754e-01
4.512e-01	4.438e-01	4.521e-01	4.315e-01	4.421e-01	4.440e-01	4.427e-01	4.419e-01
4.954e-01	4.902e-01	4.961e-01	4.790e-01	4.884e-01	4.892e-01	4.887e-01	4.875e-01
2.553e-01	2.527e-01	2.557e-01	2.504e-01	2.521e-01	2.523e-01	2.521e-01	2.517e-01
2.493e-01	2.399e-01	2.508e-01	2.124e-01	2.353e-01	2.364e-01	2.361e-01	2.327e-01

For Peer Review Only

Table 3: Standard error estimates (SEE) using SuppMM algorithm.

θ^0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
5.000e-01	1.948e-01	9.720e-02	3.460e-02	1.110e-02	7.900e-03	6.700e-03	6.500e-03
3.413e-01	1.194e-01	7.580e-02	2.740e-02	1.070e-02	6.300e-03	5.600e-03	5.400e-03
2.469e-01	9.250e-02	7.850e-02	2.440e-02	9.600e-03	5.600e-03	5.200e-03	5.000e-03
4.118e-01	1.065e-01	8.250e-02	2.660e-02	1.210e-02	6.900e-03	6.200e-03	6.200e-03
1.699e-01	4.220e-02	9.420e-02	2.210e-02	6.300e-03	4.400e-03	3.900e-03	3.600e-03
3.760e-01	1.391e-01	9.240e-02	3.560e-02	1.010e-02	7.100e-03	6.100e-03	5.700e-03
4.540e-01	1.074e-01	1.351e-01	3.760e-02	1.100e-02	7.600e-03	6.600e-03	6.200e-03
4.636e-01	7.580e-02	1.210e-01	2.960e-02	1.170e-02	8.800e-03	7.100e-03	6.900e-03
8.520e-02	2.760e-02	2.470e-02	1.160e-02	4.400e-03	3.300e-03	2.700e-03	2.600e-03
4.512e-01	6.670e-02	1.106e-01	2.640e-02	1.080e-02	7.900e-03	6.500e-03	6.200e-03
4.954e-01	1.897e-01	1.331e-01	2.670e-02	1.290e-02	8.600e-03	7.700e-03	6.700e-03
2.553e-01	1.288e-01	5.240e-02	1.750e-02	8.600e-03	5.600e-03	5.200e-03	4.500e-03
2.493e-01	1.508e-01	8.700e-02	1.640e-02	8.200e-03	5.500e-03	4.900e-03	4.200e-03

Table 4: Standard error estimates (SEE) using RES algorithm.

θ^0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
5.000e-01	7.610e-02	5.080e-02	1.610e-02	5.900e-03	4.100e-03	3.500e-03	3.300e-03
3.413e-01	9.560e-02	5.580e-02	2.030e-02	7.600e-03	4.500e-03	4.000e-03	3.900e-03
2.469e-01	7.220e-02	5.710e-02	1.780e-02	6.800e-03	4.000e-03	3.700e-03	3.600e-03
4.118e-01	8.580e-02	6.080e-02	1.970e-02	8.600e-03	5.000e-03	4.500e-03	4.500e-03
1.699e-01	3.460e-02	6.770e-02	1.590e-02	4.500e-03	3.100e-03	2.800e-03	2.600e-03
3.760e-01	1.077e-01	6.620e-02	2.550e-02	7.200e-03	5.100e-03	4.400e-03	4.100e-03
4.540e-01	8.560e-02	9.680e-02	2.700e-02	7.800e-03	5.400e-03	4.700e-03	4.400e-03
4.636e-01	1.133e-01	8.760e-02	2.180e-02	8.300e-03	6.300e-03	5.000e-03	4.900e-03
8.520e-02	2.470e-02	1.790e-02	8.800e-03	3.200e-03	2.400e-03	1.900e-03	1.900e-03
4.512e-01	8.670e-02	8.090e-02	1.990e-02	7.700e-03	5.700e-03	4.700e-03	4.500e-03
4.954e-01	1.117e-01	9.820e-02	2.010e-02	9.200e-03	6.200e-03	5.500e-03	4.800e-03
2.553e-01	8.380e-02	3.810e-02	1.350e-02	6.200e-03	4.100e-03	3.700e-03	3.300e-03
2.493e-01	9.430e-02	6.400e-02	1.280e-02	5.900e-03	4.000e-03	3.500e-03	3.100e-03

For Peer Review Only

Table 5: Standard error estimates (SEE)using GP algorithm.

θ^0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
5.000e-01	9.700e-02	7.410e-02	5.070e-02	2.480e-02	1.300e-02	1.300e-02	1.120e-02
3.413e-01	1.040e-01	8.490e-02	2.510e-02	1.150e-02	6.400e-03	6.000e-03	5.900e-03
2.469e-01	1.719e-01	1.006e-01	3.980e-02	1.600e-02	9.100e-03	8.100e-03	8.000e-03
4.118e-01	1.607e-01	1.065e-01	3.650e-02	1.400e-02	8.200e-03	7.500e-03	7.100e-03
1.699e-01	1.887e-01	8.940e-02	3.410e-02	1.060e-02	6.800e-03	6.000e-03	5.700e-03
3.760e-01	1.206e-01	1.929e-01	4.520e-02	1.280e-02	8.700e-03	7.700e-03	7.100e-03
4.540e-01	1.620e-01	1.364e-01	4.250e-02	1.270e-02	8.800e-03	7.700e-03	7.100e-03
4.636e-01	1.105e-01	1.034e-01	2.690e-02	1.070e-02	7.700e-03	6.400e-03	6.100e-03
8.520e-02	1.351e-01	1.605e-01	1.488e-01	8.760e-02	5.460e-02	4.790e-02	4.210e-02
4.512e-01	7.510e-02	5.850e-02	1.540e-02	5.900e-03	4.400e-03	3.600e-03	3.500e-03
4.954e-01	6.370e-02	3.860e-02	9.700e-03	4.200e-03	2.800e-03	2.500e-03	2.200e-03
2.553e-01	7.690e-02	3.810e-02	1.450e-02	5.500e-03	4.100e-03	3.400e-03	3.300e-03
2.493e-01	9.670e-02	7.880e-02	1.740e-02	9.700e-03	6.000e-03	5.600e-03	4.800e-03

Table 6: Means number of iterations and CPU times required for MM and GP algorithms

	MM algorithm		GP algorithm	
	Numb itera	CPU	Numb itera	CPU
B_1	2.96	0.0109	3	0.0304
B_2	3	0.0135	3	0.0444
B_3	3	0.0538	3	0.1288
B_4	3	0.2303	3	0.5830
B_5	3	0.4785	3	1.1340
B_6	3	0.5949	3	1.4231
B_7	3	0.7161	3	1.6737

For Peer Review Only
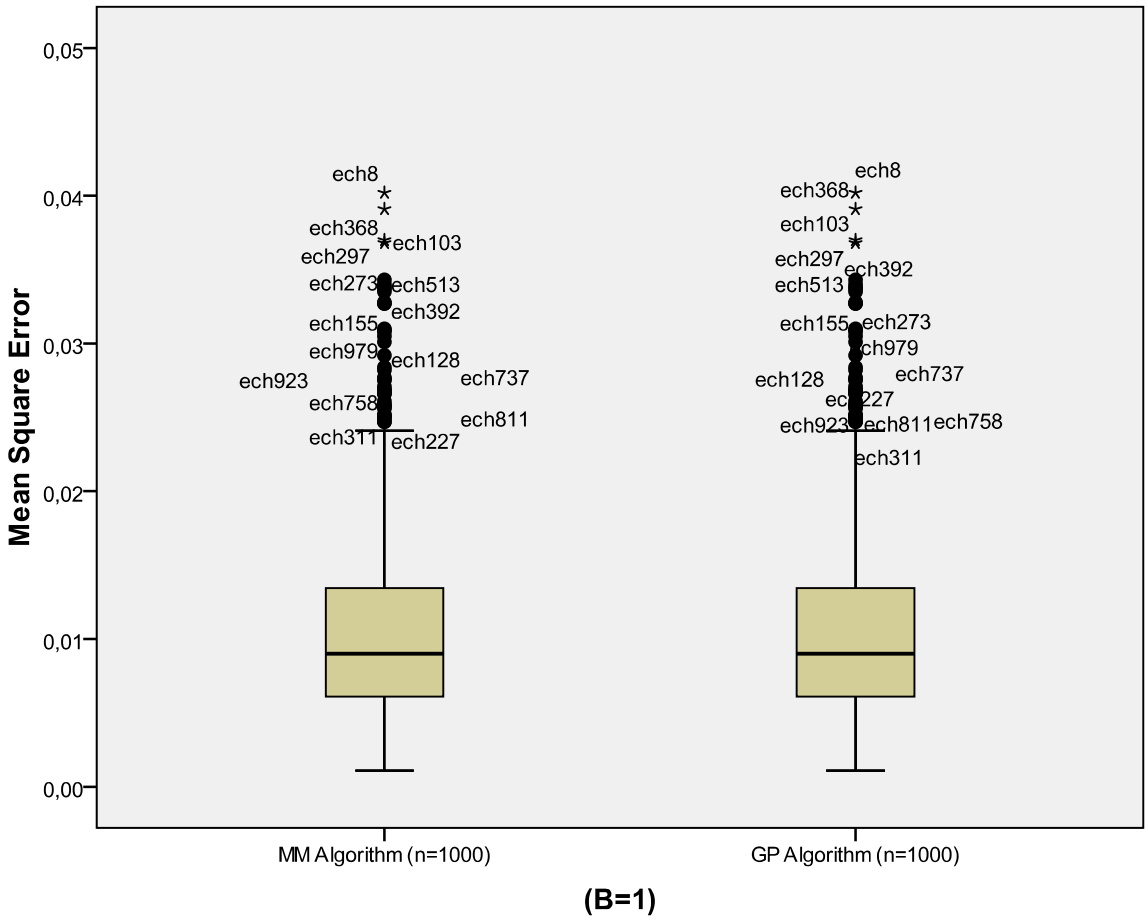
Figure 1 : MSE for n = 1000 and bracket B_1