



HAL
open science

One-step statistical estimation method for generalized linear models

Alexandre Brouste, Lilit Hovsepyan, Irene Votsi

► **To cite this version:**

Alexandre Brouste, Lilit Hovsepyan, Irene Votsi. One-step statistical estimation method for generalized linear models. 2025. hal-04866513

HAL Id: hal-04866513

<https://hal.science/hal-04866513v1>

Preprint submitted on 6 Jan 2025

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.

One-step statistical estimation method for generalized linear models

Alexandre Brouste¹, Lilit Hovsepyan¹, Irene Votsi^{2*}

^{1*}Laboratoire Manceau de Mathématiques - LMM EA 3263, Institut du Risque et de l'Assurance, Le Mans Université, Le Mans, 72000, France.

^{2*}LIEC, CNRS, Université de Lorraine, Metz, F-57000, France.

*Corresponding author(s). E-mail(s): Eirini.Votsi@univ-lorraine.fr;

Contributing authors: Alexandre.Brouste@univ-lemans.fr;

Lilit.Hovsepyan.Etu@univ-lemans.fr;

Abstract

In this article, the one-step estimation procedure is presented for generalized linear models. In these models, the maximum likelihood estimator, which is asymptotically efficient, has no closed-form and gradient-descent methods are generally used for its numerical computation. Nevertheless, when the amount of data is large and/or the number of explanatory variables is high, then the computations can be very consuming. To overcome this difficulty, the one-step estimation procedure is used, which is based on an initial (inefficient) guess estimator and a single step of the Fisher scoring. The main advantage of this procedure is that only one iterative step is required to achieve the asymptotic efficiency. The results are validated numerically by means of Monte-Carlo simulations. The estimation procedure is used to fit generalized linear models for climate risk insurance data.

Keywords: Asymptotic efficiency, maximum likelihood estimator, Fisher scoring, Monte-Carlo simulations, insurance data.

1 Introduction

Since their introduction in [24, 26], generalized linear models (GLMs) have been very popular as they provide a unified approach to many statistical models applied in different scientific fields. In the GLMs regression setting, the distribution of the response

variable is of exponential type. The exponential family of probability distributions include the most popular distributions such as Gaussian, Poisson, Binomial, Gamma and Inverse Gaussian.

The default method to estimate the parameters of GLMs from a finite number of observations n is the maximum likelihood estimation method. In the case of incomplete data due to partially missing explanatory variables on some observations, the maximum likelihood estimators are obtained by means of the Expectation-Maximization (EM) algorithm [14]. The main idea of this algorithm is the maximization of the expectation of the complete log-likelihood (instead of the likelihood that depends on unobserved explanatory variables), conditionally on the observations and a current parameter vector. The MLEs are obtained under the assumption that the missing variables follow a discrete distribution with a finite support. An EM algorithm for maximum likelihood estimation in GLMs with overdispersion was presented in [1]. Later on, an EM algorithm was presented to estimate mixtures of Poisson regressions when the dataset is sizeable and the number of components can be greater than 10 [27]. Compared to their frequentist counterparts, fewer articles have focused on Bayesian inference in generalized linear models. An efficient computational algorithm that works for an arbitrary set of linear inequality constraints for GLMs was introduced in [16] and shown to converge. The maximum likelihood estimators (MLE) of the parameters and their asymptotic properties in the large sample setting ($n \rightarrow \infty$) were studied in [12, 13]. In particular, the authors presented mild general conditions to assure the weak and the strong consistency of the estimators as well as their asymptotic normality. Moreover, they established the asymptotic efficiency of the MLE using the normalisation by the square root of the information matrix rather than any normed linear functional of the MLE.

From a numerical point of view, the MLE can be approximated by using the gradient descent methods, such as the Newton or the Fisher scoring. The MLE are generally obtained by means of the iterative weighted least square algorithm (IWLS, [11]). However, when the datasets are large or the number of predictors is high, the calculation of MLE can be time consuming and different computational issues arise. These difficulties can be overcome by using the one-step estimation procedure [22]. The main idea of this procedure is the construction of an optimal estimator based on an initial, fast-computable guess estimator (which is not asymptotically efficient) and its correction with one single step of the Fisher scoring algorithm to reach efficiency.

This procedure has been used to estimate the parameters of several stochastic models. In a high-frequency setting, the procedure was described for the fractional Gaussian noise model in [6], where an initial sequence of quadratic generalized variations-based estimators was employed. In the same setting, efficient estimation of a non-Gaussian stable Lévy process with drift and symmetric jumps observed at high frequency was considered in [5]. On the other side, in a large sample setting, the procedure has been studied for Markov processes and inhomogeneous Poisson processes in [20] and [10], respectively. More recently, an asymptotically efficient estimator was obtained for a first-order fractional autoregressive model in [7]. The method was also applied for Hawkes processes in [9]. Concerning multi-step estimator processes for ergodic diffusion processes and backward stochastic differential equations in the Markovian case,

we refer the interested reader to [18] and [19], respectively. From a computation point of view, the procedure is available in the i.i.d. setting, in the R package [4].

It is worth mentioning that the one-step method was first introduced for GLMs in [23]. The first main assumption is that the second moments of the second and third partial derivatives of the log-likelihood are uniformly bounded on a compact neighborhood of the parameter. The second assumption is that the covariates matrix X satisfies $XX^T/n \rightarrow C$ as $n \rightarrow \infty$. In the case that the explanatory variables are categorical, explicit fast-computable initial guess estimators were defined and a one-step procedure was introduced recently (see [3] and the reference therein).

The objective of this paper is to present the procedure in the case of GLMs with non-categorical response variables in a more general setting than [23]. In this work, we consider as initial guess estimator the MLE obtained in a subsample and then a correction is made with a single step of Fisher scoring. This procedure leads to an asymptotically efficient estimator. Numerical examples are used to validate the theoretical results. A real-world application to drought risk is also presented.

The paper is organized as follows: Section 2 contains the description of GLMs and the MLE method along with the corresponding regularity conditions. Section 3 includes the extended one-step procedure. Section 4 describes the estimation of the dispersion parameter. Numerical examples are available in Section 5 along with a drought risk insurance application.

2 Notation and preliminaries

2.1 Generalized Linear Model

Let us denote by $\mathbf{Y}^{(n)} = (Y_1, \dots, Y_n)^*$ the response vector composed by n independent but not identically distributed \mathbb{R} -valued random variables. We further denote by $\mathbb{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^*$ the set of deterministic explanatory variables valued in \mathbb{R}^p (or design matrix), where $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$ for $i \in \{1, \dots, n\}$. The random variable Y_i , $i \in \{1, \dots, n\}$, is supposed to belong to the exponential family of distributions, which means that the probability density function has the form

$$f_{Y_i}(y_i, (\vartheta, \phi)) = \begin{cases} \exp\left(\frac{y_i \lambda_i(\vartheta) - b(\lambda_i(\vartheta))}{\phi} + c(y_i, \phi)\right), & \text{if } y_i \in \mathbb{Y} \subset \mathbb{R}, \\ 0, & \text{elsewhere,} \end{cases} \quad (1)$$

where $\vartheta \in \Theta \subset \mathbb{R}^p$ and ϕ are unknown parameters, $b : \Lambda \rightarrow \mathbb{R}$ and $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ are known functions. We consider that Λ is a convex set of \mathbb{R} , and in the interior Λ^0 of Λ ($\Lambda^0 \neq \emptyset$), all derivatives of $b(\cdot)$ and all moments of Y_i exist for any $i \in \{1, \dots, n\}$. We can easily note that $\mathbf{E}(Y_i) = b'(\lambda_i(\vartheta)) = \mu(\vartheta)$ and $\mathbf{Var}(Y_i) = \phi b''(\lambda_i(\vartheta))$ (see Appendix A), which is supposed to be positive in Λ_0 .

Moreover, $\lambda_i(\cdot)$, $i \in \{1, \dots, n\}$, is a function of an unknown parameter $\vartheta \in \Theta \subset \mathbb{R}^p$ defined by

$$\lambda_i = \lambda_i(\vartheta) = h(\langle \mathbf{x}_i, \vartheta \rangle),$$

where h is properly fixed and $\langle \cdot, \cdot \rangle$ represents the standard scalar product in \mathbb{R}^p . Consequently, for $i \in \{1, \dots, n\}$, we have

$$g(\mathbf{E}(Y_i)) = \langle \mathbf{x}_i, \vartheta \rangle,$$

where $g = h^{-1} \circ \mu^{-1}$. The function g is called the link function.

Here we focus on the canonical case where h is the identity function. We further denote by M the image $\mu(\Lambda^0)$ of Λ^0 . In the sequel, we make some necessary regularity conditions:

1. Θ is open in \mathbb{R}^p and convex,
2. $\langle \mathbf{x}_i, \vartheta \rangle \in g(M)$, $i \in \{1, \dots, n\}$, for all $\vartheta \in \Theta$,
3. The function g is twice continuously differentiable,
4. The matrix $\mathbb{X}^* \mathbb{X}$ has full rank for $n \geq n_0$.

It is worth mentioning that for one parameter distributions such as the distributions of Bernoulli and Poisson, the dispersion parameter ϕ is assumed to be known and equal to 1. However, for the Normal and Gamma distributions, ϕ needs to be estimated (see Table 1).

Table 1: Table of distributions of the exponential family.

Distribution	$b(\lambda)$	$c(y, \phi)$	Link function name	Canonical link
Poisson	$\exp(\lambda)$	$-\log(!y)$	Log	$g(m) = \log(m)$
Bernoulli	$\log(1 + \exp(\lambda))$	0	Logit	$g(m) = \log(m/(1 - m))$
Normal	$\lambda^2/2$	$-(y^2/\phi + \log(2\pi\phi))/2$	Identity	$g(m) = m$
Gamma	$-\log(-\lambda)$	$\log(y/\phi)/\phi - \log(y) - \log(\Gamma(1/\phi))$	Inverse	$g(m) = -m^{-1}$

2.2 Maximum likelihood estimation

Given a realisation of $\mathbf{Y}^{(n)}$, the log-likelihood function associated to the statistical experiment is given by

$$\begin{aligned} \mathcal{L}_n(\vartheta, \phi) &= \sum_{i=1}^n \log f_{Y_i}(Y_i, (\vartheta, \phi)) \\ &= \frac{1}{\phi} \sum_{i=1}^n (Y_i \lambda_i(\vartheta) - b(\lambda_i(\vartheta))) + \sum_{i=1}^n c(Y_i, \phi). \end{aligned}$$

In the sequel, we denote by ϑ_0 and ϕ_0 the true, unknown parameters. The Maximum Likelihood Estimator (MLE) $(\hat{\vartheta}_n, \hat{\phi}_n)$ is defined by

$$(\hat{\vartheta}_n, \hat{\phi}_n) = \arg \max_{(\phi, \vartheta) \in \Phi \times \Theta} \mathcal{L}_n(\vartheta, \phi).$$

The MLE satisfies the score equations

$$\frac{\partial}{\partial \vartheta} \mathcal{L}_n(\widehat{\vartheta}_n, \widehat{\phi}_n) = 0 \quad \text{and} \quad \frac{\partial}{\partial \phi} \mathcal{L}_n(\widehat{\vartheta}_n, \widehat{\phi}_n) = 0.$$

The first equation also writes

$$\frac{\partial}{\partial \vartheta} \ell_n(\widehat{\vartheta}_n) = 0$$

where

$$\ell_n(\vartheta) = \sum_{i=1}^n (Y_i \lambda_i(\vartheta) - b(\lambda_i(\vartheta))). \quad (2)$$

which does not depend on ϕ . When the parameter ϕ has also to be estimated it can be infer with the MLE procedure by solving the second equation that writes

$$-\frac{1}{\widehat{\phi}_n^2} \ell_n(\widehat{\vartheta}_n) + \sum_{i=1}^n \frac{\partial}{\partial \phi} c(Y_i, \widehat{\phi}_n) = 0. \quad (3)$$

In the following, we consider the estimation of θ . The estimation of ϕ is considered in Section 4.

2.3 Asymptotic properties of the MLE of ϑ

Direct computations of the score function, denoted S_{ϑ}^n , lead to

$$S_{\vartheta,j}^n = \frac{\partial}{\partial \vartheta_j} \ell_n(\vartheta) = \sum_{i=1}^n \frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_j} (Y_i - b'(\lambda_i(\vartheta))),$$

for $j = 1, \dots, p$. In the canonical case, we have

$$\frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_j} = x_{i,j}, \quad j = 1, \dots, p,$$

which does not depend on the parameter ϑ . Direct calculations on the Hessian matrix lead to

$$\begin{aligned} H_{i,j}^{n,\vartheta} &= \frac{\partial^2}{\partial \vartheta_j \partial \vartheta_k} \ell_n(\vartheta) \\ &= \sum_{i=1}^n \frac{\partial^2 \lambda_i(\vartheta)}{\partial \vartheta_j \partial \vartheta_k} (Y_i - b'(\lambda_i(\vartheta))) - \sum_{i=1}^n \frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_j} \frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_k} b''(\lambda_i(\vartheta)). \end{aligned}$$

It can be noticed that

$$\mathbf{E} \left(H_{i,j}^{n,\vartheta} \right) = - \sum_{i=1}^n \frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_j} \frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_k} b''(\lambda_i(\vartheta)).$$

In the canonical case

$$\mathbf{E} \left(H_{i,j}^{n,\vartheta} \right) = - \sum_{i=1}^n x_{i,j} x_{i,k} b''(\lambda_i(\vartheta)).$$

Then the symmetric Fisher information matrix $\mathcal{I}_n(\vartheta)$ is defined by

$$\mathcal{I}_n(\vartheta)_{i,j} := -\mathbf{E} \left(H_{i,j}^{n,\vartheta} \right). \quad (4)$$

The sequence of maximum likelihood estimators $\widehat{\vartheta}_n$ has been shown to be consistent and asymptotically normal in [12, 13]. To obtain these asymptotic properties, some additional assumptions are necessary. In the following, we denote by $A^{\frac{1}{2}}$ (respectively $A^{\frac{*}{2}}$) a left (respectively a right) square root matrix of the positive definite matrix A , i.e. $A^{\frac{1}{2}} A^{\frac{*}{2}} = A$. Then we define, for $\vartheta_0 \in \Theta$, $\delta > 0$ and $n \geq 1$, a subset of the parameter space Θ

$$N_n(\delta) = \{ \vartheta \in \Theta \mid \| \mathcal{I}_n(\vartheta_0)^{\frac{*}{2}} (\vartheta - \vartheta_0) \|_2 \leq \delta \}.$$

We denote by $\lambda_{min}^n(\vartheta)$ ($\lambda_{max}^n(\vartheta)$) the smallest (largest) eigenvalue of $\mathcal{I}_n(\vartheta)$.

In addition, we suppose that:

1.

$$\lim_{n \rightarrow \infty} \lambda_{min}^n(\vartheta_0) = \infty.$$

2. For all $\delta > 0$, the matrix $\mathcal{I}_n(\vartheta) - c\mathcal{I}_n(\vartheta_0)$ is positive semi-definite for all $\vartheta \in N_n(\delta)$ and all $n \geq n_1$ for some constants $n_1(\delta)$ and $c > 0$ independent of δ .

3. For all $\delta > 0$,

$$\lim_{n \rightarrow \infty} \max_{\vartheta \in N_n(\delta)} |V_n(\vartheta) - I_p| = 0,$$

where

$$V_n(\vartheta) = \varphi_n(\vartheta_0)^* \mathcal{I}_n(\vartheta) \varphi_n(\vartheta_0) \quad (5)$$

is the normed information matrix with

$$\varphi_n(\vartheta) = \mathcal{I}_n(\vartheta)^{-\frac{*}{2}} \quad (6)$$

and I_p is the identity matrix of order p .

4. There is a neighborhood $U \subset \Theta$ of ϑ_0 such that, for all $n \geq n_1$,

$$\lambda_{min}^n(\vartheta) \geq c(\lambda_{max}^n(\vartheta))^{\frac{1}{2} + \delta},$$

for some constants $c > 0$ and $\delta > 0$.

Under the previous assumptions, the sequence of maximum likelihood estimators $(\widehat{\vartheta}_n, n \geq 1)$ is consistent and asymptotically normal [12, 13], that is

$$\mathcal{I}_n(\vartheta_0)^{\frac{*}{2}} \left(\widehat{\vartheta}_n - \vartheta_0 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \phi I_p). \quad (7)$$

3 One-step procedure

According to [17], the one-step procedure can be adapted to a non-rate efficient sequence of estimators if the Fisher information presents more regularity (see for instance [20] for recursive Markov chains).

The one-step procedure is based on the following idea: Given a consistent but possibly inefficient initial estimator, $\tilde{\vartheta}_n$, we first form the Fisher scoring one-step estimator defined by

$$\bar{\vartheta}_n = \tilde{\vartheta}_n + \mathcal{I}_n(\tilde{\vartheta}_n)^{-1} \frac{\partial}{\partial \vartheta} \ell_n(\tilde{\vartheta}_n). \quad (8)$$

Second, under regularity conditions, we show that this estimators have the same asymptotic properties as the MLE.

In our context, we choose as initial guess estimator, the MLE computed on a small sample of size $N = \lfloor n^\delta \rfloor$ (with $\frac{1}{2} < \delta \leq 1$), namely

$$\tilde{\vartheta}_n = \hat{\vartheta}_N. \quad (9)$$

Let us denote $\eta_n(\vartheta) = \varphi_{\lfloor n^\delta \rfloor}(\vartheta) = \varphi_N(\vartheta)$. From (7), the estimator is $\varphi_N(\vartheta_0)$ -consistent, more precisely

$$\eta_n(\vartheta_0) \left(\tilde{\vartheta}_n - \vartheta_0 \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \phi I_p). \quad (10)$$

We further assume that:

- (i) the function $V_n(\vartheta)$ defined in (5) is Lipschitz continuous,
- (ii) the quantity

$$\|\eta_n^*(\vartheta_0) \varphi_n(\vartheta_0)^{-1} \eta_n(\vartheta_0)\| \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Theorem 1. *Under all previous assumptions, the sequence of one-step estimator $(\bar{\vartheta}_n, n \geq 1)$ defined in (8) and (9) is consistent and asymptotically normal, i.e., as $n \rightarrow \infty$,*

$$\mathcal{I}_n(\vartheta_0)^{\frac{1}{2}} (\bar{\vartheta}_n - \vartheta_0) \longrightarrow \mathcal{N}(0, \phi I_p) \quad \text{in law.}$$

Proof. First, starting from the one-step estimator

$$\begin{aligned} \bar{\vartheta}_n &= \tilde{\vartheta}_n + \mathcal{I}_n(\tilde{\vartheta}_n)^{-1} \frac{\partial}{\partial \vartheta} \ell_n(\tilde{\vartheta}_n) \\ &= \tilde{\vartheta}_n + \varphi_n(\vartheta_0) \cdot \varphi_n(\vartheta_0)^{-1} \mathcal{I}_n(\tilde{\vartheta}_n)^{-1} \varphi_n(\vartheta_0)^{-*} \cdot \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\tilde{\vartheta}_n) \\ &= \tilde{\vartheta}_n + \varphi_n(\vartheta_0) V_n(\tilde{\vartheta}_n)^{-1} \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\tilde{\vartheta}_n), \quad n \geq 1, \end{aligned}$$

where $(\tilde{\vartheta}_n, n \geq 1)$ is the initial guess estimator. The renormalized statistical error is then

$$\varphi_n(\vartheta_0)^{-1} (\bar{\vartheta}_n - \vartheta_0) = \varphi_n(\vartheta_0)^{-1} (\tilde{\vartheta}_n - \vartheta_0) + V_n(\tilde{\vartheta}_n)^{-1} \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\tilde{\vartheta}_n), \quad n \geq 1. \quad (11)$$

Second, according to the mean value theorem for the score, there exists $v \in [0, 1]$ such that

$$\begin{aligned} \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\tilde{\vartheta}_n) &= \varphi_n(\vartheta_0)^* \int_0^1 \frac{\partial^2}{\partial \vartheta^2} \ell_n(\vartheta_0 + v(\tilde{\vartheta}_n - \vartheta_0)) dv (\tilde{\vartheta}_n - \vartheta_0) + \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\vartheta_0) \\ &= \varphi_n(\vartheta_0)^* \int_0^1 \frac{\partial^2}{\partial \vartheta^2} \ell_n(\vartheta_0 + v(\tilde{\vartheta}_n - \vartheta_0)) dv \varphi_n(\vartheta_0) \cdot \varphi_n(\vartheta_0)^{-1} (\tilde{\vartheta}_n - \vartheta_0) \\ &\quad + \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\vartheta_0) \\ &= - \int_0^1 V_n(\vartheta_0 + v(\tilde{\vartheta}_n - \vartheta_0)) dv \cdot \varphi_n(\vartheta_0)^{-1} (\tilde{\vartheta}_n - \vartheta_0) + \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\vartheta_0). \end{aligned}$$

The last equality is valid since we consider natural link functions. By substituting in (11) and reorganizing the terms, we obtain that

$$\varphi_n(\vartheta_0)^{-1} (\bar{\vartheta}_n - \vartheta_0) = A + B,$$

where

$$A := V_n(\tilde{\vartheta}_n)^{-1} \left(V_n(\tilde{\vartheta}_n) - \int_0^1 V_n(\vartheta_0 + v(\tilde{\vartheta}_n - \vartheta_0)) dv \right) \varphi_n(\vartheta_0)^{-1} (\tilde{\vartheta}_n - \vartheta_0)$$

and

$$B := V_n(\tilde{\vartheta}_n)^{-1} \varphi_n(\vartheta_0)^* \frac{\partial}{\partial \vartheta} \ell_n(\vartheta_0).$$

Following Lemma 1 in [13] (under conditions (2),(4)) we have that

$$B \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \phi I_p).$$

As for the first term, we have that

$$\begin{aligned} A &= \underbrace{\left(\mathbf{I} - V_n(\tilde{\vartheta}_n)^{-1} \int_0^1 V_n(\vartheta_0 + v(\tilde{\vartheta}_n - \vartheta_0)) dv \right)}_{A.1} \eta_n^{-*}(\vartheta_0) \\ &\quad \times \underbrace{\eta_n^*(\vartheta_0) \varphi_n(\vartheta_0)^{-1} \eta_n(\vartheta_0)}_{A.2} \cdot \underbrace{\eta_n^{-1}(\vartheta_0) (\tilde{\vartheta}_n - \vartheta_0)}_{A.3}, \end{aligned}$$

where $\eta_n(\vartheta) = \varphi_n^s(\vartheta)$.

To prove that the term A.1 is bounded in probability, first we notice that

$$\int_0^1 V_n \left(\vartheta_0 + v(\tilde{\vartheta}_n - \vartheta_0) \right) dv \xrightarrow{p} I_p,$$

according to the proof of Theorem 3 in [13]. Second, we consider the η_n consistency of $\tilde{\vartheta}_n$, and assume that V_n is Lipschitz.

The term A.3 is bounded in probability due to the η_n consistency of $\tilde{\vartheta}_n$. The second term A.2 tends to zero by assumption.

Then by means of Slutsky's theorem we conclude that the term A tends to zero in probability, which leads to the final result. \square

Remark 1. *The condition*

$$\|\eta_n^*(\vartheta_0)\varphi_n(\vartheta_0)^{-1}\eta_n(\vartheta_0)\| \longrightarrow 0 \quad \text{as } n \rightarrow \infty$$

is fulfilled for regular design matrices (see for instance [25, 28]) and $\delta > 1/2$. Indeed, the design matrix \mathbb{X} is called regular if

$$\frac{\mathbb{X}\mathbb{X}^T}{n^\alpha} \longrightarrow C, \quad \alpha > 0, \quad \text{as } n \rightarrow \infty,$$

where C is a symmetric positive definite matrix. Consequently, we can show that

$$\frac{\mathcal{I}_n(\vartheta_0)}{n^\alpha} \longrightarrow C',$$

where C' is a symmetric positive definite matrix. In this case, since

$$\eta_n^*(\vartheta_0)\varphi_n(\vartheta_0)^{-1}\eta_n(\vartheta_0) \sim (C')^{-1/2}n^{(1-2\delta)\alpha/2},$$

the norm of the matrix tends to zero for $\delta > 1/2$.

4 Estimation of ϕ

In several GLM models, for instance Bernoulli and Poisson, the dispersion parameter ϕ is equal to 1 and does not require estimation. When the parameter ϕ needs to be estimated, it can be obtained by solving the equation

$$-\frac{1}{\hat{\phi}_n^2}\ell_n(\hat{\vartheta}_n) + \sum_{i=1}^n \frac{\partial}{\partial \phi} c(Y_i, \hat{\phi}_n) = 0. \quad (12)$$

In the Gaussian and Inverse Gaussian cases, the equation(12) has a closed-form solution. In particular, in the case of Gaussian model, we have $c(y, \phi) = -y^2/(2\phi) -$

$\log(2\pi\phi)/2$. Hence the MLE of ϕ is defined by

$$\hat{\phi}_n = \frac{-2\ell_n(\hat{\vartheta}_n) + \sum_{i=1}^n Y_i^2}{n}. \quad (13)$$

On the other side, in the case of Inverse Gaussian model, $b(x) = -\sqrt{-2x}$ and $c(y, \phi) = -\log(2\pi\phi y^3)/2 - 1/(2\phi y)$. Then the MLE of ϕ is defined by

$$\hat{\phi}_n = \frac{-2\ell_n(\hat{\vartheta}_n) + \sum_{i=1}^n 1/Y_i^2}{n}. \quad (14)$$

In the two previous examples, the MLE $\hat{\vartheta}_n$ could be replaced by the one-step estimator $\bar{\vartheta}_n$ in (13) and (14) respectively, leading to faster and asymptotically optimal estimation.

For several other models, the equation (12) has no closed-form solution. For instance, in the case of Gamma model, the MLE $\hat{\phi}_n$ of ϕ is the solution of the equation

$$\ell_n(\hat{\vartheta}_n) + n \left(1 - \psi^{(1)}(1/\hat{\phi}_n) - \log(\hat{\phi}_n) \right) + \sum_{i=1}^n \log(Y_i) = 0$$

with $\psi^{(1)}(\alpha) = \frac{d}{d\alpha} \log \Gamma(\alpha)$ is the digamma function.

5 Numerical examples

5.1 Simulated data

5.1.1 Gamma distribution

In the following we present a numerical example to illustrate the theoretical results. In particular, we consider a generalized linear model with a bi-dimensional explanatory variable ($p = 2$) and a response variable that follows the Gamma distribution with dispersion parameter equal to $\phi = 0.1$. To illustrate the asymptotic normality of the estimators, we make 1000 Monte-Carlo simulations. Each simulation corresponds to a realization of random variables of length $n = 2000$.

For a generic estimator ϑ_n^* , the renormalized statistical error is defined by

$$\mathcal{I}_n(\vartheta_0)^{\frac{*}{2}} (\vartheta_n^* - \vartheta_0).$$

The Figure 1 presents the asymptotic distribution of the renormalized statistical errors for the maximum likelihood estimator $\hat{\vartheta}_n$ (MLE), the initial estimator $\bar{\vartheta}_n$ (subMLE with $\delta = 0.8$) as well as the one-step maximum likelihood estimator $\bar{\vartheta}_n$ (OSMLE).

We observe that contrary to the initial estimator, the one-step maximum likelihood estimator is asymptotically efficient.

Moreover, if we compare the times to compute the estimators, we notice that the calculation of the MLE necessitates 2.520 mins, whereas the calculation of the OSMLE needs 12.479 secs. The calculations have been made in R (version 4.0.5), the PC used

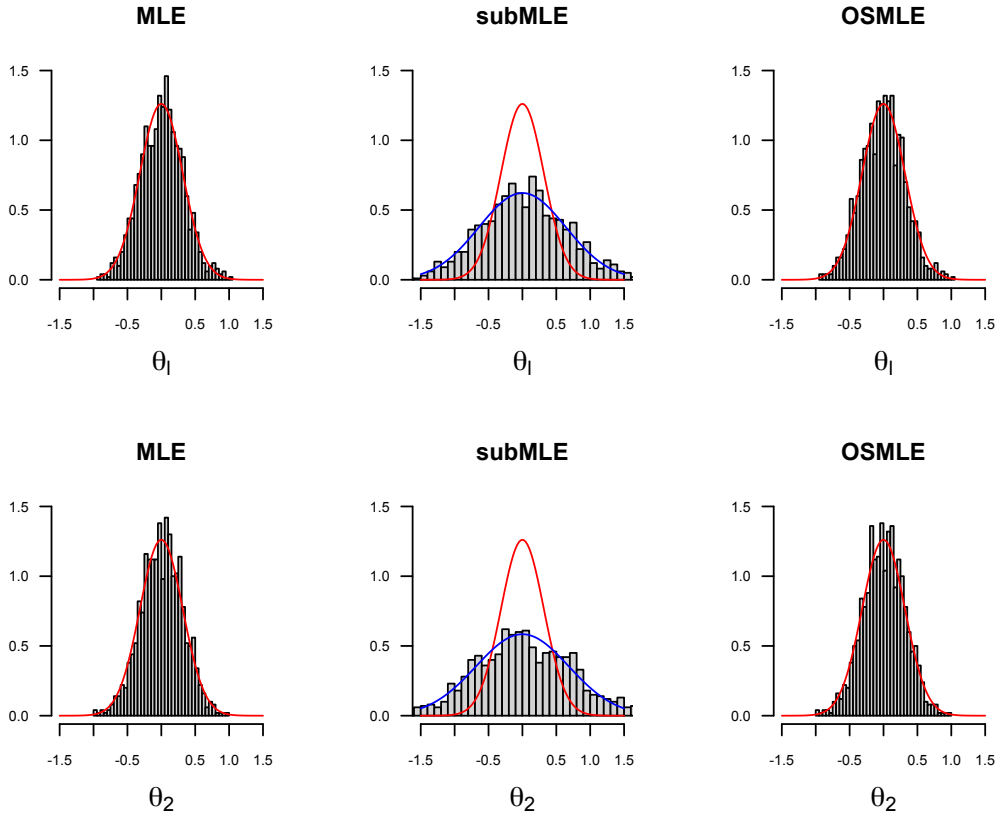


Fig. 1: Renormalized statistical errors of MLE, subMLE and OSMLE for a GLM with a Gamma-distributed response variable. Red and blue lines correspond to the theoretical asymptotic densities of the MLE and the subMLE, respectively.

is an HP ZBook 15u G3, with a memory of 16 Go and processor Intel(R) Core(TM) i7-6500U CPU . 2.50 GHz.

We further present in Figure 2 the mean time (in secs) needed to compute the MLE and the OSMLE (for a sample size $n = 500$) as a function of the number of predictors p for different values of δ . In order to calculate these mean times, 200 Monte-Carlo simulations were made for every number of predictors.

5.1.2 Poisson distribution

In the following we present a numerical example where the response variable is discrete and in particular follows the Poisson distribution, whereas the explanatory variable is bi-dimensional ($p = 2$). To illustrate the asymptotic normality of the estimators, we make 1000 Monte-Carlo simulations, where each simulation includes the generation of random variables of length $n = 2000$ (Fig. 3).

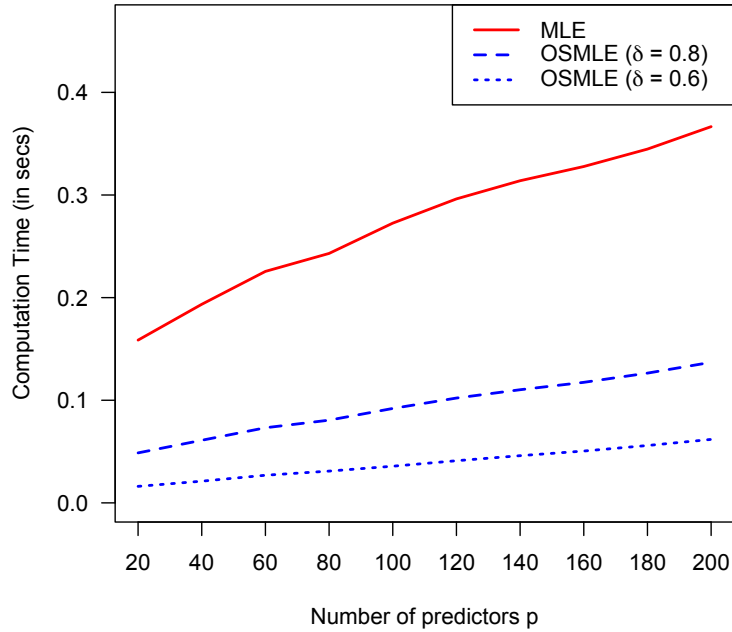


Fig. 2: Computation time of MLE and OSMLE versus number of predictors for different values of δ .

5.2 Real data

In this section, we present a comparative analysis between the One-Step Maximum Likelihood Estimator and the traditional Maximum Likelihood Estimator using a real-world dataset coming from the insurance sector. The response variables concern the claim amounts in a portfolio of drought insurance policies and the connection of incidents with the french regime of natural hazards. Indeed, if an incident belongs to this regime, it may be refunded.

The dataset includes $p = 22$ variables, 18 of them are quantitative, and comprises no missing data. Categorical variables and identifiers are excluded, whereas the quantitative variables are observed for $n = 10,884$ claims from 2006 to 2023. The dataset comes from the insurance company Groupama and the consolidation work of the Institute of Risk and Insurance in Le Mans.

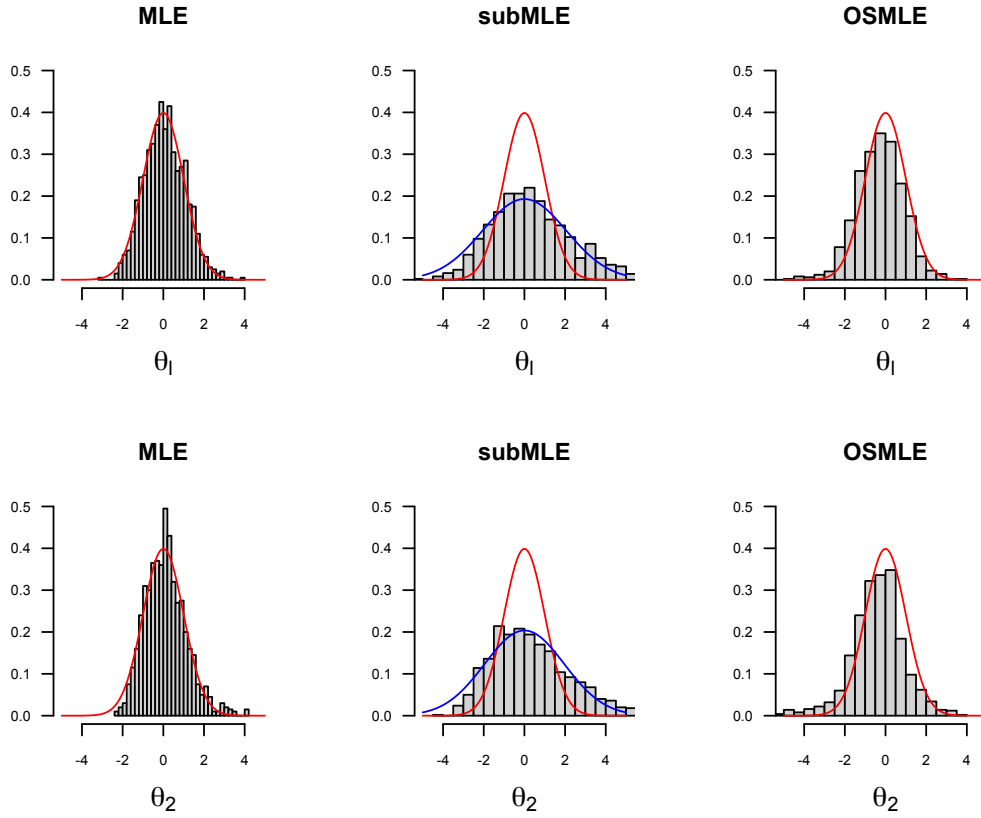


Fig. 3: Renormalized statistical errors of MLE, subMLE and OSMLE for a GLM with a Poisson-distributed response variable. Red and blue lines correspond to the theoretical asymptotic densities of the MLE and the subMLE, respectively.

5.3 Modeling the activation of a claim

Our objective is to explain the insurance claim variable Y_i ($i = 1, \dots, n$) defined by

$$Y_i = \begin{cases} 1, & \text{if the } i\text{-th insurance claim is active,} \\ 0, & \text{otherwise,} \end{cases}$$

by means of the aforementioned explanatory variables. This categorical variable captures the ongoing status of claims, offering insights into the factors affecting the duration or resolution of insurance claims.

In the logistic regression model, the link is considered to be canonical, which leads to the equation

$$\log \left(\frac{\mathbf{E}(Y_i)}{1 - \mathbf{E}(Y_i)} \right) = \sum_{j=1}^p \vartheta_j x_{i,j}, \quad (15)$$

for $i = 1, \dots, n$.

Among the 22 explanatory variables, there are two variables that are statistically significant (significance level $\alpha = 0.05$):

- Year of claim opening (\mathbf{x}_1): year that the insurance claim was initiated. It contributes to the identification of temporal trends or changes in claim processing.
- Duration of claim (\mathbf{x}_2): number of days from the opening to the closure of the claim. It measures the active duration of the claim.

The OSMLE ($\delta = 0.9$) and the MLE are presented in Table 2. First, we observe that the values of MLE and OSMLE are close to each other, with a difference equal to zero (approximated to 3 decimals).

	$\vartheta_1 \times 10^3$	$\vartheta_2 \times 10^3$
MLE	-2.5631	3.9239
subMLE	-2.5637	3.9400
OSMLE	-2.5630	3.9237

Table 2: OSMLE and MLE of the parameters of the logistic model.

Second, each parameter has the same sign for both MLE and OSMLE. The negative sign of the estimated ϑ_1 implies that as the year of the claim opening increases, it is less likely that the claim is active. On the other side, the odds ratio associated to the parameter ϑ_2 is greater than 1, which means that as the duration of a claim increases, it is more likely that the claim is active. The duration of a claim favors the probability that it is still active.

5.4 Modeling the cost of a claim

In this section, we build a Gamma glm to explain the cost of insurance claims. The link function is considered to be the canonical (i.e. inverse), and the model is described by the equation

$$\frac{1}{\mathbf{E}(Y_i)} = \sum_{j=1}^p \vartheta_j x_{i,j},$$

for $i = 1, \dots, n$.

It turns out that there are two statistically significant explanatory variables:

- Duration of claim (\mathbf{x}_1): number of days from the opening to the closure. It measures the active duration of the claim.
- Land registry urbanization index (\mathbf{x}_2): urbanization index of the area where the building is located. Values closer to 1 indicate higher levels of urbanization.

As shown in Table 3, the MLE and OSMLE ($\delta = 0.95$) of the parameters are very close to each other, which reflects the robustness of the estimation methods.

	$\vartheta_1 \times 10^6$	$\vartheta_2 \times 10^6$
MLE	0.018	4.056
subMLE	0.017	4.097
OSMLE	0.018	4.056

Table 3: OSMLE and MLE of the parameters of the Gamma GLM.

6 Concluding remarks

Over time, as more data becomes available, the development of statistical methods that allow to handle complex and large-scale problems presents a growing interest. Here we focus on GLMs and present the one-step estimation procedure. The procedure is based on the construction of a new estimator of the parameters, the one-step MLE, which has the same asymptotic properties as the MLE. This procedure presents computational advantages, as demonstrated by Monte-Carlo simulations.

In this work we consider that the response variable is univariate. From a methodological perspective, the extension of this work to the multivariate case and multivariate models is of particular interest. Moreover, the situation where a non-canonical form is considered, could be addressed in the future. Another type of estimators that could be studied for GLMs are the multi-step estimators [17], where the subsample is n^δ with $\delta \leq 1/2$. Further work includes the development of the procedure for generalized additive models which extend GLMs in the sense that the response variable depends linearly on unknown smooth functions of some explanatory variables. In a parametric context, these functions have a specified parametric form, and thus the one-step procedure could be employed to estimate the corresponding parameters.

From application point of view, the one-step procedure could be further used to other applications coming from the insurance industry. Since GLMs are widely used in many scientific fields, the proposed method could have straightforward applications in agriculture and biology.

A First two central moments for GLM

Since $f_{Y_i}(\cdot, (\vartheta, \phi))$ is the density of Y_i defined in (1), we have

$$\int_{\mathbf{Y}} f_{Y_i}(u, (\vartheta, \phi)) du = 1. \quad (16)$$

By derivating (16) with respect to θ_1 , we get

$$\frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_1} \int_{\mathbf{Y}} \frac{(u - b'(\lambda_i(\vartheta)))}{\phi} f_{Y_i}(u, (\vartheta, \phi)) dy = 0$$

and

$$\int_{\mathbf{Y}} (u - b'(\lambda_i(\vartheta))) f_{Y_i}(u, (\vartheta, \phi)) dy = 0, \quad (17)$$

which leads to $\mathbf{E}(Y_i) = b'(\lambda_i(\vartheta))$. Moreover, derivating (17) with respect to θ_1 gives

$$\frac{\partial \lambda_i(\vartheta)}{\partial \vartheta_1} \int_{\mathbb{Y}} \left(\frac{(u - b'(\lambda_i(\vartheta)))^2}{\phi} - b''(\lambda_i(\vartheta)) \right) f_{Y_i}(u, (\vartheta, \phi)) dy = 0,$$

which results to $\mathbf{Var}(Y_i) = \phi b''(\lambda_i(\vartheta))$.

References

- [1] Aitkin, M. (1996) A general maximum likelihood analysis of overdispersion in generalized linear models, *Statistics and Computing*, 6, 251-262.
- [2] Bensoussan, A., Bertrand, P. and Brouste, A. (2015) Estimation Theory for Generalized Linear Models in "Future Perspectives in Risk Models and Finance" eds. A. Bensoussan, D. Guegan and C. Tapiero, 1–70, Springer-Verlag.
- [3] Brouste, A., Dutang, C., Hovsepian, L. and Rohmer, T. (2023) One-step closed-form estimator generalized linear model with categorigal explanatory variables, *Statistics and Computing*, 33(6), 138 doi.org/10.1007/s11222-023-10313-4.
- [4] Brouste, A., Dutang, C. and Noutsu Meniedou, D. (2021) One Step: Le Cam's one-step estimation procedure, *R J.*, 13(1), 366–377.
- [5] Brouste, A. and Masuda, H. (2018) Efficient estimation of stable Lévy process with symmetric jumps, *Statistical Inference for Stochastic Processes*, 21, 289–307.
- [6] Brouste, A., Soltane, M. and Votsi, I. (2020) One-step estimation for the Fractional Gaussian noise at high-frequency, *ESAIM: Probability and Statistics*, 24, 827–841.
- [7] Ben Hariz, S., Brouste, A., Cai, C. and Soltane, M. (2024) Fast and asymptotically-efficient estimation in an autoregressive process with fractional type noise, *Journal of Statistical Planning and Inference*, 232, 106148.
- [8] Ben Hariz, S., Brouste, A., Esstafa, Y., and Soltane, M. (2023). Fast calibration of weak FARIMA models, *ESAIM : Probability and Statistics*, 27, 156–173.
- [9] Brouste, A. and Farinetto, Ch. (2023). Fast and asymptotically efficient estimation in the Hawkes processes, *Japanese Journal of Statistics and Data Science*, 6, 361-379.
- [10] Dabye, A.S., Gounoug, A.A. and Kutoyants, Y.A. (2018) Method of Moments estimators and Multi-step MLE for Poisson Processes, *J. Contemp. Mathemat. Anal.*, 53, 237–246.
- [11] Dutang, Ch. (2017) Some explanations about the IWLS algorithm to fit generalized linear models. hal-01577698.

- [12] Fahrmeir, L. and Kaufmann, H. (1986) Correction – Consistency and asymptotic normality of the maximum likelihood estimator in Generalized Linear Models, *The Annals of Statistics*, 14(4), 1643.
- [13] Fahrmeir, L. and Kaufmann, H. (1985) Consistency and asymptotic normality of the maximum likelihood estimator in Generalized Linear Models, *The Annals of Statistics*, 13(1), 342–368.
- [14] Ibrahim, G.J. (1990) Incomplete Data in Generalized Linear Models, *Journal of the American Statistical Association*, 85(411), 765–769.
- [15] Ibragimov, I. and Khasminski, R. (1981) Statistical estimation. Asymptotic theory. Springer-Verlag.
- [16] Ghosal, R. and Ghosh, S.K. (2022) Bayesian inference for generalized linear model with linear inequality constraints, *Computational Statistics and Data Analysis*, 166, 107335.
- [17] Kamatani, K. and Uchida, M. (2015) Hybrid multi-step estimators for stochastic differential equations based on sampled data, *Statistical inference for stochastic processes*, 18(2), 177–204.
- [18] Kutoyants, Y.A. (2017) On the multi-step MLE-process for ergodic diffusion, *Stochastic Processes and Their Applications*, 127(7), 2243–2261.
- [19] Kutoyants, Y.A. (2016) On approximation of BSDE and multi-step MLE-processes, *Probability, Uncertainty and Quantitative Risk*, 1(4), doi 10.1186/s41546-016-0005-0.
- [20] Kutoyants, Y.A. and Motrunich, A. (2015) On multi-step MLE process for Markov sequences, *Metrika*, 79, 705–724.
- [21] Kutoyants, Y.A. (2014) On approximation of the backward stochastic differential equation. Small noise, large samples and high frequency cases, *Proc Steklov Inst Math*, 287, 133–154.
- [22] Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypothesis, Proceedings of the 3rd Berkeley Symposium, 1, 355–368.
- [23] Lumley, Th. (2019) Fast Generalized Linear models by Database Sampling and one-step polishing, *Journal of Computational and Graphical Statistics*, 28(4).
- [24] McCullagh, P. and Nelder, J.A. (1989) Generalized linear models, Chapman & Hall/CRC Monographs on Statistics and Applied Probability.
- [25] Montfort, A. (1997) Cours de Statistique Mathématiques, Economica.

- [26] Nelder J. and Wedderburn R. (1972) Generalized Linear Models, *Journal of the Royal Statistical Society. Series A*, 3(135), 370–384.
- [27] Papastamoulis, P., Martin-Magniette, M.-L. and Maugis-Rabusseau, C. (2016) On the estimation of mixtures of Poisson regression models with large number of components, *Computational Statistics and Data Analysis*, 93, 97–106.
- [28] Wu, C-F. (1981) Asymptotic theory of nonlinear least squares estimation, *The Annals of Statistics*, 9(3), 501–513.

Acknowledgements

This work was partially supported by the ANR Project “Efficient inference for large and high-frequency data” (ANR-21-CE40-0021), the Chairs “Efficience et Sobriété Numériques”, “Risques Emergents ou atypiques en Assurance” and “Impact de la Transition Climatique en Assurance”, under the aegis of Institut Louis Bachelier, joint initiatives by Le Mans University and EREN Groupe, MMA company, member of Covea group, Groupama Centre-Manche company, member of Groupama group respectively. A special thanks go to Bertille Tierny, Iker Tardio from ILB Datalab and to Guillaume Rondeau, Guilbon Romain and Arthur Garrec from Groupama for the preparation of the database.