Bol. Soc. Paran. Mat.
©SPM –ISSN-2175-1188 ON LINE
SPM: www.spm.uem.br/bspm

(3s.) **v. 2024 (42)** : 1–10.
ISSN-0037-8712 IN PRESS
doi:10.5269/bspm.63433

Sequential Estimation of Generalized Linear Model Coefficients

Ali Labriji, Safae Msellek and Abdelkrim Bennar

ABSTRACT: Generalized Linear Models (GLM) allows us to model the relationship between a response variable and one or more predictor variables while taking into account the distribution of the response variable. It is a useful tool for modeling data that do not follow a normal distribution and can be applied to a wide range of data types and problem settings. As data becomes increasingly relevant in our daily lives, the use of these models is becoming more important. However, this increase in importance also implies an increase in the complexity of estimation due to the volume of data that must be processed. In contrast, when dealing with laboratory experiments or other situations, we may have limited observations making it challenging to obtain robust estimates. To address these challenges, this paper proposes a simple and efficient method for estimating the coefficients of GLM and provides mathematical proof for the almost sure convergence of this method towards the desired solution. The proposed method is also validated on real-world data, reinforcing its utility and effectiveness.

Key Words: GLM, Stochastic approximation, Conditional expectation, Stochastic gradient.

Contents

1. Introduction

The generalized linear model (GLM), initially developed in 1972 by Nelder and Wedderburn (cited in [1]), and further detailed in Nelder and Mc Cullagh (1983) ([2]), Agresti (1990) ([3]), and Antoniadis et al. (1992) ([4]), is a generalization of the multiple linear regression model in which the variable that we seek to explain no longer follows a normal distribution, but instead has a probability density that belongs to the exponential family of densities. Therefore, it is often more appropriate to build a linear model for the transformation of the expected value of the response, such as for $U = (U_1, U_2, ..., U_n)'$ and for the observation i^{th} we have $V_i = (V_{i1}, V_{i2}, ..., V_{ip})$ and let θ be a vector of unconditional parameters with $\theta = (\theta_1, \theta_2, ..., \theta_p)'$ the model can be written as:

$$l^{-1}(E(U|V)) = V\theta. \tag{1.1}$$

The function l s the link function, and the right side of the equation is the linear component of the model. The GLM model can be specified by the following three elements:

- the probability density of U wwhich belongs to the exponential family of densities, including the Gaussian, Poisson, Binomial, Gamma, and other distributions,
- the link function, which allows us to use a linear model to model non-normal data,
- the linear component.

2010 *Mathematics Subject Classification:* 62J12, 62L20.
Submitted May 02, 2022. Published December 29, 2022

There are two main methods for estimating the θ parameter that maximizes the likelihood, the iteratively reweighted least squares and the Fisher scoring method. The iteratively reweighted least squares method involves reformulating the maximization problem by adding weights at each iteration, as proposed in [1] (for more information, see [5]). The Fisher scoring method can be written as follows:

$$\theta_{n+1} = \theta_n - Jac\ S(\theta_n)^{-1} S(\theta_n), \tag{1.2}$$

with $S(\theta_n) = \frac{\partial L(\theta_n)}{\partial \theta_n}$ and $L()$ is the log-likelihood of the GLM for the observations $U_1, U_2, ..., U_n$.

We propose a simple method for estimating the θ parameter for a class of generalized linear model without requiring a specification of the exact distribution of the phenomenon. For example, this method allows us to avoid distinguishing between the Poisson and negative binomial distributions when estimating the GLM regression, which can help us ignore overdispersion phenomena that are commonly observed in random counting variables. Moreover, our method is less computationally intensive, especially in the era of big data when data sets can be massive and classical iterative estimation can be time-consuming. To achieve this, we use the stochastic approximation method, an iterative method for finding the root of an equation using a random process. These algorithms were introduced by Herbert Robbins and Sutton Monro in 1951 [6]. The goal is to find the unique root θ^* of the function $M(\theta) = \alpha$, where $M(\theta)$ cannot be directly observed. However, we can observe the variable $Y(\theta)$ such that $E[Y(\theta)] = M(\theta)$. Under the assumption that $\sum_{n=1}^{\infty} a_n = \infty$ and $\sum_{n=1}^{\infty} a_n^2 < \infty$, the following processes converge to the unique root of $M(\theta) = \alpha$,

$$\theta_{n+1} = \theta_n - a_n(M(\theta_n) - \alpha). \tag{1.3}$$

The method involves starting with an initial estimate of θ and iteratively updating the estimate using the observed variable $Y(\theta)$.

In our case, we are interested in the work of Bennar et al. [7], who established a theorem that identifies the almost sure convergence conditions for a stochastic gradient type process to estimate the θ parameter. We have chosen these results as the foundation of our work because the stochastic gradient process performs sampling at each iteration to complete the estimates without relying on all the available data.

We first present the convergence results elaborated by Bennar et al., then demonstrate that these results are also valid in the framework of certain types of generalized linear model. Next, we conduct a case study on real data to apply the proposed process, and finally, we conclude the work by discussing potential avenues for further development.

2. Preliminaries

Let us consider an observable random variable U and a random variable V with values in $\mathbb{R}^k$ and a law μ. We aim to estimate the parameter $\theta \in \mathbb{R}^p$ such that $\phi(V, \theta)$ approaches $E[U|V]$ in the least squares sense.

Let f be the real positive function defined in $\mathbb{R}^p$ by:

$$f(\theta) = E\left[(E[U|V] - \phi(V, \theta))^2\right], \tag{2.1}$$

we seek the value of θ^* that minimizes the function f.

Let us define the real positive function g in $\mathbb{R}^p$ by:

$$g(\theta) = E\left[(U - \phi(V, \theta))^2\right]. \tag{2.2}$$

We have:

$$g(\theta) = f(\theta) + E\left[(U - E[U|V])^2\right]. \tag{2.3}$$

Therefore, the problem comes down to looking for θ that minimize the function g. We have:

$$\bigtriangledown_\theta g(\theta) = 2E[\nabla_\theta \phi(V, \theta)(\phi(V, \theta) - U)]. \tag{2.4}$$

To estimate θ sequentially, we use a stochastic gradient algorithm that considers a random value θ_n in $\mathbb{R}^p$ and updates it according to the following equation:

$$\theta_{n+1} = \theta_n - a_n \nabla_\theta \phi(V_n, \theta_n)(\phi(V_n, \theta_n) - U_n), \tag{2.5}$$

with:

- (a_n) is a sequence of positive real numbers,
- $(U_1, V_1), (U_2, V_2), ..., (U_n, V_n)$ is a sample of n pairs of independent random variables that have the same probability distribution as (U, V),
- $\phi(.,.)$ is a real known measurable function in $\mathbb{R}^k \times \mathbb{R}^p$.

In this equation, $\nabla_\theta \phi(V_n, \theta_n)$ is the gradient of the function ϕ with respect to θ, evaluated at the current values of θ_n and V_n. The difference between the current value of ϕ and the target value U_n is used to adjust the estimate of θ, and the step size a_n determines how much the value of θ should be adjusted at each iteration.

In the following, $\langle ., . \rangle$ and $\|.\|$ denote the usual scalar product and norm in $\mathbb{R}^k$, respectively. A' denotes the transpose of the matrix A, and $\lambda_{\min(A)}$ represents the smallest eigenvalue of A. The abbreviation "a.s." stands for almost sure convergence.

2.1. Almost Sure Convergence

Bennar et al. consider the following hypotheses:

(H_1) $a_n > 0, \sum_{n=1}^{\infty} a_n^2 < \infty,$

(H_1') $a_n > 0, \sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} a_n^2 < \infty,$

(H_2) there exist a and b such that for all $\theta = (\theta_1, \theta_2, ..., \theta_p)' \in \mathbb{R}^p$,

$$Var\left[\frac{\partial \phi(V, \theta)}{\partial \theta_i}(\phi(V, \theta) - U)\right] < a g(\theta) + b, \text{ for all } i = 1, 2, ..., p, \tag{2.6}$$

(H_3) there exist $K > 0$ such that for all $\theta = (\theta_1, \theta_2, ..., \theta_p)' \in \mathbb{R}^p$,

$$\frac{\partial^2 g(\theta)}{\partial \partial \theta_j} < K, \text{ for all } i = 1, 2, ..., p, \tag{2.7}$$

(H_4) θ^* is a local minimum of g,

$$\exists\ \alpha > 0 :\ (\theta \neq \theta^*,\ \|\theta - \theta^*\| < \alpha \Rightarrow (g(\theta^*) < g(\theta)), \tag{2.8}$$

(H_5) θ^* is the unique stationary point of g,

$$\forall \theta \in \mathbb{R}^p,\ (\theta \neq \theta^*) \Leftrightarrow \bigtriangledown_\theta g(\theta) \neq 0). \tag{2.9}$$

Lemma 2.1. *Under hypotheses, H_1', H_2, H_3, H_4, H_5, we have:*

$$\theta_n \to \theta^* \ a.s \ or \ \|\theta_n\| \to +\infty \ a.s.$$

Proof: See [7]. □

3. Main results

In our case, we aim to estimate the parameters θ of a Generalized Linear Model (GLM) using one of the distributions listed in Table 1. These distributions are useful in various fields, including insurance and medicine, and are tailored to specific types of data. Further details can be found in references [8,9,10,11,12].

Table 1: Common distributions and canonical link functions.

Distribution	Support of distribution	Link name	Link function
Normal	$\mathbb{R}$	Identity	$V\theta = \mu$
Exponential	$\mathbb{R}_+^*$	Negative inverse	$V\theta = -\mu^{-1}$
Gamma	$\mathbb{R}_+^*$	Negative inverse	$V\theta = -\mu^{-1}$
Inverse Gaussian	$\mathbb{R}_+^*$	Inverse squared	$V\theta = \mu^2$
Poisson	$\mathbb{N}$	Log	$V\theta = \ln(\mu)$

However, in practice, we do not limit ourselves to these specific link functions. Let l be a link function such that:

$$l^{-1}(U) = (l^{-1}(U_1), l^{-1}(U_2), ..., l^{-1}(U_n))',$$

and let $\rho_1, \rho_2, ...\rho_p$ be p functions of q measurable, real, and known variables. We note:

$$\rho = (\rho_1, \rho_2, ...\rho_p)'.$$

To estimate θ^*, which minimizes $E[(E[l^{-1}(U)|V] - \theta'\rho(V))^2]$, we use the following stochastic approximation process (θ_n) in $\mathbb{R}^p$:

$$\theta_{n+1} = \theta_n - a_n\rho(V_n)(\rho(V_n)'\theta_n - l^{-1}(U_n)). \tag{3.1}$$

Where $(U_1, V_1), (U_2, V_2), ..., (U_n, V_n)$ is a sample of (U, V) consisting of independent random variables that are identically distributed.

We make the following assumptions:

- (H_6) The functions $\rho_1(V_1), \rho_2(V_2), ..., \rho_p(V_p)$ and $\nabla_\theta\phi(V, \theta)$ are linearly independent,
- (H_7) The fourth moment of the vector $(\rho_1(V_1), \rho_2(V_2), ..., \rho_p(V_p), U)$ exists,
- (H_8) The random variable θ_1 satisfies $[\|\theta_1\|^2] < \infty$,
- (H_9) The variables U and V are observed in finite quantities, such that $U < \infty$ and $V < \infty$.

3.1. Process convergence proof

Theorem 3.1. *Under assumptions, $H_1^{'}$, H_6, H_7, H_8, H_9 we have:*

$$\theta_n \to \theta \ a.s.$$

Proof: Let ϕ be a real-valued function of $\mathbb{R}^q \times \mathbb{R}^p$ defined by:

$$\phi(V, \theta) = \theta'\rho(V) = \sum_{j=1}^{p} \theta_j\rho_j(V_j). \tag{3.2}$$

For $j = 1, 2, ..., p$, $\frac{\partial\phi(V,\theta)}{\partial\theta_i} = \rho_j(V)$, we have $\bigtriangledown_\theta\phi(V, \theta) = \rho(V)$.

We note, $A = E[\rho(V)\rho'(V)]$.

Under (H_{10}), the matrix A is symmetric and positive definite, and therefore invertible.

Then, θ^* is a unique solution of the equation:

$$\bigtriangledown_\theta g(\theta) = 2E[\rho(V)(\rho(V)'\theta - l^{-1}(U))] = 0, \tag{3.3}$$

with:

$$\theta^* = A^{-1}E[\rho(V)l^{-1}(U)]. \tag{3.4}$$

Then, θ^* is the unique solution of the equation $\nabla_\theta g(\theta) = 2E[\rho(V)(\rho(V)'\theta - l^{-1}(U))] = 0$, with

$$\theta^* = A^{-1}E[\rho(V)l^{-1}(U)]. \tag{3.5}$$

Let us prove that assumption 2 is verified.
For $i = 1, 2, ..., p$, we have:

$$\begin{aligned} Var\left[\frac{\partial\phi(V,\theta)}{\partial\theta_i}(\phi(V,\theta) - l^{-1}(U))\right] &= Var\left[\rho_i(V_i)(\theta'\rho(V) - l^{-1}(U))\right], \\ &\leq E\left[\rho_i^2(V_i)(\theta'\rho(V) - l^{-1}(U))^2\right]. \end{aligned} \tag{3.6}$$

By (H_9), $\exists\ K$ such that:

$$\begin{aligned} Var\left[\frac{\partial\phi(V,\theta)}{\partial\theta_i}(\phi(V,\theta) - l^{-1}(U))\right] &= Var\left[\rho_i(V_i)(\theta'\rho(V) - l^{-1}(U))\right], \\ &\leq E\left[\rho_i^2(V_i)(\theta'\rho(V) - l^{-1}(U))^2\right], \\ &\leq K, \\ &\leq Ag(\theta) + B\ (A = 0,\ B \geq K). \end{aligned} \tag{3.7}$$

Let us prove that assumption 3 is verified.
For $i = 1, 2, ..., p$, we have $\frac{\partial g(V,\theta)}{\partial\theta_i} = 2E[\rho_i(V_i)(\rho(V)'\theta - l^{-1}(U))]$
Thus, for $i, j = 1, 2, ..., p$,we have $\frac{\partial^2 g(V,\theta)}{\partial\theta_i\partial\theta_j} = 2E[\rho_i(V_i)\rho_j(V_j)]$, regardless of θ.
The assumptions of Lemma 2.1 are verified, therefore,

$$\theta_n \to \theta^*\ a.s\ or\ \|\theta_n\| \to +\infty\ a.s.$$

It remains to be proven that we cannot have $\|\theta_n\| \to +\infty\ a.s.$
In fact, as Bennar et al. ([7], Lemma 1) proved that $\sum_{n=1}^{\infty} a_n \|\nabla_\theta g(\theta_n)\| < \infty$ a.s., therefore there exists a subsequence of integers (n_l) such that $\|\nabla_\theta g(\theta_{n_l})\| \to 0$ a.s.
Moreover, $\nabla_\theta g(\theta) = 2E[\rho(V)(\rho(V)'\theta - l^{-1}(U))] = 2A(\theta - \theta^*)$.
Therefore, $\|\nabla_\theta g(\theta_n)\|^2 \geq 4\lambda^2_{\min(A)} \|\theta_n - \theta^*\|^2\ (\lambda_{\min(A)} > 0)$.
Thus, if $\|\theta_n\| \to +\infty\ a.s$ then $\|\nabla_\theta g(\theta_n)\| \to +\infty\ a.s$ Which is absurd. □

Therefore, using a stochastic gradient algorithm guarantees the estimation of the parameters of the GLM model due to the almost sure convergence established under the mentioned assumptions.

In the next section of the paper, we conduct a data analysis and modeling to demonstrate the empirical interest of our process. To do this, we use the Ornstein data set, which is a collection of observations or measurements on the number of interlocks within enterprises, obtained from [13]. By analyzing and modeling this data set, we can test and validate our process, and show that it is capable of producing meaningful and accurate results when applied to real-world data. This will help us to prove the empirical interest of our process, and to establish its validity and usefulness for further research and development.

4. Case study: Interlocks modeling

The table that we used in our analysis contains four variables, two of which are continuous and two of which are categorical. For the purposes of simplifying data visualization, we focused on the continuous variables in our analysis. The descriptive table (Table 2) provides summary statistics for these two continuous variables. Additionally, Figure 1 illustrates the relationship between these two continuous variables, through the use of a scatterplot.

Table 2: Descriptive table of variables.

Statistic	N	Mean	St. Dev.	Min	Max
Assets	248	5,978.440	16,712.480	62	147,670
Interlocks	248	13.581	16.083	0	107

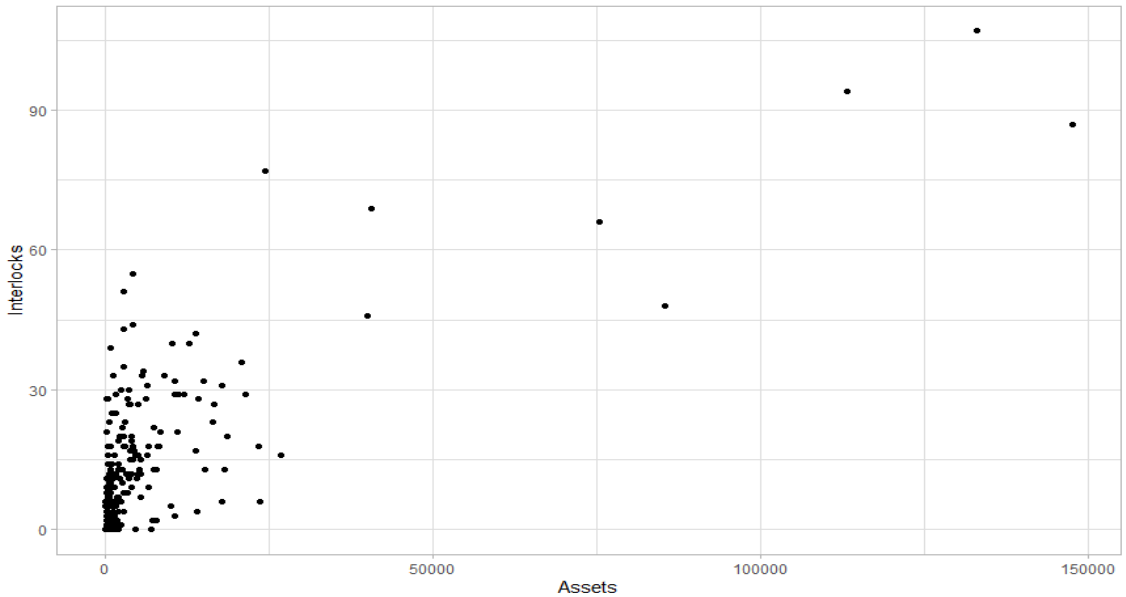

Figure 1: Assets VS Interloks.

The assets variable refers to a quantitative measure or characteristic of some entity, such as a company or individual. In our analysis, we found that this variable is highly dispersed, meaning that its values are spread out over a wide range. To address this issue and make the data more manageable, we applied a logarithmic transformation to the assets variable, using the log base 10 function. This transformation reduces the dispersion of the data and makes it easier to visualize and analyze. The resulting transformed data is plotted in Figure 2, which shows the distribution of the transformed assets variable.

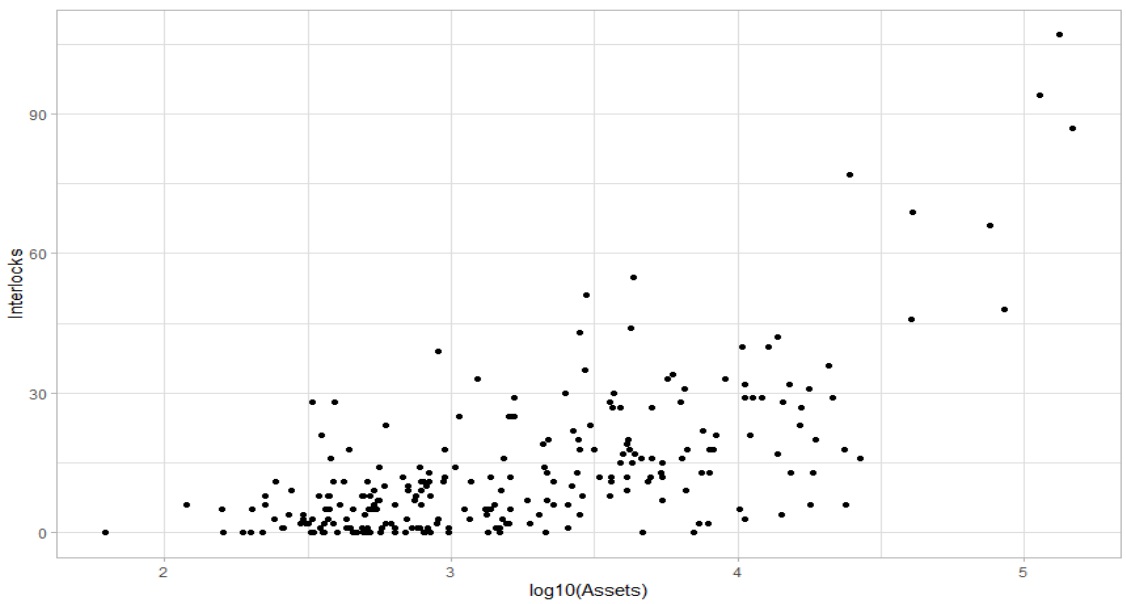

Figure 2: $\log_{10}$(Assets) VS Interloks.

To conduct our modeling, we first randomly selected 70% of the available observations to create the

learning sample. The remaining 30% of observations were set aside as the test sample. We used the learning sample to fit our statistical model. The test sample was then used to evaluate the performance of the model and assess its accuracy and predictive power. The results of the modeling are presented in Table 3, which provides summary statistics and other relevant information about the model fit and performance.

Table 3: Poisson regression results.

	Dependent variable:
	Interlocks
Log10(Assets)	1.065^{***}
	(0.025)
Constant	-1.031^{***}
	(0.098)
	Observations
Observations	178
Null deviance	3029.2 (df=177)
Residual deviance	1397.0 (df=176)
Log Likelihood	−1,024.377
Akaike Inf. Crit.	2,052.753
Number of Fisher Scoring iterations	5
Note:	*p<0.1; **p<0.05; ***p<0.01

The residual deviation to degrees of freedom ratio, also known as the dispersion parameter, is calculated as the ratio of the residual deviation to the degrees of freedom. In our case, this ratio is reported as 1397.0/176, or 7.9375. A ratio that is largely superior to 1 indicates the presence of over dispersion, which means that the data exhibit more variability than expected based on the statistical model being used. In this case, it would be necessary to use a different error structure in the regression model to account for the over dispersion.

One possible approach to addressing over dispersion is to use a negative binomial error structure in the regression model. This error structure is suitable for modeling data with excess variability and can improve the fit and accuracy of the model. The results of estimating the model using a negative binomial error structure are presented in Table 4, which provides summary statistics and other relevant information about the model fit and performance.

We can observe an improvement in the quality of the estimate when comparing the results of this model to the first model. The dispersion parameter, which measures the level of overdispersion in the data, is now closer to 1, indicating that the model is better able to capture the variability in the data. Additionally, the Akaike information criterion (AIC) is lower in this model, indicating that it provides a better fit to the data while using fewer parameters. Overall, these results suggest that the negative binomial error structure is a more appropriate choice for modeling the data given the presence of overdispersion.

In our estimation using stochastic approximation, we employed the log link function and selected a sequence of values for the learning rate, defined as $a_n = \frac{1}{n^2}$. We also specified the starting point for the estimation process as $Constant = 0$ and $\theta = 0$. To assess the goodness of fit of the model, we used the mean absolute error (MAE) as a performance metric. The MAE is calculated as the average absolute difference between the predicted values ($\hat{u}_i$) for each observation and the corresponding actual values (u_i), therefore $MAE = \frac{\sum_{i=1}^{n} |\hat{u}_i - u_i|}{n}$. A lower MAE indicates a better fit of the model to the data.

It is important to recall that the proposed process involves randomly selecting a pair of observations from the available data at each iteration, and using these observations to perform the necessary calcu-

Table 4: Negative binomial regression results.

	Dependent variable:
	Interlocks
Log10(Assets)	1.084^{***}
	(0.106)
Constant	-1.097^{***}
	(0.354)
	Observations
Observations	178
Null deviance	331.15 (df=177)
Residual deviance	211.02 (df=176)
Log Likelihood	−613.685
Dispersion parameter for Negative Binomial	1.309^{***} (0.168)
Akaike Inf. Crit.	1,231.370
Number of Fisher Scoring iterations	1
Note:	*p<0.1; **p<0.05; ***p<0.01

lations and update the model estimates. As a result, the number of observations used in the estimation process is equal to the number of iterations that were carried out. Table 5 presents a summary of the estimation process, including details such as the number of iterations, the number of observations used, and yhe MAE. It should be noted that the calculations were performed using the R programming language.

Table 5: Estimation process summary.

Iteration	MAE	Constant	θ
1	13.581	0	0
2	96,380.290	-2.641	3.487
3	10.933	-2.415	1.047
4	7.322	-2.401	1.347
5	7.532	-2.405	1.322

The estimation process converged after 5 iterations, during which it used a total of 5 observations to achieve an MAE of 7.53 on the set of available observations. However, when we applied the negative binomial model distribution to the test sample, we obtained an MAE of 7.37. This suggests that the negative binomial model may be a slightly better fit for the data, at least in terms of the MAE metric. Figure 3 provides a visual comparison of the two models, with the negative binomial model (NB) represented in red and the result of the stochastic approximation (SA) in blue.

From the visual comparison in Figure 3, we can see that there is some discrepancy in the predictions made by the negative binomial model and the stochastic approximation. One possible reason for this discrepancy is that the stochastic approximation method is not as sensitive to outliers as the negative binomial model. Outliers are observations that are significantly different from the majority of the data, and can have a strong influence on the estimates produced by statistical models. In the case of the stochastic approximation, the small number of observations used in the estimation process may not include these particular outliers, leading to estimates that are more closely aligned with the mean of the data rather than being influenced by the outliers.

In addition to potentially producing different estimates, the use of stochastic approximation has the advantage of being more computationally efficient, as it requires fewer observations and may not require

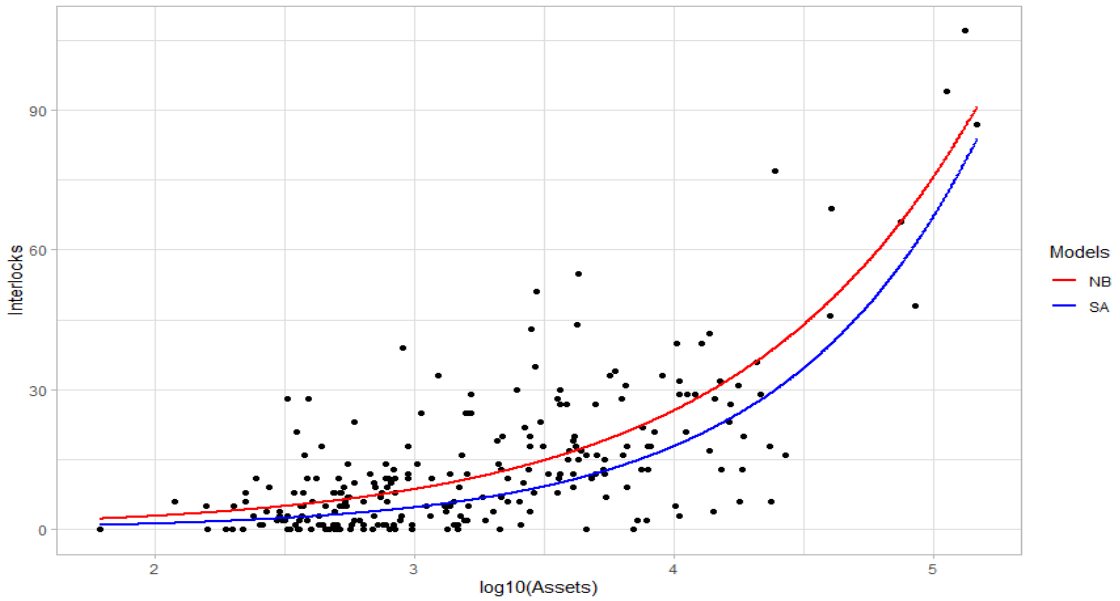

Figure 3: Negative binomial (NB) versus stochastic approximation (SA) modeling.

as much preprocessing of the data. This can be a useful consideration when dealing with large data sets or when time is limited.

The results of the comparison between the negative binomial model and the stochastic approximation highlight the advantages of using this type of estimation in certain situations. The quality of the estimate, as measured by the MAE, is one important factor to consider, as it reflects the accuracy and precision of the model. In addition, the speed of convergence and the number of observations used are other important considerations, particularly when the cost of obtaining additional observations is high or when the available data is very large.

In experimental environments where these factors are of concern, the use of stochastic approximation can be particularly beneficial. By using a small number of observations and achieving convergence quickly, the method can provide accurate and efficient estimates while minimizing the cost and time required for data collection and analysis.

5. Conclusion

In this study, we successfully demonstrated the almost certain convergence of the stochastic gradient process, which is a statistical method for estimating model parameters using a sequence of observations selected at random from a larger data set. One major advantage of this result is the potential for significant time and computational savings, especially when working with large data sets or when the number of available observations is limited. By basing the estimation process on a sample of the population rather than the entire data set, the stochastic gradient process can provide accurate estimates in a more efficient manner.

While the choice of $a_n = \frac{1}{n^2}$ for the learning rate sequence produced satisfactory results in this study, it is possible that other sequences could lead to even faster convergence. Further research could focus on identifying the optimal sequence of a_n to achieve the most efficient and accurate estimates using the stochastic gradient process.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

1. Nelder, J. A., & Wedderburn, R. W., *Generalized linear models*, Journal of the Royal Statistical Society: Series A (General) 135(3), 370-384, (1972).
2. McCullagh, P., & Nelder, J. A., *Generalized linear models*. 37, (1983).
3. Agresti, A., *Categorical Data Analysis*. (1990).
4. Antoniadis, A., Berruyer, J., & Carmona, R., *Régression non linéaire et applications*, Economica. (1992).
5. Madsen, H., & Thyregod, P., *Introduction to general and generalized linear models*, CRC Press. (2010).
6. Robbins, H., & Monro, S., *A stochastic approximation method*, The annals of mathematical statistics 400-407, (1951).
7. Bennar, A., Bouamaine, A., & Namir, A., *Almost sure Convergence and in quadratic mean of the gradient stochastic process for the sequential estimation of a conditional expectation*, Applied Mathematical Sciences 2(8), 387-395, (2008).
8. De Jong, P., & Heller, G. Z., *Generalized linear models for insurance data*, Cambridge Books. (2008).
9. Adesina, O. S., *Models for Zero Truncated Count Data in Medicine and Insurance*. (2021).
10. Gao, Z., & Hastie, T., *DINA: Estimating Heterogenous Treatment Effects in Exponential Family and Cox Models*, arXiv preprint arXiv:2103.04277. (2021).
11. Chen, X., Aravkin, A. Y., & Martin, R. D., *Generalized linear model for gamma distributed variables via elastic net regularization*, arXiv preprint arXiv:1804.07780. (2018).
12. Sudarwanto, S., Ambarwati, L., & Hadi, I., *Development of rental property insurance models with Generalized Linear Models (GLM)*, Journal of Physics: Conference Series, IOP Publishing 1402, (7), 077104, (2019).
13. Fox, J., Friendly, G. G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., ... & Suggests, M. A. S. S., *The car package*, R Foundation for Statistical Computing 1109, (2007).

Ali Labriji,
Department of Mathematics and Computer Science,
Faculty of Sciences Ben M'Sik Hassan II University
Morocco.
E-mail address: alilabriji@gmail.com

and

Safae Msellek,
Department of Mathematics and Computer Science,
Faculty of Sciences Ben M'Sik Hassan II University
Morocco.
E-mail address: msellek.safae@gmail.com

and

Abdelkrim Bennar,
Department of Mathematics and Computer Science,
Faculty of Sciences Ben M'Sik Hassan II University
Morocco.
E-mail address: bennar1@yahoo.fr