МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ

UDC 004.852
DOI: 10.20535/SRIT.2308-8893.2023.2.10

ESTIMATION OF THE PARAMETERS OF GENERALIZED LINEAR MODELS IN THE ANALYSIS OF ACTUARIAL RISKS

R.S. PANIBRATOV, P.I. BIDYUK

Abstract. Methods of estimating the parameters of generalized linear models for the case of paying insurance premiums to clients are considered. The iterative-recursive weighted least squares method, the Adam optimization algorithm, and the Monte Carlo method for Markov chains were implemented. Insurance indicators and the target variable were randomly generated due to the problem of public access to insurance data. For the latter, the normal and exponential law of distribution and the Pareto distribution with the corresponding link functions were used. Based on the quality metrics of model learning, conclusions were made regarding their construction quality.

Keywords: actuarial risk, generalized linear models, simulation modeling, exponential family of distributions, iterative-recursive weighted least squares method, Adam method, Monte Carlo method for Markov chains.

INTRODUCTION

Actuarial risk is classified as the risk that the assumptions implemented by actuaries in the model for estimating the prices of insurance policies may be inaccurate or incorrect. This term is also identified as “insurance risk”. The level of actuarial risk is directly proportional to the reliability of the assumptions implemented in the pricing models used by insurance companies to set premiums. Probability estimates are used to set the price of insurance policies, allowing insurers to implement payments subject to normal business operations. If the proposed hypotheses are wrong, the events that were not considered become the reasons for increasing the frequency of payments, which, in turn, leads to serious financial consequences for the insurer.

Monograph [1] provides useful methodology for system analysis of complex processes. The methodology was applied by the authors to analysis of actuarial processes. In studies [2; 3], mathematical modeling issues related to complicated non-stationary processes and systems are discussed.

Generalized Linear Models (GLM) allow making explicit assumptions about the nature of the insurance data and their relationship with the predicted variables. Furthermore, GLM provides statistical diagnostics that helps in selecting only significant variables and testing model assumptions. This approach is widely recognized as a standard way of pricing for different types of insurance in different markets of countries.

The GLM consists of a wide variety of models, including the linear regression model as a special case. Assumptions for the latter, usually including normal distribution, constant variance, and additivity of effects, are rejected. For example, the target variable can be taken from an exponential family of distributions [4].

The exponential family of distributions has the following general form:

$$f_i(y_i, \theta_i, \varphi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\varphi)} + c(y_i, \varphi)\right\},$$

where $a_i(\varphi)$, $b(\theta)$, and $c(y_i, \varphi)$ are functions, that are defined at the beginning; θ_i is a parameter related to the mean value; φ is a scale parameter related to variance.

The variance can vary together with the mean of the distribution. The influence of explanatory variables is assumed to be additive on an another scale. The following assumptions are made for GLM:

Stochastic component: each component of Y is independent and is taken from the one distribution of exponential family.

Systematic component: p covariates (explanatory variables) form linear predictor η:

$$\eta = X\beta.$$

Link function: the relationship between the random and systematic components is established through a link function that is differentiable and monotonic:

$$E[Y] = \mu = g^{-1}(\eta).$$

PROBLEM STATEMENT

The purpose of the study is to implement methods for estimating GLM parameters for the analysis of actuarial risks, using different distribution laws and link functions for the predicted variable.

METHODS OF ESTIMATING PARAMETERS OF GENERALIZED LINEAR MODELS

GLM parameter estimation is a rather important issue and deserves appropriate attention. The following algorithms were used to evaluate parameters for the purpose of comparing methods: iterative-recursive weighted least squares method; Adam optimization algorithm, Monte Carlo method for Markov chains.

Iterative-recursive weighted least squares method

The heteroskedastic model can be adjusted using the weighted least squares method (WLS),

$$\tilde{\beta} = (X^T W X)^{-1} X^T W y,$$

where y is a target variable; $W = Diag\left[Var(y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 \right]^{-1}$ is a diagonal matrix of weights; X is a matrix of covariates.

We will use Fisher's scoring [5]:

$$\beta^{(t+1)} = (X^T W X)^{-1}(X^T W \beta^{(t)} + X^T A(y - \mu)),$$

where $A = Diag\left[Var(y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) \right]^{-1}$.

The matrices A and W are related in a next way

$$A = W\left(\frac{\partial \eta}{\partial \mu}\right) = W\, Diag\left(\frac{\partial \eta_i}{\partial \mu_i}\right).$$

Then equation for estimating parameters of GLM can be rewritten as follows:

$$\beta^{(t+1)} = (X^T W X)^{-1}(X^T W \beta^{(t)} + X^T W z),$$

where $z = \eta + \left(\frac{\partial \eta}{\partial \mu}\right)(y - \mu)$.

At each step of algorithm:

1. The current estimate of β is used to calculate a new working variable z and a set of weights W.

2. Update (regress) z using X and W to obtain new values.

Iteratively reweighted least squares with random effects (IRWLSR), which was proposed in [6], successfully overcomes the challenge posed by high dimensional intractable integrals in fitting GLMMs for non-normal data. IRWLSR is used for Maximum Likelihood estimations of generalized linear mixed effects models (GLMMs). The benefit is that a working linear mixed effects model (WWLMM) for normal data performs the bulk of the calculation and prediction. Only the working responses and weights in the WWLMM are updated using the GLMM distribution and link function. The complete algorithm can be applied with ease even if high-dimensional intractable integrals are present in the likelihood function because it does not require any numerical evaluations of intractable integrals.

Adam (Adaptive moment estimation)

The adaptive movement estimation algorithm or Adam is an extension of the gradient optimization algorithm. It was created to speed up the optimization process to improve the ability of the optimization algorithm. This is achieved by increasing the step size for each input parameter being optimized. Each step size is automatically adapted through a search process based on partial derivatives or gradients for each input variable. This involves computing the first and second moments of the gradient as an exponentially decreasing first moment and second moment for the input variables.

First, the moment vector and exponentially weighted norm at infinity are adjusted for each optimization parameter, which is denoted as *m* and *v*. Their initial values are equal to zero.

The algorithm consists of the following steps [7]:

The gradient of the target function for the current step is calculated:

$$g(k) = f(x(k-1)).$$

1. Next, the first moment is updated using the gradient and the hyperparameter β_1:

$$m(k) = \beta_1 m(k-1) + (1-\beta_1) g(k).$$

2. The second moment is then updated using the gradient square and the hyperparameter β_2:

$$v(k) = \beta_2 * v(k-1) + (1-\beta_1) * g^2(k).$$

The last two values are biased because their initial values are equal to zero.

3. The first and second moments are adjusted according to the following formulas:

$$\tilde{m}(k) = \frac{m(k)}{1-\beta_1^k}; \quad \tilde{v}(k) = \frac{v(k)}{1-\beta_2^k}.$$

4. The optimum point is calculated by the expression:

$$x(k) = x(k-1) - \alpha * \frac{\tilde{m}(k)}{\sqrt{\tilde{v}(k)} + \varepsilon},$$

where α is the hyperparameter of step size; ε is a small value that ensures there will not be zero division error.

According to [8], the Adam optimization algorithm combining adaptive coefficients and composite gradients based on randomized block coordinate descent is proposed, which improves the algorithm's performance. Starting with the improvement of the Adam algorithm to accelerate the convergence speed, accelerate the search for the global optimal solution, and enhance the high-dimensional data processing ability. The suggested approach improves the Adam optimization algorithm's performance to some degree, but it ignores the effects of second-order momentum and various learning rates on the performance of the original algorithm.

Quasi-Hyperbolic Momentum (QHM) and Quasi-Hyperbolic Adam (QHAdam) are computationally affordable, easily understood, and straightforward to use. They were implemented in [9] by the authors. In many situations, they can be great substitutes for momentum and the Adam algorithm. They specifically make use of high exponential discount factors possible by employing rapid discounting.

As objective functions, the log-likelihood functions of the distributions of the predicted variables were used with their link functions.

Bayesian approach for parameter estimation

The flexibility of the Bayesian technique to adapt, or the capacity to estimate model parameters iteratively in the situation where new measurements are continually coming in, is one of its benefits. The Monte Carlo technique for Markov chains (MCMC) may be used to estimate the model's parameters to the data at a particular step *i*, and in general, the process can be expressed as follows:

$$\tilde{\theta}_i = \tilde{\theta}_{i-1} + \alpha\Delta\tilde{\theta}(i),$$

where $\tilde{\theta}_i$ is an estimation of some parameter θ of the model at step *i*; α is a weight factor; $\Delta\tilde{\theta}(i)$ is the increment in the value of the estimate to be evaluated.

The Markov condition for representing stochastic processes using Markov chains is satisfied since the estimate of the parameter at the subsequent stage solely depends on the estimate at the preceding phase. The Monte Carlo approach is used to generate data from the suitable distribution of the parameter, which is required in order to estimate the rise in the estimate. Therefore, the Monte Carlo approach for Markov chains gets its broad name.

This architecture has concerns with re-estimating the parameter distribution in response to fresh observations and producing data from the proper distribution for estimating unknown parameters. The MCMC method may be used to answer these queries.

Typically, the Bayesian method for creating a regression model looks like follows:

$$y_k \sim N(a_1 x_{k1} + \ldots + a_n x_{kn}, \sigma), \quad k = 1, \ldots, m,$$

where N is the Normal distribution with appropriate parameters.

In general, the distribution type can be chosen based on how the data are actually distributed. It can be rewritten in the matrix form:

$$y \sim N(X\alpha, \sigma_e I), \alpha \sim N(\alpha_0, K_{\alpha,0}),$$

where α_0 is a vector of mean values at zero step; $K_{\alpha,0}$ is the corresponding covariance matrix.

The zero step refers to the initial estimation that was made and will be improved as new data and slow adaptation become available.

Ordinary Least Squares (OLS) can be used when the measured factors and the dependent variable are both normally distributed random variables. The convenience of MCMC, however, comes from the ability to update the values of a few distribution parameters without performing a full recalculation of all matrix operations. This is made possible by the so-called property of conjugate prior distributions from the family of normal, gamma, and their inverse distributions.

The regression model has the following form:

$$y = a + bx + \varepsilon.$$

Primary estimation of vector α_0 can be written in the next form:

$$\alpha_0 = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$$

and covariance matrix:

$$K_{\alpha,0} = \begin{bmatrix} \sigma^2_{a,0} & 0 \\ 0 & \sigma^2_{b,0} \end{bmatrix}.$$

Applying the aforementioned property to each new pair of observations, it is possible to arrive at a formula for the updated values (t_1, s_1), (t_2, s_2) of the distribution parameters:

$$\alpha_1 = K_{\alpha,1}(K_{\alpha,0}^{-1}\alpha_0 + K_y^{-1}\mu_y);$$

$$K_{\alpha,1} = (K_{\alpha,0}^{-1} + K_y^{-1})^{-1};$$

$$\mu_y = \left[\frac{t_1 s_2 - t_2 s_1}{t_1 - t_2},\ \frac{s_1 - s_2}{t_1 - t_2}\right]^T;$$

$$K_y = \begin{bmatrix} \left[\left(\frac{t_1}{t_1 - t_2}\right)^2 + \left(\frac{t_2}{t_1 - t_2}\right)^2\right]\sigma_e^2 & 0 \\ 0 & 2\left(\frac{1}{t_1 - t_2}\right)^2 \sigma_e^2 \end{bmatrix}.$$

However, in the majority cases, when working with non-stationary processes, there is a case of probabilistic uncertainty, or uncertainty about the distribution, which means it is impossible to guarantee an unchanged normal distribution of variables and parameters, making it impossible to always use the above property. It is also theoretically incorrect to continue using OLS because the conditions required for its use are broken. Even in situations when the kind of distribution sought is unknown, MCMC offers a method for approximating the unknown distribution by producing data for the posterior distribution.

Let's say that an unknown parameter's distribution has to be calculated. To do this, it is recommended that a large number of points be produced from this distribution, from which the required distributional parameters may then be estimated using the law of large numbers. Nevertheless, as the posterior distribution of the parameters is unknown, a tool for such creation is required. The Metropolis-Hastings method is a potent tool for resolving the given issue. The algorithm's primary steps may be summarized as follows:

1. The method begins by working with an initial value α_0, which can initially be any random integer.

2. The next step is to choose a new value for α^* that is suggested as being produced from the chosen distribution. Some propositional function $g(\alpha^*, \alpha_0)$ serves as the foundation for the production of α^*. The function g is often selected to be Gaussian and must be symmetric. The Bayes theorem is used to account for the effect of the newly acquired observations in this process.

3. Then calculations are made to determine the resulting value's acceptance rate β:

$$\beta = \frac{g(y \mid \alpha^*)g(\alpha^*)}{g(y \mid \alpha_0)g(\alpha_0)}.$$

4. If $u \leq \beta$, or the newly created point α^*, is accepted and becomes α_1, a random number $0 \leq u \leq 1$ is chosen; if not, the value is rejected, and the procedure stays at the previous point.

5. Until there are sufficient observations, steps 2, 3, 4 are repeated.

The Metropolis-Hastings method asserts that the remaining points originate from the posterior distribution $f(\alpha|y,X)$ after certain initial values are discarded since the values created in the first stages are not yet stable.

The computation of new posterior distributions is repeated using the considered technique when new data is introduced since the prior posterior distributions, according to the Bayes scheme, become a priori for the following iteration.

In [10], authors analyzed insurance claim data and represented the total claims during a specific time frame as a random sum of individual claims' individual positive random variables. According to the data, the variances of both the random total and the random claim size are as big as cubic powers of their respective means. They used natural exponential family modeling to match distributions with cubic variance functions to the insurance data. The authors took into account three discrete counting factors for the random sum and three positive continuous distributions for the claim size, both of which were derived from natural exponential families. In order to create sampling algorithms, they provided a comprehensive study of the nontrivial discrete counting variables. They ran Monte Carlo simulations to compute tail probabilities, particularly for big losses, using samples from the overall claim distribution. Two methods increased these models' effectiveness. The first is that convolutions belong to the same family as the individual distribution because the claim size distributions fulfill the reproducibility condition. The use of importance sampling is the second enhancement.

In [11], authors presented a variety of machine learning regression approaches ranging from multivariate adaptive regression splines and kernel regression to ordinary and generalized least-squares regression variants over GLM and generalized additive model (GAM) approaches for high-dimensional variable selection applications, such as the calibration step in the least-squares Monte Carlo (LSMC) framework. Regression algorithms proved to be suitable machine learning methods for proxy modeling of life insurance companies in their slightly disguised real-world example and given LSMC setting, with potential for both performance and computational efficiency gains by fine-tuning model hyper-parameters and implementation designs. After all, none of the author's tried and true methods could entirely remove the bias seen in the validation figures and produce findings that were consistent throughout the validation sets. The range of suggested regression techniques can be further reduced by looking at whether these findings are systematic for the approaches, the consequence of the Monte Carlo error, or a mix of the two. Although though such assessments would be quite expensive computationally, they would be extremely helpful in revealing new ways to improve the quality of approximations.

NUMERICAL EXPERIMENT

Insurance data is not always publicly available, so it was decided to generate insurance indicators and target variables randomly. The data consisted of the following variables:

- age (range from 18 to 64 years);
- sex;
- body mass index (normal distribution was used);

- number of children (range from 0 to 5);
- smoker status;
- region (generated from sample view [‘north’, ‘south’, ‘east’, ‘west’, ‘center’]);
- charges.

The last variable is the target and the following distribution laws with the corresponding link functions were used for it:

- normal distribution with a known variance σ and a logarithmic link function;
- exponential distribution with the identity link function;
- pareto distribution with a known scale parameter x_m and a link function of the form $f(x) = -1 - x$.

Gaussian noise with variable variance, which is a linear function, was added to the predicted variable.

RESULTS OF THE EXPERIMENTS

The following metrics were used to estimate the quality of the models:

Mean squared error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2 .$$

Root mean squared error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2} .$$

Mean absolute error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| .$$

The results of the estimation of GLM parameters using three methods are presented in Tables 1, 2, 3.

Table 1. Results of GLM construction for a target variable with a Gaussian distribution with known variance and a logarithmic link function

Metric	MCMC	ADAM	IRWLS
MSE	4410.78	686.65	2458.21
RMSE	66.41	26.20	49.58
MAE	51.88	25.81	49.58

Table 2. Results of GLM construction for the target variable with a Pareto distribution with a known scale parameter and a negative linear link function

Metric	MCMC	ADAM	IRWLS
MSE	52725.56	82188.07	638121.57
RMSE	229.62	286.69	798.83
MAE	154.02	205.31	773.33

Table 3. Results of GLM construction for the target variable with an exponential distribution and an identity link function

Metric	MCMC	ADAM	IRWLS
MSE	213.52	148.95	288.53
RMSE	14.61	12.20	16.98
MAE	11.17	3.64	15.66

According to the results of GLM construction for three cases, it can be seen that in most cases, the Adam method showed quite good results. The MCMC method also showed good results for the case of the Pareto distribution. The final choice of parameter estimation method is made after practical application of models to solve the problem of predicting possible losses.

CONCLUSIONS

The methods of estimating GLM parameters for the analysis of actuarial risks in case of payment of insurance premiums to clients are considered. The following three methods are implemented: the iterative-recursive weighted least squares method, the Adam optimization algorithm, and the Monte Carlo method for Markov chains. Since insurance data are often not publicly available, the insurance indicators and the target variable were randomly generated: age, gender, body mass index, number of children, smoking status, region, and charges. The latter was generated using: a normal distribution with a known variance and a logarithmic link function; exponential distribution with the identity link function; of the Pareto distribution with a known scale parameter and a link function of the form of a negative linear function. According to the results of the experiments, it can be concluded that the Adam optimization algorithm demonstrated good results in most cases. The Monte Carlo method for Markov chains also showed good results for the case of the Pareto distribution.

REFERENCES

1. M.Z. Zgurovsky and N.D. Pankratova, *System Analysis: Problems, Methodology, Applications.* Kyiv: Naukova Dumka, 2011, 727 p.
2. O.G. Nakonechny and S.V. Demydenko, “Minimax mean square estimates of trend in the regression problems,” *Tavria Visnyk of Informatics and Mathematics*, no. 1, pp. 23–30, 2009.
3. O.G. Nakonechny, M.M. Veres, and B.V. Stashchuk, “Guaranteed estimates of linear equations solutions in Hilbert space with random parameters,” *Journal of Computational and Applied Mathematics*, vol. 104, no. 1, pp. 85–96, 2011.
4. D. Anderson et al., *A Practitioner’s Guide to Generalized Linear Models – a foundation for theory, interpretation and application*; 3rd edition. Towers Watson, 2007, 122 p.
5. P. McCullagh and J. Nelder, *Generalized Linear Models*; 2nd edition. Chapman & Hall, 1989, 532 p.
6. T. Zhang, “Iteratively reweighted least squares with random effects for maximum likelihood in generalized linear mixed effects models,” *Journal of Statistical Computation and Simulation*, vol. 91, no. 16, pp. 3404–3425, 2021. doi: 10.1080/00949655.2021.1928127.

7. D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014. doi: 10.48550/arXiv.1412.6980.
8. M. Liu et al., “An Improved Adam Optimization Algorithm Combining Adaptive Coefficients and Composite Gradients Based on Randomized Block Coordinate Descent,” *Computational Intelligence and Neuroscience*, vol. 2023, 14 p., 2023. doi: 10.1155/2023/4765891.
9. J. Ma and D. Yarats, “Quasi-hyperbolic momentum and Adam for deep learning,” *arXiv preprint arXiv:1810.06801*, 2018. doi: 10.48550/arXiv.1810.06801.
10. S. Bar-Lev and A. Ridder, “Monte Carlo methods for insurance risk computation,” *arXiv preprint arXiv:* 1807.02201, 2018. doi: 10.48550/arXiv.1807.02201.
11. A.S. Krah, Z. Nikolić, and R. Korn , “Machine learning in least-squares Monte Carlo proxy modeling of life insurance companies,” *Risks*, vol. 8, no. 1, 21 p., 2020. doi: 10.3390/risks8010021.

Received 26.01.2023

INFORMATION ON THE ARTICLE

Roman S. Panibratov, ORCID: 0000-0002-8604-4420, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: roman.panibratov@gmail.com

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: pbidyuke_00@ukr.net

ОЦІНЮВАННЯ ПАРАМЕТРІВ УЗАГАЛЬНЕНИХ ЛІНІЙНИХ МОДЕЛЕЙ В АНАЛІЗІ АКТУАРНИХ РИЗИКІВ / П.І. Бідюк, Р.С. Панібратов

Анотація. Розглянуто методи оцінювання параметрів узагальнених лінійних моделей для аналізу актуарних ризиків у випадку виплат страхових премій клієнтам. Було реалізовано ітеративно-рекурентно зважуваний метод найменших квадратів, алгоритм оптимізації Adam та метод Монте-Карло для ланцюгів Маркова. Страхові показники та цільова змінна генерувалися випадковим чином у зв’язку з проблемою публічного доступу страхових даних. Для останньої використовувався нормальний та експоненціальний закон розподілу і розподіл Парето з відповідними функціями зв’язку. На основі метрик якості навчання моделей були зроблені висновки щодо їх якості побудови.

Ключові слова: актуарний ризик, узагальнені лінійні моделі, імітаційне моделювання, експоненційна множина розподілів, ітеративно-рекруентно зважуваний метод найменших квадратів, метод Adam,метод Монте-Карло для марківських ланцюгів.