پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

A Comparison of Maximum Likelihood Estimation and Particle Swarm Optimization for Parameter Estimation in Logistic and Probit Regression Models

Seyed Mohsen Mohammadi[1], Reza Etesami[2], Mohsen Madadi[3*]

[1]M.Sc. Student, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.
mohsenmohammadi@math.uk.ac.ir

[2]Ph.D. Student, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.
rezaetesamii@math.uk.ac.ir

[3]Associate Professor, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.
madadi@uk.ac.ir

Abstract

This study compares Maximum Likelihood Estimation (MLE) and Particle Swarm Optimization (PSO) for parameter estimation in logistic and probit regression models, using the Iris and Titanic datasets. After preprocessing and model evaluation, PSO demonstrated superior performance over MLE in both models. The findings demonstrate that metaheuristic algorithms like PSO can enhance the accuracy of classification models, offering valuable insights for researchers choosing optimization methods in statistical modeling.

Keywords: Maximum Likelihood Estimation (MLE), Particle Swarm Optimization (PSO), Logistic Regression, Probit Regression, Parameter Estimation.
Mathematics Subject Classification (2020): 62J12, 62F10, 68T20, 90C59.

1 Introduction

Parameter estimation plays a fundamental role in the development and application of regression models used for classification tasks (Bunke et al, 1977). Traditional statistical methods, such as Maximum Likelihood Estimation (MLE), have long been the standard approach for estimating parameters in logistic and probit regression models due to their theoretical soundness and widespread familiarity (Vasisht, 2007). However, these classical techniques often face limitations when dealing with complex, high-dimensional, or nonlinear problems, which can lead to issues like convergence to local optima or computational inefficiency (Aziz et al., 2016).

*Corresponding author: madadi@uk.ac.ir

In recent years, the advent of artificial intelligence and optimization algorithms has provided alternative methods capable of overcoming some of these challenges (Zga et al., 2025). Among these, metaheuristic algorithms, notably Particle Swarm Optimization (PSO), have gained popularity owing to their ability to efficiently explore complex search spaces and identify optimal or near-optimal solutions. PSO, inspired by the social behavior of flocks of birds or schools of fish, has been successfully applied to various optimization problems in engineering, science, and data science (Sadik et al., 2025).

This study addresses the lack of comparative evaluations between traditional and modern optimization methods by systematically comparing Maximum Likelihood Estimation (MLE) and Particle Swarm Optimization (PSO) for parameter estimation in logistic and probit regression models. Using the Iris and Titanic datasets to represent univariate and multivariate scenarios, the study assesses model performance through mean squared error (MSE) and mean absolute error (MAE). The findings highlight PSO's potential to outperform MLE, underscoring the value of intelligent optimization techniques in statistical modeling.

2 Literature Review

Binary regression is a crucial branch of regression modeling that is applied when the dependent variable assumes only two possible outcomes, such as "yes/no", "success/failure", "buy/not buy", or "diseased/healthy". The central objective in binary regression is to estimate the probability of the occurrence of one of the two possible outcomes, conditional on a set of explanatory variables (Harris, 2021). These independent variables can be either quantitative or qualitative. Based on the values of these variables, binary regression models compute the likelihood of a particular event occurring. Such models are widely used across various scientific disciplines including medicine, management, social sciences, and marketing (Wilson et al., 2024).

Two of the most prominent binary regression models are *logistic regression* and the *probit model*. Logistic regression transforms a linear combination of input variables into probabilities using the logistic (sigmoid) function, providing interpretability and computational convenience. This model is extensively employed in data mining and machine learning tasks (Williams and Jorgensen, 2023).

A central challenge in fitting binary regression models such as logistic or probit regression lies in accurately and efficiently estimating model parameters. Despite using different distributional assumptions (logistic and normal), both models share similar optimization frameworks and estimation methodologies. In this study, we investigate two widely-used parameter estimation methods: Maximum Likelihood Estimation (MLE) and the Particle Swarm Optimization (PSO) algorithm. These methods are compared and evaluated in a tabular format.

2.1 Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation is one of the most widely-used techniques for estimating model parameters in statistical modeling, including logistic and probit regression models (Kasap and Faouri, 2025). The method seeks to maximize the likelihood function, denoted as $L(\boldsymbol{\beta})$. Given a dataset $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$, where y_i is the binary response variable and $\boldsymbol{x}_i$ is the vector of predictors, the likelihood function is defined as:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}), \tag{1}$$

where $f(y_i \mid \boldsymbol{x}_i; \boldsymbol{\beta})$ is the conditional probability density (or mass) function of y_i given $\boldsymbol{x}_i$ and parameter vector $\boldsymbol{\beta}$. For computational simplicity, the log-likelihood function is often used:

$$\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log f(y_i \mid \boldsymbol{x}_i; \boldsymbol{\beta}). \tag{2}$$

The goal is to find the parameter vector $\hat{\boldsymbol{\beta}}$ that maximizes the log-likelihood function:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}). \tag{3}$$

Due to the nonlinearity and non-convexity of the likelihood function in many practical cases, an analytical solution for $\hat{\boldsymbol{\beta}}$ is often intractable. Thus, numerical optimization techniques such as the Newton-Raphson or Fisher Scoring algorithms are employed. These iterative methods update the parameter estimates as follows:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \mathbf{H}^{-1}(\boldsymbol{\beta}^{(t)})\mathbf{s}(\boldsymbol{\beta}^{(t)}), \tag{4}$$

where $\mathbf{s}(\boldsymbol{\beta})$ is the score vector (the gradient of the log-likelihood), and $\mathbf{H}(\boldsymbol{\beta})$ is the Hessian matrix (the matrix of second-order partial derivatives). However, these methods may converge slowly or get trapped in local optima, especially in cases where the likelihood surface has multiple peaks. In such cases, metaheuristic optimization methods like Particle Swarm Optimization can be advantageous.

2.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a nature-inspired, population-based metaheuristic algorithm originally developed to simulate the social behavior of bird flocking or fish schooling. It is particularly suitable for solving complex, non-linear, and non-convex optimization problems, including those involving multi-modal objective functions where classical techniques face convergence difficulties (Abualigah, 2025).

In PSO, each particle represents a potential solution in the parameter space. Over successive iterations, particles adjust their positions and velocities based on their own experience and that of their neighbors, aiming to converge toward the global optimum. The algorithm is governed by the following equations:

Velocity update equation:

$$\boldsymbol{v}_i^{(t+1)} = w\boldsymbol{v}_i^{(t)} + c_1 r_1(\boldsymbol{pbest}_i - \boldsymbol{x}_i^{(t)}) + c_2 r_2(\boldsymbol{gbest} - \boldsymbol{x}_i^{(t)}), \tag{5}$$

Position update equation:

$$\boldsymbol{x}_i^{(t+1)} = \boldsymbol{x}_i^{(t)} + \boldsymbol{v}_i^{(t+1)}, \tag{6}$$

where w is the inertia weight, c_1 and c_2 are cognitive and social acceleration coefficients, respectively, and r_1, r_2 are uniformly distributed random numbers in the interval $[0, 1]$. $\boldsymbol{pbest}_i$ denotes the best position discovered by the i-th particle, while $\boldsymbol{gbest}$ represents the best global position found by the entire swarm (Gad, 2022).

The PSO algorithm proceeds until a stopping criterion is met, such as reaching a maximum number of iterations or achieving convergence in the objective function value.

PSO has been successfully applied in estimating parameters of complex statistical models, including differential equations and nonlinear regressions. Its robustness against multimodal likelihood surfaces and multicollinearity makes it a powerful alternative to classical methods, particularly in noisy and high-dimensional datasets (Shami et al., 2022).

3 Research Methodology

This study compares two prominent methods for estimating the parameters of binary regression models, namely logistic and probit regression. The first method is the classical *Maximum Likelihood Estimation* (MLE), a well-established statistical technique for parameter estimation

in probabilistic models. The second is the *Particle Swarm Optimization* (PSO) algorithm which has demonstrated considerable effectiveness in solving complex nonlinear optimization problems. The primary goal is to evaluate the accuracy and efficiency of these methods in estimating the coefficients of binary regression models.

3.1 Binary Regression Models Under Study

3.1.1 Logistic Regression

Logistic regression is a nonlinear model used when the dependent variable is binary, i.e., $y_i \in \{0, 1\}$. The model estimates the conditional probability of the outcome being 1 using the logistic (sigmoid) function:

$$P(Y = 1|X) = \frac{1}{1 + \exp(-(\beta_0 + \beta^T X))} \tag{7}$$

where X is the vector of independent variables, β_0 is the intercept, and β is the vector of model coefficients (Das, 2024).

MLE of Parameters in Logistic Regression For n independent observations, the likelihood function is given by:

$$L(\beta) = \prod_{i=1}^{n} [P(Y_i = 1|X_i)]^{Y_i} [1 - P(Y_i = 1|X_i)]^{1-Y_i} \tag{8}$$

The log-likelihood is:

$$\ell(\beta) = \sum_{i=1}^{n} [Y_i \log(P(Y_i = 1|X_i)) + (1 - Y_i) \log(1 - P(Y_i = 1|X_i))] \tag{9}$$

The MLE seeks to maximize $\ell(\beta)$:

$$\hat{\beta}_{\text{MLE}} = \arg\max_{\beta} \ell(\beta) \tag{10}$$

Since this optimization problem lacks a closed-form solution, numerical methods such as Newton-Raphson or Fisher Scoring are typically used (Nagarajah and Wijekoon, 2015).

PSO of Parameters in Logistic Regression The PSO algorithm minimizes the negative log-likelihood, which serves as the fitness function:

$$f(\beta) = -\ell(\beta) = -\sum_{i=1}^{n} \left[Y_i \log\left(\frac{1}{1 + \exp(-(\beta_0 + \beta^T X_i))} \right) + (1 - Y_i) \log\left(1 - \frac{1}{1 + \exp(-(\beta_0 + \beta^T X_i))} \right) \right] \tag{11}$$

PSO attempts to minimize $f(\beta)$ by iteratively updating particles representing potential solutions in the search space.

3.1.2 Probit Regression

Probit regression is another method for modeling binary outcomes, but it uses the standard normal cumulative distribution function $\Phi(\cdot)$ instead of the logistic function. The model is derived from a latent variable formulation (Hsiao, 1996):

$$Y_i^* = \beta_0 + \beta^T X_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1) \tag{12}$$

The observed binary response is:

$$Y_i = \begin{cases} 1, & \text{if } Y_i^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Thus, the probability of success is:

$$P(Y=1|X) = \Phi(\beta_0 + \beta^T X) = \int_{-\infty}^{\beta_0+\beta^T X} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz \tag{14}$$

MLE of Parameters in Probit Model The likelihood for n independent observations is:

$$L(\beta) = \prod_{i=1}^{n} \left[\Phi(\beta_0 + \beta^T X_i)\right]^{Y_i} \left[1 - \Phi(\beta_0 + \beta^T X_i)\right]^{1-Y_i} \tag{15}$$

The log-likelihood is:

$$\ell(\beta) = \sum_{i=1}^{n} \left[Y_i \log \Phi(\beta_0 + \beta^T X_i) + (1 - Y_i) \log(1 - \Phi(\beta_0 + \beta^T X_i))\right] \tag{16}$$

The MLE estimate is then:

$$\hat{\beta}_{\text{MLE}} = \arg\max_{\beta} \ell(\beta) \tag{17}$$

PSO of Parameters in Probit Regression Similar to the logistic model, the negative log-likelihood is used as the cost function in PSO:

$$f(\beta) = -\ell(\beta) = -\sum_{i=1}^{n} \left[Y_i \log \Phi(\beta_0 + \beta^T X_i) + (1 - Y_i) \log(1 - \Phi(\beta_0 + \beta^T X_i))\right] \tag{18}$$

3.2 Data Collection and Preprocessing

The datasets used in this research include the well-known `Iris` (Mithy et al., 2022) and `Titanic` (Grandhe et al., 2019) datasets. For the Iris dataset, 100 samples were selected from two balanced classes. In the Titanic dataset, 714 valid samples were extracted after removing missing entries and ensuring class balance. Both datasets were split randomly into 70% training and 30% testing sets. Data preprocessing involved categorical-to-numeric conversion, handling of missing values, and randomization. All preprocessing was conducted in `R` version 4.1.0.

3.3 Performance Evaluation

To comprehensively evaluate the estimation accuracy of MLE and PSO methods, two well-known error metrics were employed:

3.3.1 Mean Squared Error (MSE)

This metric measures the average of the squared differences between the true coefficients (β) and the estimated coefficients ($\hat{\beta}$):

$$\text{MSE} = \frac{1}{K} \sum_{i=1}^{K} (\beta_i - \hat{\beta}_i)^2 \tag{19}$$

A lower MSE indicates higher estimation precision and sensitivity to large deviations or outliers.

3.3.2 Mean Absolute Error (MAE)

MAE evaluates the average absolute difference between the true and estimated parameters:

$$\text{MAE} = \frac{1}{K}\sum_{i=1}^{K}|\beta_i - \hat{\beta}_i| \tag{20}$$

Unlike MSE, MAE is more robust to outliers and offers a more stable measure of average estimation error (Ghareib, 1996).

4 Discussion and Results

In this study, data analysis was conducted on two benchmark datasets: `Iris` and `Titanic`, using both Maximum Likelihood Estimation (MLE) and Particle Swarm Optimization (PSO) methods. For the `Iris` dataset, logistic and probit regression models were fitted using sepal length as the sole predictor. In contrast, for the `Titanic` dataset, age, passenger class, and gender were used as predictors for survival.

The results indicated that the PSO algorithm consistently yielded more stable parameter estimates than the MLE approach in both regression models. Table 1 presents the detailed findings for the `Iris` dataset, including estimated coefficients and error metrics, while Table 2 summarizes the corresponding results for the `Titanic` dataset.

Table 1: Comparison of MLE and PSO methods on the `Iris` dataset

Model	**Intercept** β_0	**Sepal Length Coefficient**	**MSE**	**MAE**
Logistic (MLE)	-35.2310	6.4082	0.1636	0.2615
Probit (MLE)	-19.9398	3.6261	0.1662	0.2650
Logistic (PSO)	-33.2379	6.0748	0.1599	0.2621
Probit (PSO)	-32.2310	5.9007	0.1628	0.2251

Table 2: Comparison of MLE and PSO methods on the `Titanic` dataset

Model	**Intercept** β_0	**Age**	**Class**	**Gender**	**MSE**	**MAE**
Logistic (MLE)	7.1370	-0.0376	-1.1668	-2.4521	0.1297	0.2702
Probit (MLE)	4.1103	-0.0208	-0.6559	-1.4447	0.1305	0.2731
Logistic (PSO)	8.1592	-0.0457	-1.4083	-2.7115	0.1291	0.2538
Probit (PSO)	6.5213	-0.0376	-1.1537	-2.0961	0.1291	0.2209

5 Conclusion

This research aimed to compare the accuracy of parameter estimation in binary classification regression models, specifically logistic and probit models, using both the traditional Maximum Likelihood Estimation (MLE) method and the population-based metaheuristic Particle Swarm Optimization (PSO) algorithm.

The findings revealed that PSO provided meaningful improvements in performance metrics compared to MLE. In the `Iris` dataset, the probit model estimated using PSO achieved the lowest Mean Squared Error (MSE) of 0.1628, outperforming its MLE counterpart (MSE = 0.1662). Similarly, in the `Titanic` dataset, the use of PSO led to a 1.3% reduction in MSE for the probit model.

Additionally, across all models and estimation techniques, the gender variable was consistently the most influential predictor of survival. These outcomes suggest that PSO is a

more robust and efficient alternative for estimating parameters in binary regression models, especially when model stability and error minimization are of concern.

Given these advantages, future research is encouraged to investigate the effectiveness of PSO in other regression structures and across more diverse and complex datasets.

References

Bunke, H., Henschke, K., Strüby, R. and Wisotzki, C. (1977), Parameter estimation in nonlinear regression models, *Statistics: A Journal of Theoretical and Applied Statistics*, **8**(1), 23–40.

Vasisht, AK. (2007), Logit and probit analysis, *IASRI, Library Avenue, New Delhi–110*, **12**.

Aziz, NH Abdul, Ibrahim, Zuwairie, Razali, Saifudin, and Aziz, Nor Azlina Ab. (2016), Estimation-based metaheuristics: a new branch of computational intelligence, *Proc. 3rd National Conf. Postgraduate Research*, pp. 469–476.

Zga, Adel, Zitouni, Farouq, Harous, Saad, Sallam, Karam, Almazyad, Abdulaziz S., Xiong, Guojiang, and Mohamed, Ali Wagdy. (2025), A comparative study of the performance of ten metaheuristic algorithms for parameter estimation of solar photovoltaic models, *PeerJ Computer Science*, **11**, e2646.

Sadik, Laith, Samui, Pijush, Keawsawasvong, Suraparb, Al-Jeznawi, Duaa, and Samui, Ritaparna. (2025), Harnessing Metaheuristics and Probabilistic Machine Learning for Uncertainty-Aware Bearing Capacity Estimation of Shallow Foundations, *Transportation Infrastructure Geotechnology*, **12**(1), 1–31.

Kasap, P., and Faouri, A.O. (2025), Maximum likelihood estimation for the two-parameter Maxwell distribution, *Journal of the National Science Foundation of Sri Lanka*, **52**(4).

Abualigah, L. (2025), Particle Swarm Optimization: Advances, Applications, and Experimental Insights, *Computers, Materials & Continua*, **82**(2).

Shami, T.M., El-Saleh, A.A., Alswaitti, M., Al-Tashi, Q., Summakieh, M.A., and Mirjalili, S. (2022), Particle Swarm Optimization: A Comprehensive Survey, *IEEE Access*, **10**, 10031–10061.

Gad, A.G. (2022), Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review, *Archives of Computational Methods in Engineering*, **29**(5), 2531–2561.

Harris, Jenine K. (2021), Primer on binary logistic regression, *Family Medicine and Community Health*, **9**(Suppl 1), e001290.

Wilson, Jeffrey R., Lorenz, Kent A., and Selby, Lori P. (2024), Introduction to binary logistic regression, *Modeling Binary Correlated Responses: Using SAS, SPSS, R and STATA*, Springer, pp. 3–18.

Williams, Richard and Jorgensen, Abigail. (2023), Comparing logit & probit coefficients between nested models, *Social Science Research*, **109**, 102802.

Das, Abhik. (2024), Logistic regression, In: *Encyclopedia of Quality of Life and Well-Being Research*, pp. 3985–3986, Springer.

Nagarajah, Varathan and Wijekoon, Pushpakanthie. (2015), Stochastic restricted maximum likelihood estimator in logistic regression model, *Open Journal of Statistics*, **5**(7), 837–851.

Hsiao, Cheng. (1996), Logit and probit models, in *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Springer, pp. 410–428.

Ghareib, Ahmed Hamdy. (1996), Evaluation of logit and probit models in mode-choice situation, *Journal of Transportation Engineering*, **122**(4), 282–290.

Grandhe, Padmaja, Damarla, Vishnu Priya, and Mohammad, Shaziya. (2019), Extensive data set analysis & prediction using R, *Journal of Physics: Conference Series*, **1228**(1), 012048.

Mithy, S. A., Hossain, S., Akter, S., Honey, U., and Sogir, S. B. (2022), Classification of iris flower dataset using different algorithms, *International Journal of Scientific Research In.*