Application of maximum likelihood estimation in various mathematical models

Le Fang

Stony Brook Institute, Anhui University, Anhui, 230039, China

r22114037@stu.ahu.edu.cn

Abstract. Maximum likelihood estimation is a breakthrough in the history of statistics, which overcomes the main weakness of Bayesian estimation and has been widely used in various fields, such as language and image processing, and system identification, etc. This paper analyzes the application of maximum likelihood estimation on different mathematical models. It is proved that the universality of maximum likelihood estimation plays an important role in promoting the continued in-depth research on maximum likelihood estimation. This paper also analyzes and summarizes the application of maximum likelihood estimation to specific parameters in different mathematical models. In addition, this work conducts research on the application conditions of maximum likelihood estimation and its main properties, such as variability, consistency, and asymptotic normality, etc. In different mathematical models, such as the annealing furnace efficiency system, gamma environmental factors and adaptive algorithm, etc. Therefore, this paper finds the reason for why the maximum likelihood estimate is widely used in various fields.

Keywords: Maximum likelihood estimation, the annealing furnace efficiency system, gamma environmental factors, adaptive echo cancellation algorithm.

1. Introduction

Maximum likelihood estimation is a parameter estimation method widely used in statistics. It mainly involves the training process of probability models, especially parameter estimation. In the training of probabilistic models, determining parameters is a crucial step, because the selection of parameters will directly affect the performance of the model [1]. The origin of maximum likelihood estimation can be traced back to 1822, when it was first proposed by the German mathematician Gauss when dealing with the normal distribution. It became widely used after British statistician R. A. Fisher demonstrated its correlation properties in 1921. In the 20th century, Wald discussed the asymptotic properties of maximum likelihood estimation in detail in his works, providing a range of theoretical support for the application of maximum likelihood estimation in large sample situations.

In addition to the field of statistics, maximum likelihood estimation is also widely used in other fields, such as machine learning, natural language processing, image processing, etc. For example, in the field of machine learning, maximum likelihood estimation is widely used in parameter estimation of various models. Christopher M. Bishop introduced in detail the application of maximum likelihood estimation in machine learning models such as logistic regression, naive Bayes classifiers, and hidden Markov models in his 2006 book "Pattern Recognition and Machine Learning". It introduces in detail the application of maximum likelihood estimation in machine learning models such as logistic regression,

naive Bayes classifier, and hidden Markov model. In 2012, "Econometric Analysis" written by William H. Greene discussed the application of maximum likelihood estimation in economic models such as time series analysis and regression analysis, as well as related statistical properties and hypothesis testing. At present, with the widespread use of maximum likelihood estimation, there is insufficient research on the applicability range of using it to estimate parameters, which is of great significance to the development of maximum likelihood estimation. The author explores the applicability of parameter estimation and application process through multiple application examples of maximum likelihood estimation.

The aim of this paper is to explore the application of maximum likelihood estimation in mathematical models in different fields. The author will firstly present that in order to improve the online performance of the annealing furnace, he established a mathematical model of the heat balance of the annealing furnace and constructed a data diagnosis method using the idea of maximum likelihood estimation. Next, it also includes the application in annealing furnace online energy efficiency monitoring system, estimating gamma distribution environmental factors, and block-sparse least mean square algorithm [2]. Through the exploration of three mathematical models, this paper finds details of the application of maximum likelihood estimation in different mathematical models and discover the reasons for its widespread use.

2. Principle of preparation

2.1. Basic definitions

Definition 1 [1]: Suppose the population X be a discrete random variable with a probability distribution column $P\{X = x\} = p(x; \theta)$, where $\theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ is an unknown parameter. Suppose $(X_1, X_2, \ldots, X_n)$ Is the sample of size n derived from the population, so that the joint distribution law of $(X_1, X_2, \ldots, X_n)$ is $\prod_{i=1}^n p(x_i; \theta)$. Then suppose a set of observations of $(X_1, X_2, \ldots, X_n)$ is $(x_1, x_2, \ldots, x_n)$, so that the probability of observations derived is

$$L(\theta) = L(x_1, x_2, \ldots, x_n) = \prod_{i=1}^n p(x_i; \theta). \tag{1}$$

It is a function of parameter θ, which is noted as L(θ). And it is called as Likelihood function.

Definition 2 [2]: Suppose $(x_1, x_2, \ldots, x_n)$ is certain, to find the value of θ, which makes

$$L(x_1, x_2, \ldots, x_n; \hat{\theta}) = \max L(x_1, x_2, \ldots, x_n; \theta), \tag{2}$$

where $\hat{\theta}(x_1, x_2, \ldots, x_n)$ is noted as Maximum likelihood estimates, and the corresponding statistic $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is noted as Maximum likelihood estimator, it can also be called $\hat{\theta}$ or MLE.

2.2. Basic theorems

Theorem 1 (Consistency). Suppose the sample of the population X is $X_1, X_2, \ldots, X_k$, and set θ_* be the real value of θ . Then define $M_k(\theta) = \frac{1}{k} \sum_i^k \log \frac{f(X_i; \theta)}{f(X_i; \theta_*)}$, and $M_k(\theta) = -D(\theta_*, \theta)$. Assume that $\sup |M_k(\theta) - M(\theta)| \xrightarrow{q} 0$, and for any $\varepsilon > 0$, $\sup_{\theta: |\theta_* - \theta| \geq \varepsilon} M(\theta) < M_k(\theta)$, then order $\hat{\theta}_k$ represent the maximum likelihood estimation, such that $\hat{\theta}_k \xrightarrow{q} \theta$ [3].

Proof: In order to make $M_k(\theta)$ be maximum, one can get $M_k(\hat{\theta}_k) \geq M_k(\theta_*)$, by $\hat{\theta}_k$ is the maximum likelihood estimate. So that

$$M(\theta_*) - M(\hat{\theta}_k) = M(\theta_*) - M(\hat{\theta}_k) + M_k(\theta_*) - M_k(\theta_*)$$

$$\leq M(\theta_*) - M(\hat{\theta}_k) + M_k(\hat{\theta}_k) - M_k(\theta_*) \leq \sup |M_k(\theta) - M(\theta)| + M(\theta_*) - M_k(\theta_*) \xrightarrow{q} 0. \tag{3}$$

According to $\sup |M_k(\theta) - M(\theta)| \xrightarrow{q} 0$, so that to $\exists\ \delta > 0$, there exists $P(M(\hat{\theta}_k) < M(\theta_*) - \delta) \to 0$, and to $\forall\ \varepsilon > 0$, by $\sup\limits_{\theta:|\theta_* - \theta| \geq \varepsilon} M(\theta) < M_k(\theta)$, there exists $\delta > 0$, $|\hat{\theta}_k - \theta_*| \geq \varepsilon$, it sets $M(\theta) < M(\theta_*) - \delta$, then there exists

$$P(|\hat{\theta}_k - \theta_*| \geq \varepsilon) \leq P(M(\hat{\theta}_k) < M(\theta_*) - \delta) \to 0. \tag{4}$$

Theorem 2 (Asymptotic normality). Assume that $\hat{\theta}_k = \hat{\theta}(X_1, X_2, \ldots, X_k)$ is an estimate of θ, under appropriate regular conditions, set $\sigma_k = \sqrt{V(\hat{\theta}_k)}$, the following equation holds that when $\sigma_k \approx \sqrt{1/I_k(\theta)}$, then $\hat{\theta}_k - \theta/\sigma_k \xrightarrow{L} N(0,1)$. In the other hand, when $\sigma_k = \sqrt{1/I_k(\theta)}$, then $\hat{\theta}_k - \theta/\widehat{\sigma}_k \xrightarrow{L} N(0,1)$ [3].

Theorem 3 (Covariance). Assume $\hat{\theta}$ be the maximum likelihood estimate of θ. If the continuous function of θ is $g(\theta)$, then the maximum likelihood estimates of $g(\theta)$ is $g(\hat{\theta})$. Particularly, when $g'(\theta) \neq 0$, $g(\hat{\theta})$ is remain the maximum likelihood estimate of $g(\theta)$ [3].

Proof: In order to make $g(\theta)$ be monotonous, assuming $g'(\theta) = 0$, then one can address its extreme point, because $g(\theta)$ is the continuous function of θ. Assuming there exist n extreme points, which is $\theta_1, \theta_2, \theta_3, \ldots, \theta_n$. So that $g(\theta)$ is monotonous on $[\theta_1, \theta_2), [\theta_2, \theta_3), \ldots, [\theta_{n-1}, \theta_n)$, then there exists inverse function of $g(\theta)$. Let $\tau = g(\theta)$, then $\theta = g^{-1}(\tau)$. According to $L(\theta) \leq L(\hat{\theta})$, there exists

$$L[g^{-1}(\tau)] = L(\theta) \leq L(\hat{\theta}) = L[g^{-1}(\hat{\tau})] \tag{5}$$

on $[\theta_1, \theta_2), [\theta_2, \theta_3), \ldots, [\theta_{n-1}, \theta_n)$. So that $\hat{\tau}$ is the maximum likelihood estimate of τ, and $g(\hat{\theta})$ is also the maximum likelihood estimate of $g(\theta)$.

3. Applications

3.1. Annealing furnace online energy efficiency monitoring system
In modern society, the production of cold-rolled and hot-dip galvanized strips is developing with a rapid speed. The strips are heat treated in a hot-dip galvanizing continuous annealing furnace according to certain annealing process requirements to achieve the required physical and chemical properties of the material [4]. The energy consumption of the annealing furnace accounts for more than 30% of the energy consumption of the cold rolling process. Research on the energy efficiency of the annealing furnace will help reduce the energy consumption of the cold rolling process and improve the core competitiveness of steel companies. This application derived an online monitoring system for continuous furnace energy efficiency based on heat balance analysis and maximum likelihood estimation method [5,6].
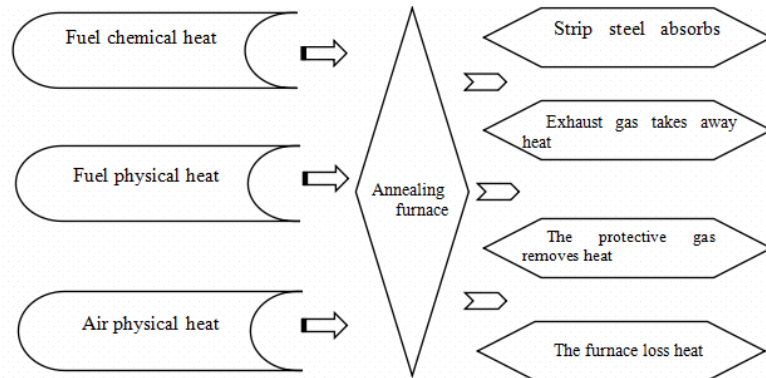
Figure 1. The overall energy balance system of the annealing furnace online.

This application establishes an energy balance model for each part based on the annealing furnace process and equipment characteristics. Based on the annealing furnace equipment characteristics and

online operation conditions, the energy efficiency monitoring and analysis focuses on the Section of jet preheating. Section of heating/soaking section and Section of slow cooling/rapid cooling section and other parts that have a greater impact on the overall energy consumption. The overall energy balance system of the annealing furnace online as shown in the Figure 1.

In this application, because the energy consumption data of the annealing furnace calculated based on online heat balance data samples fluctuates greatly, it is impossible to simply use a single data as the optimal value for diagnosis, so a diagnostic model based on the maximum likelihood method is established. Analyze the fluctuations of heat balance data samples to determine the current thermal status of the annealing furnace.

Assuming that the sample size is sufficiently large, this paper assumes that the sample follows a normal distribution. Suppose the measured value of an indicator is x_k, which is satisfied normal distribution with mean μ_k and variance σ_k. It is also assumed that the difference $x_k - \mu_k$ between the measured value and the true value follows a normal distribution with mean 0 and variance σ_k. So that $x_k \sim N(\mu_k, \sigma_k)$, or $x_k - \mu_k \sim N(0, \sigma_k)$. Then the likelihood function is that

$$L(\mu_1, \mu_2, \ldots, \mu_k) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left[-\frac{(x_k - \mu_k)^2}{2\sigma_k^2}\right]. \tag{6}$$

Next, assuming the absolute error of x_k is δ_k, and its confidence interval probability is 95%. Then, integrate the probability density function to get the mean square error of x_k as $\sigma_k = \frac{\sigma_k}{1.96}$; After these steps, calibration value can be acquired as $\overline{x_k}$. So that when $|\frac{x_k - \overline{x_k}}{\sigma_k}| > 1.96$, the confidence probability of x_k is less than 95%. On the contrary, it can be trusted when the confidence interval is ($\overline{x_k} - 1.96\sigma_k$, $\overline{x_k} + 1.96\sigma_k$).

The author with the above content can obtain that the diagnostic method outlined in this article incorporates the concept of maximum likelihood estimation. It takes the energy efficiency data derived from the heat balance model as the subject of testing and utilizes historical vertical furnace energy efficiency figures as the benchmark. Furthermore, it also establishes an appropriate confidence probability and performs a retrospective analysis to evaluate the efficiency of the data.

3.2. Estimating gamma distribution environmental factors

In different working environments, people may encounter data conversion problems between environments. To solve such problems, how to determine the values of different environmental factors is the key. In addition to using environmental factors to convert data, this type of problem also relies on the conversion principle between life spans in different environments, and the most common of which is gamma distribution [7].

This model assumes that the probability density function of the two-parameter gamma distribution is that

$$f(x) = \begin{cases} \frac{\lambda^k}{\Gamma(k)} y^{k-1} e^{-\lambda}, y \geq 0 \\ 0 \qquad\qquad , y < 0 \end{cases}. \tag{7}$$

And this model assumes that k is known, and there is no replacement for the environmental factor u under the definite truncation model. Assuming that the product life under two environments is $Y_{ij} \sim \Gamma(k_i, \lambda_j), i = 1, 2$. k and λ are scale parameters and shape parameters respectively. The failure mechanism of the product is the same in two environments, so that when $k_1 = k_2 = k$, the gamma distribution environmental factor is that [8]

$$u = \frac{\lambda_2}{\lambda_1}. \tag{8}$$

Suppose the first q_i minimum observation values in a random sample with capacity m_i from the gamma distribution are $y_{i1} \ll y_{i2} \ll \ldots \ll y_{iq_i}$, and set up the vector $y_i = (y_{i1}, y_{i2}, \ldots, y_{iq_i})$, then the likelihood function based on Y_i is that

$$L_i(y_i \mid \lambda_i) = \frac{m_i!}{(m_i - q_i)!} \prod_{j=1}^{q_i} [f(y_{ij}; k, \lambda)] F^{m_i - q_i}(y_{iq_i}; k, \lambda_i). \tag{9}$$

By substituting Eq. (8) into Eq. (9), then

$$L_i(y_i|\pi_i) = C_i \lambda_i^{kq_i} e^{-H_i \lambda_i}, \tag{10}$$

where $C_i = \frac{m! \prod_{j=1}^{q_i} y_{ij}^{k-1}}{(m_i - q_i)!(\Gamma(k))^{q_i}}$ and $H_i = \sum_{j=1}^{q_i} y_{ij} + (m_i - q_i) y_{ij}$, $i = 1, 2$. Then the joint likelihood function of the test results in the two environments is $L = L_1 L_2$, with [9]

$$L(\lambda_1, \lambda_2) = \prod_{i=1}^{2} C_i \lambda_i^{kq_i} e^{-H_i \lambda_i}. \tag{11}$$

By performing logarithmic operations on both sides of the joint likelihood function simultaneously, then

$$\ln L(\lambda_1, \lambda_2) = \sum_{i=1}^{2} C_i + \sum_{i=1}^{2} kq_i \ln \lambda_i - \sum_{i=1}^{2} H_i \lambda_i \tag{12}$$

By taking partial derivatives of λ_1 and λ_2 respectively, then

$$\frac{\partial \ln L}{\partial \lambda_i} = \frac{kq_i}{\lambda_i} - H_i = 0, i = 1, 2. \tag{13}$$

Therefore, $\widehat{\lambda_{1L}} = \frac{kq_1}{H_1}$, $\widehat{\lambda_{1L}} = \frac{kq_2}{H_2}$. From these, when k is known, the maximum likelihood estimation of environmental factors under constant censoring model without replacement u is that

$$\widehat{u_L} = \frac{\widehat{\lambda_{2L}}}{\widehat{\lambda_{1L}}} = \frac{H_1 q_2}{H_2 q_1}, \tag{14}$$

where $H_i = \sum_{j=1}^{q_i} y_{ij} + (m_i - q_i) y_{ij}$, q_i is censored data, i = 1, 2.

The author with the above content can obtain that in the above-mentioned constant censoring model without replacement, on the premise that the relevant conditions of the environmental factor u (k is known) and the likelihood function of the sample have been assumed, the estimated value of the parameter u is obtained by using maximum likelihood estimation. In this mathematical model, the basic definition of maximum likelihood estimation has been fully and reasonably used, but the relevant properties have not been brought into exert. Maximum likelihood estimation greatly improves the efficiency and accuracy of determining environmental factors in different environments [10].

3.3. Block-sparse least mean square algorithm

The echo cancellation algorithm is the core of front-end acoustic signal processing and intelligent voice terminal equipment, and is of great significance to improving user intelligibility, auditory experience, and quality of life. In order to break through the limitations of fixed numerical settings and actual engineering requirements, maximum likelihood estimation algorithms based on robust statistical ideas have emerged [11,12]. Based on the idea of robust statistics, Hampel three-segment function is introduced to establish a block sparse adaptive filtering algorithm based on maximum likelihood estimation.

Echo cancellation ensures that far-end user A will not hear an echo of his or her own voice when talking to near-end speaker B. The system is equipped with an echo cancellation module at near-end B, which integrates a control unit, adaptive algorithm, and residual echo suppression function. In addition,

in order to enhance the clarity and comfort of speech, the system adds comfort noise to optimize the user's listening experience. The specific principle is shown in Figure 2. The maximum likelihood estimation based on the idea of robust statistics replaces the original minimum mean square error criterion. This mathematical model establishes a robust block sparse normalized minimum mean square algorithm and corrects the cost function to be that [13],

$$H(n) = E[\varphi(e(k))] + \lambda||\omega(k)||_{2.0}, \tag{15}$$

where $e(k) = d(k) - \mu^T(k)\omega(k), d(k) = y(k) + v(k), y(k) = \mu^T(k)\omega(k)$, and $\mu(k)$ is voice signal sent by remote user A, $\omega(k)$ is unknown echo path, $y(k)$ is echoe signal, $v(k)$ is background noise. $\varphi(e(k))$ is maximum likelihood estimate to suppress impulses noise.

Then suppose $\varrho, \Upsilon_1, \Upsilon_2$ are the threshold parameters that control the compression amplitude of impulse noise. Then define the Hampel three-segment function as follows

$$\varphi(e) = \begin{cases} \dfrac{e^2}{2} & ,0 \le |e| \le \varrho \\[2mm] \varrho|e| - \dfrac{\varrho^2}{2} & , \varrho \le |e| \le \Upsilon_1 \\[2mm] \dfrac{\varrho}{2}(\Upsilon_1 + \Upsilon_2) - \dfrac{\varrho^2}{2} + \dfrac{\varrho}{2}\dfrac{(|e| - \Upsilon_2)^2}{\Upsilon_1 - \Upsilon_2} & ,\Upsilon_1 \le |e| \le \Upsilon_2 \\[2mm] \dfrac{\varrho}{2}(\Upsilon_1 + \Upsilon_2) - \dfrac{\varrho^2}{2} & , \Upsilon_1 \le |e| \end{cases} \tag{16}$$

Through the negative gradient steepest descent method, the filter weight vector ω of the Eq. (15) is derived, and the new weight coefficient update formula is obtained as follows

$$\omega(k+1) = \omega(k) + \dfrac{\kappa\dfrac{\partial\varphi(e(k))}{\partial\omega(k)}e(k)\mu(k)}{\mu(k)\mu^T(k) + \phi} + \xi f(\omega(k)) + \Lambda f(\omega(k)), \tag{17}$$

where $f(\omega) = [f_1(\omega), f_2(\omega), \ldots, f_n(\omega)]^T, f_i(\omega) \triangleq 2\beta^2\varpi_i - \dfrac{2\beta\varpi_i}{||\omega([d/c])||}$ or zero. β is a constant, κ is the step parameter that adjusts the balance between steady-state error and convergence speed. ϕ is a positive constant that prevents division by zero in normalization processing, and $\Lambda = \kappa\lambda/2$.
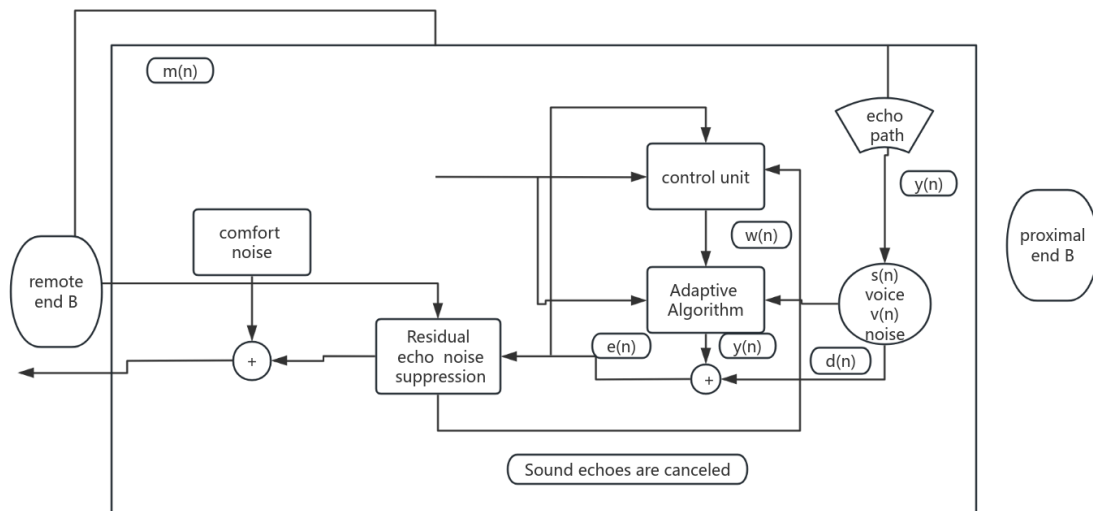
Figure 2. Acoustic echo system model schematic diagram

The author with the above content can obtain that the output error is small in order to make the cost function of the adaptive echo algorithm have a low-order norm. When the output error is large, the

derivative of the cost function approaches 0. Thus, a nonlinear bounded function Hampel three-segment function is introduced. The segment function is the cost function, and combined with the idea of maximum likelihood estimation, the required weight coefficient update formula is obtained by derivation. This mathematical model makes full use of the core idea of maximum likelihood estimation and involves related properties.

4. Conclusion

Throughout the paper, the author explores three mathematical models, which are parameter estimation of the annealing furnace efficiency system, estimation of gamma environmental factors, and function likelihood improved by adaptive echo cancellation algorithm. It is found that the reason why maximum likelihood estimation is widely used is because of its unique calculation cogitation, which is finding a special value by taking logarithms and derivation. This special value also has the core characteristics of the model, so that it is the most appropriate estimate. As for the relevant properties of maximum likelihood estimation, it is not used very frequently and may be only used under specific conditions. What is more, it is the most widespread to combine the idea of maximum likelihood estimation to build mathematical models or algorithm functions. As the parameter estimation of the annealing furnace efficiency, the idea of maximum likelihood estimation and the method of posterior probability are combined to find the estimated value of the parameter. Meanwhile, in the adaptive echo cancellation algorithm, the ideas of robust statistics and maximum likelihood estimation are combined, the Hampel three-segment function is introduced, and a new filtering algorithm is established to meet the needs of the model.

References

[1] Zhao Chunxuan, Xu Wei. (2002). Mathematical Statistics. Beijing: Science Press.
[2] Morns H. DeGroot, Mark J. Schervish. (2012). Probability and Statistics. China Machine Press.
[3] Hu Yanli, Liu Miao. (2021). Theoretical research on the main properties of maximum likelihood estimation, Journal of Yili Normal University (Natural Science Edition). 15 (1), 14-17.
[4] Zhang Guozhi, Chen Guojun, Su Fuyong. (2023). Development and application of online energy efficiency monitoring system for annealing furnace based on maximum likelihood estimation, Metallurgical Automation. 172-176.
[5] Steinboeck A, Wild D, Kugi A. (2013). Energy-efficient control of continuous reheating furnaces. IFAC Proceedings Volumes. 46(16), 359.
[6] Chakravarty K, Kumar S. (2020). Increase in energy efficiency of a steel billet reheating furnace by heat balance study and process improvement. Energy Reports. (6), 343.
[7] Hu Shasha, Wei Chengdong, Qiu Yan. (2001). Maximum likelihood estimation and Bayes estimation of gamma distribution environmental factors. Journal of Guangxi Teachers Education University. 28 (2), 25-28.
[8] Wang Hongzhou, Ma Baohua, et al. (1991). Estimation of gamma distribution environmental factors. Chinese Space Science and Technology, 5, 8-15.
[9] Wang Bingxing. (1998). Definition of environmental factors and their statistical inference. Intensity and Environment, 4, 24-30.
[10] Wei Dandan, Zhou Yi, Zhao Yu. (2023). Block-Sparse Least Mean Square Algorithm Based on M-Estimator Against Impulsive Interference. Journal of Southwest University. 45(5), 210-218.
[11] Vega L R, Rey H, Benesty J, et al. 2008. A new robust variable step-size NLMS algorithm. IEEE Transactions on Signal Processing. 56(5), 1878-1893.
[12] Zhou Y, Liu H, Chan S C. (2014). New partial update robust kernel least mean square adaptive filtering algorithm. International Conference on Digital Signal Processing.
[13] Chan S C, Zhou Y, Ho K L. (2010). A new sequential block partial update normalized least mean m-estimate algorithm and its convergence performance analysis. Journal of Signal Processing Systems. 58(2), 173-191.