World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Using Monte Carlo method to Upgrade Zhentong Gao method

Jiajin Xu *

L and Z International Leasing Co. Ltd., Guangzhou, China.

World Journal of Advanced Research and Reviews, 2024, 22(02), 633–640

Publication history: Received on 31 March 2024; revised on 06 May 2024; accepted on 08 May 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.2.1391

Abstract

In recent decades, it has been found that the three-parament Weibull distribution plays an increasingly important role in the fields such as structural fatigue and reliability. However, the complexity of estimating these three parameters greatly affects the application of the Weibull distribution. In the past three years, the author has successively proposed the Zhentong Gao Method and the Generalized Zhentong Gao Method, which have effectively solved this problem. However, it has also been found that there is room for improvement in these two methods. That's what this paper is about, how to use the characteristics of the Monte Carlo and Metropolis-Hastings methods to upgrade the Zhentong Gao method and the Generalized Zhentong Gao method, so that these two upgraded methods can not only efficiently estimate the three parameters of the Weibull distribution, but also provide their confidence intervals, and can be used in more fields.

Keywords: Three-Parameter Weibull Distribution; MC Method; MH method; Zhentong Gao Method; Generalized Zhentong Gao Method

1. Introduction

Monte Carlo (MC) method[1] is essentially a statistical simulation method, or a collective term for ideas or methods, which mainly uses random numbers to estimate computational problems. Due to the popularity of computers, this method has been widely applied as never before. And the Metropolis-Hastings (MH) method[2] has been widely used since it was proposed in 1953 [3]-[6]. The author found that MC and MH method are available to upgrade the Zhentong Gao method(ZT Gao). ZT Gao method is a relatively simple algorithm proposed by the author in 2021 to estimate the three parameters of the Weibull distribution[7]. However, there is a problem with it that the confidence intervals for these three parameters cannot be determined simultaneously, and requires a different method to do so[8]. In addition, in 2023, the author proposed the Generalized Zhentong Gao(G_ZT Gao) method to solve the problem of Maximum Likelihood Estimation(MLE) estimating the three parameters of the Weibull distribution[9]. However, there is still a problem that the confidence intervals of the three parameters cannot be determined simultaneously, and it requires a relatively long CPU time. For this reason, the author considers using the characteristics of MC and MH methods to upgrade the ZT Gao and the G_ZT Gao methods, so that the upgraded ZT Gao and the G_ZT Gao methods can not only efficiently estimate the three parameters of the Weibull distribution, but also estimate their corresponding confidence intervals. More importantly, it can also be used in solving equations, finding extrema, and other fields.

2. Introduction to Zhentong Gao Method and Generalized Zhentong Gao Method

The so-called Zhentong Gao method, was originally used to estimate the three parameters of the Weibull distribution that fit the fatigue life of the structure, avoiding cumbersome derivation and directly obtaining the estimation of these three parameters using the "brute force method". The actual approach is[7],[10]"

* Corresponding author: Jiajin Xu

- Enter the raw data. If the raw data is not sorted properly, sort it first;
- Using the scipy in Python, one can directly traverse the possible value interval of x_0 (0, x_{min}) with a given precision interval to obtain the x_0 that maximizes the correlation coefficient, which is x_{0max}.
- Notice that when calculating the correlation coefficient in the scipy it is in fact a matter of first finding the corresponding coefficients *b* and *d* of the linear equation in least squares then you can get $\lambda = \exp(-d/b)$. Therefore, once x_{0max} is determined, the corresponding parameter *b* and λ of the Weibull distribution are obtained at the same time."

Its advantage lies in utilizing the understanding of the physical meaning of the "positional parameters" of the Weibull distribution[7],[10] and the characteristics of computers, especially the convenience of using Python to estimate the three parameters of the Weibull distribution relatively easily. The Generalized Zhentong Gao Method [9] is based on the ZT Gao Method and follows the same approach to extend the ZT Gao Method to solve the problem of estimating three parameters of Weibull distribution using MLE as the basis. Its advantage lies in avoiding the tedious formula derivation and transcendental equation system in MLE. However, the computational workload will greatly increase. Moreover, neither of these methods can provide confidence intervals for these three parameters. Therefore, there is room for improvement and upgrading.

3. Upgrading Zhentong Gao Method with MC and MH Method

In the previous section, it was pointed out that the essence of the so-called ZT Gao method is to use computers to estimate the 3 parameters of the Weibull distribution through the "brute force method", which is simple in logic and reliable in method. The ZT Gao method is used to determine the location parameter by Least Square Method(LSM), and then it is easy to determine the shape and scale parameters using the two coefficients of linear regression obtained from LSM [7]. But it is not possible to estimate the confidence intervals for these 3 parameters. However, if the MC method is utilized, i.e., a randomized approach instead of the brute force approach, as in the case of the Bootstrap[11] it is possible to estimate the confidence intervals. Of course, it is also necessary to utilize the feature of the MH method here so that suitable random values can be accumulated quickly to find suitable extreme values. So, we can consider this feature to upgrade the ZT Gao method. The second point of ZT Gao method will be changed to "uniformly sample in the interval (0, x_{min}) and estimate the x_0 that maximizes the relative coefficient according to the characteristics of MH." Also because of the randomness of the MC, the three estimates calculated each time will not be strictly the same, so the calculation can be repeated, for example, 100 times, and the resulting distribution of each parameter can be regarded as a normal distribution, from which the mean and the corresponding confidence interval of each parameter can be obtained. To obtain the upgraded ZT Gao method, there are two steps: 1) first complete the upgrade of ZT Gao method and compare the results with the original ZT Gao method; 2) then perform multiple calculations on the upgraded ZT Gao method, and according to the central limit theorem, obtain the Gaussian distribution of three parameters, thereby obtaining their confidence intervals.

Now let's look at the results of the first step. Here we still take the data from Example 8.1-2 of P118 in [10]. For the Upgraded Zhentong Gao Method (abbreviated as ZT Gao_MC Method), you can modify the Python code of the original ZT Gao Method to get the following result:

N= [350, 380, 400, 430, 450, 470, 480, 500, 520, 540, 550, 570, 600, 610, 630, 650, 670, 730, 770, 840]

ZT Gao method: number of calculations= 2000 ,*b*= 2.039, λ= 320.90 ,*N0*= 276.67 ,r= 0.99922

Time required to run ZT Gao method= 0.990334 sec.

ZT Gao_MC method: number of calculations= 80 ,*b*= 1.992 , λ= 316.29 ,*N0*= 281.00 ,r= 0.99919

Time required to run the ZT Gao_MC method= 0.046988 sec.

It seems that the upgraded Zhentong Gao method works well, the running time is almost 1/21 of the original one, while the accuracy is almost the same. This is indeed a bit unexpected. It is precisely because of the characteristics of using MC and MH methods that the Zhentong Gao method has been improved, so it is called "upgrading".

The random method is indeed good, but there is a problem that the results of each run seem to be random, that is, the results of obtaining three parameters are not completely consistent. This is a "disadvantage", but it is also an "advantage" , because after running this code a few more times, according to the central limit theorem, the mean and corresponding confidence interval of these three parameters can be obtained. Here are the results obtained from running 100 times:

γ= 0.95 ,overall sampling number= 100

The confidence interval for *b* is (1.954 , 2.126)

The confidence interval for *λ* is (312.529 , 329.589)

The confidence interval for *N0* is (268.514 , 284.532)

ZT Gao_MC: *b*=2.040 ,*λ*=321.06 ,*N0*=276.52 ,*r*=0.99914 ,*R^2*=0.99824

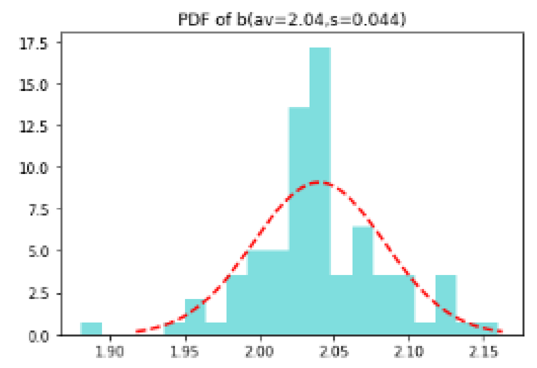

Figure 1 Histogram of Shape Parament *b*

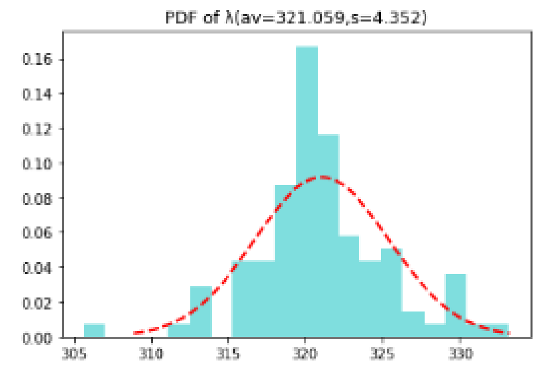

Figure 2 Histogram of Scale Parament *λ*

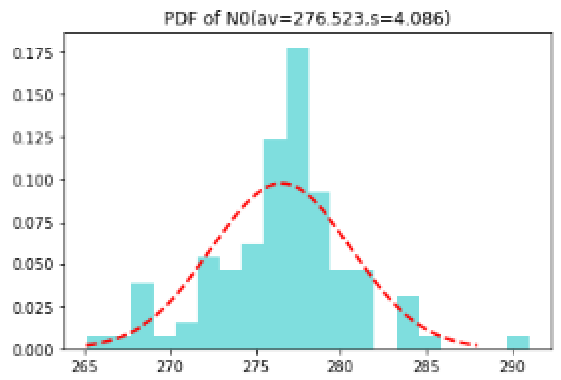

Figure 3 Histogram of Location Parament *N0*

4. Upgrading the Generalized Zhentong Gao Method with MC and MH Methods

The essence of using MLE to determine the three parameters of a Weibull distribution is that these three parameters can make the likelihood function of the Weibull distribution reach its extremum. It is not difficult to obtain the likelihood function of the Weibull distribution after taking the logarithm[9],

$$LL=\ln L=n\ln(b/\lambda)+(b-1)\Sigma_{i=1}^{n}\ln[(x_i-x_0)/\lambda]-\Sigma_{i=1}^{n}[(x_i-x_0)/\lambda]^b \quad (4.1)$$

The traditional approach is to take the partial derivatives of *LL* with respect to each of the three parameters b,λ,x_0 and then make them zero to solve the system of nonlinear equations with these three unknowns[12]. Instead of taking partial derivatives and of course solving the system of nonlinear equations, the generalized Zhentong Gao method proposed by the authors "directly calculates the value of *LL* within the "reasonable range" of the three parameters b, λ, x_0 by the brute-force method, finds out the maximal value of *LL*, and then gives the corresponding values of the three parameters b, λ, x_0, The authors call this method the Generalized Zhentong Gao method because its idea is the same as the ZT Gao method, and it also needs the ZT Gao method to provide the estimation of the three parameters in order to get its "reasonable range""[10]. However, one of the problems with this method is that it is computationally intensive. Therefore, similar to the upgrading of the ZT Gao method, the G_ZT Gao method is upgraded with random sampling instead of the brute force method. The key here is the range of values of the three parameters, or to use the ZT Gao method to find out the value of the three parameters and then give a "reasonable range", for example, for b can be taken as ± 5%, and λ can be taken as ± 10%, which in a sense can also be seen as a kind of a priori distribution[1], but of course this percentage can be adjusted according to the specific circumstances. adjusted according to the specific situation. The range of N_0 remains (0, x_{min}).

Another problem that must be solved is how to determine the confidence intervals of the three parameters of the Weibull distribution by using the upgraded G_ZT Gao method. It seems that we have to utilize the stochastic characteristics of MC, that is, the estimates of the three parameters are slightly different each time due to randomness, and each calculation is regarded as a sampling, so that according to the Central Limit Theorem, the mean is the estimate, and the mean square deviation can be used to estimate the confidence intervals. A modification of the original G_ ZT Gao method along these lines is made to make a comparison with an example 3 in [13]. After modifying the Python code, it runs as follows.

N= [3956.42, 4004.18, 4091.61, 4355.05, 4355.4, 4376.01, 4391.79, 4487.68, 4487.68, 4736.67, 4736.67, 4939.85, 4963.62, 5220.19, 5353.41, 5372.72, 5418.04, 5444.11, 5603.17, 5698.1, 5746.17, 5843.52, 6175.14, 6197.41, 6249.69, 6279.76, 6279.76, 6572.74, 6740.48, 6887.65, 7183.09, 7209, 7209, 7209, 7209, 7366.4, 7581.64, 7581.64, 7581.64, 7645.59, 8246, 8599.7, 8713.97, 8936.34, 9044.22, 9197.45, 9511.73, 9754.47, 9967.45, 10136.31, 10172.88, 10172.88, 10308.04, 10395, 10609.23, 10609.23, 10788.97, 10879.97, 10971.75, 11594.41, 11990.59, 12237.31, 12400.31, 12400.31, 12550.01, 13198.73, 13947.78, 15557.12, 17646.12, 19848.23, 23199.07]

γ(confidence level)= 0.95

G_TZ Gao_MC： b= 1.136 ,λ= 4661.77 ,$N0$= 3937.82 ,LL= -666.814

bs= 0.021 ,λs= 90.83 ,$N0s$= 8.41 ,LLs= 0.02957

The confidence interval for b is (1.094 , 1.177)

The confidence interval for λ is (4483.756 , 4839.791)

The confidence interval for N_0 is (3921.329 , 3954.303)

Time taken to run the program= 0.113149 sec.

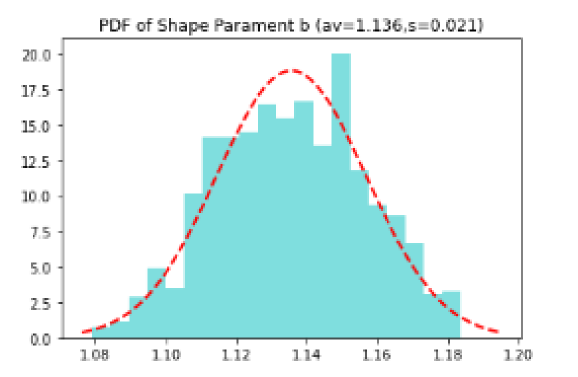

Figure 4 Histogram of Shape Parament b

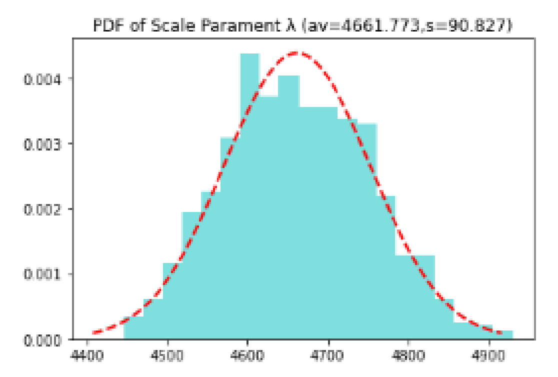

Figure 5 Histogram of Scale Parament λ

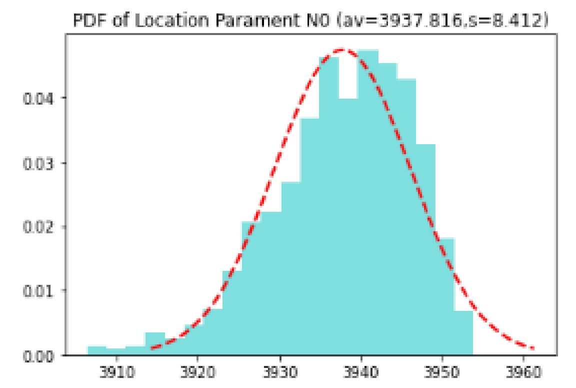

Figure 6 Histogram of Location Parament *N0*

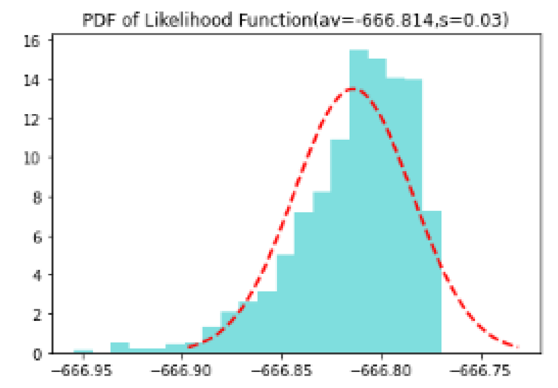

Figure 7 Histogram of Likelihood Function *LL*

Looking back at the results obtained from the generalized Zhentong Gao method that has not been upgraded,

ZT Gao:,b=1.147 ,λ=4958.76 ,$N0$=3812.01 ,r=0.99699 ,R^2=0.99392 ,LL=-667.741

MLE:,b=1.139 ,λ=4936.10 ,$N0$=3821.50 ,r=0.99691 ,R^2=0.99375 ,LL=-667.701

G_ZT Gao: b=1.190 ,λ=4800.00 ,$N0$=3877.29 ,r=0.99688 ,R^2=0.99297 ,LL=-667.054Time taken to run the program = 22 sec.

The running time ratio of the two is 22/0.111=198, which is about 200 times, and the results obtained are even better than the G_ZT Gao method without upgrading, and even better than traditional methods. More importantly, the G_ZT Gao_MC method also provides confidence intervals for each parameter..

5. Further Application of Upgraded Zhentong Gao and Generalized Zhentong Gao Methods

For the upgraded ZT Gao and G_ZT Gao Method, they can not only be used to estimate the three parameters of Weibull distribution, but also for finding zeros, extremal points, and other fields. Here is a brief list:

5.1. ZT Gao_MC method can be used to solve complex univariate equations

For example, when encountering a transcendental equation in [10]:

$$\Gamma(1+1/x)^x = \ln 2 \qquad (5.1)$$

Its numerical solution can be obtained using Newton's method or fixed-point method[10]. Now, according to the upgrade Zhentong Gao method, modifying the code in [10] yields the following result.,

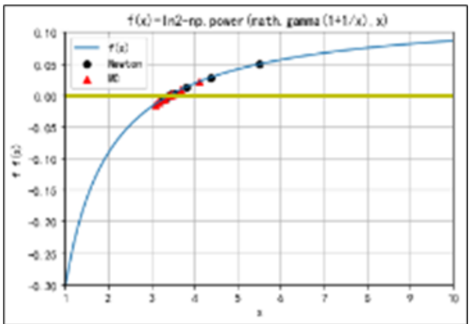

Figure 8 Comparison chart of solving transcendental equations using ZT Gao Method and ZT Gao_MC

Newton's method (bisection degree)=14, E (accuracy)=1e-05, b=3.43951

ZT Gao_MC: (number of sample)=17, E (accuracy)=1e-05, b=3.43957

It can be seen that the results of the two methods are indistinguishable from each other.

5.2. Find the extreme points and their extreme values of the function

In fact, the upgraded ZT Gao method solves the problem of finding the extremum of a univariate function and the corresponding extremum, while the upgraded G_ZT Gao method solves the problem of finding the extremum of a ternary function and the corresponding value. However, there is a noteworthy issue here that the range of parameters is very important, and the position interval of the Weibull distribution is limited by the minimum value of the given data. Therefore, there is no problem using the ZT Gao method and the upgraded ZT Gao method; For the G_ZT Gao method, there is no problem with the estimated value obtained first with the ZT Gao method. Therefore, for general functions, it is important to pay attention to this point when calculating the extremum and extremum points, otherwise errors may occur.

Example 1. An example from reference [13] , there is an iron plate of width 24, which is folded so as to maximize its cross-sectional area, what should x and α be taken?

It is not difficult to give its cross-sectional area as,

$$A=[(24-2x)+(24-2x+2x\cos\alpha)](1/2)\, x*\sin\alpha$$

i.e. $$A=(24-2x+x\cos\alpha)x*\sin\alpha,(0<x<12, <\alpha\leq\pi/2)$$

Of course, one can follow the solution in [13] to take the partial derivatives of x and α for A respectively, then make them zero and solve for x and α. You can get it,

$$\partial A/\partial x=(12-2x+x*\cos\alpha)\sin\alpha=0$$

$$\partial A/\partial\alpha=[12\cos\alpha-x*\cos\alpha+x(\cos^2\alpha-sjin^2\alpha)/2]x=0$$

Then solving this system of equations yields: $x=8$; $\cos\alpha= 1/2$, $\alpha= \pi/3=60^\circ$.

This result is certainly good, but it is not as pleasant when encountering complex functions. Therefore, the author wants to use the upgraded ZT Gao method to solve it. Write Python code according to the meaning of the question, and the following results can be obtained:

k(Number of sample)= 100 ,E= 1e-10 ,γ= 0.95

A_ av=82.795; x_av=7.965; α_ av=60.149

The confidence interval for A is (82.058 , 83.532),

The confidence interval for x is (7.111 , 8.818)

The confidence interval for α is (53.858 , 66.44)

A_max=83.13844

A_max(Theoretical)=83.13737

It should be noted that the so-called "A_max" refers to the value of A obtained by substituting the values of x_av and α_av into the formula of A, which seems to be closer to the theoretical value of A_max.

5.3. Nonlinear programming and Lagrange multiplier method

In fact, nonlinear programming problems[14] can be solved through the Lagrange multiplier method[15]. Now use an example to illustrate it. Still using the example in [13]:

Example 2. Find the largest cube with a surface area of $100cm^2$.

Solution: Assume that the three sides of the required rectangle are x,y,z. The problem then becomes: V=xyz, (x,y,z>0); V is extremely large under the condition Φ(x,y,z)=2xy +2yz+2zx-100=0.

This is actually a nonlinear programming problem that can be solved by the Lagrange multiplier method. Introducing the Lagrange function:

L=V-λΦ=xyz-λ(2xy+2yz+2zx-100) (5.2)

Taking the partial derivatives of x,y,z and making them zero gives:

yz-2λ(y+z)=0; zx-2λ(z+x)=0; and xy-2λ(x+y)=0

It follows that x=y=z

This result is naturally expected. However, we still want to use the G_ZT Gao_MC method to solve it, on the one hand, it is to add a method of solving; but more importantly, for the more complex nonlinear planning problems, it is more convenient to use the G_ZT Gao_MC method to solve them, because there is no need to be partial derivatives and can be solved directly. At this point, just note that as long as two of the three variables are determined then the third is also determined. That is, z = (100-2xy)/(x + y). Again, the Python code can be obtained as per the question and the result of the run is:

k(number of samples) = 151 ,E= 1e-10 ,

V_{max}=68.03807 ;X=4.03980 ;Y=4.11971 ;Z=4.08813

Exact : X=Y=Z=4.08248 ,V_{max}=68.04138

And the results obtained are still satisfactory. This immediately reminds the author in [16] to use nonlinear programming to find the three parameters of the Weibull distribution can actually be carried out by the upgraded generalized Zhentong Gao method. As for the other Boltzmann annealing method in [16], it is essentially the same as the G_ZT Gao_MC method.

6. Conclusion

The following conclusions can be drawn from the above discussion:

- The use of MC and MH methods can upgrade the Zhentong Gao and the Generalized Zhentong Gao methods, and the results are quite good. Not only does it save running time, but it can also provide confidence intervals for various parameters simultaneously;
- The upgraded Zhentong Gao and Generalized Zhentong Gao methods can also be used in fields such as finding zeros, extremum problems, and requiring optimization, which has considerable application prospects and is worth exploring.
- If viewed from a philosophical perspective, ZT Gao and G_ZT Gao methods represent exhaustive law in a sense, that is, as long as all possible paths are exhausted, the goal can always be achieved. But for a large number of "paths" (such as NP problems), theoretically it may seem exhaustible but in reality it is impossible. The upgraded ZT Gao and G_ZT Gao methods rule reflects the MC method. The fact is that in many cases, people do not need to exhaust all paths to obtain more suitable results, which is not only feasible but also necessary.

Compliance with ethical standards

Acknowledgments

Author would like to thank Chinese Academician Zhentong Gao, Mr. Weihao Wan and Dr. Rihua Li for theirs support of this article and related research work.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Goodfellow I. and etc. Deep Learning, Beijing: POSTS & TELECOM PRESS; 2017. p43,p359

[2] Metropolis N. and etc. Equation of state calculation by fast computing machines; Journal of Chemical Physics. 1953, 21: 1087-1091

[3] Mao SS. and etc. Advanced Mathematical Statistics. Second Edition .Beijing: Higher Education Press; 2006.p45

[4] Banterle M. and etc. Accelerating Metropolis-Hastings algorithms by Delayed Acceptance. Statistics. March, 2015: DOI: 10.3934/ FODS.2019005

[5] Lucci S. & Kopec D. Artificial Intelligence in the 21st Century， Second Edition. Beijing: POSTS & TELECOM PRESS. 2018. p339

[6] Chen P. & Xu RX. Metropolis-Hastings adaptive algorithms and their applications. Systems Engineering-Theory & Practice, 2018, No.1. DOI: 10.3321/j.issn:1000-6788.2008.01.015

[7] Xu JJ. The Gao Zhentong Method in the Intelligentization of Fatigue Statistics，*Journal of Beijing University of Aeronautics and Astronautics*, 2021, 47(10): 2024-2033.Doi: 10. 13700/j.bh. 1001-5965. 2020. 0373

[8] Xu JJ. & Gao ZT. Further Research on Fatigue Statistics Intelligence, *Acta Aeronautical et Astronautical Sinica*. 2022, 43(8), Doi: 10. 7527/S1000.6893.2021.25138225138

[9] Xu J. Generalized Zhentong Gao Method for Estimating Three Parameters of Weibull Distribution. SCIREA Journal of Computer. 2023, Volume 8, Issue 2. Doi:10.54647 /computer520345

[10] Gao ZT. & Xu JJ. *Intelligent Fatigue Statistics*, Beijing: Beihang University Press. 2022. p116

[11] Efron B. &Hastie T. Computer Age Statistical Inference. Beijing: China Machine Press. 2019. p103

[12] Yan XD. and etc. Comparison of Three-Parameter Weibull Distribution Parameter Estimation Methods. Journal of Ningbo University. 2005, V18(3)

[13] Department of Mathematics of Tongji University. Higher Mathematics, 7th ed. Beijing: Higher Education Press. 2014. p115,p118

[14] Cormen TH. and etc. Introduction to Algorithm. Beijing: China Machine Press. 2016. p495

[15] Li H. Statistical Learning Methods. Beijing: Tsinghua university pres. 2012.,p103

[16] Xu JJ. From Zhentong Gao Method to Generalized Zhentong Gao Method. J0urnal of Physics.: Conference Series 2650 (2023)012013, doi:10.1088/1742-6596/2650/1/012013

Authors short biography

Jiajin Xu is a researcher specializing in the fields of semiconductors, laser physics, nonlinear optics, and computational physics. He worked as a visiting scholar at the Optical Research Center, University of Arizona, United States, between 1988 and 1991. During the period 1982-1992, he co-authored more than ten papers in national and international journals, including Journal of Physics, Chinese Science , Physical Review Letter and so on. In recent years,he has collaborated with Mr. Gao Zhentong, an academician of the Chinese Academy of Sciences, in the field of "Intelligent Fatigue Statistics" and have published together or independently several papers and one monograph.