Dhaka Univ. J. Sci. 71(1): 13-16, 2023 (January)
ISSN 1022-2502, E-ISSN 2408-8528
DOI: https://doi.org/10.3329/dujs.v71i1.65267

Assessing Goodness of Approximate Distributions for Inferences about Parameters in Nonlinear Regression Model

Md. Jamil Hasan Karami*

Department of Statistics, University of Dhaka, Dhaka-1000, Bangladesh

(Received: 3 August 2022; Accepted: 31 January 2023)

Abstract

It is often crucial to make inferences about parameters of a nonlinear regression model due to a dependency of Fisher information on the parameter being estimated. Here, the distribution of the relevant test statistic is not exact, but approximate. Therefore, similar conclusion, based on the values of different test statistics, may not be reached. This study shows, in this circumstance, how to come up with a nonlinear regression model that can be used for forecasting and other related purposes. The goodness of the approximate distributions, F and χ^2 , has been assessed to reach a correct decision. The simulation results show that the simulated probability of committing a type I error is very close to its true value in case of F distribution corresponding to F statistic. However, the χ^2 distribution does not do a similar job for the LRT statistic since the simulated type I error is quite larger.

Keywords: Nonlinear regression, MLE, Test statistic, ANOVA, Simulation

I. Introduction

In regression analysis, one of the main purposes is to find a relationship between a response variable and covariates. This relationship is expressed through some models.The general framework of regression model is that observed response (Y) is a linear combination of function of covariates ($f(X,\ \boldsymbol{\theta})$ where X denote covariates and $\boldsymbol{\theta}$ denote parameters of the model) and a random error (ε) term:

$$Y = f(X, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}. \qquad (1)$$

Therefore, the random response is dependent on the covariate $X = (x_1,\ x_2,\ \ldots,\ x_p)^{\mathrm{T}}$, Twhere p is the number of covariates. Here, the mean function $E(Y|X) = f(X, \boldsymbol{\theta})$ is assumed to have a known functional form although it contains unknown parameters $\boldsymbol{\theta}$**.** It is interesting to note that if the function of covariates $f(X, \boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$ we consider linear regression model. However, if $f(X, \boldsymbol{\theta})$ is a nonlinear function of $\boldsymbol{\theta}$ we consider a nonlinear regression model[1]. In this research, we use nonlinear regression models which are often encountered in chemical reactions, in biology, clinical trials[2], reliability and life testing[3].

We may know whether the relationship between response and predictor is nonlinear either by looking at the scatterplot or by assessing the functional form of a model. For nonlinear regression models, likelihood based inferences are usually recommended[4]. Iterative methods are used to deal with score function and Fisher information upon which the inference procedure is based. We can use F-test and likelihood ratio test (LRT) for testing the adequacy of the model. We use the corresponding approximate distributions of the test statistic to calculate p-values. However, for non-linear regression models, all distributions about the test statistic are just approximate, not exact. The approximation will be good if the sample size is large and the number of parameters is small compared to the sample size[1]. In this research, along with an exploration of inference procedures for nonlinear regression model, we would like to investigate how good the approximation is to the true distribution of a test statistic.

II. Methods

Let us consider a dataset $(y_i; x_{ij})$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$. Now, the parameters under the model (1) can be estimated by using the technique of least square estimates (LSE). We know that theLSE of θ is defined as

$\hat{\theta} = arg\min_\theta \sum_{i=1}^{n} \varepsilon_i^2 = arg\min_\theta \sum_{i=1}^{n}[y_i - f(x_i, \theta)]^2$.The idea of LSE of linear model can be extended to nonlinear model. However, if the distribution function is known the maximum likelihood estimation (MLE) can be applied to estimate parameters of the model. A likelihood function for the observed dataset is

$$L(\theta, \sigma^2;\ y_1, y_2, \ldots, y_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left(\frac{-\sum_{i=1}^{n}\{y_i - f(x_i, \theta)\}^2}{2\sigma^2}\right).$$

Then the log likelihood function is given by

$$l(\theta, \sigma^2) = \frac{-\sum_{i=1}^{n}\{y_i - f(x_i, \theta)\}^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2). \qquad (2)$$

By minimizing the objective function

$S(\theta) = \sum_{i=1}^{n}\{y_i - f(x_i, \theta)\}^2$, we get similar estimate $\hat{\theta}_{LSE} = \hat{\theta}_{MLE}$. However, we often do not find the close form of θ by solving the equation $\frac{\partial S(\theta)}{\partial \theta} = 0$. In this situation, we use Gauss-Newton algorithm to get numerical solution[1]. We can approximate $f(x_i, \theta)$ by a linear function:

* Author of correspondence, e-mail: karimi.stat@du.ac.bd

$$f(x_i,\theta)\approx f\left(x_i,\theta^{(0)}\right)+\left\{\frac{\partial f(x_i,\theta)}{\partial\theta}|_{\theta=\theta^{(0)}}\right\}^T\left(\theta-\theta^{(0)}\right),$$

where $\theta^{(0)}$ is the given initial value of θ. Using this approximation the model in (1) can be re-written as $y_i - f(x_i,\theta^{(0)}) \approx \left\{\frac{\partial f(x_i,\theta)}{\partial\theta}|_{\theta=\theta^{(0)}}\right\}^T (\theta-\theta^{(0)}) + \varepsilon_i$, which is approximately a linear regression model. Then in matrix notation, the model can be written as

$$Y-\mu(\boldsymbol{\theta}^{(0)}) \approx V(\boldsymbol{\theta}^{(0)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})+\boldsymbol{\varepsilon},$$ where,

$$V(\boldsymbol{\theta})=\begin{pmatrix}\left\{\frac{\partial f(x_1,\theta)}{\partial\theta}\right\}^T\\ \left\{\frac{\partial f(x_2,\theta)}{\partial\theta}\right\}^T\\ \vdots\\ \left\{\frac{\partial f(x_n,\theta)}{\partial\theta}\right\}^T\end{pmatrix} \text{and } \mu(\boldsymbol{\theta})=\begin{pmatrix}f(x_1;\theta)\\ f(x_2;\theta)\\ \vdots\\ f(x_n;\theta)\end{pmatrix}.$$

Therefore, the objective function takes the form,

$$S(\boldsymbol{\theta})=\{Y-\mu(\boldsymbol{\theta}^{(0)})-V(\boldsymbol{\theta}^{(0)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})\}^T\{Y-\mu(\boldsymbol{\theta}^{(0)}) - V(\boldsymbol{\theta}^{(0)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})\}.$$

Then using the results for the linear model, we get $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{(0)} \approx \left\{V(\boldsymbol{\theta}^{(0)})^T V(\boldsymbol{\theta}^{(0)})\right\}^{-1} V(\boldsymbol{\theta}^{(0)})^T\{Y-\mu(\boldsymbol{\theta}^{(0)})\}$. Thus the value of θ is updated according to the following expression:

$$\boldsymbol{\theta}^{(r+1)}=\left\{V(\boldsymbol{\theta}^{(r)})^T V(\boldsymbol{\theta}^{(r)})\right\}^{-1} V(\boldsymbol{\theta}^{(r)})^T\{Y-\mu(\boldsymbol{\theta}^{(r)})\},$$

$r = 0, 1, 2, 3, \ldots.$

Therefore, theoretically we can stop the iteration if $\boldsymbol{\theta}^{(r+1)} \approx \boldsymbol{\theta}^{(r)}$. Using the estimates of the parameters the residual vector is obtained,

$$\boldsymbol{r}=\begin{pmatrix}y_1-f(x_1;\hat{\theta})\\ y_2-f(x_2;\hat{\theta})\\ \vdots\\ y_n-f(x_n;\hat{\theta})\end{pmatrix}=Y-\widehat{Y}$$. Then the LSE and MLE of σ^2 are respectively given by $\hat{\sigma}^2_{LSE}=\frac{\sum_{i=1}^n r_i^2}{n-q}$, where q is the number of parameters in the model and $\hat{\sigma}^2_{MLE}=\frac{\sum_{i=1}^n r_i^2}{n}$.

For nonlinear regression model, we do not have close form of $\widehat{\boldsymbol{\theta}}$, which makes the study of properties of $\widehat{\boldsymbol{\theta}}$ quite hard. Suppose the true value of $\widehat{\boldsymbol{\theta}}$ is $\boldsymbol{\theta}^*$ (usually unknown). Making Taylor expansion on $f(x,\boldsymbol{\theta})$ around $\boldsymbol{\theta}=\boldsymbol{\theta}^*$, we approximate $f(x,\boldsymbol{\theta})$ by

$$f(x_i,\boldsymbol{\theta}) \approx f(x_i,\boldsymbol{\theta}^*)+\left\{\frac{\partial f(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}^T(\boldsymbol{\theta}-\boldsymbol{\theta}^*). \quad (3)$$

We expand $f(x,\boldsymbol{\theta})$ around $\boldsymbol{\theta}^*$ because we need measure how close the estimate is to the true value. Repeating the same procedure, as before, for getting the updating formula for Gauss-Newton algorithm, we have

$$\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^* \approx \{V(\boldsymbol{\theta}^*)^T V(\boldsymbol{\theta}^*)\}^{-1}V(\boldsymbol{\theta}^*)^T\{Y-\mu(\boldsymbol{\theta}^*)\}$$
$$=\{V(\boldsymbol{\theta}^*)^T V(\boldsymbol{\theta}^*)\}^{-1}V(\boldsymbol{\theta}^*)^T\boldsymbol{\epsilon}.$$

Using (3), the residuals are $r_i=y_i-f(x_i;\widehat{\boldsymbol{\theta}}) \approx$

$$y_i-f(x_i,\boldsymbol{\theta}^*)-\left\{\frac{\partial f(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}^T(\boldsymbol{\theta}-\boldsymbol{\theta}^*).$$

In matrix notation,

$$\boldsymbol{r} \approx [I-V(\boldsymbol{\theta}^*)\{V(\boldsymbol{\theta}^*)^T V(\boldsymbol{\theta}^*)\}^{-1}V(\boldsymbol{\theta}^*)^T]\boldsymbol{\epsilon}.$$

Therefore, the estimate of σ^2 is $\hat{\sigma}^2=\frac{\boldsymbol{r}^T\boldsymbol{r}}{n-q}$

$$=\frac{\boldsymbol{\epsilon}^T[\boldsymbol{I}-V(\boldsymbol{\theta}^*)\{V(\boldsymbol{\theta}^*)^T V(\boldsymbol{\theta}^*)\}^{-1}V(\boldsymbol{\theta}^*)^T]\boldsymbol{\epsilon}}{n-q}.$$

When sample size is large enough, it can be shown that $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*$ is approximately distributed as multivariate normal with mean vector $\mathbf{0}$ and dispersion matrix $\sigma^2\{V(\boldsymbol{\theta}^*)^T V(\boldsymbol{\theta}^*)\}^{-1}$; $\widehat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ are independent, and $\frac{(n-q)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-q}$.

We can define several test statistics to conduct hypotheses testing about parameters of the model. It is interesting to note that if there are replicates available (that is, repeated measurements at the same value of covariate), then we can assess the mean function with these tests. This can be done by comparing the chosen nonlinear regression model to a more general analysis of variance (ANOVA) model. Note that the specification of a one-way ANOVA is analogous to a regression analysis. The only difference is that the covariate needs to be a factor and not a numeric variable[5].

The one-way ANOVA model imposes no restrictions on how the response changes due to change of values of the covariate, as there will be one parameter for each distinct value of the covariate. Consequently, it is a more general model than the nonlinear regression model, or, in other words, the nonlinear regression model is a submodel of the ANOVA model[6]. Here we are interested in testing the null hypothesis that the ANOVA model (i.e., full model) can be simplified to the nonlinear regression model (i.e., reduced model). We may adopt two approaches, namely, the F-test and the likelihood ratio test (LRT) for testing the hypothesis[1, 7].

In comparing the two models (full model and reduced model) we shall use extra sum of squares principle. Based on this principle, the F-test statistic is defined as

$$F=\frac{(SSE_{reduced}-SSE_{full})/df\ changes}{SSE_{full}/(n-q)}, F_{df\ changes,(n-q)}$$

which follows distribution approximately under the null hypothesis. Here, SSE stands for error sum of square, df indicates degrees of freedom and q is the number of parameters in the model. Therefore, for a given significance level α (probability of committing a type I error), if $F \geq F_{df\ changes,(n-q)}(1-\frac{\alpha}{2})$ (or, p-value is smaller) then we reject the null hypothesis. Here for the given model, either reduced or full model, where r_i is the residual for i-th observation based on the given model.

Similarly, the above mentioned purpose can be served using the likelihood ratio test where the maximum likelihood function under full model and reduced model are compared. Using equation (2) the likelihood ratio statistic is

$$R_n = 2\{\max\log likelihood\ (full\ model) - \max\log likelihood\ (reduced\ model)\}.$$

If $R_n > \chi^2_{df\ changes,(1-\alpha)}$ then we reject the null hypothesis.

We use F -test and LRT to test the adequacy of the model. In order to calculate the p-values for the above test, we need to use the corresponding approximate distributions. But the distribution about F -test statistic and LRT are just approximate, not exact. If the conclusion based on F test and LRT differ substantially (i.e., not similar) then we can check goodness of the approximation of the distribution of the relevant test statistic. This can be done by implementing the following steps under the assumption that the fitted model is true and the calculated values of the test statistic are obtained from their true distributions.

Step 1: Generate M samples of size n.

Step 2: Determine the value of the desired test statistic (F or LRT statistic) for each of the M random samples.

Step 3: Count the number of the statistic that exceeds $(1-\alpha)100\%$ quantile of the corresponding distribution (F or χ^2 distribution) (let this number is G when we use F distribution or K when we use χ^2 distribution).

Step 4: If the simulated probability of type I error, $\frac{G}{M} \approx \alpha$ then F distribution provides a good approximation to the true distribution of F statistic. If, otherwise, $\frac{K}{M} \approx \alpha$, then χ^2 distribution provides a good approximation to the true distribution of LRT statistic.

III. Results and Discussion: Data Analysis and Simulation

Example Data Analysis

In this research we will use a dataset named 'Chwirut2' from R package NISTnls[8]. These data are the result of a study conducted by National Institute of Standards and Technology (NIST) involving ultrasonic calibration[9]. The dataset contains measurements from an experiment examining how ultrasonic response depends on the metal distance. The response variable is ultrasonic response, and the covariate is metal distance. A preliminary plot of the data is shown in Figure 1[6, 9]. We see that the linear relationship assumption is not reasonable here and the plot seems to exhibit some kind of exponential decay. An exponential class model with three parameters suggested for this data by NIST is given below:

$$y_i = f(x_i, \theta_1, \theta_2, \theta_3) + \epsilon_i = \frac{exp(-\theta_1 x_i)}{\theta_2 + \theta_3 x_i} + \epsilon_i, \qquad (4)$$

where, y and x represent ultrasonic response and metal distance respectively, and $\theta_1, \theta_2, \theta_3$ are the three parameters of the model[5].

The parameters are estimated based on the ultrasonic data. From Table 1[6], we see that all three parameters of the model are significant (since p-values are very small). Therefore, ultrasonic response can be predicted for different values of metal distance using the estimated model that was suggested by NIST.

There are some replications of covariate values in the dataset as evident from Figure 1. We can test whether the model suggested by NIST is adequately fitted to the data by comparing the distribution of F statistic and LRT statistic. That is, we would like to compare the suggested nonlinear regression model to a more general ANOVA model.
From Table 2, we see that the p-value corresponding to F statistic is approximately 0.1912. This indicates that we cannot reject the null hypothesis. That is, we do not have enough evidence to reject the model proposed by NIST. On the other hand, a small p-value of 0.025 is obtained based on the LRT (see Table 3). This leads to reject the null hypothesis. That is, we have strong evidence to reject the model suggested by NIST. Clearly, we have reached a contradiction.

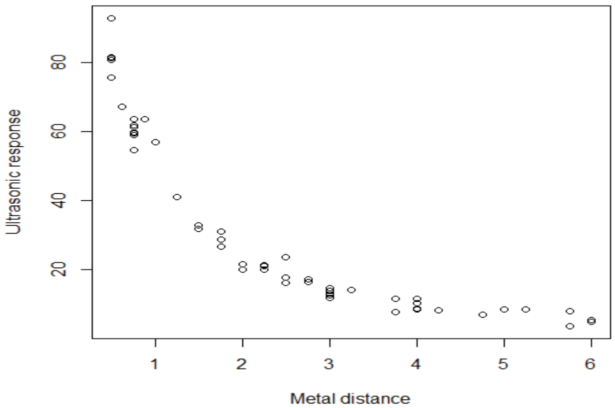

Fig. 1. A plot of ultrasonic response data[6, 9]

Table 1. Estimates and Inferential statistics for the model parameters[6]

Parameters	Estimated values	Standard error	Test statistic, t	p-values
θ_1	0.1666	0.0383	4.349	<0.001
θ_2	0.0052	0.0007	7.753	<0.001
θ_3	0.0122	0.0015	7.939	<0.001

Table 2. Model comparison using ANOVA

Model	Residual df	Residual sum of square	df	Sum of square	F-value	p-value
Reduced model	51	513.05				
Full model	32	279.38	19	233.67	1.4087	0.1912

Table 3. Model comparison using LRT

LRT statistic value	df	p-value
32.822	19	0.025

Table 4. Simulated probability of type I error following four steps of the methods section

Test statistic	Distribution	Simulated probability of type I error
F statistic	F	0.049
LRT statistic	χ^2	0.276

Simulation Study

We conduct a simulation study to make a reasonable decision. In order to reach a more convincing conclusion we should check which approximation is better, the F distribution to the F statistic or the χ^2 distribution to the LRT as discussed in the methods section.

The simulation study is based on the ultrasonic data of size $n = 54$. For the simulation, we consider $M = 20000$ and $\alpha = 0.05$. Note that we have assumed the estimated model (see Table 1) as the true model.The response variable is simulated according to the model mentioned in equation (4), where the error ϵ's are considered to follow a normal distribution with mean 0 and standard deviation 3.172, which is actually the residual standard error obtained by fitting the model suggested by NIST. The steps mentioned in the methods section are implemented through R programming and only relevant output is reported in Table 4.

It is observed from the Table 4 that the simulated probability of making a type I error obtained by using the F distribution corresponding to the F statistic is very close to α, the true probability of making a type I error. Therefore, we say that the F distribution provides a very good approximation to the true distribution of the F statistic. On the other hand, the simulated probability of making a type I error obtained by using the χ^2 distribution corresponding to the LRT statistic is far away from α, the true probability of making a type I error. This indicates that the true quantile of LRT statistic is larger than that of χ^2 distribution. Therefore, the actual p-value for LRT might be much higher than 0.025 as reported in Table 3.

IV. Conclusion

For the given data and model, the F distribution has provided a very good approximation to the true distribution of the F statistic. However, the χ^2 distribution does not do a similar job for the LRT statistic.Through our simulation study we can tell that the result based on F statistic is more convincing. Therefore, we do not have evidence to reject the model suggested by NIST.

Acknowledgements

I am grateful to University of Dhaka for providing all kinds of support in preparing this manuscript.

References

1. Bates, D.M. and D.G. Watts, 1988. Nonlinear Regression Analysis and Its Applications. John Wiley & Sons, New York, 32-65.
2. Begg, C.B., and L.A. Kalish, 1984. Treatment Allocation for Nonlinear Models in Clinical Trials: The Logistic Model. Biometrics. **40**, 409-420.
3. Meeker, W.Q., 1984. A Comparison of Accelerated Life Test Plans for Weibull and Lognormal Distributions and Type I Censoring. Technometrics, **26**, 157-171.
4. Seber, G. A. F., and C. J. Wild, 1989. Nonlinear Regression. John Wiley and Sons, New York, 32-37.
5. Dalgaard, P., 2008. Introductory Statistics with R. Springer-Verlag, New York, 126-133.
6. Ritz, C. and J. C. Streibig, 2008. Nonlinear regression with R. Springer, New York, 93-107.
7. Huet, S., A. Bouvier, M.A. Poursat and E. Jolivet, 2004. Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples. 2nd ed., Springer-Verlag, New York, 166-177.
8. Bates, D. 2012. Nonlinear least squares examples from NIST. Original from National Institutes for Standards and Technology (NIST). https://cran.r-project.org/web/packages/NISTnls/NISTnls.pdf, 3-4.
9. NIST, 1979. Ultrasonic Reference Study Block (D. Chwirut). Technical report. http://www.itl.nist.gov/div898/strd/nls/data/chwirut2.shtml.