Statistika, Vol. 7 No. 2, 33 – 37
Nopember 2007

Selection for Parameter λ by Using Newton-Raphson Method

WAN MUHAMAD AMIR W AHMAD[1] , NYI NYI NAING[2]
AND MOHD TENGKU ARIFF RAJA HUSSEIN[3]

[1, 2]Unit Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, USM, Healthy Campus, 16150 Kubang Kerian, Kelantan, Malaysia.
[3]Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, USM, Healthy Campus, 16150 Kubang Kerian, Kelantan, Malaysia.

ABSTRACT

Many problems in mathematics or statistics involve, at some point or another, solving an equation for an unknown quantity. To solve the problem many of methods are applied as such Maximum Likelihood Method, Least Square Method and Newton-Raphson Method and much more. One of the famous methods is Newton Method. The Newton-Raphson method or Newton–Fourier method is an efficient algorithm for finding approximations to the zeros (or roots) of a real-valued function. As such, it is an example of a root-finding algorithm. It can also be used to find a minimum or maximum value of such a function, by finding a zero in the function's first derivative. In this paper, we will work with Newton-Raphson method to estimate the value of λ for the use of transformation. The value of λ, determine the kind of transformation should be done according to the data. These papers emphasize the Newton Method in finding the value of λ by using C++ software.

Keywords: Newton-Raphson Method and Parameter λ .

1. Introduction to Newton Method

Newton's method was described by Isaac Newton. Basically, the idea of the Newton-Raphson method is as follows: one starts with an initial guess which is reasonably close to the true root, then the function is approximated by its tangent line and one computes the x-intercept of this tangent line. This x-intercept will typically be a better approximation to the function's root than the original guess and the method can be iterated. Suppose $f:[a,b]\to R$ is a differentiable function defined on the interval $[a,b]$ with values in the real numbers R. The formula for converging on the root can be easily derived. Suppose we have some current approximation x_n. Then we can derive the formula for a better approximation, x_{n+1} . We know from the definition of the derivative at a given point that it is the slope of a tangent at that point. That is

$$f'(\lambda)=\frac{f(\lambda_n)-0}{\lambda_n-\lambda_{n+1}}=\frac{0-f(\lambda_n)}{(\lambda_{n+1}-\lambda_n)}$$

Here, f' denote the derivative of the function f. Then by simple algebra we can derive $\lambda_{n+1}=\lambda_n-\frac{f(\lambda)}{f'(\lambda)}$ and it is equivalent to $\lambda_{n+1}=\lambda_n-\frac{f'(\lambda)}{f''(\lambda)}$.

For the Newton Method, we usually start the process off with some arbitrary initial value x_0 (Usually closer to the zero is better - in the absence of any intuition about where the zero might lie, a "guess and check" method might narrow the possibilities to a reasonably small interval by appealing to the intermediate value theorem). The method will usually converge provided this initial guess is close enough to the unknown zero. Furthermore, for a zero of multiplicity 1, the convergence is at least quadratic in a neighbourhood of the zero, which intuitively means that the number of correct digits roughly at least doubles in every step.

2. Transformation to Normality

Data screening is the most important technique to check the nature of the data. One of the methods to screen the data is the transformation method. There are a lot of transformations but the famous one is the Box-Cox Transformation. The Box-Cox family of transformation is a well-known approach to make data behave accordingly to assumption of linear regression and ANOVA. Moore and Tukey (1954) initiated the study of family of transformations. They introduced the simple family of transformations

$$\rho_{\lambda_1,\lambda_2}(X) = \begin{cases} (X+\lambda_1)^{\lambda_2}, & 0 \neq \lambda_2 \leq 1, \\ \log(X+\lambda_1), & \lambda_2 = 0, \end{cases} \quad \text{for } X+\lambda_1 > 0.$$

In a practical application, the selection of λ_1 and λ_2 depends on what is desired by transformation. Moore and Tukey (1954) select λ_1 and λ_2 so that interactions are not significant and additivity holds in two-way layouts. Later Box and Cox (1964) modified the simple family of transformation and considered the following transformation:

$$\rho_\lambda(X) = \begin{cases} \dfrac{X^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(X), & \lambda = 0, \end{cases} \quad \text{for } X > 0.$$

To obtain approximate normality from symmetric long-tailed distributions, John and Draper (1980) propose another family called the family of modulus transformations

$$\rho_\lambda(X) = \begin{cases} \operatorname{sgn}(X)\{[(|X|+1)^\lambda - 1]/\lambda\}, & \lambda \neq 0, \\ \operatorname{sgn}(X)\log(|X|+1), & \lambda = 0, \end{cases}$$

Where $\operatorname{sgn}(\cdot)$ is the sign function. The modulus family is monotonic, continuous in λ and is applicable in the presence of negative data. Note that when the observations are all positive, the transformation reduces to the Box-Cox power transformation family. The basic idea is to apply the same power transformation to both tails of a distribution symmetric about zero. Another modification to the Box-Cox transformation by Bickel and Doksum (1981) is

$$P_{BD\lambda}(X) = [X^\lambda \operatorname{sgn}(X) - 1]/\lambda,\ \lambda > 0,\ X \in (-\infty, \infty)$$

for an unlimited support.

3. An Approach of Newton Method to Determine the Parameter λ

The normal distribution and non-normal distribution arises in many areas of statistics. For example, the sampling distribution of the sample mean is approximately normal, even if the distribution of the population from which the sample is taken is not normal. Non-normality is a way of life, since no characteristic will have exactly a normal distribution. One strategy to make non-normal data resemble normal data is by using a transformation strategy. There is no dearth of transformations in statistics; the issue is which one to select for the situation at hand. Unfortunately, the choice of the "best" transformation is generally not obvious.

Since the seminal by Box and Cox (1964), the Box-Cox types of power transformation have generated a great deal of interests, both in theoretical work and in practical applications. The Box-Cox family of transformation has become a widely used tool to make data behave according to a linear regression model. Box and Cox (1964) modified the simple family of transformation and considered the following transformation:

$$\rho_\lambda(X) = \begin{cases} \dfrac{X^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(X), & \lambda = 0, \end{cases} \quad \text{for } X > 0 \tag{1}$$

We focus our attention on transforming a random sample from a distribution to near normality. We also assume that, for some λ, the transformed observation variables can be treated as normally distributed with some mean μ and variance σ^2. Given the vector of data observations $x = (x_1, \ldots, x_n)'$, one way to select the power λ is to use the λ that maximizes the logarithm of the likelihood function

$$l_n(\theta \mid x) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{\xi(\lambda, x_i) - \mu\}^2 + (\lambda - 1)\sum_{i=1}^{n} sgn(x_i)log(|x_i| + 1) \quad (2)$$

where $\theta = (\lambda, \mu, \sigma^2)$.To calculate the λ value, we use the approached of Newton-Raphson method to estimate the value of λ. In this research, we are using the Newton-Raphson Method to find optima λ and also using C++ programming language to build the program. The flowchart in Figure 1.1 summarizes the flow of calculation by C++ program.

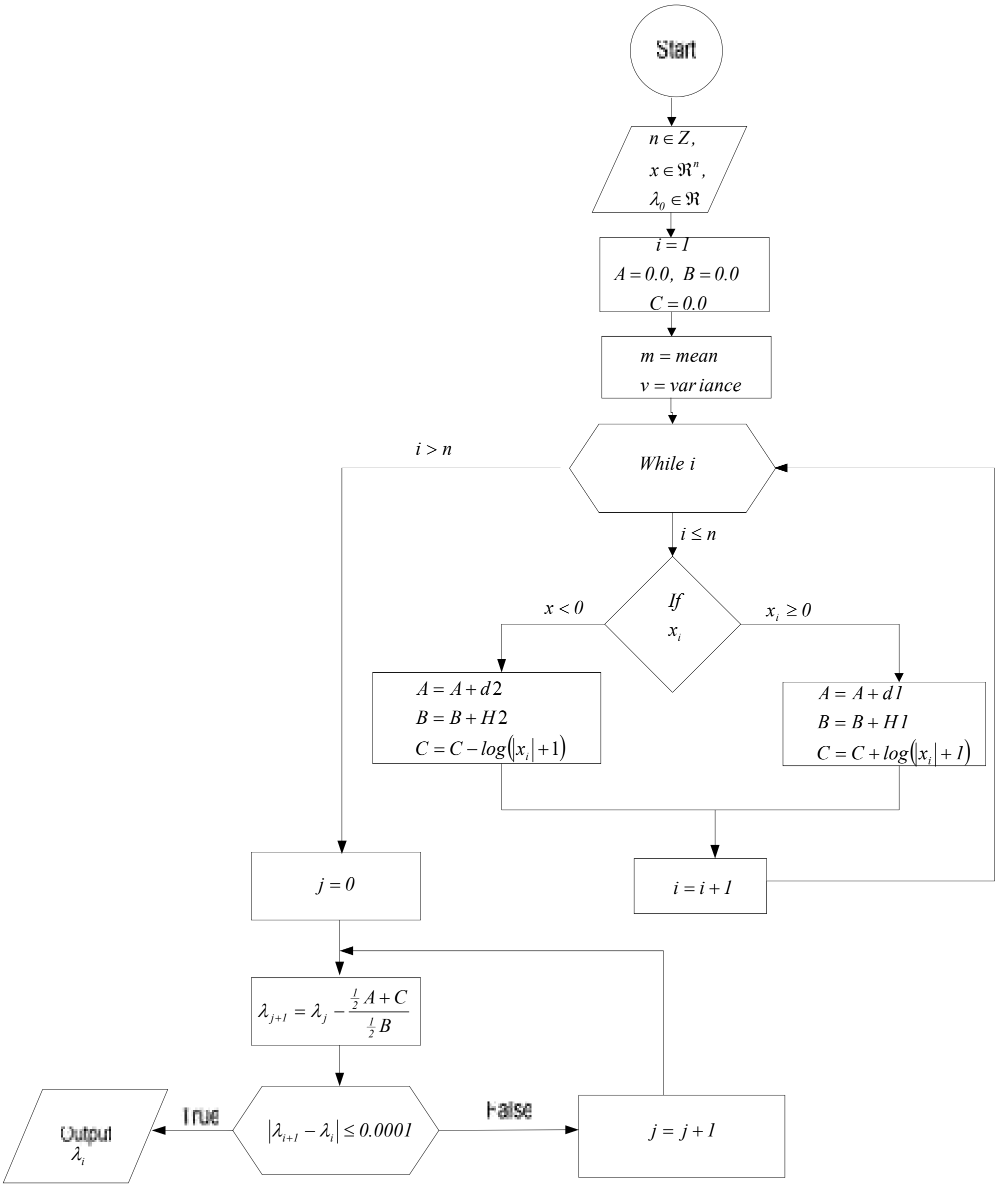

Figure 1.1 Flowchart of Calculation Using Newton-Raphson Method

Notes:

$$A = A + d1 = A + 2\left(\frac{(x_i+1)^{\lambda}-1}{\lambda} - m\right)\left(\frac{(x_i+1)^{\lambda}\ln(x_i+1)}{\lambda} - \frac{(x_i+1)^{\lambda}-1}{\lambda^2}\right)$$

$$A = A + d2 = A + 2\left(-\frac{(-x_i+1)^{2-\lambda}-1}{2-\lambda} - m\right)\left(\frac{(-x_i+1)^{2-\lambda}\ln(-x_i+1)}{2-\lambda} - \frac{(-x_i+1)^{2-\lambda}-1}{(2-\lambda)^2}\right)$$

$$B = B + H1 = B + 2\left(\frac{(x_i+1)^{\lambda}\ln(x_i+1)}{\lambda} - \frac{(x_i+1)^{\lambda}-1}{\lambda^2}\right)^2 + 2\left(\frac{(x_i+1)^{\lambda}-1}{\lambda} - m\right)\left(\frac{(x+1)^{\lambda}\ln(x_i+1)^2}{\lambda} - \frac{2(x_i+1)^{\lambda}\ln(x_i+1)}{\lambda^2} + \frac{2((x_i+1)^{\lambda}-1)}{\lambda^3}\right)$$

$$B = B + H2 = B + 2\left(\frac{(-x_i+1)^{2-\lambda}\ln(-x_i+1)}{2-\lambda} - \frac{(-x_i+1)^{2-\lambda}-1}{(2-\lambda)^2}\right)^2 + 2\left(-\frac{(-x_i+1)^{2-\lambda}-1}{2-\lambda} - m\right)\left(-\frac{(-x+1)^{2-\lambda}\ln(x_i+1)^2}{2-\lambda} + \frac{2(-x_i+1)^{2-\lambda}\ln(-x_i+1)}{(2-\lambda)^2} - \frac{2((-x_i+1)^{2-\lambda}-1)}{(2-\lambda)^3}\right)$$

The procedure of this Newton method can be described as follows.

Data: $\lambda_0 \in \Re, \varepsilon \in \Re$

1. Set $i = 0$
2. do $\lambda_{i+1} = \lambda_i - \dfrac{l'(\theta \mid x)}{l''(\theta \mid x)}$
3. While $\|\lambda_{i+1} - \lambda_i\| \geq \varepsilon,$ do
   - 3.1 $i = i + 1$
   - 3.2 $\lambda_{i+1} = \lambda_i - \dfrac{l'(\theta \mid x)}{l''(\theta \mid x)}$
4. $\lambda_k = \lambda_{i+1}$

4. Numerical Results

We are using the data from the "Effects of Cross and Self-Fertilization in the Vegetable Kingdom" Study by Darwin (1876). In his study he used fifteen pair of seedling of the same age, one produced by cross-fertilization and the other by self-fertilization, were grown together so that the members of each pair were reared under nearly identical conditions. The differences between the final heights of plants in each pair after a fixed period of time were 6.1, -8.4, 1.0, 2.0, 0.7, 2.9, 3.5, 5.1, 1.8, 3.6, 7.0, 3.0, 9.3, 7.5 and -6.0. When we applied the equation in (2) by using the Newton-Raphson to the data above, the obtain value of estimate $\hat{\lambda} = 0.970946$. Figure 4.1 summarize the results of λ according to different method.

Method	Maximum Likelihood	Newton Raphson	Method that minimizes the standard deviation of data
Value of λ	1.0305	0.970946	0.52

5. Conclusion

The Box-Cox algorithm provides a simple method to determine the best way to transform the data for reducing heterogeneity of errors if all the data is in positive value condition. Box-Cox algorithm cannot compute if negative value is included. To calculate the value of λ we have to modify the formula given so that it is valid for all the values from negative infinity to positive infinity. After getting the value of lambda, the next step can be done that is transformation by substituting the value of λ to the modified formula of Box-Cox.

Reference

[1] Box, G.E.P., and D.R. Cox (1964). An Analysis of Transformation. J.R. Stat Soc. B 26:211-252.
[2] Moore, P.G. and Tukey, J.W (1954). Answer to Query 112. Biometrics, 10, 562-568.
[3] Draper, N. R. & John, J. A. (1980). Testing For Three or Fewer Outliers in Two-Way Tables. Technometrics, 22, 9 -15
[4] Bartlett, M.S. (1947). The Use of Transformation. Biometrical 3 39-52.
[5] Bickel, P. J. & Doksum, K. A. (1981). An Analysis of Transformations Revisited. Journal of the American Statistician Association, 76, 269-311.

[6] In-Kwon Yeo and Johnson, R.A (2000). A New Family of Power Transformation to Improve Normality or Symmetry. Biometrika, 87, 4, pp 954-959.
[7] Goh K.W. and Ismail Mohd (2007). An Alternative Newton-like Exact Line Search For Steepest Descent Method. Proceedings of The Third Conferences On Research And Education In Mathematics. page 222-227