2004

Newton-Raphson Versus Fisher Scoring Algorithms in Calculating Maximum Likelihood Estimates

Andrew Schworer

Peter Hovey

Electronic Proceedings of Undergraduate Mathematics Day
2004, No. 1, http://www.udayton.edu/~mathdept/EPUMD/EPMUD.htm

Newton-Raphson Versus Fisher Scoring Algorithms in Calculating Maximum Likelihood Estimates

Andrew Schworer and Dr. Peter Hovey
University of Dayton Research Institute
Dayton, OH 45409
Email: {schworap | peter.hovey}@notes.udayton.edu

Abstract: In this work we explore the difficulties and the means by which maximum likelihood estimates can be calculated iteratively when direct solutions do not exist. The Newton-Raphson algorithm can be used to do these calculations. However, this algorithm has certain limitations that will be discussed. An alternative algorithm, Fisher scoring, which is less dependent on specific data values, is a good replacement. The Fisher scoring method converged for data sets available to the authors, that would not converge when using the Newton-Raphson algorithm. An analysis and discussion of both algorithms will be presented. Their real world application on analysis of jet engine part inspection data will also be discussed.

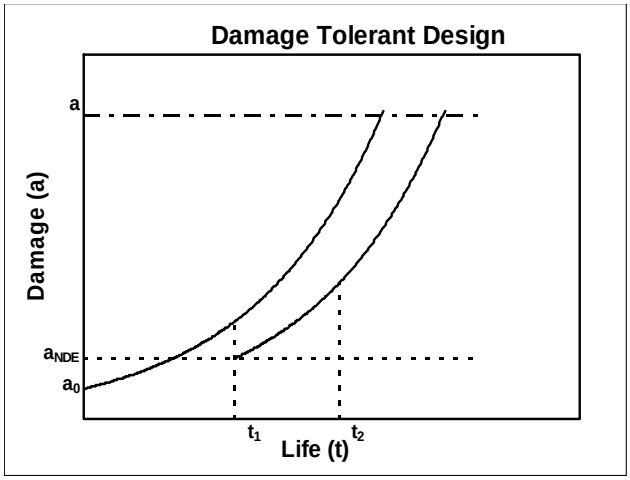
This research was initiated with the end goal of improving the safety and reliability of aircraft through understanding the capabilities of current nondestructive inspection techniques. The main emphases in designing "safe" aircraft are to prevent failures and to ensure that the failure of one component will not result in a loss of the aircraft. In the safe-life concept individual aircraft are retired from service when they have flown a specified number of flight hours or a specified number of missions. This predetermined length of time is called the aircraft's design life. The design life is based

on mission requirements and conservative predictions of crack growth are used to establish structural requirements.

Fail-safe structures utilize redundant load paths so that the aircraft will not fail catastrophically when one structural element fails. The Air Force has combined the safe life and fail-safe concepts into what is called Damage Tolerant Design (1,2). This design concept uses redundant load paths in the structure and requires planned inspections at regular intervals, to catch and repair damage before a structural element fails.

Figure 1 shows the graph of a typical inspection cycle for a given aircraft part. The figure is graphed as the Life of the part in flight hours or missions versus the Damage size or crack size. The first curve of this graph begins at a_0. This value is the assumed crack size after manufacturing. No manufacturing methods are perfect, so it is assumed that crack growth from manufacturer's defects is bounded by crack growth that starts at a_0.

Figure 1. Damage Tolerant Design Graph



The crack size grows until it reaches size a, at which the structure fails. In this graph non-destructive inspections are done at times t_1 and t_2. These inspections are done at half the time it would take the crack to grow to failure. When an inspection is done it reduces

the assumed crack size to the level of which the inspection is capable of detecting, designated a_{NDE} in the figure. The process is then repeated using a_{NDE} as the starting crack length to determine subsequent inspection times.

In order to analyze an inspection process it is vital to understand how large a defect the process might not detect. The United States Air Force and commercial airlines currently use probability of detecting a crack as a measure of inspection capability to use in determining inspection intervals. It is important that the probability of detection be related to flaw size because Damage Tolerant Design uses the flaw size in life prediction and management of the inspection intervals.

The main purpose of non-destructive inspection is to find flaws so that they can be repaired and eliminated as a potential source of a failure. The link between aircraft safety and non-destructive inspection capability is the probability of detecting, and then eliminating, the flaw. Ideally the probability of detection would be a step function, but that is unrealistic. The lognormal distribution function has shown to be a good fit as a model of the probability of detection in a wide variety of applications (3). Therefore, we use this function to fit the inspection data to the probability of detection curve.

Figure 2 is an illustration of the data collected for eddy current inspections on retired aircraft parts. The graph is a depiction of crack size versus the measured probability of detection. Each point represents the results of 60 inspections of the same crack performed by different inspectors at different military inspection depots around the country. It can be gathered from the graph that the larger the crack the higher probability of detection. An interesting feature that was not recognized prior to this study, but is

evident from the graph is that different flaws of the same size are detected at different rates.

Figure 2. "Have Cracks Will Travel" Data - Example

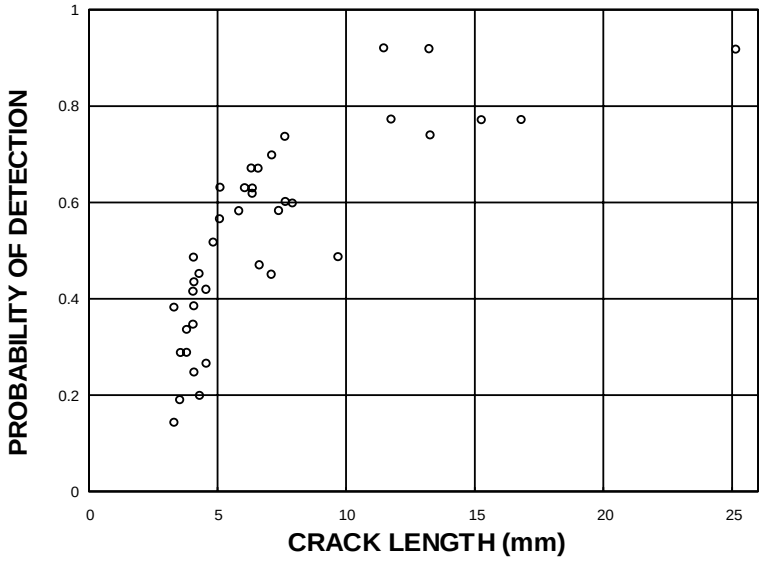


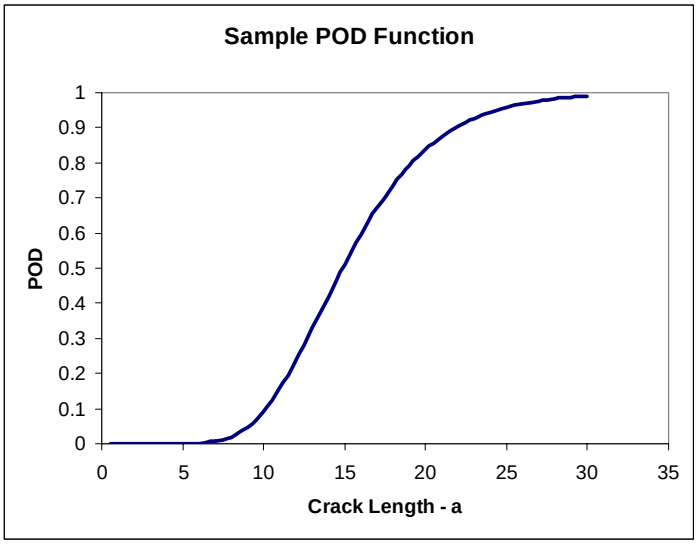
Figure 3 is a depiction of the probability of detection function using the lognormal distribution function shown in equation 1. The parameter μ is the logarithm of the crack length that will be detected 50 percent of the time and the parameter σ is a measure of how steeply the central portion of the curve rises.

Equation 1. Lognormal Distribution Function

$$POD(a) = \Phi\left(\frac{\ln(a) - \mu}{\sigma}\right)$$

where Φ(z) is the cumulative standard normal distribution function.

Figure 3. Sample Probability of Detection Function



The data that was used to fit to the probability of detection curve came from Pass/Fail inspection results. In the Pass/Fail analysis a set of specimens with known crack sizes is inspected and the inspection result is whether or not the crack was detected.

The likelihood function is given in equation 2, where $Y_{i,}$ which is either 1 or 0, and a_i is the crack size is the result of the inspection. When the crack was found, Y_i is 1, and the likelihood function resolves to $POD(a_i)$. If the crack was not found, Y_i is 0, and the function becomes $1\text{-}POD(a_i)$.

Equation 2. Likelihood equation

$$L(\mu,\sigma) = \prod_{i=1}^{n} POD(a_i)^{Y_i} (1 - POD(a_i))^{1-Y_i}$$

Estimates of the μ and σ parameters are found by setting the derivatives of the log of the likelihood function equation to zero and solving for μ and σ respectively. Equations 3 and 4 are the maximum likelihood equations for estimating μ and σ.

Equation 3. Maximum Likelihood Equation

$$0 = \sum \frac{(Y_i - \hat{\mu})}{P_i(1-P_i)} \phi\left(\frac{\ln(a_i) - \hat{\mu}}{\hat{\sigma}}\right)$$

Equation 4. Maximum Likelihood Equation

$$0 = \sum \frac{(Y_i - \hat{\mu})}{P_i(1-P_i)} \phi\left(\frac{\ln(a_i) - \hat{\mu}}{\hat{\sigma}}\right)(\ln(a_i) - \hat{\mu})$$

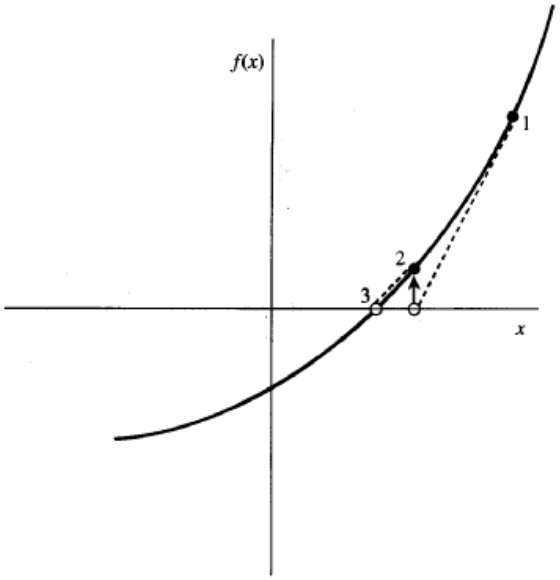
where $\phi(z)$ is the standard normal density function. $\left(\Phi(z) = \int_{-\infty}^{z} \phi(x)\partial x\right)$

The Newton-Raphson algorithm was used initially used to solve these equations for the $\hat{\mu}$ and $\hat{\sigma}$ parameters. The Newton-Raphson algorithm, shown in equation 5, is an iterative procedure for finding the zero of a function (4,5). Starting with an initial guess per the zero of the function the Newton-Raphson algorithm approximates the function with the tangent line. The zero of the tangent line replaces the initial guess and the process repeats until convergence is found or not. The Newton-Raphson function is shown in Equation 7. This iterative process is graphically shown in Figure 4.

Equation 5. Newton-Raphson

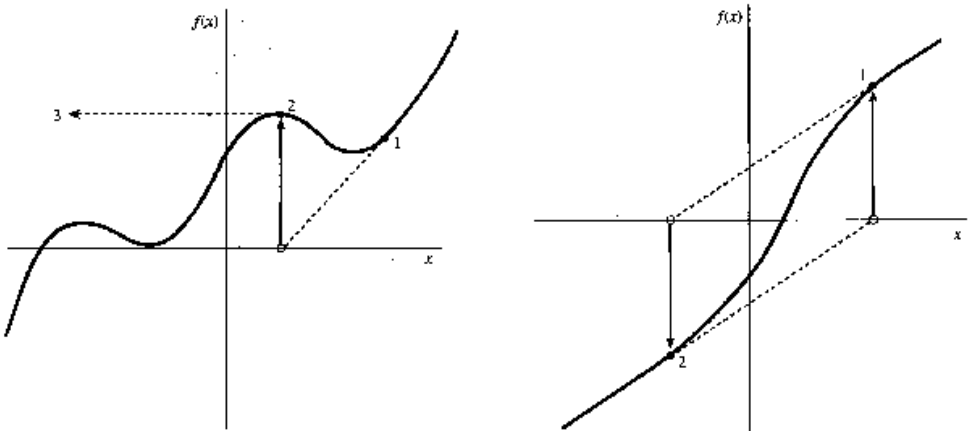
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Figure 4. Newton-Raphson Algorithm



This is an elegant and simple way to determine the roots of a function. However, the Newton-Raphson algorithm can fail in some cases. Two cases where this occurs are shown in Figure 5. If the algorithm encounters a local minimum or maximum the function will evaluate to infinity and never find a root. Similarly, if the function is malformed the algorithm could put itself into an infinite loop, evaluating points over and over again.

Figure 5. Newton-Raphson Failure Cases



The equations 6 and 7 below are the functions for the incremental changes in the parameter estimates (deltas) for the next iteration of the Newton-Raphson algorithm for estimating the probability of detection function parameters. An important point to make at this point in the analysis of the Newton-Raphson algorithm is to realize that the second derivatives of these delta functions are taken directly from the Pass/Fail analysis data. This data, as stated earlier, is binary in nature; meaning it is either a '1' or a '0'. Taking the derivates from this type of data creates undesirable results.

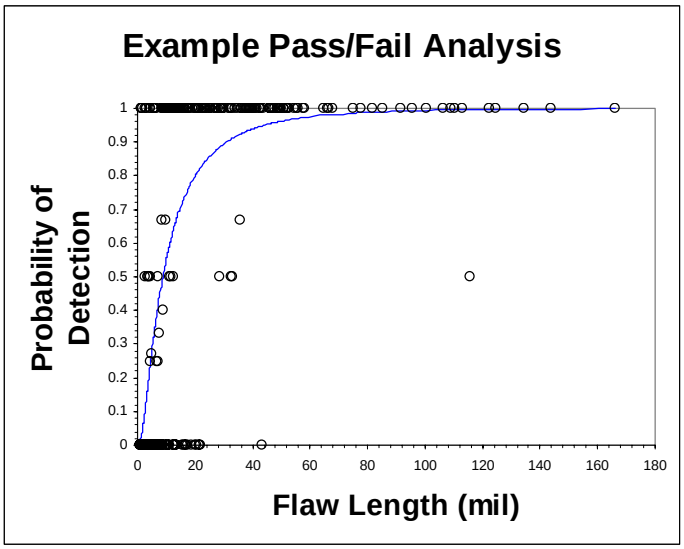
Equations 6 & 7. Delta Equations from Newton-Raphson

$$\frac{\partial \ln(L)}{\partial \mu} + \Delta\mu \frac{\partial^2 \ln(L)}{\partial \mu^2} + \Delta\sigma \frac{\partial^2 \ln(L)}{\partial \mu \partial \sigma} = 0$$

$$\frac{\partial \ln(L)}{\partial \sigma} + \Delta\mu \frac{\partial^2 \ln(L)}{\partial \mu \partial \sigma} + \Delta\sigma \frac{\partial^2 \ln(L)}{\partial \mu^2} = 0$$

These equations are much too dependent on the values of Y_i. This can cause convergence problems. The actual distribution of the data plot can be seen in figure 6. The significant data points are in the scatter area near the middle of the graph. An alternative approach is to use the expected values of the second derivative of the ln(L). This is called Fisher scoring. The Fisher scoring method is less dependent on individual Y_i values and provides more stable convergence (4).

Figure 6. Probability of Detection Graph



Fisher scoring is also known as Iteratively Reweighted Least Squares estimates. The Iteratively Reweighted Least Squares equations can be seen in equation 8. This is basically the Sum of Squares function with the weight (w_i) being accounted for. The further away the data point is from the middle scatter area of the graph the lower the weight, and the closer the data point is to the middle scatter area the higher the weight. This algorithm gives a better fit because it takes into account the relative weight of the data with respect to the middle scatter area.

Equation 8. Iteratively Reweighted Least Squares

$$\Sigma(y_i - \hat{y})^2 w_i$$

Equations 9 and 10 show the new delta functions found from the new algorithm (6). Although these functions of the deltas look complicated they are quite simple. They are linear functions of μ and σ. Moreover, an important point here is that the second derivatives have been replaced with the expected values of those derivatives.

Equations 9 & 10. Delta Equations from Iteratively Reweighted Least Squares

$$\Delta\hat{\mu}\sum\frac{1}{\hat{\sigma}_k^2 P_k(1-P_k)}\left(\phi\left(\frac{\ln(a_i)-\hat{\mu}_k}{\hat{\sigma}_k}\right)\right)^2+\Delta\hat{\sigma}\sum\frac{-(\ln(a_i)-\hat{\mu}_k)}{\hat{\sigma}_k^3 P_k(1-P_k)}\left(\phi\left(\frac{\ln(a_i)-\hat{\mu}_k}{\hat{\sigma}_k}\right)\right)^2$$
$$=\sum\frac{Y_i-P_k}{\hat{\sigma}_k P_k(1-P_k)}\phi\left(\frac{\ln(a_i)-\hat{\mu}_k}{\hat{\sigma}_k}\right)$$

$$\Delta\hat{\mu}\sum\frac{-(\ln(a_i)-\hat{\mu}_k)}{\hat{\sigma}_k^3 P_k(1-P_k)}\left(\phi\left(\frac{\ln(a_i)-\hat{\mu}_k}{\hat{\sigma}_k}\right)\right)^2+\Delta\hat{\sigma}\sum\frac{(\ln(a_i)-\hat{\mu}_k)^2}{\hat{\sigma}_k^4 P_k(1-P_k)}\left(\phi\left(\frac{\ln(a_i)-\hat{\mu}_k}{\hat{\sigma}_k}\right)\right)^2$$
$$=\sum\frac{-(Y_i-P_k)(\ln(a_i)-\hat{\mu}_k)}{\hat{\sigma}_k^2 P_k(1-P_k)}\phi\left(\frac{\ln(a_i)-\hat{\mu}_k}{\hat{\sigma}_k}\right)$$

The advantage of Fisher scoring was evaluated with simulated data sets and depends to a large extent on sample size. In the simulations, 50 data sets were generated with a sample of size 100 and 50 with a sample size of 300. The Newton-Raphson method failed on one of the sample size 100 data sets but worked for all of the sample size 300 data sets. The Fisher scoring method did work on all data sets in the study. In his comments on Fisher scoring, Knight (4) noted that the Newton-Raphson algorithm generally converges more quickly, while Fisher scoring is more robust and will converge when Newton-Raphson doesn't.

References:

1) "Aircraft Structural Integrity Program, Airplane Requirements," Military Standard MIL-STD-1530A, 1975.

2) "Airplane Damage Tolerance Requirements," Military Specifications MIL-A-83444, 1974.

3) Hovey, P. W. and A. P. Berens, "Evaluating Crack Detection Capability for Aircraft Risk Analysis," The 1997 Proceedings of the Section on Physical and Engineering Sciences of the American Statistical Association, Washington, D. C., 1998.

4) Knight, Keith, Mathematical Statistics, Chapman & Hall/CRC, Boca Raton, 2000.

5) Press, William H., S. A. Teuklolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed., Cambridge University Press, New York, 1992.

6) Finney, D. J., Probit Analysis, Cambridge University Press, London, 1964.

Acknowledgement: The authors would like to thank the United States Air Foce. This work conducted as part of the University of Dayton Research Institute Turbine Engine Sustainment Initiative supported by the USAF Contract number #F42620-00-D-0039/RZ02.