Otterbein University

Digital Commons @ Otterbein

Mathematics Faculty Scholarship   Mathematical Sciences

2009

Reliability Confidence Intervals for Oil Spills in the Gulf of Mexico

William V. Harper
Otterbein University, wharper@otterbein.edu

Thomas R. James
Otterbein University, TJames@otterbein.edu

Ted G. Eschenbach
TGE Consulting

Follow this and additional works at: https://digitalcommons.otterbein.edu/math_fac

🎨 Part of the Mathematics Commons, and the Statistics and Probability Commons

Reliability Confidence Intervals for Oil Spills in the Gulf of Mexico
William V. Harper[1](wharper@otterbein.edu), Ted G. Eschenbach[2], Thomas R. James[1]
[1]Mathematical Sciences, Otterbein College, 1 Grove Street, Westerville, OH 43081-2006
[2]TGE Consulting, 4376 Rendezvous Circle, Anchorage, AK 99504

Abstract
An extensive study [Eschenbach and Harper (2006)] of offshore oil spills in the Gulf of Mexico
with extensions to the northern seas of Alaska involved the estimation of the likelihood of oil
spill volumes in the Gulf of Mexico for both pipeline and platform spills. This paper develops
both maximum likelihood based reliability and percentile confidence intervals for the 3-
parameter Weibull distribution. The statistical aspects are discussed along with applications of
developed Excel VBA functions. The Excel routines are available free on the web at
http://faculty.otterbein.edu/WHarper/. The functions are illustrated with Gulf of Mexico oil spill
data.

Introduction
One of the challenges of a literature search is keeping track of the Weibull parameter notation
and terminology. Below are both the pdf and cdf used in this paper. The 3rd Weibull parameter
γ goes by a variety of names such as location, minimum life, threshold, shift, and origin. The
parameter α is generally called the scale and β is either shape or slope (typically in probability
paper or rank regression based approaches).

$$pdf: f(x) = \beta \alpha^{-\beta}(x-\gamma)^{\beta-1}e^{-((x-\gamma)/\alpha)^\beta} \ for \ x > \gamma; \ 0 \text{ otherwise.} \ \ cdf: F(x) = 1 - e^{-((x-\gamma)/\alpha)^\beta}$$

Maximum likelihood estimation (MLE) has much to offer (consistency, asymptotic normality,
and asymptotic efficiency) but such properties are based on large samples. In practice large
samples may not be available and thus the often stated advantages of maximum likelihood
estimation may not be applicable to moderate sized samples. Nonetheless, maximum likelihood
is a very common method for difficult parameter estimation and is the method used in the
developed Excel VBA code. Below is the development of the key aspects of MLE based
confidence intervals for the Weibull distribution.

Likelihood Based Derivation of Weibull Confidence Intervals
For n independent and identically distributed three parameter Weibull random variables, the
likelihood function L is defined below. The more standard two parameter Weibull has γ = 0.

$$L = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \beta \alpha^{-\beta}(x_i - \gamma)^{\beta-1}e^{-((x_i-\gamma)/\alpha)^\beta}$$

Since finding the general three parameter Weibull MLEs is difficult, Lawless (2003) suggests
maximizing the profile log-likelihood function. A profile log-likelihood fixes one or more
parameters of the distribution and then maximizes the log-likelihood for the remaining
parameters. This is performed for the three parameter Weibull by assessing potential values for
the threshold and then maximizing the log-likelihood at this fixed threshold. This in essence is
just solving the two parameter MLE problem for $x_i^* = x_i - \hat{\gamma}$ where $x_i, \hat{\gamma}$ are the original data
values and current value of the threshold, respectively. This is an iterative process and how the

Weibull MLE VBA code works as described in Harper, James, Eschenbach, and Slauson (2008). The natural log (ln) of the Likelihood function, ln (L), is as follows where $x_i^* = x_i - \hat{\gamma}$.

$$\ln(L) = \sum_{i=1}^{n}\left[\ln\beta + (\beta-1)\ln(x_i^*) - \beta\ln(\alpha) - \left(\frac{x_i^*}{\alpha}\right)^{\beta} \right]$$

While maximizing ln(L) is essential to finding the maximum likelihood parameter estimates, it is not the main purpose in this paper. Instead the focus is on confidence interval development for percentiles and reliability. Using the profile log-likelihood approach to find an estimate of γ, we treat it as known with zero variability once it is found. Solving the full three parameter maximum likelihood problem directly is currently only tried to the best of our knowledge by two software vendors (SAS, Reliasoft) with interesting results as documented in Harper, Eschenbach, and James (submitted to *The American Statistician* in 2009). To develop the asymptotic maximum likelihood based confidence intervals it is necessary to find the Fisher Information matrix F below.

$$F = \begin{bmatrix} -\dfrac{\partial^2 \ln(L)}{\partial\beta^2} & -\dfrac{\partial^2 \ln(L)}{\partial\beta\,\partial\alpha} \\[2ex] -\dfrac{\partial^2 \ln(L)}{\partial\beta\,\partial\alpha} & -\dfrac{\partial^2 \ln(L)}{\partial\alpha^2} \end{bmatrix}$$

The variances of the estimated parameters are then defined by the inverse of the Fisher Information matrix given below. These quantities are used later in the confidence interval development once the necessary partial derivatives have been computed.

$$F^{-1} = \begin{bmatrix} \mathrm{var}(\hat{\beta}) & \mathrm{cov}(\hat{\beta},\hat{\alpha}) \\ \mathrm{cov}(\hat{\beta},\hat{\alpha}) & \mathrm{var}(\hat{\alpha}) \end{bmatrix}$$

Recalling that the derivative of a^x equals $[\ln(a)]a^x$, the first partial of ln(L) with respect to β is derived below. Subsequent partial derviatives follow to develop the Fisher Information matrix F.

$$\frac{\partial \ln(L)}{\partial\beta} = \sum_{i=1}^{n}\left[\frac{1}{\beta} + \ln(x_i^*) - \ln(\alpha) - \left[\ln\left(\frac{x_i^*}{\alpha}\right)\right]\left(\frac{x_i^*}{\alpha}\right)^{\beta} \right]$$

$$\frac{\partial^2 \ln(L)}{\partial\beta^2} = \sum_{i=1}^{n}\left[\frac{-1}{\beta^2} - \left[\ln\left(\frac{x_i^*}{\alpha}\right)\right]^2\left(\frac{x_i^*}{\alpha}\right)^{\beta} \right]$$

$$\frac{\partial^2 \ln(L)}{\partial \beta \, \partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \left[\frac{1}{\beta} + \ln(x_i^*) - \ln(\alpha) - \left[\ln\left(\frac{x_i^*}{\alpha}\right) \right] \left(\frac{x_i^*}{\alpha}\right)^{\beta} \right]$$

$$= \sum_{i=1}^{n} \left[\frac{-1}{\alpha} - \left[\ln(x_i^*) - \ln(\alpha) \right] \left(x_i^* \right)^{\beta} \left(-\beta \alpha^{-(\beta+1)} \right) + \frac{1}{\alpha} \left(\left(x_i^* \right)^{\beta} \alpha^{-\beta} \right) \right]$$

$$= \sum_{i=1}^{n} \left[\frac{-1}{\alpha} + \frac{1}{\alpha} \left(\left(\frac{x_i^*}{\alpha} \right)^{\beta} \right) \left[\beta \ln\left(\frac{x_i^*}{\alpha} \right) + 1 \right] \right]$$

$$\frac{\partial^2 \ln(L)}{\partial \alpha^2} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \left[-\frac{\beta}{\alpha} - \left(x_i^* \right)^{\beta} (-\beta) \alpha^{-(\beta+1)} \right] = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \left[-\frac{\beta}{\alpha} - \frac{\beta \left(x_i^* \right)^{\beta}}{\alpha^{(\beta+1)}} \right]$$

$$= \sum_{i=1}^{n} \left[\frac{\beta}{\alpha^2} - (\beta+1) \frac{\beta \left(x_i^* \right)^{\beta}}{\alpha^{(\beta+2)}} \right]$$

Reliability Confidence Intervals

Confidence intervals for reliability can be found using the expressions below [Dodson and Nolan (2002); Dodson (2006)]. The reliability $R(x^*) = 1-F(x^*)$ where the cdf $F(x^*) = 1 - \exp(-(x^*/\alpha)^{\beta})$. To solve for the reliability for the user's input x in the original non-shifted units, the VBA code computes $x^* = x - \hat{\gamma}$ to get $R(x^*)$ which is the estimated probability that the three parameter Weibull random variable exceeds x in the non-shifted (by γ) units. If x is time for say the lifetime of a part, then $R(x^*)$ represents the likelihood that the part will last for at least x time (in the non-shifted time units). The corresponding confidence interval is adjusted for the original units.

$$\exp\left[-\exp\left(\hat{u} + z_{a/2} \sqrt{\text{var}(\hat{u})} \right) \right] \leq R(x^*) \leq \exp\left[-\exp\left(\hat{u} - z_{a/2} \sqrt{\text{var}(\hat{u})} \right) \right]$$

$$\hat{u} = \hat{\beta} \left[\ln(x^*) - \ln(\hat{\alpha}) \right]$$

$$\text{var}(\hat{u}) \approx \beta^2 \left[\left(\frac{\text{var}(\hat{\alpha})}{\hat{\alpha}^2} \right) + \left(\frac{\hat{u}^2 \, \text{var}(\hat{\beta})}{\hat{\beta}^4} \right) - \left(\frac{2\hat{u} \, \text{cov}(\hat{\beta}, \hat{\alpha})}{\hat{\beta}^2 \hat{\alpha}} \right) \right]$$

An example application for pipeline oil spill data in the Gulf of Mexico follows. This data may be found on the web along with the software. Eschenbach and Harper (2006) describe this n = 36 data set in detail. Before computing either reliability or percentile confidence intervals in the VBA software, it is necessary to first compute the Weibull MLEs using the function below. This is described in more detail in Harper, James, Eschenbach, and Slauson (2008) as well as in the user documentation on the web. The Excel VBA function Compute3ParmWeibull as seen below returns the 3 Weibull MLEs.
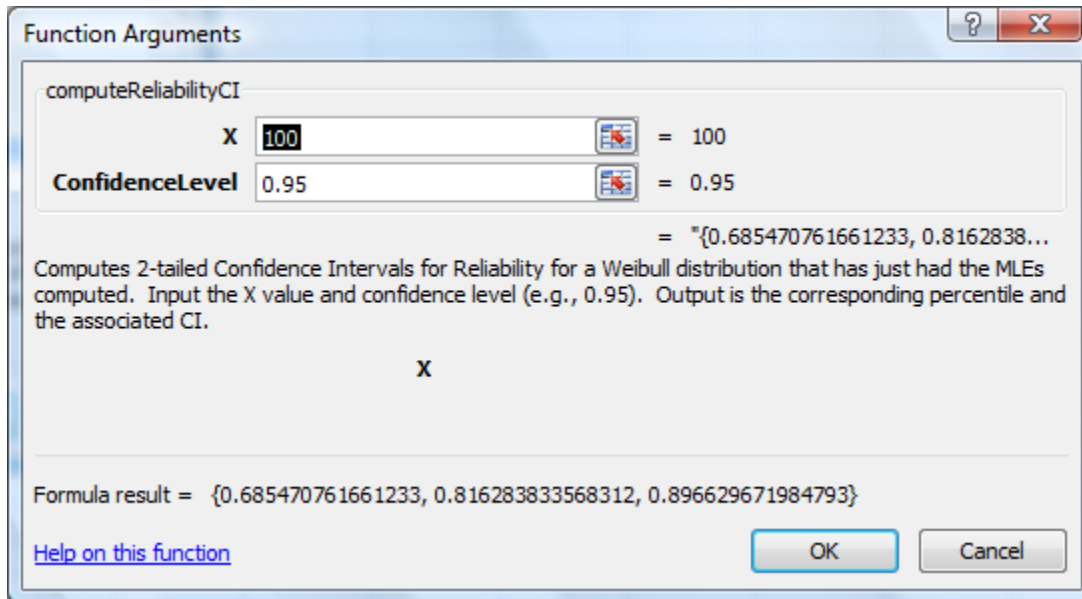
=Compute3ParmWeibull('4 data sets'!C2:C37)
Shape = 0.473800956334852, Scale = 1497.37633567834, Threshold = 48.2741719996721

Once this step is complete the user may compute reliability confidence intervals by either directly typing in the function name or more easily using the Insert Function option in Excel as shown below in Excel 2007. Two images are shown with the first showing the initial Insert Function dialogue box. The VBA functions are found in the User Defined category. Comments have been added to help describe what the computeReliabilityCI VBA function does.
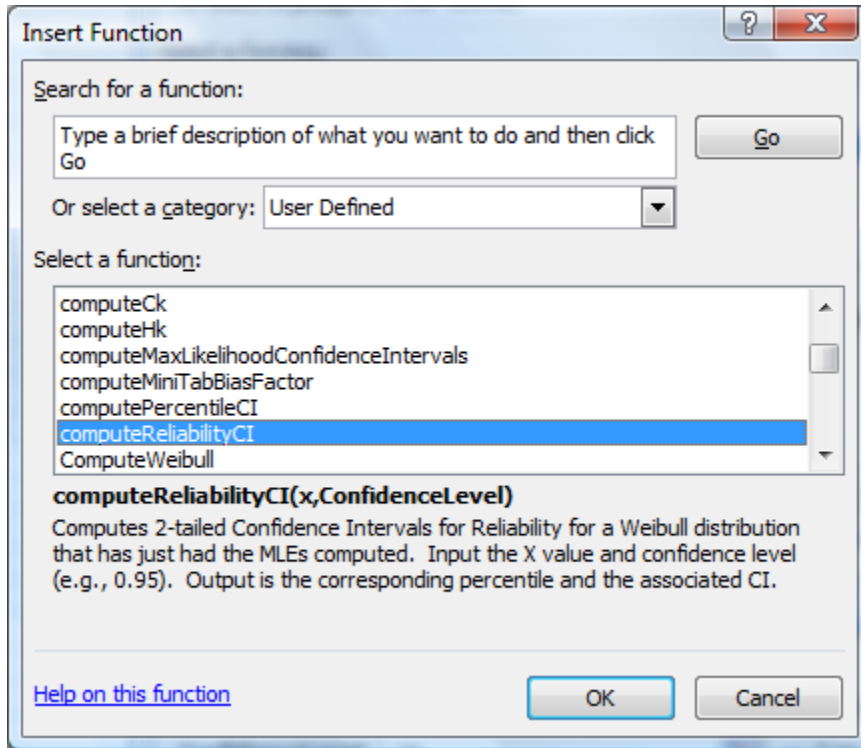
The screen dump image below shows the function and its inputs. Variable names that hopefully will make sense to the user were used. The results of the function that will appear in an Excel cell are also shown. In this particular case of oil spill data in the Gulf of Mexico, X = 100 barrels and a confidence level of .95 were input. The output triplet has the estimated reliability of approximately .816 implying that the estimated likelihood of a spill being greater than 100 barrels is 81.6%. The first and third numbers in this triplet are the associated 95% confidence interval of (68.5%, 89.7%).

Function Arguments dialog showing:
- computeReliabilityCI
- **X** 100 = 100
- **ConfidenceLevel** 0.95 = 0.95
- = "{0.685470761661233, 0.8162838...
- Computes 2-tailed Confidence Intervals for Reliability for a Weibull distribution that has just had the MLEs computed. Input the X value and confidence level (e.g., 0.95). Output is the corresponding percentile and the associated CI.
- X
- Formula result = {0.685470761661233, 0.816283833568312, 0.896629671984793}
- Help on this function   OK   Cancel

Percentile Confidence Intervals

Confidence intervals for a percentile p can be found using the expressions below [Dodson and Nolan (2002); Dodson (2006)]. These are computed for the two parameter Weibull with $x^* = x - \hat{\gamma}$ and then appropriately adjusted for the original x units. The hat on x* indicates that these are estimates of the percentiles for the given Weibull distribution.

$$e^{y_L} \leq \hat{x}^* \leq e^{y_U}$$

$$\hat{x}^* = \hat{\alpha}\left[-\ln(1-p)\right]^{1/\hat{\beta}}$$

$$y_L = \ln\hat{\alpha} + \frac{\ln\left[-\ln(1-p)\right]}{\hat{\beta}} - K_\alpha\sqrt{\text{var}(\hat{y})}$$

$$y_U = \ln\hat{\alpha} + \frac{\ln\left[-\ln(1-p)\right]}{\hat{\beta}} + K_\alpha\sqrt{\text{var}(\hat{y})}$$

$$\text{var}(\hat{y}) = \frac{var(\hat{\alpha})}{\hat{\alpha}^2} + \frac{\left\{\ln\left[-\ln(1-p)\right]\right\}^2 var(\hat{\beta})}{\hat{\beta}^4} - \frac{2\left\{\ln\left[-\ln(1-p)\right]\right\}\text{cov}(\hat{\alpha},\hat{\beta})}{\hat{\beta}^2\hat{\alpha}}$$

As with the reliability confidence intervals shown above, the user must first use the VBA software to generate the Weibull MLEs. Then the function computePercentileCI provides the desired percentile estimate with the associated confidence interval. This is shown in two ways with the initial display below showing how this would be input directly into an Excel cell by typing in the function name. The associated output is on the subsequent line. In this case the input is 0.99 for the desired percentile with a 95% confidence level. The output triplet gives the 95% confidence interval with the middle number being the estimated percentile. Thus the estimated 99th percentile is 37,647 barrels of oil with a 95% confidence interval of 14,953 to 94,892 barrels of oil. As before the following two screen dumps show the use of this function using the Insert Function Excel option.

=computePercentileCI(0.99,0.95)
{14953.3733135367, 37646.8408859048, 94891.804263018}
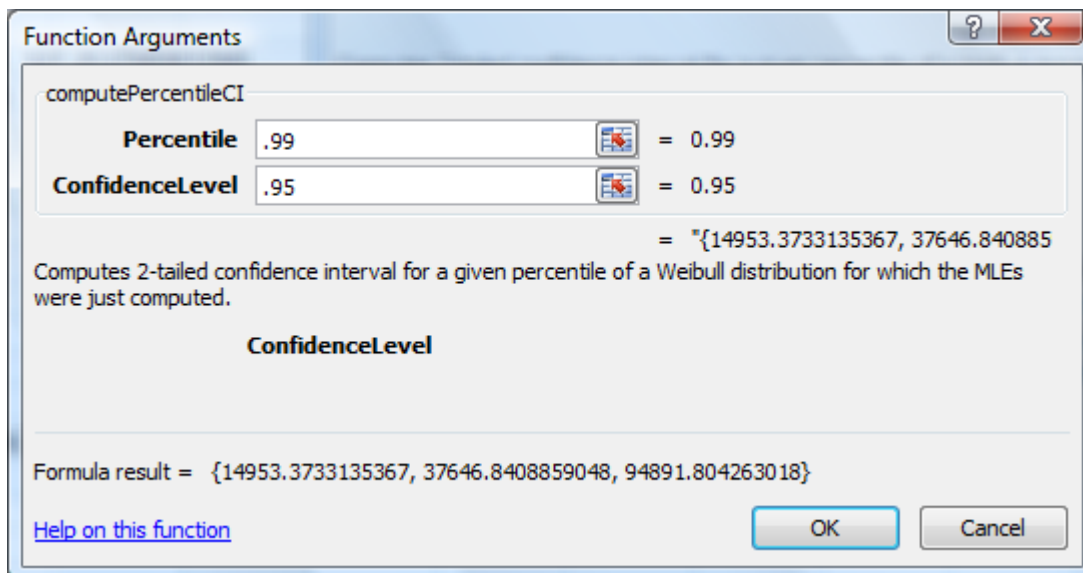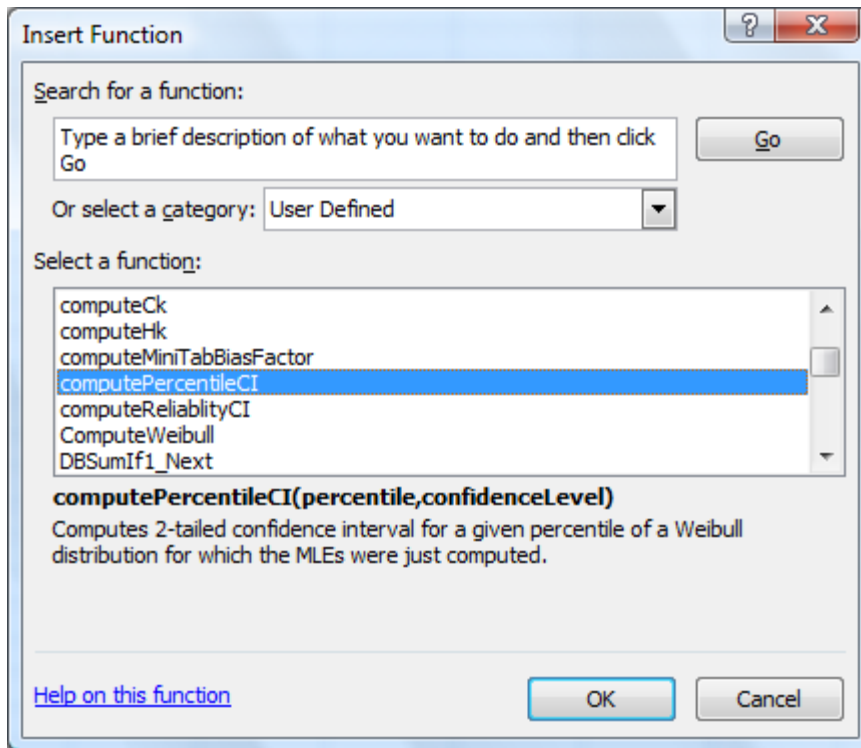
Summary
This paper extends the results developed earlier in Harper, James, Eschenbach, and Slauson (2008). The ability to estimate both percentiles and reliability is an important component in

assessing the likelihood of varying volumes of oil spills in the Gulf of Mexico. Additionally the quantification of the associated uncertainty with MLE based confidence intervals provides a more realistic and complete view of the risks of future spills.

References

Dodson, Bryan and Dennis Nolan (2002), *Reliability Engineering Handbook*, CRC Taylor & Francis, ISBN 0-8247-0364-2. Note: good book, but a fair number of typos.

Dodson, Bryan (2006), *The Weibull Analysis Handbook, 2ⁿᵈ Ed.,* ASQ Quality Press, Milwaukee, Wisconsin, ISBN 978-0-87389-667-2.

Eschenbach, Ted G., and Harper, William V. (2006), Alternative Oil Spill Occurrence Estimators for the Beaufort/Chukchi Sea OCS (Statistical Approach), OCS Study MMS 2006-059, http://www.mms.gov/alaska /reports/2006rpts/2006_059.pdf.

Harper, William V., Thomas R. James, Ted G. Eschenbach, Leigh Slauson (2008), "Maximum Likelihood Estimation Methodology Comparison for the Three-Parameter Weibull Distribution with Applications to Offshore Oil Spills in the Gulf of Mexico," JSM Proceedings, Risk Analysis Section, pp. 1231-1238, American Statistical Association, Denver, CO**.**

Harper, William V., Thomas R. James, Ted G. Eschenbach, (submitted May 2009), "Maximum Likelihood Estimation Statistical Software Comparison for the Three-Parameter Weibull Distribution", submitted to *The American Statistician.*

Lawless, Jerald F., *Statistical Models and Methods for Lifetime Data, 2ⁿᵈ Ed*., 2003, John Wiley & Sons, New York.