METHODS FOR ESTIMATING THE PARAMETERS OF THE WEIBULL DISTRIBUTION

Sidita Duli, Božo Krstajić

Faculty of Natural Sciences, University of Shkoder, Albania
Department of Electrical Engineering, University of Montenegro

siditaduli@yahoo.com
bozok@ac.me

ABSTRACT

The Weibull distribution is used in meteorology as representing the wind speed frequency distribution. There are different methods for estimating the two Weibull parameters (scale parameter c and shape parameter k) from wind statistics. The algorythm of these methods can be implemented in a computer by using a programming language so it can estimate the Weibull parameters for different givven samples. This article presents the Weibull main methods for estimating the two Weibull parameters. It also shows the performance of a C application that makes the calculations.

INTRODUCTION

Weibull Distribution is highly used in meteorology to model the wind speed. This statistical tool tells us how often winds of different speeds will be seen at a location with a certain average wind speed.This distribution is used also wind energy applications. In this article are shown the reasons why is this distribution used for this purpose and how can the parameters estimated. Based on the mathematical calculations, it is build and tested a serial computer application that estimates these parameters. The variation in wind speed are best described by the Weibull probability distribution function 'h' with two parameters, the shape parameter 'k', and the scale parameter 'c'. The probability of wind speed being v during any time interval is given by the following:

$$h(v)=\left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^{k}} \quad \text{for } 0 < v < \infty \qquad (1)$$

In the probability distribution chart, h is plotter against v over a chosen time period, where :

h=fraction of time wind speed is between v and (v + Δv) / Δv

is the plot of h versus v for three different values of k. The curve on the left with k = 1 has a heavy bias to the left, where most days are windless (v=0). The curve on the right with k = 3 looks more like a normal bell shape distribution, where some days have high wind and equal number of days have low wind. The curve in the middle with k = 2 is a typical wind distribution found at most sites. In this distribution, more days have lower than the mean speed, while few days have high wind. The value of k determines the shape of the curve, hence is called the 'shape parameter'. The Weibull distribution with k = 1 is called the exponential distribution which is generally used in the reliability studies. For k>3, it approaches the normal distribution, often called the Gaussian or the bell-shape distribution. For greater values of c, the curves shift right to the higher wind speeds. That is, the higher the c, the more number of days have high winds. Since this shifts the distribution of hours at a higher speed scale, the c is called the scale parameter. (C.G.Justus, W.R. Hardgraves, A.Mikhail, D.Graber, 1977)

Advantages of the Weibull distribution are that :

1. it is a two-parameter distribution, depending only in c and k,
2. in a wide number of cases the Weibull seems to give a reasonable fit to observed distributions. 3- with Weibull c and k values known at one height, a consistent methodology can be used to adjust these paramters to another desired height.

METHODS

There are several methods which can be used to estimate the Weibull parameters c and k, depending on which wind statistics are available and what level of sophistication in data analysis one wishes to employ. (Rinne, Horst,2009)

Least –squares

Least square method is used to calculate the parameters in a formula when modeling an experiment of a phenomenon and it can give an estimation of the parameters. When using least square method, the sum of the squares of the deviations s which is defined as below, should be minimized

$$S=\sum_{i=1}^{n} w_i^2\left[y_i-g\left(x_i\right)\right]^2 \qquad (2)$$

in the equation, xi is the wind speed, yi is the probability of the wind speed rank, so (xi,yi) mean the data plot, wi is a weight value of the plot and n is a number of the data plot. The estimation technique we shall discuss is known as Linear Least Square Method (LLSM) which ia a computational approach to fitting a mathematical or statistical model to data. Ii is applied in engineering and mathematics problem that is often not thought of as an estimation problem. The linear least square method (LLMS) is a special case for the least square method with a formula which consists of some linear functions and it is easy to use. An in the more special case that the formula is line, the linear least square method is much easier.

Maximum likelihood

The method of maximum likelihood [2] is a commonly used procedure because it has very desirable properties. Let x_1 , x_2 , ..., x_n be a random sample of size *n* drawn from a probability density function $f_x(x;\theta)$ where θ is an unknown parameter. The likelihood function of this random sample is the joint density of the *n* random variables and is a function of the unknown parameter. Thus

$$L=\prod_{i=1}^{n} fxi\left(xi,\theta\right) \qquad (3)$$

is the likelihood function. The maximum likelihood estimator (MLE) of θ , say $\hat{\theta}$, is the value of θ that maximizes *L* or, equivalently, the logarithm of *L* . Often, but not always, the MLE of θ is a solution of

$$\frac{d\log L}{d\theta}=0 \qquad (4)$$

where solutions that are not functions of the sample values x_1, x_2, ..., x_n are not admissible, nor are solutions which are not in the parameter space.

Applying the MLE to estimate the Weibull parameters, namely the shape parameter and the scale parameter. Consider the weibull probability density function given in (1) then likelihood function will be

$$L\left(x_1,x_2,\ldots,x_n,k,c\right)=\prod_{i=1}^{n}\left(\frac{k}{c}\right)\left(\frac{x_i}{c}\right)^{k-1} e^{-\left(\frac{x_i}{c}\right)^k} \qquad (5)$$

on taking the logarithms of (5), differentiating with respect to k and c, in turn and equating to zero, we obtain the estimating equations

$$\frac{\partial \ln L}{\partial k}=\frac{n}{k}+\sum_{i=1}^{n}\ln x_i-\frac{1}{c}\sum_{i=1}^{n}x_i^k\ln x_i=0 \qquad (6)$$

$$\frac{\partial \ln L}{\partial c}=\frac{-n}{c}+\frac{1}{c^2}\sum_{i=1}^{n}x_i^k=0 \qquad (7)$$

On eliminating c between these two above equations and simplifying, we get

$$\frac{\sum_{i=1}^{n}x_i^k\ln x_i}{\sum_{i=1}^{n}x_i^k}-\frac{1}{k}-\frac{1}{n}\sum_{i=1}^{n}\ln x_i=0 \qquad (8)$$

which may be solved to get the estimate of k. This can be accomplished by newton raphson method, which can be written in the form :

$$x_{n-1}=x_n-\frac{f(x_n)}{f'(x_n)} \qquad (9)$$

where

$$f(k)=\frac{\sum_{i=1}^{n}x_i^k\ln x_i}{\sum_{i=1}^{n}x_i^k}-\frac{1}{k}-\frac{1}{n}\sum_{i=1}^{n}\ln x_i \qquad (10)$$

and

$$f'(k)=\sum_{i=1}^{n}x_i^k(\ln x_i)^2-\frac{1}{k^2}\sum_{i=1}^{n}x_i^k(k\ln x_i-1)-\left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right)\left(\sum_{i=1}^{n}x_i^k\ln x_i\right) \qquad (11)$$

once k is determined, ca can be estimated using equation (7) as

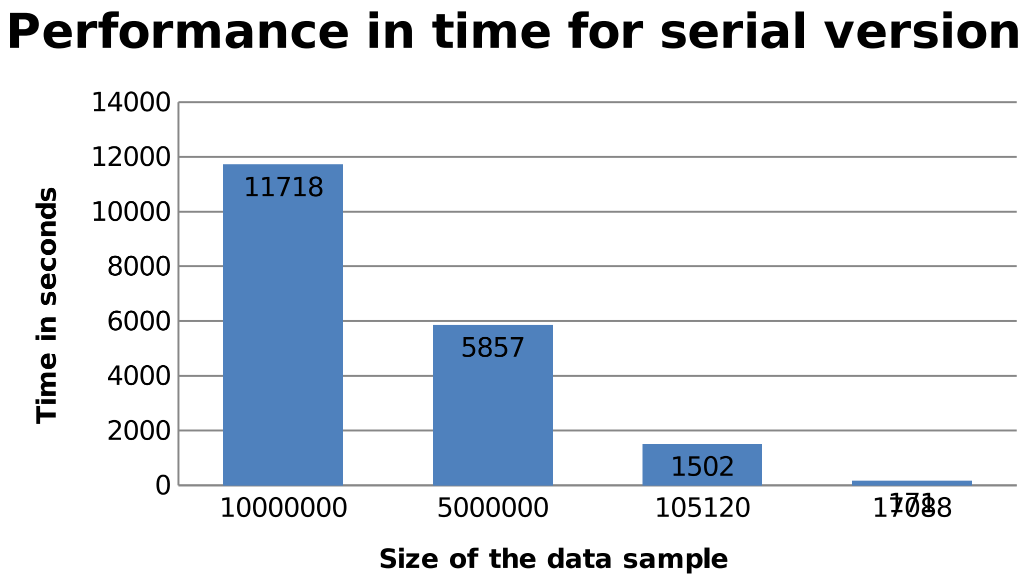
$$c=\frac{\sum_{i=1}^{n}x_i^k}{n} \qquad (12)$$

There is build an C program, a serial version, that based on a sample written in a text file, are calculated the two parameters of Weibull distribution. The program takes a censored set of data, which might be ordered or unordered, taken from a sample of N data. The location parameter a is given, and the algorithm makes an estimation of the scale parameter c and shape parameter k of the equation (1) of the Weibull distribution. The method used for this estimation is the maximum likelihood estimation. To calculate the root of (4) where the L is the likelihood function and is computed via a Newton-Raphson.

RESULTS AND DISCUSSION

Different tests are made to measure the time needed to calculate the parameters for different data samples. There are used two types of data : random data and concrete data. Random data samples are wide in dimension, containing over 1 million data, all random numbers. The purpose of taking these samples is testing how long in time can perform the execution of a wide data sample. Tests are performed with two samples, one containing 10 million data, and another containing 5 million data. Two other concrete data samples are tested. One data sample contains concrete numeric values of wind speed measured in a weather station in Podgorica, Montenegro, This sample contains 105122 data. These are values are collected from the weather station during the whole years 2010-2011, measuring the wind speed every 10 minutes and saving these data in a text file. Another data sample contains values of the wind speed measured in the weather station located in University of Shkodra, Albania. This sample contains 17088 data. This station measures the wind speed every 30 minutes, and saves the results in a text file too. The data used from this station are for the period from 1 June 2010 to 20 May 2011.

Table 1:performance in time for the serial version for four tested samples



Testing the performance of this application results that with increasing the size of the sample, increases the time of calculating the Weibull parameters for the given sample. The time varies from 171 seconds when the sample is 17088 data, up to 11718 seconds for a wide sample containing 10 million data. Also tests shows that there is no difference if the sample contains random data or concrete data.

CONCLUSIONS

This paper introduces the methods used for the Weibull distribution estimating parameters. Based on one of them, the maximum likelihood method, is built an application in C that calculates the parameters for a given sample. Different tests are made to calculate the performance of this application. Results show that the time needed to get the result depends on the size of the sample. When the sample is at a very large size, it takes too long to have the result. In this case, the performance of the serial version is not at all optimal. For small samples, the application can have a faster result, which can vary to hundreds of seconds. In order to get faster the result, may be the implementation of the algorithm in a parallel architecture.

REFERENCES

Methods for estimating wind speed frequency distributions, C.G.Justus, W.R. Hardgraves, A.Mikhail, D.Graber, Journal of applied meteorology, 1977

The Weibull distribution, a handbook, Rinne, Horst. Chapman & Hall/CRC,2009