International Journal of Applied Mathematics

Volume 26 No. 6 2013, 671-683

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v26i6.3

BAYESIAN ANALYSIS OF WEIBULL AND LOGNORMAL SURVIVAL MODELS WITH CENSORING MECHANISM

Yasmin Khan[1] §, A.A Khan[2]

Department of Statistics and O.R.
Aligarh Muslim University
Aligarh, 202002, Uttar Pratesh, INDIA

Abstract: The survival analysis is a powerful statistical methodology widely used in the medical research. In this paper, intercept as well as regression analyses are carried out for real survival data problems with censoring mechanisms only. The Bayesian approach is implemented with R and appropriate illustrations are also made.

AMS Subject Classification: 62xx, 62F15, 62F40
Key Words: Bayesian inference, censoring, package LaplacesDemon, function LaplaceApproximation, sampling importance, resampling, model comparison

1. Introduction

The analysis of survival data is a major focus of the statistics business. The main subject of this paper is the analysis of survival data using two parametric models, namely Weibull and lognormal. The Weibull and lognormal models both are very flexible and have been found to provide a good description of many types of time-to-event data. These are distributions which occupy a central role because of their demonstrated usefulness in a wide range of situations. There are many potential life time models but these two models are used quite effectively to analyze skewed data sets and to give best data fit. The Weibull distribution has two parameters: shape and scale. Its density, survival and hazard functions, are respectively:

Received: November 10, 2013

§Correspondence author

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}\exp[-(\lambda t)^{\beta}],$$
$$S(t) = \exp[-(\lambda t)^{\beta}],$$
$$h(t) = \lambda\beta(\lambda t)^{\beta-1},$$

where $\lambda > 0$ and $\beta > 0$ are the rate and shape parameters, respectively. The corresponding probability density, survival and hazard function of the lognormal model are:

$$f(t) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma t}\exp\left[-\frac{1}{2}\left(\frac{logt-\mu}{\sigma}\right)^2\right], \qquad t > 0,$$
$$S(t) = 1 - \Phi\left(\frac{logt-\mu}{\sigma}\right),$$
$$h(t) = \frac{f(t)}{S(t)}.$$

One important feature of the survival data is the presence of censoring, which create special problems in the analysis of the survival data. The lifetime data are censored when the exact failure time for a specific trial is unknown. When analyzing the censored data, the Bayesian methods have an important advantage over the classical methods. From a classical perspective, confidence interval and other inferential statements must be made with respect to repeated sampling of the data. From Bayesian perspective, only the observed censoring pattern is relevant. There are several categories of censoring, but in this paper we will discuss only right censoring mechanism.

The likelihood function for right censored data is

$$L = \prod_{i=1}^{n} Pr(t_i, \delta_i) = \prod_{i=1}^{n}[f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i}, \tag{1}$$

where δ_i is an indicator variable which takes value 1 if the observation is uncensored, and 0 otherwise. Section 2 begins with a brief discussion of the Bayesian analysis of intercept model based on the Weibull and lognormal distribution. Next Section 3 contains the analysis of regression model with censoring. The goal of regression is to understand the behaviour of response variable given covariates. There has been growth in the development and application of the Bayesian inference. The Bayesian inference enable us to fit very complex model that can not be fit by alternative frequentist methods. To fit the Bayesian models, one needs a statistical computing environment. An environment that meets these requirements is the R (R Development Core Team [6]) software.

In this paper, an attempt has been made to illustrate the Bayesian modeling by using the `R` language. The package used in the paper to get the posterior summary is `LaplacesDemon` Hall [3] which reveals the conceptual simplicity of the Bayesian approach for survival data analysis. This package have several optimization and simulation algorithms. The default optimization algorithm is LBFGS (Broyden-Fletcher-Goldfarb-Shanno). Also a very popular algorithm, i.e. the Nelder-Mead [5], is a derivative-free, direct search method which is efficient in small-dimensional problems. The `LaplacesDemon` offers numerous MCMC algorithms for simulation in Bayesian inference, and are, Random-Walk Metropolis (RWM), Metropolis-within-Gibbs (MWG), Delayed Rejection Metropolis (DRM), etc. This package provides a complete Bayesian environment to the user. Section 4 provides a model compatibility study based on Bayesian predictive information criteria (BPIC) so that the choice of these two models can be justified for the data under consideration. Finally, in the last section a brief discussion and conclusion is given.

2. Fitting of Intercept Model

Here we want to make inferences about the response and intercept. In this section we will consider two models, the Weibull and lognormal models and the data discussed here are the remession times, in weeks, for a group of 30 patients with leukemia who received similar treatments, see Lawless [4].

```
time(weeks): 1,1,2,4,4,6,6,6,7,8,9,9,10,12,13,14,18,19,24,26,
29,31,42,45,50,57,60,71,85,91
```

The intercept model is in the form of

$$y = \mu + e,$$

$$y \sim Weibull(a, b),$$

where, a and b are the shape and scale parameters and $\mu = X\beta$,

$$log(b) = X\beta$$

$$\beta \sim N(0, 1000)$$

$$a \sim halfcauchy(25).$$

Here we focus on right censored survival data since this type of data are most frequently encountered in applications.

2.1. Fitting of Weibull Model

The Weibull distribution has two parameters, shape and scale. It can be denoted as

$$y \sim W(a, b)$$

thus, the likelihood is

$$p(y|a,b) = \prod_{i=1}^{n} \frac{a}{b}\left(\frac{y_i}{b}\right)^{a-1} \exp\left[-\left(\frac{y_i}{b}\right)^{a}\right]$$

which implies log-likelihood as

$$logp(y|a,b) = \sum_{i=1}^{n} log\left(\frac{a}{b}\left(\frac{y_i}{b}\right)^{a-1} \exp\left[-\left(\frac{y_i}{b}\right)^{a}\right]\right).$$

Using Equation (1), this loglikelihood including censoring mechanism is expressed in R as:

```
loglikelihood<-sum[censor*(dweibull(x=y,shape=a,scale=b,log=T)+
(1-Censor)pweibull(x=y,shape=a,scale=b,lower.tail=F,log.p=T))]
```

Now, prior for regression coefficient β and shape parameter is normal with mean 0 and standard deviation 1000 and half-Cauchy with scale parameter 25, respectively. Consequently, logprior for β and shape are $-\left(\frac{y^2}{2*1000}\right)$ and $log\left(\frac{2*25}{\pi(scale^2)*25^2}\right)$, respectively, which can be expressed in R as

```
beta.prior<-dnorm(beta,0,1000,log=T)
shape.prior<-dhalcauchy(shape,25,log=T)
```

Thus,

```
logposterior= loglikelihood+beta.prior+shape.prior
```

Bayesian fitting of Weibull model for this data can be done in R by using the function `LaplaceApproximation`, and then with `LaplacesDemon`. Its fitting includes codes for creation of data and definition of model as discussed above. The R codes to fit Weibull model are described below.

```
library(LaplacesDemon)
options(digits=2)
y<-c(1,1,2,4,4,6,6,6,7,8,9,9,10,12,13,14,18,19,24,26,29,31,
42,45,50,57,60,71,85,91)
censor<-c(rep(1,21),0,1,0,0,1,1,0,0,1)
N<-30
J<-1
X<-matrix(1,nrow=length(y))
mon.names<-c("LP","shape")
parm.names<-as.parm.names(list(beta=rep(0,J),log.shape=0))
MyData<-list(J=J,X=X,mon.names=mon.names,parm.names=parm.names,y=y)
Initial.Values <- c(rep(0,J), log(1))
Model<-function(parm,Data)
{
beta<-parm[1:Data$J]
shape<-\exp(parm[Data$J+1])
beta.prior<-sum(dnorm(beta,0,1000,log=T))
shape.prior<-dhalfcauchy(shape,25,log=T)
mu<-tcrossprod(beta,Data$X)
scale<-\exp(mu)
LL<-sum(censor*dweibull(Data$y,shape,scale,log=T)+
(1-censor)*pweibull(Data$y,shape,scale,log.p=T,lower.tail=F))
LP<-LL+beta.prior+shape.prior
Modelout<-list(LP=LP,Dev=-2*LL,Monitor=c(LP,shape),yhat=mu,parm=parm)
return(Modelout)
}
M1<-LaplaceApproximation(Model,Initial.Values,Data=MyData,Sample=10000,
Iterations=10000)
Initial.Values<-as.initial.values(M1)
M10<-LaplacesDemon(Model,Data=MyData,Initial.Values,Status=FALSE)
```

The output obtained from `M1`, `M10` objects are summarized in Table 1, Table 2 and Table 3, respectively.

	Mode	SD	LB	UB
beta	3.37	0.24	2.89	3.86
log.shape	-0.18	0.16	-0.50	0.14

Table 1: Approximated posterior summary using `LaplaceApproximation` function with posterior mode, posterior sd and their quantiles.

	Mean	SD	LB	Median	UB
beta	3.38	0.25	2.89	3.38	3.89
log.shape	-0.23	0.17	-0.57	-0.22	0.08

Table 2: Simulated posterior summary using sampling importance resampling with posterior mean, posterior sd and their quantiles.

	Mean	SD	LB	Median	UB
beta	3.41	0.26	2.88	3.39	3.91
log.shape	-0.23	0.17	-0.56	-0.22	0.08

Table 3: Simulated posterior summary using `LaplacesDemon` function with posterior mean, posterior sd and their quantiles

2.2. Fitting of Lognormal Model

The lognormal distribution is another parametric function widely used in survival analysis. In survival analysis, if the event time `Y` is lognormally distributed, then `log Y` is normally distributed, denoted by

$$\texttt{logY} \sim N(\mu, \sigma^2).$$

Model,

$$y = \mu + e,$$

$$y \sim lognormal(\mu, \sigma),$$

where μ and σ are the location and scale parameters with $\mu = X\beta$.
Prior,

$$\beta \sim N(0, 1000),$$

$$\sigma \sim halfcauchy(25),$$

$$p(y) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma y} \exp\left[-\frac{1}{2}\left(\frac{logy - \mu}{\sigma}\right)^2\right], \qquad t > 0,$$

thus the likelihood is

$$p(y|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma y} \exp\left[-\frac{1}{2}\left(\frac{logy - \mu}{\sigma}\right)^2\right]$$

which implies log-likelihood as,

$$logp(y|\mu,\sigma) = \sum_{i=1}^{n} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma y} \exp\left[-\frac{1}{2}\left(\frac{logy-\mu}{\sigma}\right)^2\right],$$

this likelihood with censoring is expressed n R as

```
loglikelihood<-sum[censor*dnorm(x=y,location=mu,scale=sigma,
log=T)+(1-censor)*pnorm(x=y,location=mu,scale=sigma,lower.tail=
F,log.p=T)
```

Now, prior for regression coefficient β and sigma parameter is normal with mean 0 and standard deviation 1000 and half-Cauchy with scale parameter 25, respectively. Consequently, logprior for β and sigma are $-\left(\frac{y^2}{2*1000}\right)$ and $log\left(\frac{2*25}{\pi(scale^2)*25^2}\right)$, respectively, which can be expressed in R as

```
beta.prior<-dnorm(beta,0,1000,log=T)
sigma.prior<-dhalcauchy(sigma,25,log=T)
```

Thus,

```
logposterior<-loglikelihood+beta.prior+sigma.prior
```

All R codes are not reported and are skipped without loss of continuity. Approximate and simulated posterior summries are reported Table 4, Table 5 and Table 6, respectively.

	Mode	SD	LB	UB
beta	2.77	0.26	2.24	3.29
log.sigma	0.34	0.15	0.04	0.63

Table 4: Approximated posterior summary using `LaplaceApproximation` function with posterior mode, posterior sd and their quantiles.

3. Bayesian Regression Analysis with Censoring

The next step, and perhaps the most applicable for practical work, is regression analysis for censored data from a Bayesian perspective. The goal of regression is

	Mean	SD	MCSE	ESS	LB	Median	UB
beta	2.76	0.26	0.01	1000.00	2.27	2.78	3.27
log.sigma	0.36	0.13	0.00	1000.00	0.10	0.36	0.65

Table 5: Simulated posterior summary using sampling importance resampling with posterior mean, posterior sd and their quantiles.

	Mean	SD	MCSE	ESS	LB	Median	UB
beta	2.82	0.29	0.04	100.00	2.29	2.77	3.44
log.sigma	0.36	0.14	0.02	100.00	0.14	0.37	0.65

Table 6: Simulated posterior summary using `LaplacesDemon` with posterior mean, posterior sd and their quantiles.

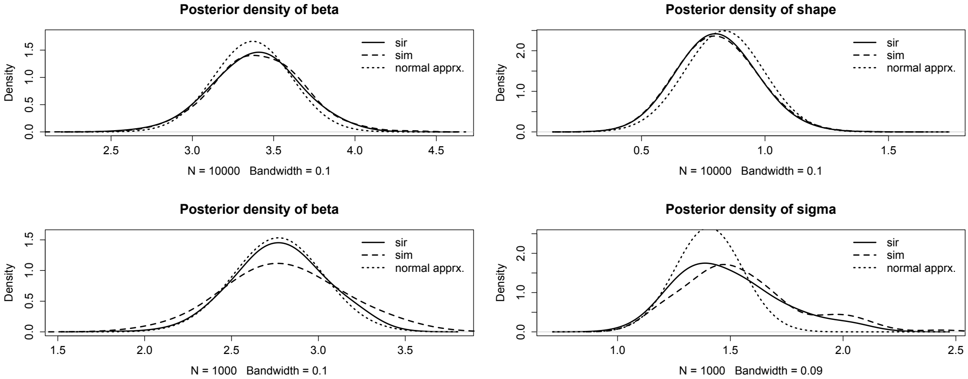

Figure 1: Graphical output of posterior density plots. Upper side of the display is the posterior density plots of parameter of Weibull model and lower side is posterior density plots of parameter of log-normal model. Weibull model shows the overlap of two density plots that is density plots generated by SIR and by simulation. However, there is a little difference in lognormal case.

to summarize observed data as simply, usefully, and elegantly as possible. Here we discuss and illustrate parametric regression models (the Weibull and lognormal regression models) for the `chemotherapy` data available in the `LearnBayes` package. The data is in the form of data frame with 26 observations on 5 variables, namely `patient`, `time`, `status`, `treat`, `age`. Edmunson et al. [2]

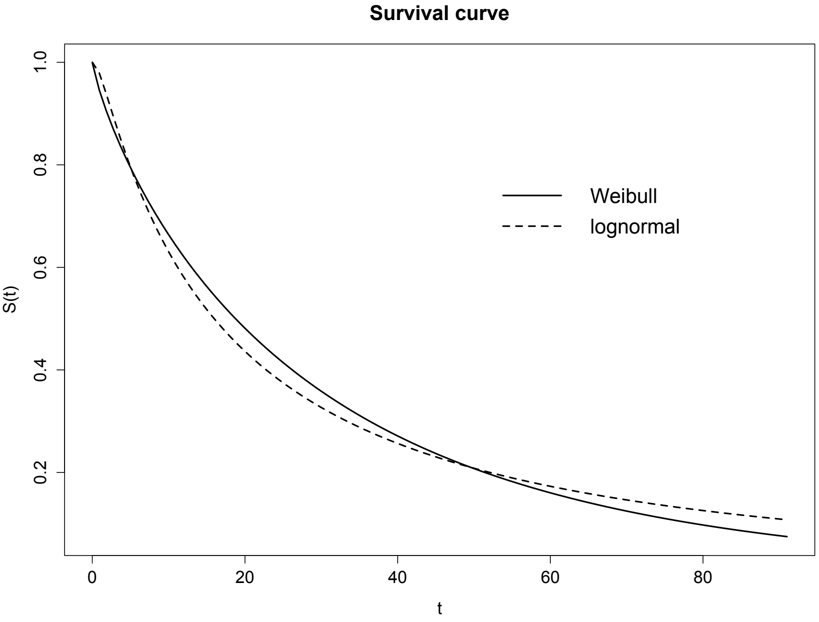

Figure 2: Estimated survival curves of Weibull and lognormal model for leukemia data. Survival time for Weibull is a little more than that of lognormal model.

studied the effect of different chemotherapy treatments following surgical treatment of ovarian cancer. Here, fitting is done between `time` as response variable and `treat` and `age` as regressor. Techniques for assessing the fit of these two parametric models is done by `LaplacesDemon` in `R`.

3.1. Fitting of Weibull Regression Model

In this section, we consider the Weibull regression model with two predictors, `treatment` and `age`. Thus, the regression model is

$$Time = \beta_0 + \beta_1 treat_i + \beta_2 age_i + e_i.$$

The approximated posterior mode and their standard deviation (in bracket) of Weibull regression model with `LaplaceApproximation` is:

$$\beta_0 = 7.393(1.687), \quad \beta_1 = 1.262(0.735),$$

$$\beta_2 = -0.043(0.021), \quad shape = 1.512(1.53).$$

Posterior Summary of Weibull regression model with posterior mean and posterior sd in bracket by sampling importance resampling is:

$$\beta_0 = 10.44(1.015), \quad \beta_1 = 0.71(0.291),$$

$$\beta_2 = -0.08(0.015), \quad shape = 1.65(0.291).$$

Posterior Summary of Weibull regression model with posterior mean and posterior sd in bracket by using `LaplacesDemon`:

$$\beta_0 = 8.22(0.4029), \quad \beta_1 = 0.50(0.1563),$$

$$\beta_2 = -0.04(0.0039), \quad shape = 1.74(0.3522).$$

3.2. Fitting of Lognormal Regression Model

Posterior summary of lognormal regression model with `LaplaceApproximation`, sampling importance resampling and `LaplacesDemon` is:

The approximated posterior mode and their standard deviation (in bracket) of lognormal regression model with `LaplaceApproximation` is:

$$\beta_0 = 7.393(1.687), \quad \beta_1 = 1.262(0.735),$$

$$\beta_2 = -0.043(0.021), \quad sigma = 1.512(1.53).$$

Posterior Summary of lognormal regression model with posterior mean and posterior sd in bracket by sampling importance resampling is:

$$\beta_0 = 10.44(1.015), \quad \beta_1 = 0.71(0.291),$$

$$\beta_2 = -0.08(0.015), \quad sigma = 1.65(0.291).$$

Posterior Summary of lognormal regression model with posterior mean and posterior sd in bracket by using `LaplacesDemon`:

$$\beta_0 = 8.22(0.4029), \quad \beta_1 = 0.50(0.1563),$$

$$\beta_2 = -0.04(0.0039), \quad sigma = 1.74(0.3522).$$

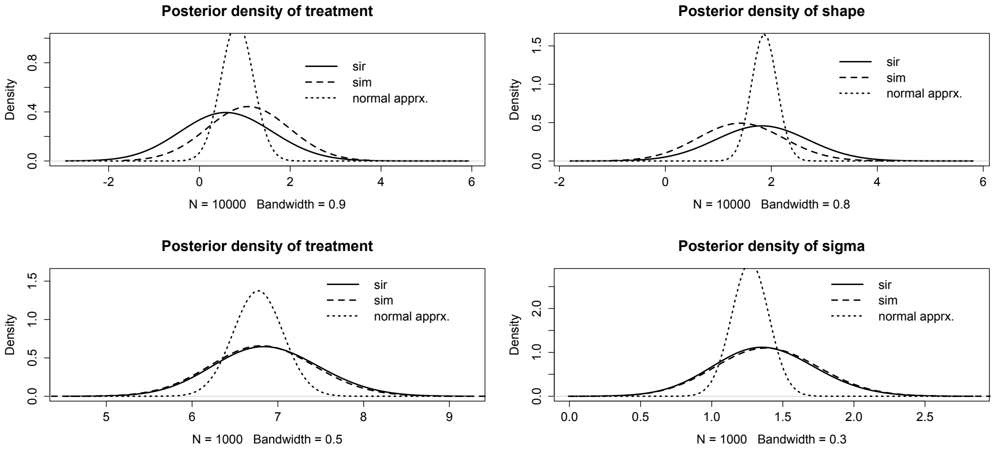

Figure 3: Posterior density plots of parameters of Weibull and lognormal regression model

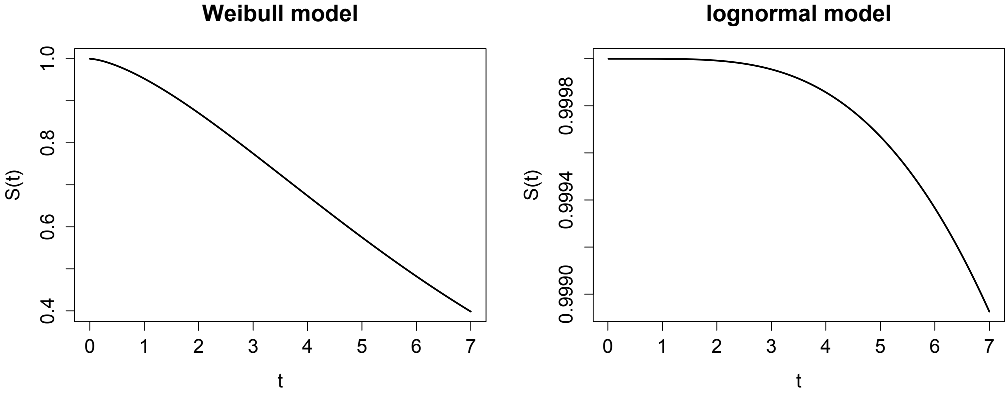

Figure 4: Survival curves for Weibull and lognormal regression model

3.3. Model Comparison

The Bayesian inference provides a model fit statistic that is to be used as a tool to refine the current model or select the better-fitting model. In this section, we will compare both models whose fitting is discussed in the above section

Model	BPIC
Weibull (intercept model)	226
lognormal (intercept model)	222
Weibull (regression model)	184
lognormal (regression model)	186

Table 7: Table for comparison of Weibull and lognormal model for both intercept and regression analysis. The table clearly shows that for the analysis of intercept model lognormal model has the small value of BPIC, hence lognormal model is the best fit for the data. However, on the other hand, Weibull model in regression analysis has low BPIC value meaning Weibull give appropriate fit for `chemotherapy` data.

and their comparison is made from Bayesian perspective. Bayesian predictive information criteria (BPIC) was introduced as a criterion of model comparison whose goal is to pick a best model for respective survival data. BPIC is a variation of DIC where the effective number of parameters is 2pD.

4. Conclusion and Discussion

In this article we are concerned with only right censored survival data. The Bayesian analysis shows that for the analysis of survival data which are generally not symmetric and are positively skewed, will performed well when Weibull or lognormal distribution is used for modeling. It has been observed that there is a very close relation between these two models. It may be noted that model comparison is made in Bayesian setup. In fitting of intercept model, lognormal model is found to be the best model as it has low value of BPIC than Weibull. Contrary to this, in regression modeling Weibull is the model which is appropriate for the analysis of `chemotherapy` data. It could be seen that, there is very small difference in BPIC values for both models in both intercept and regression analysis. Thus, it is justified that the use of Weibull and lognormal model is appropriate for these two data sets. Hence, we can say that these two models could be a good choice for the analysis of survival data. The use of Laplace approximation method made a great contribution in Bayesian framework. However, being an asymptotic approach one of the limitations of this approach is that this method is recommended for the data whose sample size is at least 5 times of the number of parameters available in a particular statistical model and it has been found in the present study that this approximation works well. Bernardo and Smith [1] note that Laplace approximation is an attractive numerical approximation algorithm, and will continue to develop.

References

[1] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, John Wiley & Sons, West Sussex (2000).

[2] J. Edmunson, T. Felming, D. Decker, G. Malkasian, E. Jorgensen, J. Jefferies, M. Webb and L. Kvols, Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal residual disease, *Cancer Treatment Reports*, **63** (1979), 241-247.

[3] B. Hall, `LaplacesDemon:` *Software for Bayesian Inference* `R` package, Version 11.12.05, URL http://cran.r-project.org/web/packages/LaplacesDemon/index.html, *Computer Journal*, **7** (2011), 308-313.

[4] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, 2nd Ed., Wiley, New York (2003).

[5] J.A. Nelder and R. Mead, A simplex method for function minimization, *The Computer Journal,* **7**, No 4 (1965), 308-313.

[6] `R` Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2011), ISBN 3-900051-07-0, http://www.R-project.org.