R Package: TBSSurvival

Adriano Polpo Cassio de Campos Debajyoti Sinha
Stuart Lipsitz Jianchang Lin

July 1, 2012

Abstract

The purpose of this text is to provide a simple manual for the `TBSSurvival` package for `R` language. In short, we give some examples on how to use the main functions of the package.

Keywords: TBSSurvival, R package, Generalized Power Transformation, reliability, error distribution, median estimation.

1 Introduction

We assume the reader to be familiar with the survival estimation / reliability problem. For more details we suggest the technical paper: *Transform Both Sides Model: A Parametric Approach*, Polpo et al. 2012. In this document we use as example a reliability analysis situation, but the methods will work as fine with survival analysis.

The problem in view can be defined by `time`, an array of times of failure (survival) or components (patients) and `delta` the event indicator, that is, `delta` equals to one if an event happened at that time, or zero in case of right-censoring. Furthermore, covariates can be present in the data set, which can be used in a regression analysis, in the very much same way as done by well-known functions such as `survreg`. First, we exemplify the use of the methods without covariates.

Example 1 In this example we simulated data set with right-censored failure time. The problem is about an experiment with 30 machines. We observe the failure time of these machines, but our experiment has a maximum time of 6 week. After that, all machines that have not failed are considered to be censored with respect to their failure time. For simplicity, we define the failure time by a random variable with Gamma distribution.

```
> time <- pmin(rgamma(30,10,2),rep(6,30))
> delta <- rep(1,30)
> for (i in 1:30) {
+   if (time[i] == 6) delta[i] <- 0
+ }
> data <- cbind(time,delta)
> data
```

```
          time delta
 [1,] 3.071913     1
 [2,] 5.246828     1
 [3,] 5.319838     1
 [4,] 3.690964     1
 [5,] 5.561896     1
 [6,] 3.905557     1
 [7,] 6.000000     0
 [8,] 3.839161     1
 [9,] 4.043061     1
[10,] 2.435072     1
[11,] 4.849857     1
[12,] 6.000000     0
[13,] 4.581559     1
[14,] 4.778270     1
[15,] 5.899024     1
[16,] 3.805206     1
[17,] 4.958891     1
[18,] 4.023899     1
[19,] 4.757719     1
[20,] 4.692185     1
[21,] 2.780329     1
[22,] 4.767577     1
[23,] 4.765207     1
[24,] 5.852793     1
[25,] 6.000000     0
[26,] 6.000000     0
[27,] 6.000000     0
[28,] 2.748356     1
[29,] 6.000000     0
[30,] 3.245861     1
```

2 TBS Model

The TBS Model is defined as

$$g_\lambda(\log(T)) = g_\lambda(\theta(\mathbf{X})) + \varepsilon, \tag{1}$$

where the function $g_\lambda(\cdot)$ is the TBS defined by

$$g_\lambda(u) = \frac{\text{sign}(u)|u|^\lambda}{\lambda}, \tag{2}$$

$\text{sign}(u) = 1$ if $u \geq 0$ and $\text{sign}(u) = -1$ if $u < 0$, $\lambda > 0$, $\theta(\mathbf{X})$ is a function of co-variables and the error has some distribution ($\varepsilon \sim F_\varepsilon$) with parameter ξ. The error distributions currently implemented in the TBSSurvival package are: normal (used with `dist = "norm"`), t-student (`dist = "t"`), Cauchy (`dist = "cauchy"`), double-exponential (`dist = "doubexp"`) and logistic (`dist = "logistic"`) distributions.

In case you have not yet done so, the first thing to do before using the functions is to install and load the library.

```
> install.packages("TBSSurvival_VERSION.tar.gz",repos=NULL,type="source")
```

```
> library("TBSSurvival")
```

Example 2 Density, reliability and hazard functiosn for TBS model are available using the commonly used notation: `dtbs, ptbs, qtbs, rtbs, htbs`. The error distribution is passed as argument, as well some other parameters such as λ (for the TBS), ξ (parameter of the error distribution) and covariates' weights $\boldsymbol{\beta}$.

```
> x <- seq(0.1,10,0.1)
> plot(x, (1-ptbs(x,lambda=1,xi=sqrt(2),beta=1,
+                 dist="norm")),
+      type="l",lwd=2,lty=1,col=2,ylim=c(0,1),xlab="t",
+      ylab="R(t)",main="Reliability",cex.lab=1.2)
```

```
> plot(x, (dtbs(x,lambda=1,xi=sqrt(2),beta=1,
+               dist="norm")),
+      type="l",lwd=2,lty=1,col=2,xlab="t",
+      ylab="f(t)",main="Density",cex.lab=1.2)
```

```
> plot(x, (htbs(x,lambda=1,xi=sqrt(2),beta=1,
+               dist="norm")),
+      type="l",lwd=2,lty=1,col=2,xlab="t",
+      ylab="h(t)",main="Hazard",cex.lab=1.2)
```

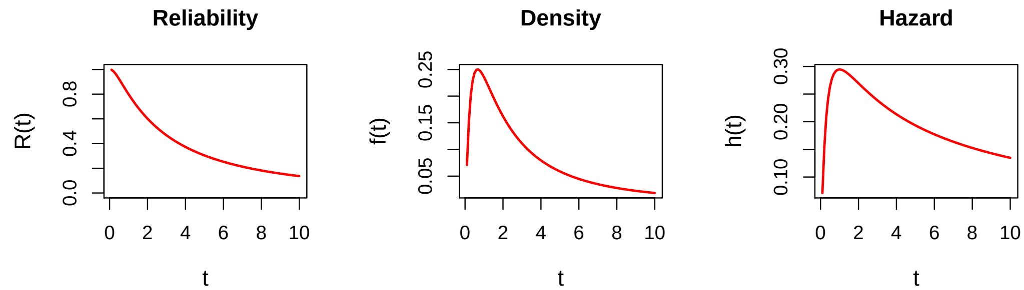

Figure 1: Density, reliability and Hazard functions
($\lambda = 1$; $\epsilon \sim \text{Normal}(0, 2^2)$; $\beta_0 = 1$).

```
> x <- seq(0.1,10,0.1)
> plot(x, (1-ptbs(x, lambda=1, xi=1, beta=1, dist="doubexp")),
+      type="l", lwd=2, lty=1, col=2, ylim=c(0,1), xlab="t",
+      ylab="R(t)", main="Reliability", cex.lab=1.2)

> plot(x, (dtbs(x, lambda=1, xi=1, beta=1, dist="doubexp")),
+      type="l", lwd=2, lty=1, col=2, xlab="t", ylab="f(t)",
+      main="Density", cex.lab=1.2)

> plot(x, (htbs(x, lambda=1, xi=1, beta=1,
+                        dist="doubexp")), type="l", lwd=2,
+      lty=1, col=2, xlab="t", ylab="h(t)",
+      main="Hazard", cex.lab=1.2)
```

2.1 Error distribution

Table 1 presents the available error distributions. Note that the meaning of the parameter changes according to the distribution being used. Moreover, for the normal distribution, the parameter is σ (standard deviation) and not σ^2 (variance).

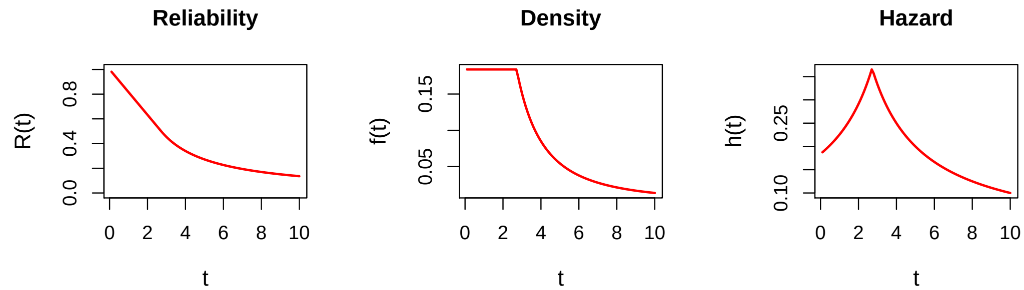

Figure 2: Density, reliability and Hazard functions
($\lambda = 1$; $\epsilon \sim$ doubexp(1); $\beta_0 = 1$).

Table 1: Error distribution.

Distribution	Parameter	Density function ($f_{\varepsilon}(\epsilon\|\xi)$)
Normal	$\xi = \sigma$	$(2\pi\sigma^2)^{-1/2} \exp\left\{-\epsilon^2/(2\sigma^2)\right\}$
DoubExp	$\xi = b$	$(2b)^{-1} \exp\left\{-\|\epsilon\|/b\right\}$
t-Student	$\xi = \eta(d.f.)$	$\frac{\Gamma((\eta+1)/2)}{\Gamma(\eta/2)\sqrt{\pi\eta}} \left(1 + \frac{\epsilon^2}{\eta}\right)^{-(\eta+1)/2}$
Cauchy	$\xi = c$	$\left[\pi c \left(1 + (\epsilon/c)^2\right)\right]^{-1}$
Logistic	$\xi = s$	$\frac{\exp\{-\epsilon/s\}}{s\left[(1+\exp\{-\epsilon/s\})^2\right]}$

The parametric space for ξ is $(0, +\infty)$ in all cases.

3 Estimation

The most important part is the parameter estimation of the TBS model. There are two procedures available: Maximum Likelihood Estimation (MLE) and Bayesian Estimation (BE). For illustration purposes, we will perform the estimation with both methods for the simulated data in Example 1. Remember that we defined the observed failure time as `time` and the censor indicator δ as `delta`.

3.1 MLE

In order to perform the estimation, following the common practice with other survival analysis packages, the user has to build a formula with time and censoring indication (delta), which is then used to call the estimation method itself:

```
> formula <- Surv(data[,1],data[,2]) ~ 1
> tbs.mle <- tbs.survreg.mle(formula,dist="norm", nstart=3,
+                             method="Nelder-Mead")
> tbs.mle

$method
[1] "Nelder-Mead"

$par
[1] 1.1147249 0.1038346 1.5437541

$std.error
[1] 1.10145998 0.08345101 0.06142687

$log.lik
[1] -46.43379

$error.dist
[1] "norm"

$AIC
[1] 98.86759

$AICc
[1] 99.79067

$BIC
[1] 103.0712

$convergence
[1] TRUE

$time
 [1] 3.071913 5.246828 5.319838 3.690964 5.561896 3.905557
 [7] 3.839161 4.043061 2.435072 4.849857 4.581559 4.778270
[13] 5.899024 3.805206 4.958891 4.023899 4.757719 4.692185
[19] 2.780329 4.767577 4.765207 5.852793 2.748356 3.245861
```

```
$error
 [1] -0.435398577  0.120181485  0.134832510 -0.247697659
 [5]  0.182101369 -0.189285047 -0.207037921 -0.153381941
 [9] -0.667833168  0.037040855 -0.022805407  0.021378584
[13]  0.244831911 -0.216225707  0.060488758 -0.158317466
[17]  0.016842275  0.002253928 -0.535929553  0.019020597
[21]  0.018497278  0.236430826 -0.547517698 -0.379429602

$run.time
[1] 0.007216667

$call
tbs.survreg.mle(formula = formula, dist = "norm", method = "Nelder-Mead",
    nstart = 3)

$formula
Surv(data[, 1], data[, 2]) ~ 1
```

Here, we compare the TBS model with the non-parametric Kaplan-Meier estimator. The result is presented in the Figure 3.

```
> # Kaplan-Meier estimation
> km <- survfit(formula)
> plot(km,ylab="R(t)", xlab="t: number of cycles (in thousands)",
+      main="Reliability function (MLE)", conf.int=FALSE, lty=1,
+      lwd=1, xlim=c(min(data[,1]),max(data[,1])))
> t <- seq(min(data[,1]),max(data[,1]),
+          (max(data[,1])-min(data[,1])-0.01)/1000)
> legend(2.6,0.2,
+    c("Kaplan-Meier",
+      expression(textstyle(paste("TBS / ",sep="")) ~ epsilon
+          ~ textstyle(paste("~",sep="")) ~ Norm)),
+    col=c(1,2),lty=c(1,1),cex=1.1,lwd=c(1,2),bg="white")
> lines(t,1-ptbs(t, lambda=tbs.mle$par[1], xi=tbs.mle$par[2],
+                beta=tbs.mle$par[3],
+                dist=tbs.mle$error.dist), type="l",
+       lwd=2, col=2, lty=1)
```

In the following, we present a simple example on how to define covariates to be used. In fact, we use the same framework of formulas as other

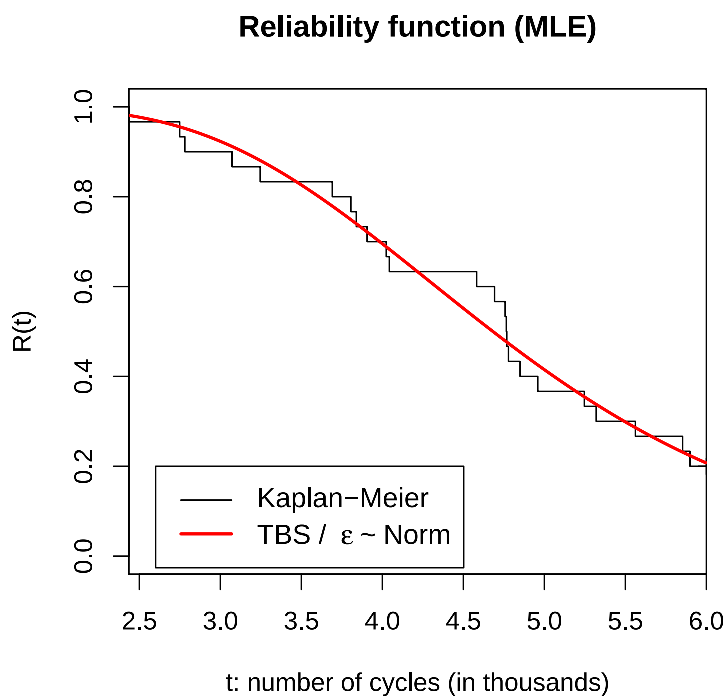

Figure 3: TBS model (MLE).

survival packages. We also point out that different optimization methods are available in order to maximize the likelihood with the TBS model. If the argument `method` is not specific, all of them will be tried, and eventually the best one will be returned. The output, saved in the variable `s` in the following example, contains some score criteria and other useful information.

```
> library(survival)
> data(colon)
> ## Running MLE on colon (from survival package) with a covariate
> colon$age60=as.numeric(colon$age>60) #threshold defined from medical papers
> s=tbs.survreg.mle(Surv(colon$time,colon$status==1) ~ colon$age60,
+        dist="norm",method=c("Nelder-Mead"),nstart=3,verbose=FALSE)
```

3.2 BE

When dealing with the Bayesian estimation, it is necessary to consider an informative choice of priors, as well as the convergence rate of the MCMC method, on which the estimation is based. However, we have experimented with some reasonable non-informative priors, which have shown adequate results. Nevertheless, the user is entitled to choose `prior.mean` and `prior.sd` for the normal distribution that is used as prior for the $boldsymbol\beta$, a `scale` parameter, and initial guesses for the previously mentioned parameters of the model. Other well known arguments that are used my the Metropolis Hastings can also be set. We refer to the technical paper for more details about the priors.

```
> tbs.be <- tbs.survreg.be(formula,dist="norm",
+                          guess.lambda=2,guess.xi=4,
+                          guess.beta=1.5,burn=1000,
+                          jump=10,size=1000,scale=0.06)
```

As a visual example, we construct the plots with the estimated reliability and the 95% credible interval of High Posterior Density of the reliability function. The result is presented in the Figure 4.

```
> aux.survival <- matrix(0,length(tbs.be$time),1000)
> for (j in 1:1000) {
+   aux.survival[,j] <-
+     c(1-ptbs(tbs.be$time,
+              lambda=tbs.be$post[j,1],
+              xi=tbs.be$post[j,2],
+              beta=tbs.be$post[j,3:length(tbs.be$post[1,])],
+              x=tbs.be$x,dist="norm"))
+ }
> survival <- matrix(0,length(tbs.be$time),8)
> for (i in 1:length(tbs.be$time)) {
+   survival[i,] <-
+     c(summary(aux.survival[i,]),
+       HPDinterval(as.mcmc(aux.survival[i,]),0.95))
+ }
> tbs.be$survival <- survival
> plot(km,ylab="R(t)",
+      xlab="t: number of cycles (in thousands)",
+      main="Reliability function (BE)", conf.int=FALSE,
```

```
+       lty=1, lwd=1, xlim=c(min(data[,1]),max(data[,1])))
> legend(2.6,0.3,
+    c("Kaplan-Meier",
+      expression(textstyle(paste("TBS / ",sep="")) ~ epsilon
+          ~ textstyle(paste("~",sep="")) ~ Logistic),
+      "95% HPD Interval"),
+    col=c(1,2,2),lty=c(1,1,2),cex=1.1,lwd=c(1,2,2),
+    bg="white")
> lines(tbs.be$time,tbs.be$survival[,3],type="l",lwd=2,col=2,
+        lty=1)
> lines(tbs.be$time,tbs.be$survival[,7],type="l",lwd=2,col=2,
+        lty=2)
> lines(tbs.be$time,tbs.be$survival[,8],type="l",lwd=2,col=2,
+        lty=2)
```

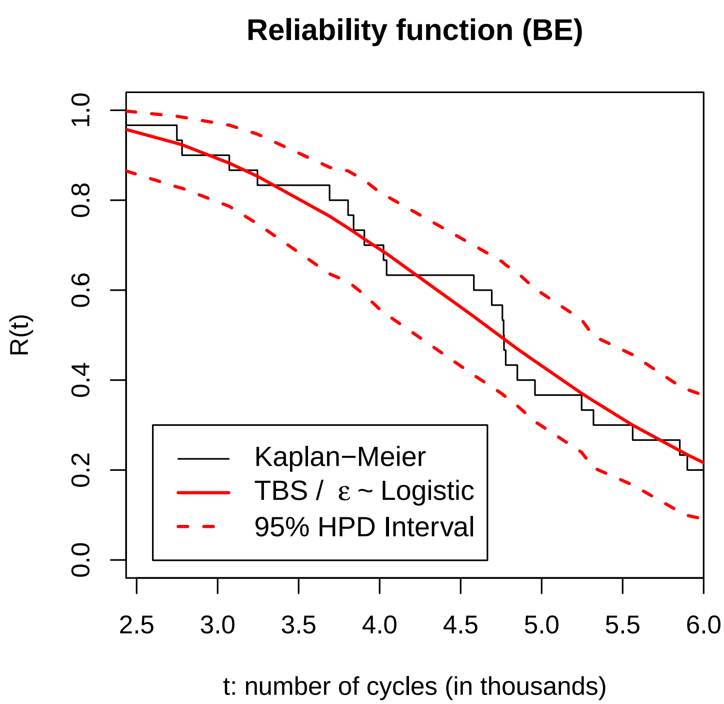

Figure 4: TBS model (BE).

4 Remarks

This "manual" describes the basis of the TBSSurvival package, which is mainly centered around the estimation methods. We invite the user to the functions' help pages (available with the package) and the technical paper mentioned in the beginning of this document.