

## Stat 511 HW#9 Spring 2004

1. An R function that will create bootstrap samples (taken from page 399 of Efron and Tibshirani) is

```
> bootstrap<-function(x,nboot,theta)
+ {data<-matrix(sample(x,size=length(x)*nboot,replace=T),nrow=nboot)
+ return(apply(data,1,theta))
+ }
```

(A fancier version that will compute BCa and ABC intervals can be had by downloading the R contributed package "bootstrap" that has been written to accompany the Efron and Tibshirani book.) Load the MASS package.

Below are times to failure (in millions of cycles) for high speed turbine engine bearings made from two different compounds (taken from "Analysis of Single classification Experiments Based on Censored Samples from the Two-Parameter Weibull Distribution: by J.I. McCool, *Journal of Statistical Planning and Inference*, 1979). (These are pretty small samples to be trying out a bootstrap technology, but for sake of exercise we will ignore this fact.)

Compound 1	3.03	5.53	5.60	9.30	9.92	12.51	12.95	15.21	16.04	16.84
Compound 2	3.19	4.26	4.47	4.53	4.67	4.69	5.78	6.79	9.37	12.75

Enter these two sets of values into R as

```
> compound1<-c(3.03,5.53,5.60,9.30,9.92,12.51,12.95,15.21,16.04,16.84)
> compound2<-c(3.19,4.26,4.47,4.53,4.67,4.69,5.78,6.79,9.37,12.75)
```

a) Make normal plots for the two samples as

```
> qqnorm(compound1)
> qqnorm(compound2)
```

Do you expect "constant variance/normal distribution" ordinary statistical methods to be reliable in the analysis of these data? Why?

b) Make normal plots of the log-lifetimes here as

```
> qqnorm(log(compound1))
> qqnorm(log(compound2))
```

Are you any more optimistic about the reliability of "constant variance/normal distribution" ordinary statistical methods in the analysis of the log lifetimes?

c) Consider inference for the median of a distribution  $F$  that one might assume to be generating the transplant Compound 2 lifetimes. The sample median Compound 2 lifetime can be had by typing

```
> median(compound2)
```

As a way of assessing the precision of the sample median (as an estimator of the median of  $F$ , you may generate, say,  $B = 10,000$  bootstrap sample medians by typing

```
> B<-10000
```

```
> comp2boot.non<-bootstrap(compound2,B,"median")
```

Based on these values, what number might you report as a standard error for the sample median?

b) Use the bootstrapped sample medians to produce a 95% (unadjusted) percentile bootstrap confidence interval for the median of  $F$ . Some R code for doing this is

```
> kl<-floor((B+1)*.025)
> ku<-B+1-kl
> sortcomp2boot.non<-sort(comp2boot.non)
> sortcomp2boot.non[kl]
> sortcomp2boot.non[ku]
```

c) A parametric bootstrap approach here might be to use a Weibull distribution to describe transplant survival times. It's easy to use maximum likelihood to fit a Weibull to these data by typing

```
> fit2<-fitdistr(compound2,"weibull")
> fit2
```

Then a set of 10,000 simulated sample medians from the fitted distribution can be simulated by typing

```
> Wboot<-function(samp,nboot,theta,shape,scale)
+ {data<-scale*matrix(rweibull(samp*nboot,shape),nrow=nboot)
+ return(apply(data,1,theta))
+ }
> comp2boot.Wei<-Wboot(10,B,"median",fit2$estimate[1],fit2$estimate[2])
```

Use these to produce a parametric bootstrap standard error for the sample median and a parametric bootstrap 95% (unadjusted) percentile bootstrap confidence interval for the median of  $F$ . How do these compare to what you found above?

d) As Koheler's Assignment #9 from Spring 2002 points out, large sample theory says that if  $f$  is the density function for  $F$  and  $\theta = F^{-1}(.5)$  (the population median) the standard deviation of the sample median is approximately

$$\frac{1}{2f(\theta)\sqrt{n}}$$

How does a (Weibull) parametric bootstrap standard error for the sample median Compound 2 lifetime compare to a plug-in version of the above (where the sample median is used in place of  $\theta$  and the fitted Weibull density is used for  $f$ )?

e) If one models the Compound 1 survival times as iid from second distribution,  $G$ , it might be of interest to estimate the difference in medians  $F^{-1}(.5) - G^{-1}(.5)$ . Modify the above to create  $B = 10,000$  (nonparametric) bootstrap sample medians from the Compound 1 cases. Then compute 10,000 differences between a bootstrap Compound 2 sample median and a bootstrap Compound 1 sample median. Based on these 10,000 differences, find 95% percentile confidence limits for the difference in underlying median lifetimes. Is this difference clearly non-zero?

2. Problem 3 of the Spring 2003 Stat 511 Final Exam is a bootstrap problem. Do the R computing required to get your own copy of the printout provided on that Exam. Then answer the Exam questions.