STAT-6494 Advanced Statistical Computing with R

Homework 7

Wenjie Wang

21 April 2016

1 Exercise: Estimating the Median with Bootstrap

Efron (1979) had an example of estimating the median Section 3. Suppose that $X_1, \ldots, X_n$, $n = 2m - 1$ from a random sample from a distribution F. Let $\theta(F)$ be the median of F and $t(\boldsymbol{X}) = X_{(m)}$ be the sample median. Having observed $\boldsymbol{X} = \boldsymbol{x}$ with $n = 13$ from $N(0, 1)$, use Monte Carlo bootstrap method to approximate the sampling distribution of

$$R(\boldsymbol{X}, F) = \frac{|t(\boldsymbol{X}) - \theta(F)|}{\sigma(F)},$$

the absolute error of the sample median relative to the population standard deviation.

2 Reproduction of Example Simulation Study

The following is the function we revised from the course notes and the corresponding simulation results generated.

```
median.boot <- function(x, nboot) {
    tx <- median(x)
    sx <- sd(x) / sqrt(nx <- length(x))
    do1rep <- function(x) {
        x.b <- sample(x, nx, replace = TRUE)
        abs(median(x.b) - tx) / sx
    }
    r.sample <- replicate(nboot, do1rep(x))
    c(mean(r.sample), sd(r.sample))
}
set.seed(421)
x <- rnorm(13)
median.boot(x, 50)
```

```
> [1] 1.97921 1.21562
```

```
sim <- replicate(10, median.boot(rnorm(13), 100))
resDat <- data.frame(t(sim))
colnames(resDat) <- c("Ave.", "S.D.")
resDat
```

```
>         Ave.      S.D.
> 1  1.4211969 1.1325893
> 2  1.0242994 1.0539495
> 3  0.9959081 0.9656427
```

```
> 4  0.7490013 1.0068949
> 5  1.3137223 1.3220628
> 6  1.5285239 1.0811221
> 7  0.6340498 0.7675287
> 8  1.2978502 1.2054278
> 9  0.9578096 0.9211110
> 10 0.9873382 1.0661605
```

```
## average of bootstrap estiamtes and sd estimates
colMeans(resDat)
```

```
>     Ave.     S.D.
> 1.090970 1.052249
```

```
## empirical sd of bootstrap estimates and sd estimates
apply(resDat, 2, sd)
```

```
>      Ave.      S.D.
> 0.2907066 0.1533219
```

The results shown in the first column resemble the second column in Table 1 on page 8 in Efron (1979). Therefore, the similar results to the example simulation study is produced successfully. Note that the only revise we did on the function `median.boot` from course notes is $\sigma(F)$, the so-called "population standard deviation", which is introduced to help the quantity of interest, $R(\boldsymbol{X}, F)$ be more numerically stable.

Reference

Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7 (1). The Institute of Mathematical Statistics: 1–26. doi:10.1214/aos/1176344552.