Prof. Donna Ankerst, Katharina Selig

June 3rd, 4th, 17th & 18th 2019

Problem 1

Consider the following observations which represent the times in hours between failures of air-conditioning equipment.

```
times <- c(3, 5, 7, 18, 43, 85, 91, 98, 100, 130, 230, 487)
```

(a) Assume n failure times are i.i.d. $\text{Exp}(\lambda)$ distributed with $f(x) = \lambda e^{-\lambda x}$. Write out the steps to calculate the maximum likelihood estimator (mle) for λ. Write `R` code to solve for the mle for the above data.

(b) Write `R` code to obtain the parametric bootstrap estimator of λ for 10 i.i.d. observations from an $\text{Exp}(0.001)$ distribution along with its bias and MSE.

(c) Using the data above, write `R` code to estimate the bias and standard error of the mle estimate via the non-parametric bootstrap.

`In R:`

(d) Execute the `R` code for (a) - (c) reporting the numerical results requested. Note that this is an opportunity to check your answers (a) - (c) are correct, an opportunity you will not have on the exam.

Problem 2

We consider a data set containing the height(cm) of 20 randomly selected German men:

$$173, 183, 187, 179, 180, 186, 179, 196, 202, 198, 197, 185, 194, 185, 191, 182, 182, 187, 184, 186$$

(a) Write down the empirical cumulative distribution function F_n considering intervals of length 5cm.

(b) Write `R` code to compute 95% bootstrap confidence intervals for the mean using the four approaches: standard normal bootstrap, basic bootstrap, bootstrap percentile and bootstrap t.

`In R:`

(c) Write `R` code for (a) using the `seq` and `sapply`. Compare it to the `ecdf` command in `R`.

(d) Implement the `R` code for (b) to obtain the four types of CI's. Overlay the four CI's on a histogram of the bootstrap replicates. Comment on the distribution of the bootstrap replicates and how it relates to

Problem 3

An experiment is to be performed to test equality of two sample means using the two-sample t-test assuming equal variances and a significance level of 0.05. The test statistic is given by

$$t = \frac{\bar{x}_A - \bar{x}_B}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \quad \text{with } s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2},$$

where n_A, n_B are the sample sizes, $\bar{x}_A, \bar{x}_B$ denote the sample means, and s_A^2, s_B^2 denote the sample variances for each group. Reject the null hypothesis if $|t| > t_{n_a+n_B-2,1-\alpha/2}$.

(a) Develop and write down an algorithm to calculate the observed significance level (Type I error rate) for the two-sample t-test. Assume the data are normal distributed with variance equal to one and $n_A = n_B = N$.

(b) Develop and write down an algorithm to calculate the empirical power for the two-sample t-test. Assume the data are normal distributed with variance equal to one, $n_A = n_B = N$, and the mean difference $\delta = 0.5$.

`In R:`

(c) Implement the algorithms of (a) and (b) in `R` using the following different simulation scenarios.

- The data are normal distributed with variance one or follow a chi-square distribution with mean and degrees of freedom 2 and $2 + \delta$.
- The sample sizes vary from 10 in each group to 100 in steps of 10.
- The mean differences δ range from 0.1 to 1 in steps of 0.1.

Present the results in graphical form.