Probabilistic inference for data science 2 Examination 19.08.2021

1. Suppose a die is thrown repeatedly 20 times. Throws are independent. It is possible that the die is biased. Define $X =$ "number obtained by the throw of the die". To estimate the probability $\mathbb{P}(X = 6)$ from a data proportion of number six is used. The proportion of throws resulting 6 is denoted by $\hat{p}_6$.

   The outcomes of 20 throws are as follows:

   $$6, 6, 6, 6, 5, 5, 5, 2, 2, 2, 3, 2, 6, 4, 6, 2, 4, 6, 1, 1$$

   Use non-parametric bootstrap with 5000 resamples to estimate $Var(\hat{p}_6)$.

2. Assume random sample $X_1, \ldots, X_n$ form the exponential distribution with unknown parameter λ, $f(x|\lambda) = \lambda e^{-\lambda x}, x > 0$ and $f(x|\lambda) = 0$ otherwise. The distribution of the maximum likelihood estimator $\widehat{\lambda}_n$ of λ follows asymptotically normal distribution $N(\lambda, \frac{\lambda^2}{n})$ i.e.

   $$\sqrt{n}(\widehat{\lambda}_n - \lambda) \xrightarrow{d} N(0, \lambda^2)$$

   Determine the asymptotic distribution of $\dfrac{1}{\widehat{\lambda}_n}$.

3. Suppose we are interested in learning about the effect of a newly developed gasoline detergent additive on automobile mileage. To gather information, 14 cars have been assembled, and their gasoline mileages (in units of miles per gallon) have been determined. For each car this determination is made both when gasoline without the additive is used (variable no.additive) and when gasoline with the additive is used (variable additive). The data can be in the file consumption.txt

   ```
   > # Read data
   > dat <- read.table("consumption.txt", sep=";", header=T)
   ```

   For instance, car 1 got 32.7 miles per gallon by using gasoline without the additive and 32.3 miles per gallon by using gasoline with the additive, whereas car 4 obtained 25.0 miles per gallon by using gasoline without the additive and 25.3 miles per gallon by using gasoline with the additive.

   We may assume that consumption without additive and with additive are samples of the same size from different normal populations having respective expected values μ_{no} and μ_{yes}. Moreover we may assume that difference

   $$\text{consumption without additive - consumption with additive}$$

   follows normal distribution with expected value $\mu_{no} - \mu_{yes}$.

   Use the method based on the likelihood ratio test to find 95% confidence interval for the $\mu_{no} - \mu_{yes}$.