(11) Nonparametric inference

Study questions

1. Write down the definition of the empirical distribution of a random variable X.
2. Write down the definition of the empirical distribution function of a random variable X.
3. State the Glivenko-Cantelli theorem.
4. Write down the definition of a histogram estimator with binwidth b for the PDF of a random variable X.
5. Write down the definition of a kernel density estimator with bandwidth b for the PDF of a random variable X.
6. Give a conceptual explanation of the formula of a kernel density estimator.
7. Give a conceptual explanation of how the bootstrap estimates the variance of a statistic.

Exercises (Theory)

1. Introduce the Kolmogorov-Smirnov test of a simple hypothesis (DeGroot and Schervish, 2012, Theorem 10.6.2, Definition 10.6.2, pp. 657 - 663).
2. Introduce the Jackknife estimator for the variance of a statistic (Wasserman, 2004, Section 8.5.1).
3. Prove the Glivenko-Cantelli theorem (Van der Vaart, 2000, Theorem 19.1).

Exercises (Programming)

1. Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$. Implement a histogram density estimator for the PDF of $X_i := X_1$. Visualize the relationship between the sample size n and the binwidth b in determining the quality of the estimate.
2. Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$. Implement a kernel density estimator for the PDF of $X_i := X_1$. Visualize the relationship between the sample size n and the bandwidth b in determining the quality of the estimate.
3. For $X_1, ..., X_n \sim N(\mu, \sigma^2)$ use the bootstrap to estimate the variance of the T-statistic

$$T = \sqrt{n}\frac{\bar{X}_n - \mu}{S_n}. \tag{9}$$

   Compare the bootstrap estimates to the analytical variance $\mathbb{V}(T) = \frac{n}{n-2}$ for varying $n > 2$ and varying number of bootstrap samples.