University of Edinburgh, School of Mathematics

Statistical Research Skills

Assignment 3 - Simulation Report

Johnny Lee

26th Apr 2022

1. Introduction

This report consists of two main parts. In the first part, we aim to compute one-shot experiment on density function, `density()` also known as the kernel density estimator. In addition to that, we will compare with its other competitors such as orthogonal series estimator and penalised kernel density estimator. In the second part, we will conduct a Monte Carlo simulation study for different sample sizes on the previously suggested methods. Thus, we will evaluate the integrated squared error (ISE) for different cases of sample sizes.

2. Preliminary Experiment

2.1 Methodology

2.1.1 Kernel Denisty Estimator

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f$. The kernel density estimator of f is defined as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

where $K_h(x) = \frac{K(x)}{h}$, K is a kernel and $h > 0$ is a parameter controlling smootheness of the estimate. This is avaiable as the base function of `R` as `density()`.

2.1.2 Orthogonal Series Estimator

With reference to (Kreyszig, 1991)[5], Kreyszig introduced the orthogonal series estimator using normalised Hermite polynomials. These polynomials form a orthonormal sequence for the univariate case as below,

$$\hat{f}(x) = \frac{1}{(2^i i! \sqrt{x})^{\frac{1}{2}}} \exp(-x^2/2) H_i(x)$$

and

$$H_0(x) = 1, \quad H_i(x) = (-1)^i \exp(x^2) \frac{d^i}{dx^i} \exp(-x^2)$$

where $H_i(x)$ is called the *Hermite polynomial of order i*. Furthermore, we can further simplify the process as Kreyszig showed its properties as below.

$$H_{i+1}(x) = 2xH_i(x) - H_i'(x), \quad H_i'(x) = 2iH_{i-1}(x)$$

Based on the properties above, we define a new function known as `OS_Hermite()` to iterate the process to obtain the series of estimation.

2.1.3 Penalised Kernel Density Estimator

(Kauermann et al, 2009)[4] introduces the penalised likelihood

$$\hat{f}(x) = \sum_{n=-m}^{m} c_n \phi_n(x),$$

where $\phi_n(x)$ is the basis densities. Then the weight c_n is parameterised as follow

$$c_n(y) = \frac{\exp(\beta_n)}{\sum_{n=-m}^{m} \exp(\beta_n)}$$

with $\beta_0 = 0$ and $\boldsymbol{\beta} = \beta_{-m}, \ldots, \beta_{-1}, \beta_1, \ldots \beta_m$ so that $\int f(x)dx = 1$.

(Deng et al, 2011)[1] then further suggested the simplified kernel apporach

$$\hat{f}(x) = \frac{1}{m} \sum_{n=1}^{m} K\left(\frac{x - \mu_n}{h}\right)$$

and introduced a package `gss`, (Gu, 2011)[3] which uses a penalised likelihood approach for nonparametric density estimation. With the functions `ssden()` and `dssden()`, we can estimate the kernel density.

2.2. Generating Data

We generated

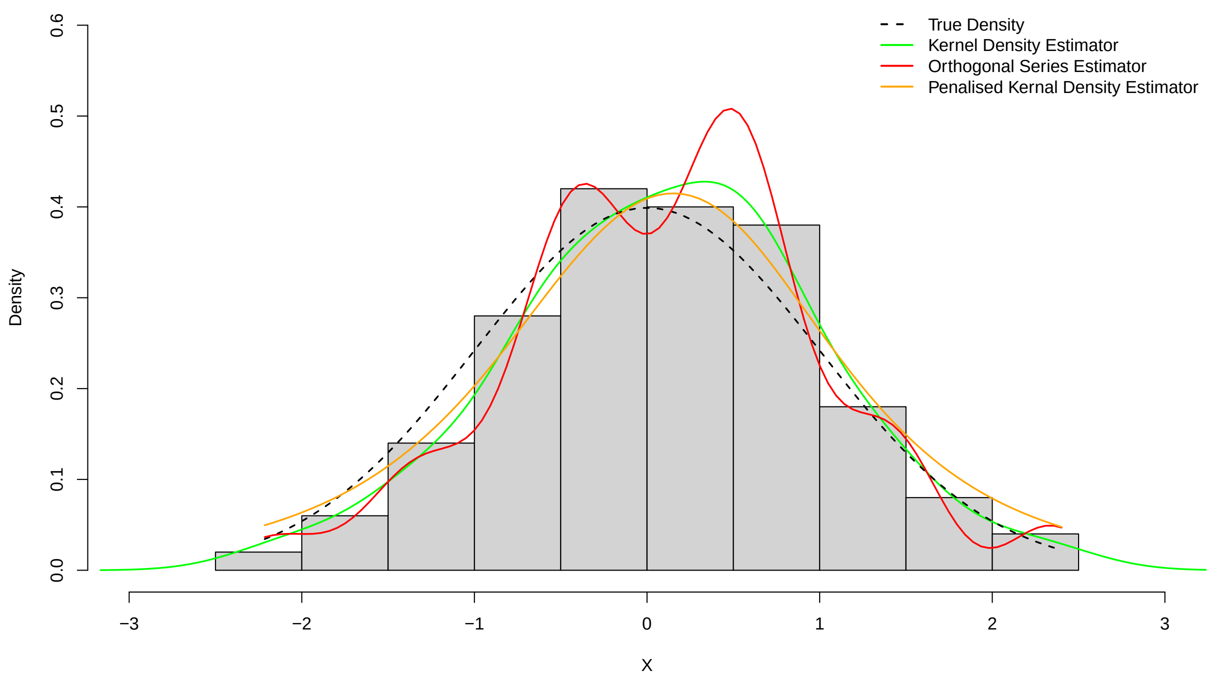

Figure 1: One-shot Experiment on Normal distribution

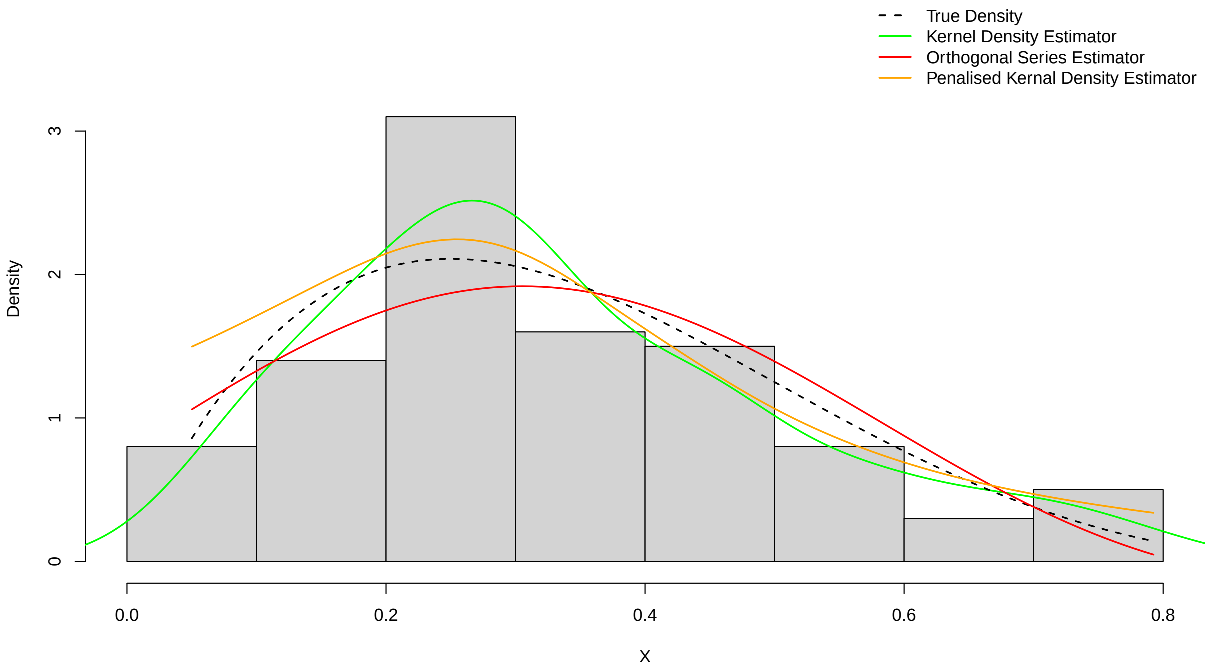

Figure 2: One-shot Experiment on Beta distribution

4. Monte Carlo Simulation Study

5. Conclusion

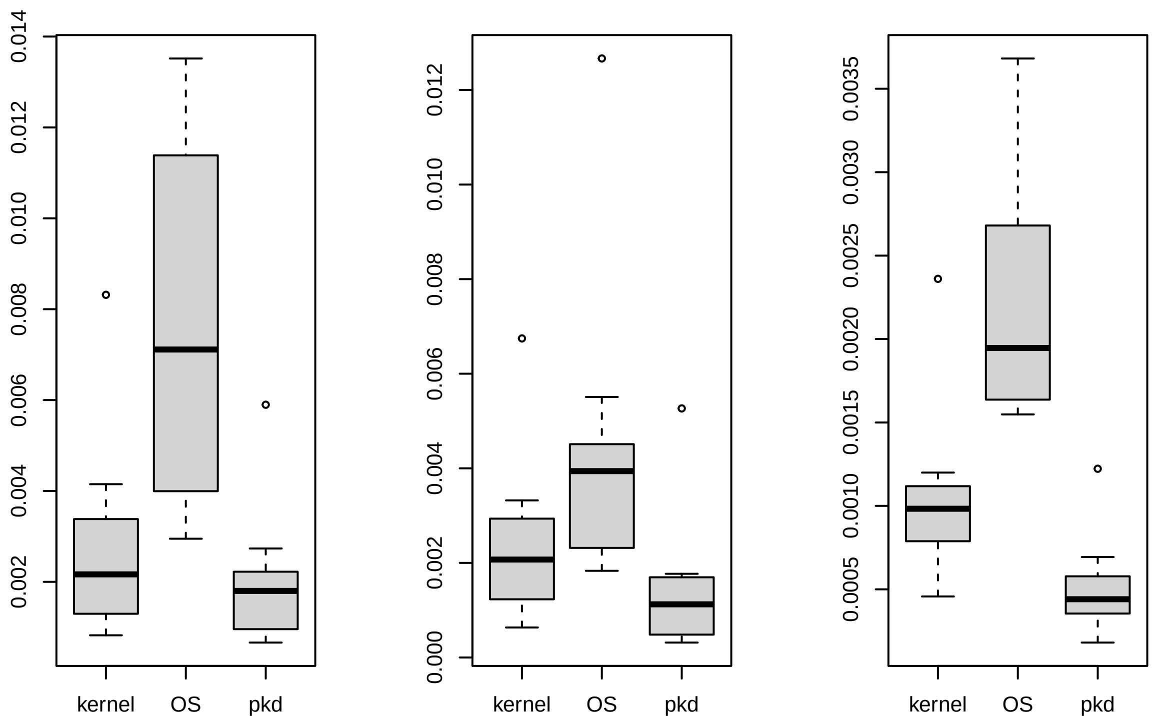

Figure 3: Boxplot of Different Estimators with Different Sample Sizes

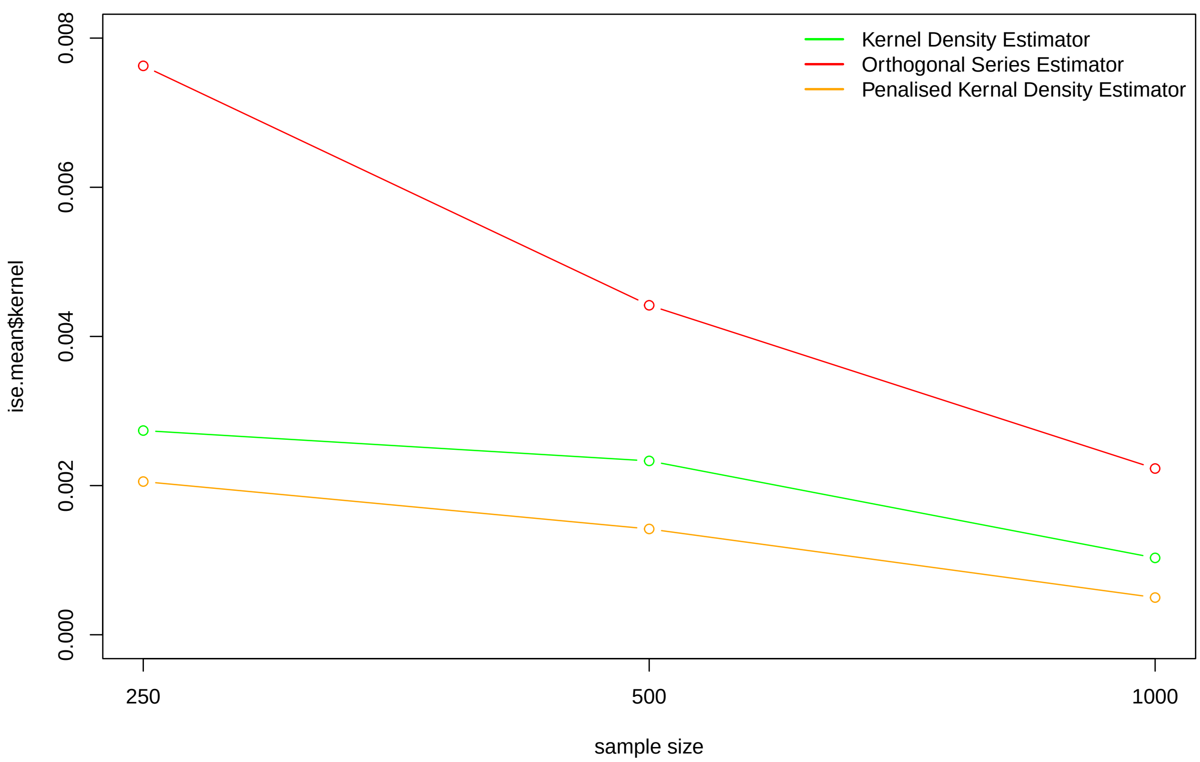

Figure 4: Mean Integrated Squared Error with Different Sample Sizes

Reference

1. Deng, H. and Wickham, H., 2011. Density estimation in R. Electronic publication.
2. Girolami, M., 2002. Orthogonal series density estimation and the kernel eigenvalue problem. Neural computation, 14(3), pp.669-688
3. Gu, C., 2011. Smoothing spline ANOVA models: R package gss. Journal of Statistical Software, 58, pp.1-25.
4. Kauermann, G. and Schellhase, C., 2019. Density Estimation with a Penalized Mixture Approach.
5. Kreyszig, E., 1991. Introductory functional analysis with applications (Vol. 17). John Wiley & Sons.