Statistical Research Skills - Assignment 3

Yile Shi

2022/4/18

1. Introduction

In this report, we compare the performance of the kernel density estimation with other density estimators. To this end, our report consists of two main sections. In the first section, we run a one-shot experiment using kernel density estimation and its competitors based specified data generating process. Moreover, we compare the estimation performances intuitively with a figure and discuss the weaknesses and strengths with the methods. In the second part, we conduct a Monte Carlo simulation study for different sample sizes and observe how the estimation accuracy changes when we increase the sample size by reporting the corresponding integrated squared errors.

2. Data Generating Processes and Preliminary Experiments

2.1 Methodology

Before any data simulation or experiment, we introduce the methodologies of the kernel density estimation, as well as its competitors first.

Kernel Density Estimation

The kernel density estimation (KSE) is a widely-used density estimator. Comparing with histogram approaches, the kernel density estimation overcomes the discreteness of them by centring a smooth kernel function at each data point then summing to get a density estimate.

Let $X_1, \cdots, X_n \sim^{iid} f$. The kernel estimator of f is defined as

$$\hat{f}_{kde}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where K is the density kernel and h is the bandwidth.

The kernel, K, is a symmetric, and usually positive function that integrates to one. Common kernel functions are uniform, triangle, Epanechnikov, quartic (biweight), tricube (triweight), Gaussian (normal), and cosine.

The bandwidth, h, is a smoothing parameter. Generally, large bandwidths produce very smooth estimates, while small values produce wiggly estimates. The bandwidth influences estimation accuracy much more than the kernel, so choosing a good bandwidth is critical to get a good estimate.

Typically, we select the bandwidth of KDE and evaluate its performance by the integrated square error (ISE):

$$ISE = \int \left[\hat{f}_{kde}(x) - f(x)\right]^2 dx$$

Lower value of ISE suggests better selection on the bandwidth and better performance of the estimator. However, this optimization problem can be difficult to solve as usually the true density function $f(x)$ is unknown. General consensus is that plug-in selectors and cross validation selectors are the most useful over a wide range of inputs.

Average Shifted Histogram

The histogram is the oldest and least sophisticated method for density estimation. One of its simple enhancement is the average shifted histogram (ASH), which is smoother and avoids the sensitivity to the choice of origin. Specifically, the premise of this approach is to take m histograms, $\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_m$, of bin width h with origins of $t_o = 0, \frac{h}{m}, \frac{2h}{m}, \cdots, \frac{(m-1)h}{m}$. Based on these, we simply define the naive ASH as:

$$\hat{f}_{ash}(x) = \frac{1}{m}\sum_{i=1}^{m}\hat{f}_i(x)$$

There are $k = 1, \cdots, m \cdot n$ bins across all histograms, each spanning $\left[k\frac{h}{m}, (k+1)\frac{h}{m}\right]$ with centre $(k+0.5)\frac{h}{m}$. The ASH can be somewhat more general by using all bins to estimate the density at each point, weighting bins closer to the data more highly. In this way, ASH is also considered as an improvement for the discreteness of the traditional histogram. The weighted ASH is defined as:

$$\hat{f}_{ash}(x) = \frac{1}{m}\sum_{k=1}^{m\cdot n}\omega(l_k - x)\hat{c}_k(x)$$

where ω is a weighting function, l_k is the centre of bin k, and $\hat{c}_k(x)$ is the number of points in that bin.

Penalized Likelihood Estimation

Penalized likelihood estimation (PLE) is another advanced approach to estimate the density, which considered the density as a mixture of m "basis" densities, compromising between estimation accuracy and model complexity. As mentioned, PLE generally approximates the density of x, $f(x)$, as a mixture of m densities:

$$\hat{f}_{pen}(x) = \sum_{i=1}^{m} c_i\phi_i(x)$$

where $\phi_i(x)$ is a "basis" density and c_i is the corresponding weight picked to ensure that $\hat{f}_{pen}$ can integrate to 1. The basis densities are weighted equally and differ only by a location parameter, μ_i. Thus, we can obtain a simplified definition which has a similar format with the kernel approach:

$$\hat{f}_{pen}(x) = \frac{1}{m}\sum_{i=1}^{m} K\left(\frac{x-\mu_i}{h}\right)$$

Note that the "basis" densities are no longer constrained to be centred on the data points inn PLE, compared with KDE.

Determining the number and location of μ_i appropriately or not decides the performance of the penalized likelihood estimation. A general way is placing a large number of μ_i at equally spaced locations along the domain of the data and then minimizing a penalized likelihood to remove μ_i with little contribution to the overall quality.

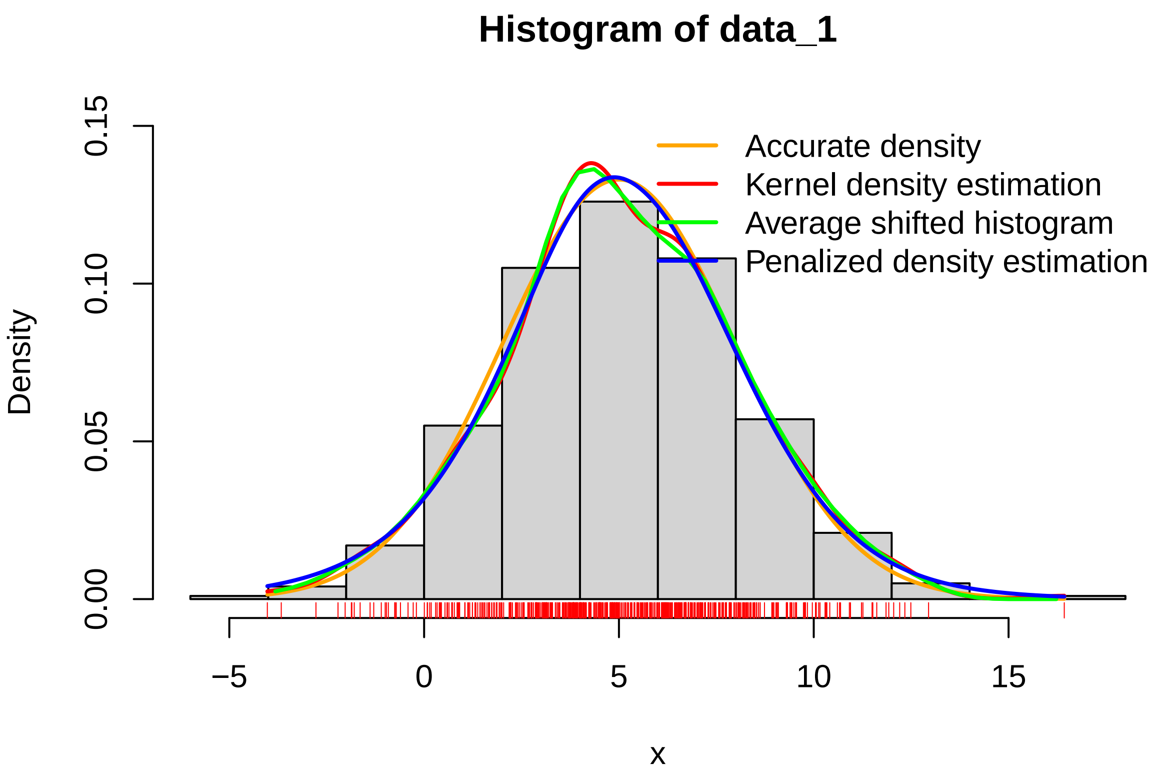
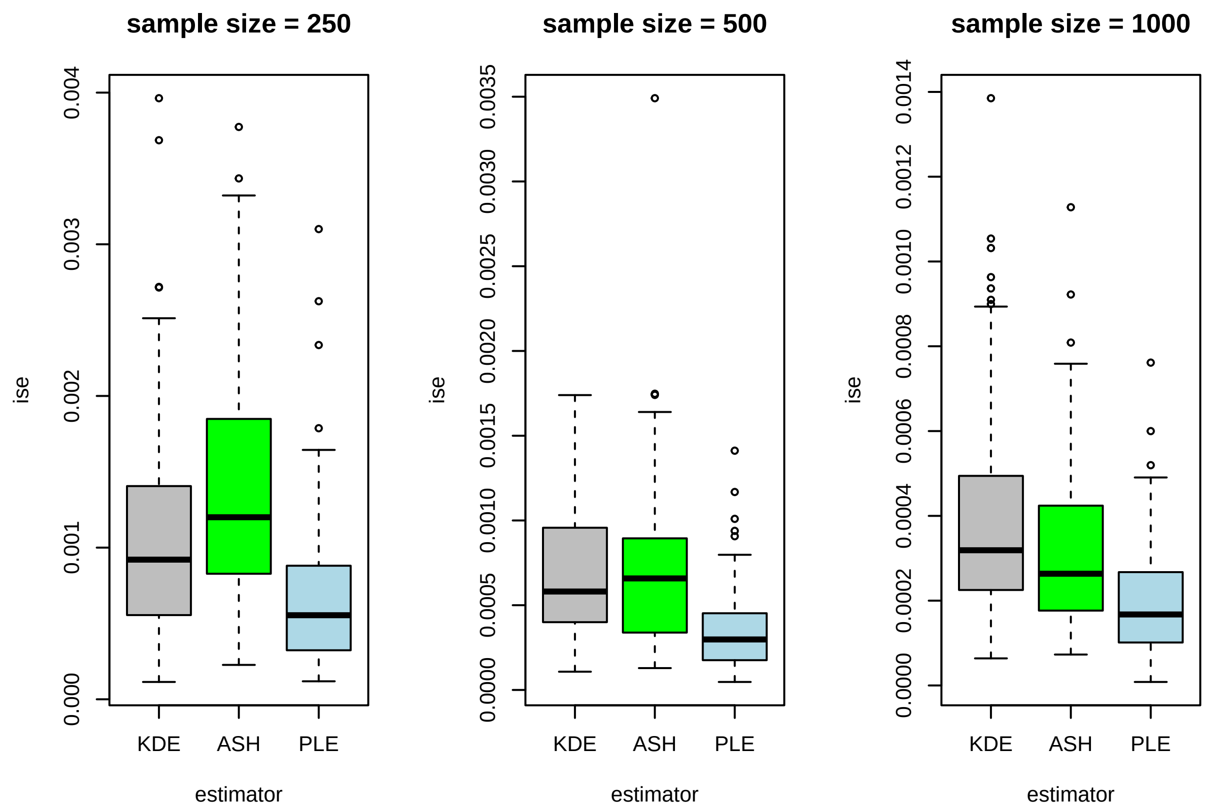
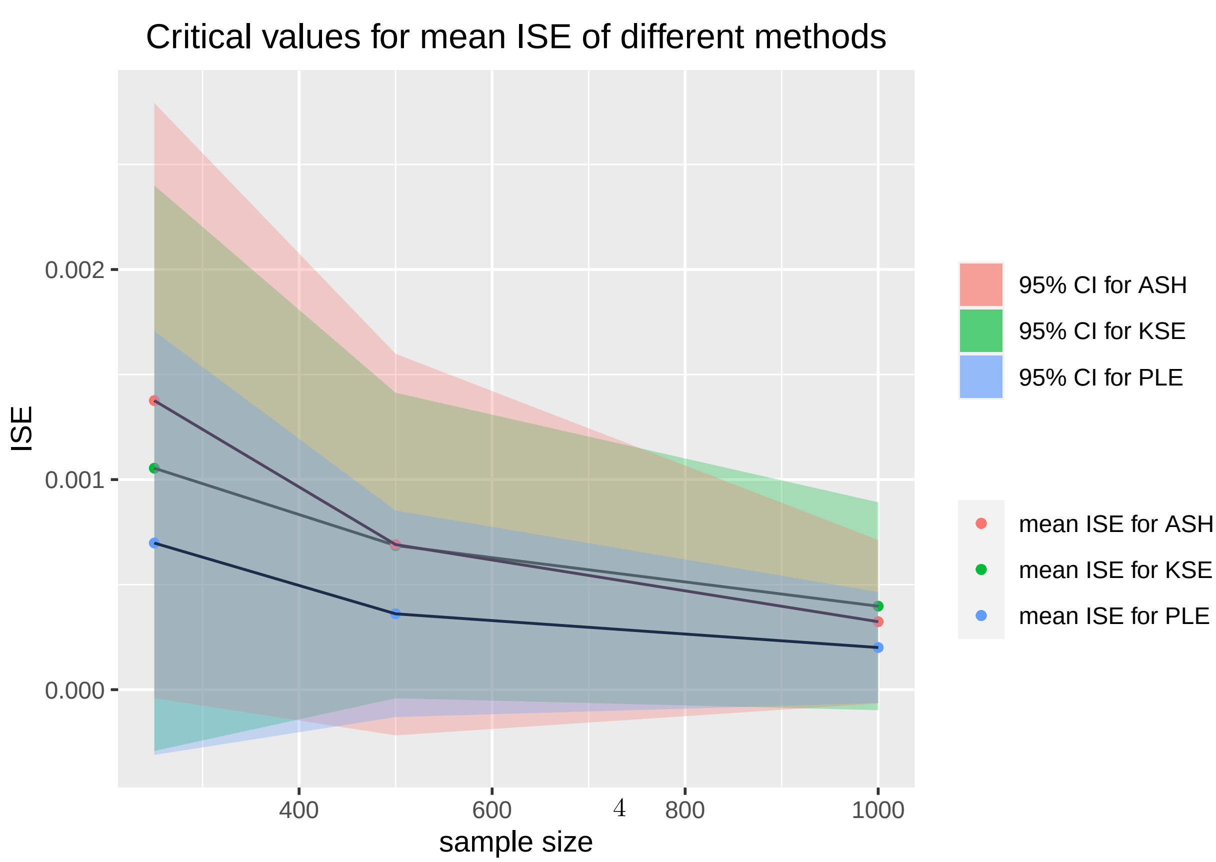


Figure 1: Estimated densities for Normal(5, 9) with different estimators

2.2 Data Generating Processes

3. Monte Carlo Simulation Study





References