Statistical Research Skills - Assignment 3

Yile Shi

2022/4/18

1. Introduction

In this report, we compare the performance of 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, including average shifted histogram and penalized likelihood estimation, based on specified data sequences. In the second part, we conduct a Monte Carlo simulation study for different sample sizes and observe how estimation accuracy changes when we increase the sample size by reporting the corresponding integrated squared errors.

2. Data Generating Processes and Preliminary Experiments

In this section, we introduce the methodologies of kernel density estimation and its competitors first. Next, we run one-shot experiments and construct density plots to compare the estimation performance, after data generating processes.

2.1 Methodology

Kernel Density Estimation

Kernel density estimation (KDE) is a widely-used density estimator, which centres a smooth kernel function at each data point then summing to get a density estimate [1].

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 [2].

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.

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) [3], 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. 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 [4], 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:

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

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.

2.2 Data Generating Processes

To test the performance of aforementioned density estimators on different distributions, we consider data simulation from both symmetric and asymmetric distributions. Specifically, we generate data sequences from the following distributions, without the loss of generality:

- Scenario 1: $Normal(5, 3^2)$, $f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-5}{3}\right)^2}$

- Scenario 2: $Beta(3, 8)$ $f(x) = \frac{x^2(1-x)^7}{B(3,8)}$, where $B(3,8) = \frac{\Gamma(3)\Gamma(8)}{\Gamma(3+8)}$ and Γ is the Gamma function

2.3 One-shot Experiments

Based on previous data simulation, we apply different methods to estimate the densities of data sequences. To keep the inquiry simple, we use the default options for each method. We illustrate the estimated densities in a figure intuitively, as well as the true density and the traditional histogram. We compare the estimation performances and further the strengths and weaknesses of each estimator.
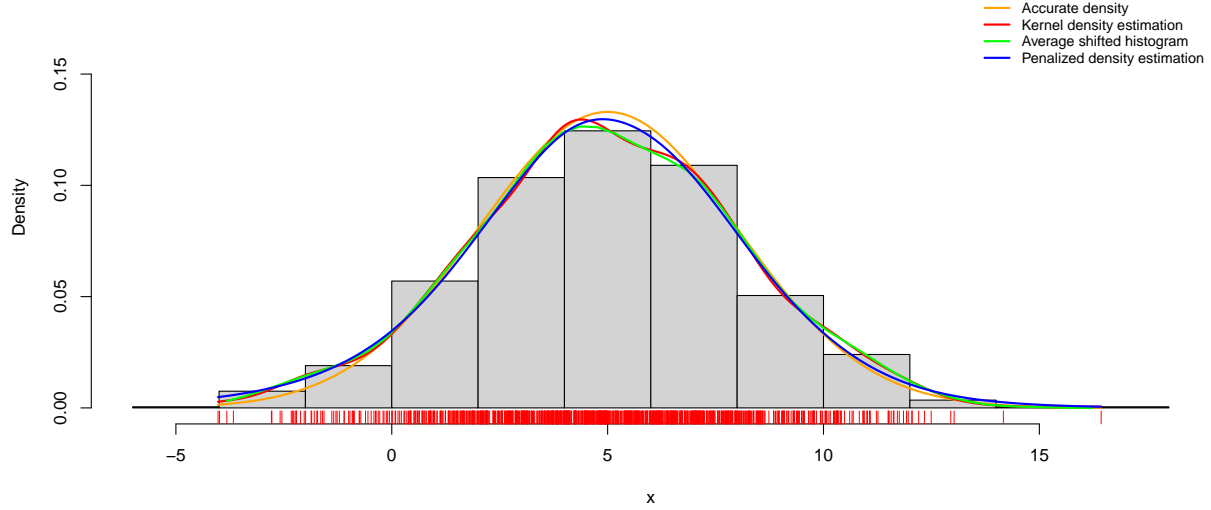
Figure 1: Estimated densities for Normal(5, 9) with different estimators

Scenario 1: Normal distribution

According to Figure 1, we consider the penalized likelihood estimation as the best fitted method in this experiment as its density curve is the smoothest and overlaps the true density most. The kernel density estimator has a quite similar performance with the average shifted histogram. Both of their density curves fall sharply after their peaks.

Scenario 2: Beta distribution

From Figure 2, in this experiment, we observe that all 3 methods don't fit as well as in the normal scenario. Density curves reach higher peaks than the true density and then drop quickly. Still, PLE performs the best due to its smoothest curve, while the other 2 curves vary at around $x = 0.4$.

Strength and Weakness

Based on one-shot experiments, we discuss the advantages and disadvantages of each method.

- `Kernel density estimation`: Comparing with histogram approaches, KDE doesn't rely on the the choice of origin and the number of bins, and produces smooth estimated densities. However, KDE depends on the choice of bandwidth much and it's not easy to determine the best bandwidth. Improper bandwidth leads to poor estimation performance. As shown in scenario 2, the default bandwidth is no longer suitable for asymmetric distributions.

- `Average shifted histogram`: As an improvement of traditional histogram, ASH also avoids the sensitivity of the choice of origin and remains computationally efficient. On the other hand, the performance of ASH is significantly affected by the choice of the number of bins, which is the common weakness of histogram approaches.

- `Penalized likelihood estimation`: The main idea of PLE which estimates the density as a mixture of multiple "basis" densities improves its quality of fit. Moreover, PLE doesn't require "basis" densities to be centred on the data points, compared with KDE, resulting in higher accuracy. As mentioned, one of the weakness of PLE is its high dependence on the location parameter μ_i.
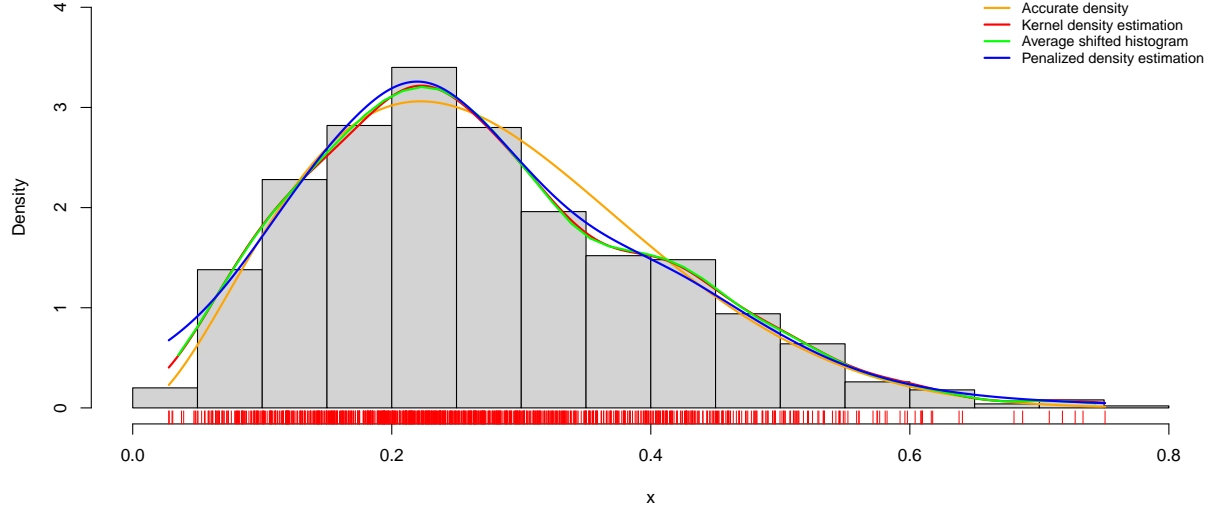
Figure 2: Estimated densities for Beta(3, 8) with different estimators

3. Monte Carlo Simulation Study

To obtain a robust conclusion on the performance of the kernel density estimator and other competitors, we conduct a Monte Carlo simulation study in this section.

Specifically, we repeat simulations from $Normal(5, 3^2)$ distribution 100 times, and in each simulation we generate 250, 500 and 1000 observations. To compare the estimation performance, for each fixed sample size, we compute the integrated square error [5]

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

for each underlying simulated set and report them with boxplots. Moreover, we calculate the overall mean ISE for each sample size with corresponding 95 confidence intervals and illustrate the values using a figure.

Generally, lower ISE suggests better performance of the estimator. From Figure 3, for each sample size, we observe that PLE over-performs other methods with the significantly lower ISEs. KDE shows better performance than ASH when sample size $n = 250$, while it is beaten by ASH when we increase the sample size to 500 and 1000. Figure 4 shows consistent results and we observe a decreasing trend of mean ISE for each methods as the sample size increases. Moreover, corresponding confidence intervals also narrow as n increases, indicating improved estimation accuracy.

4. Conclusion

In summary, we compare the performance of 3 density estimators, including kernel density estimation, average shifted histogram and penalized likelihood estimation. For both one-shot experiments and Monte Carlo study, penalized likelihood estimation, a more advanced method, shows the best performance with best fitted curves and lowest ISEs.
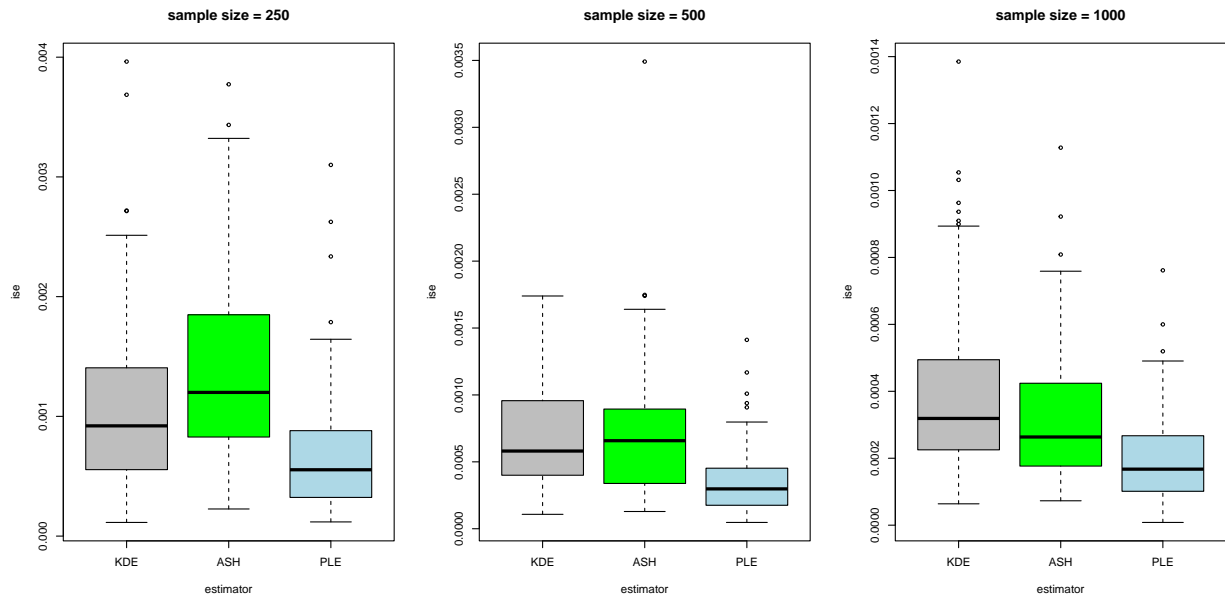
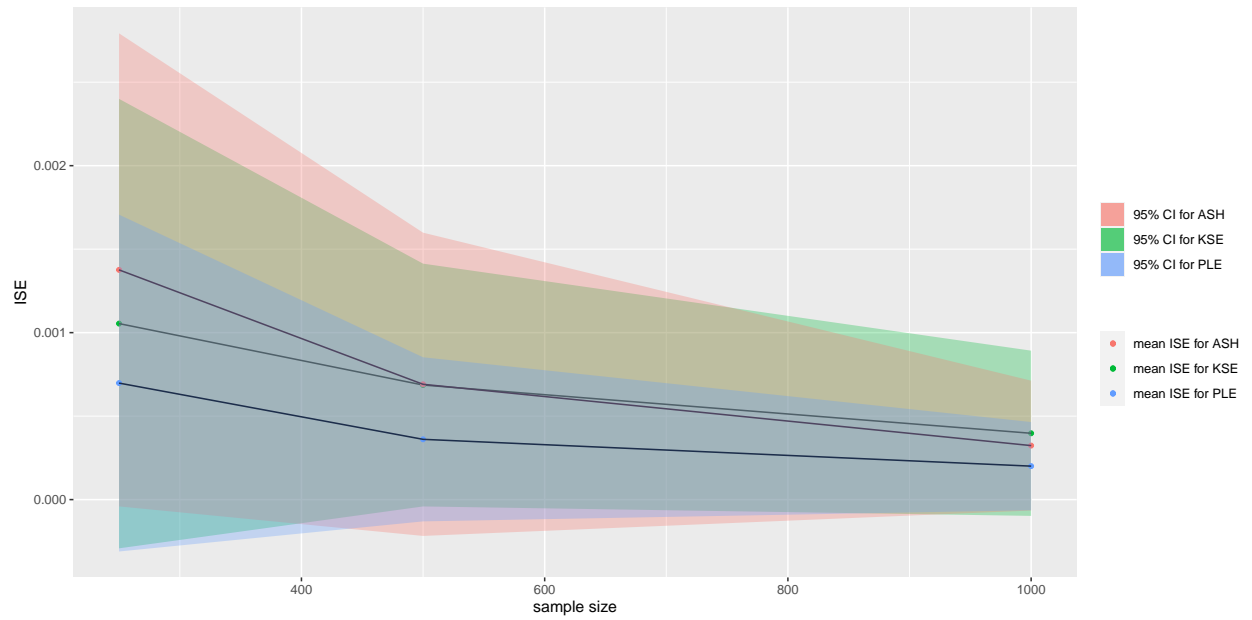Figure 3: Boxplots for ISE of 3 density estimators

Figure 4: Critical values for mean ISEs of different methods

References

[1]   Deng H, Wickham H. Density estimation in r. Electronic publication. 2011;

[2]   Scott D. Multivariate density estimation: Theory, practice, and visualization. John Wiley & Sons; 2015.

[3]   Scott D. Averaged shifted histograms: Effective nonparametric density estimators in several dimensions. The Annals of Statistics. 1985;1024–1040.

[4]   Schellhase C. Density estimation with a penalized mixture approach. 2019;

[5]   Marron JS. Exact mean integrated squared error. The Annals of Statistics. 1992;20:712–736.

[6]   Scott D. Ash: David scott's ASH routines. R package version 10-12 [Internet]. 2009; Available from: https://cran.r-project.org/web/packages/ash/index.html.

[7]   Gu C. Gss: General smoothing splines. R package version 11-7 [Internet]. 2011; Available from: https://cran.r-project.org/web/packages/gss/index.html.