Manav Prabhakar
Shiv Nadar University

Kernel Density Estimate

Manav Prabhakar

February 24, 2021

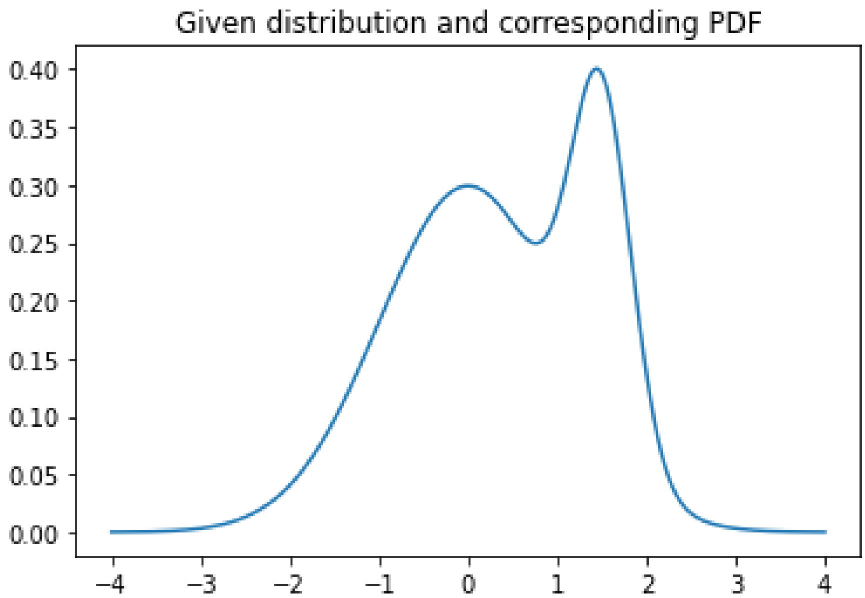
A density function $f_1(x)$ has been defined as

$$f_1(x) = \frac{3}{4}\emptyset(x) + \frac{1}{4}\emptyset_{\frac{1}{3}}\left(x - \frac{3}{2}\right) \qquad (1)$$

Here, $\emptyset(x)$ is defined as

$$\emptyset_\sigma(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \times e^{-\frac{x^2}{2\sigma^2}} \qquad (2)$$

Denoted as $N(0,\sigma^2)$ density. X_1 , X_2 , … X_n are random samples taken from a continuous univariate density f.



The above graph shows the PDF for the function $f_1(x)$

Total 30,000 samples were taken. Out of these 1000 samples were randomly chosen for the next half of the quest

Following that, the density function was calculated using Eq 1.

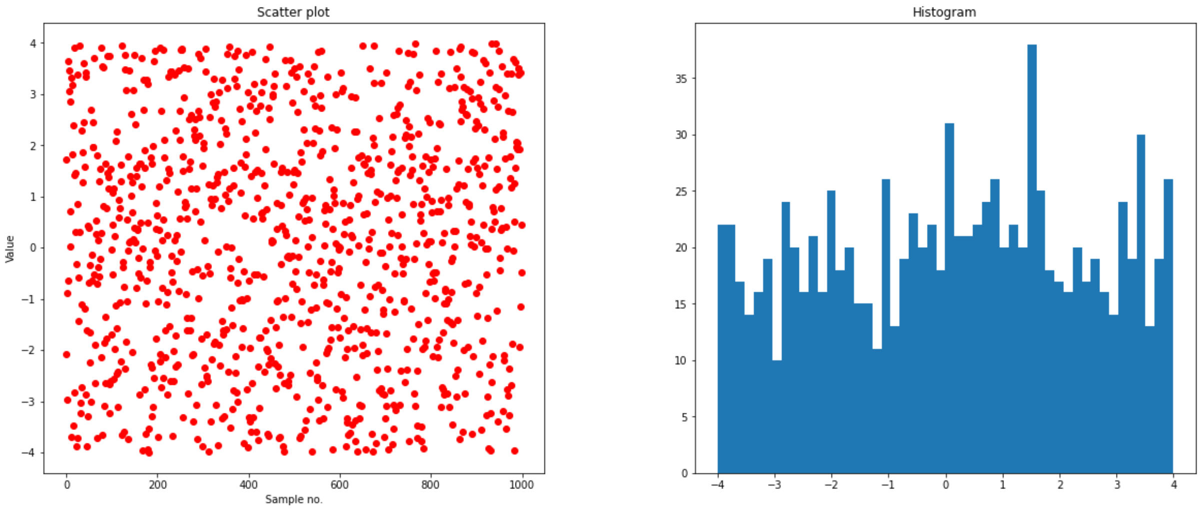


Figure 2: Scatter plot and histogram for 1000 random variables sampled from our population of 30,000 observations.

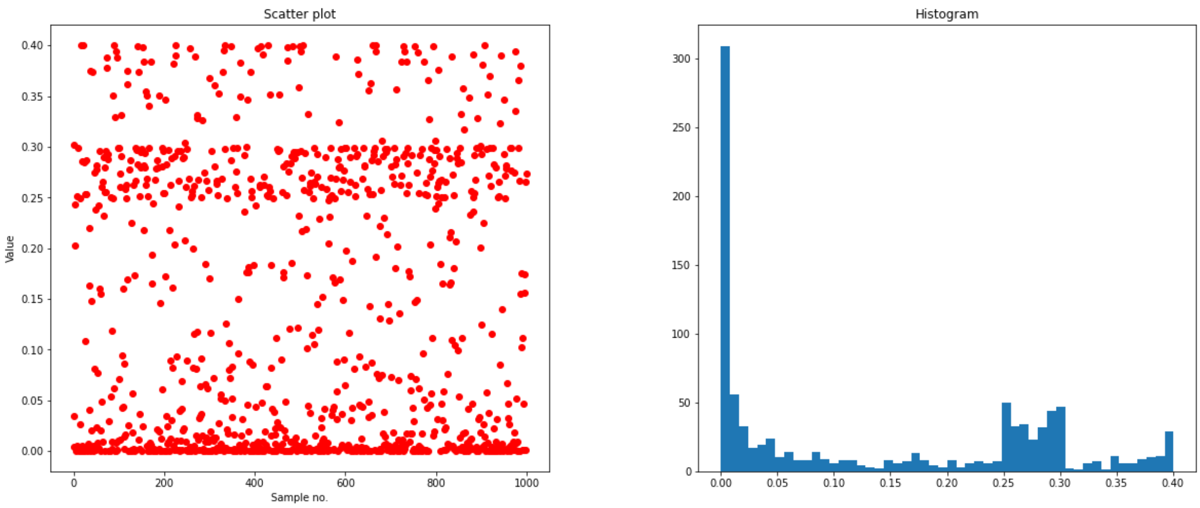


Figure 3: Scatter plot and histogram for the pdf of the 1000 random variables sampled from the population

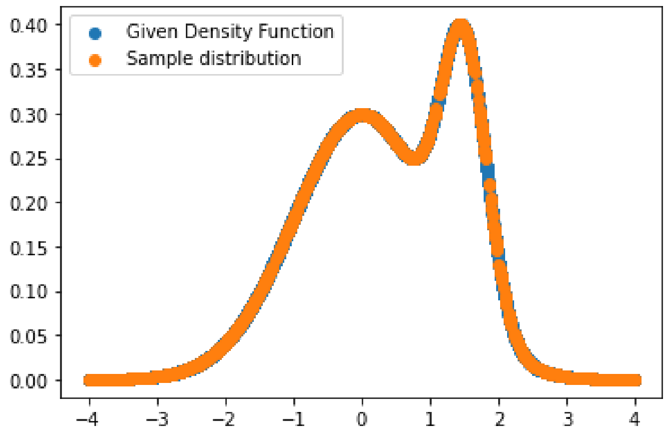


Figure 4: Probability Density function for the sample distribution and population

Manav Prabhakar
Shiv Nadar University

For Kernel Density Estimation, two types of Kernel were used:-

- Gaussian Kernel

$$K(a) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}a^2}$$

- Epanechnikov Kernel

$$K(a) = \frac{3}{4}\left(1 - \frac{a^2}{5}\right)/\sqrt{5}$$
$$for\ abs(t) < \sqrt{5}, else\ 0$$

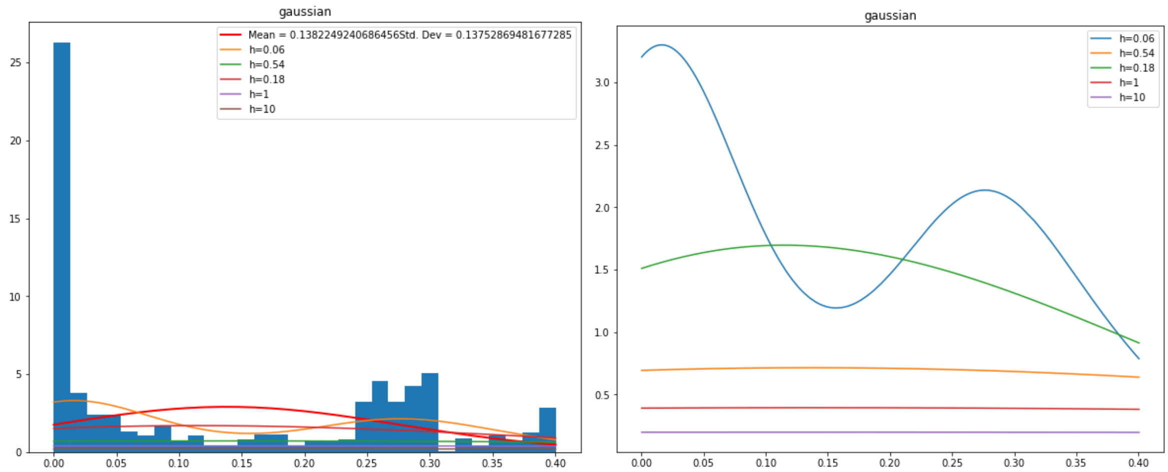


Figure 5: Kernel Density Estimate plots by using a Gaussian Kernel

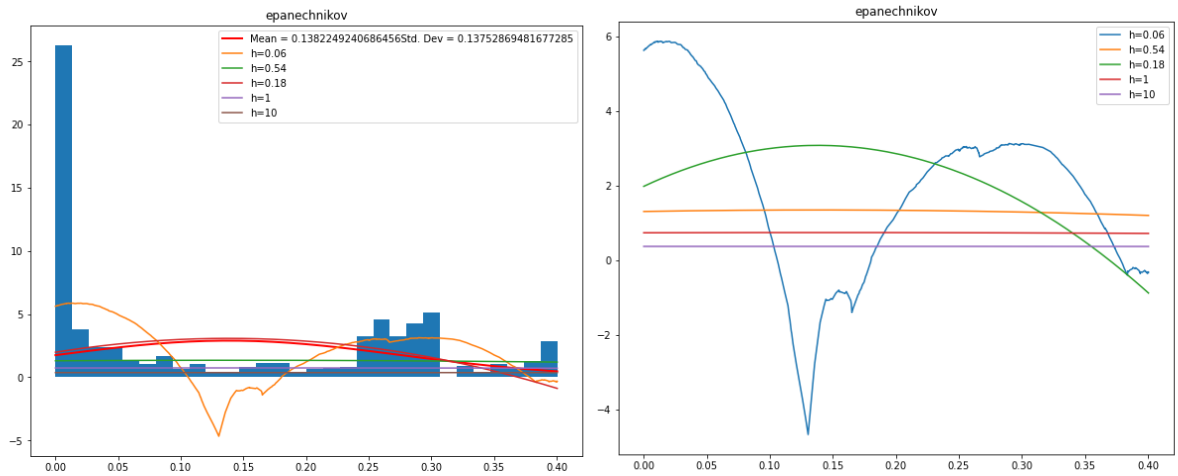


Figure 6: Kernel Density Estimate plots by using Epanechnikov Kernel

Takeaways

The smoothing factor determines how well the estimator fits in the given density function. High smoothing factor (like h=10) in this case leads to over smoothing while a low smoothing factor (like h – 0,06) may lead to under smoothing. The low smoothing factor fits in excessively along the values. However, since the values are only a sample set of the total population, the function obtained may not be the most accurate estimate of the density function.