

Homework 3

Problem 1:

a) **Solution:**

The histogram of the 1000 generated random variables is shown in Fig 1.

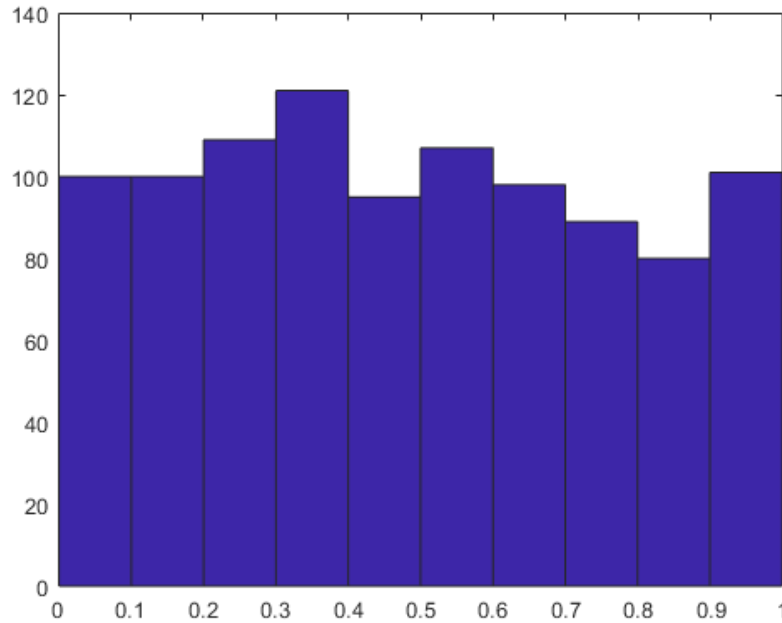


Figure 1: Histogram of 1000 random variables

According to the Law of Large Numbers, $\mathbb{E}[K_h(X - x)] \approx \frac{K_h(X-x_1)+K_h(X-x_2)+\dots+K_h(X-x_n)}{n} \approx f(x)$.

When we approximate the pdf using the Gaussian kernel,

$$K_h(X) = \frac{e^{-\frac{1}{2h}x^2}}{\sqrt{2\pi h}}.$$

The results are shown in Fig 2.

We can see from Fig 2 that when the h is large, the model underfit the pdf, which is not close enough. As the h is decreasing, the support of the random variable becomes closer to $[0, 1]$, and the approximation of pdf becomes more zigzag around 1, starts to overfit.

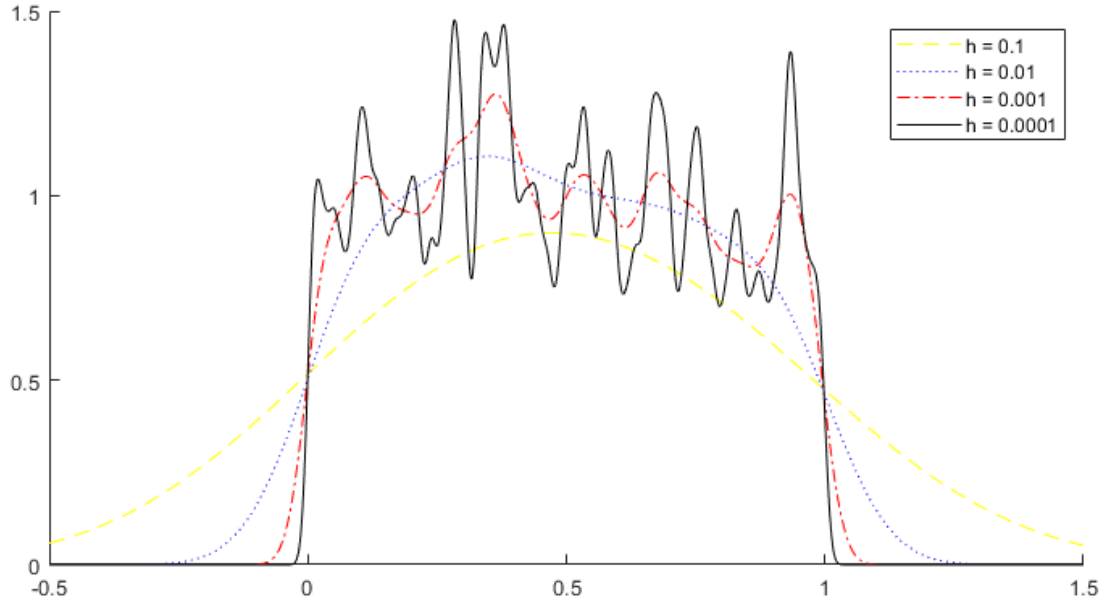


Figure 2: Approximation using Gaussian kernel under different h s.

b) **Solution:**

When we approximate the pdf using the Laplacian kernel,

$$K_h(X) = \frac{e^{-\frac{1}{2h}x^2}}{\sqrt{2\pi h}}.$$

The results are shown in Fig 3.

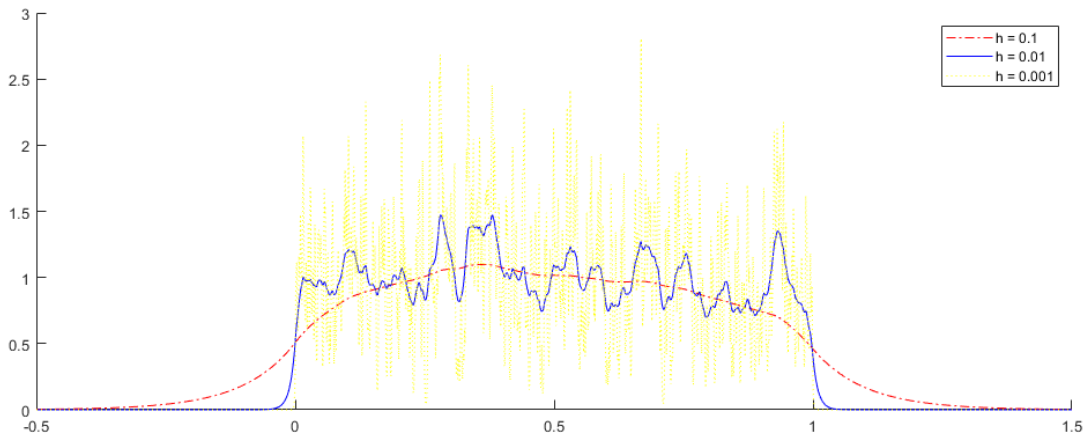


Figure 3: Approximation using Laplacian kernel under different h s.