Report on Advanced Digital Signal Processing Coursework

Kehan He

CID 00682566

Section 1, Random signals and stochastic processes

1.1 Statistical estimation

1.1.1 The theoretical mean $E\{x\} = \int xp(x)dx = \int_0^1 xdx = \frac{x^2}{2}|_0^1 = 0.5$. Comparing to the sample mean of 0.5078 that we have simulated in MATLAB, the discrepancy is due to the limited number of 1000 sample taken. The sample mean will converge to the theoretical mean of 0.5 as number of samples increases.

1.1.2 The theoretical standard deviation $\sigma = \sqrt{E\{(X - E\{X\})^2\}} = \sqrt{E\{(X - 0.5)^2\}}$ $= \sqrt{\int (x - 0.5)^2 p(x)dx} = \sqrt{\int_0^1 (x^2 - x + 0.25)dx} = \sqrt{\left(\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right)|_0^1} \approx$ 0.2887, while the standard deviation of the sample calculated by MATLAB is 0.29594. Similar to the mean estimation, the accuracy in biasness is increased as more samples are taken into simulation.
The result for 1 & 2 shown in the command window is as below

```
The sample mean is: 0.51235
The sample standard deviation is: 0.29725
```

1.1.3 Figure 1 shows the result of the standard deviation and mean of the simulated ten 1000-samples. As we stored the sample values as a 1000x10 matrix under ensemble, mean (ensemble,1) and std(ensemble,1) are used to calculate the sample mean and standard deviation of the 10 columns respectively. Result

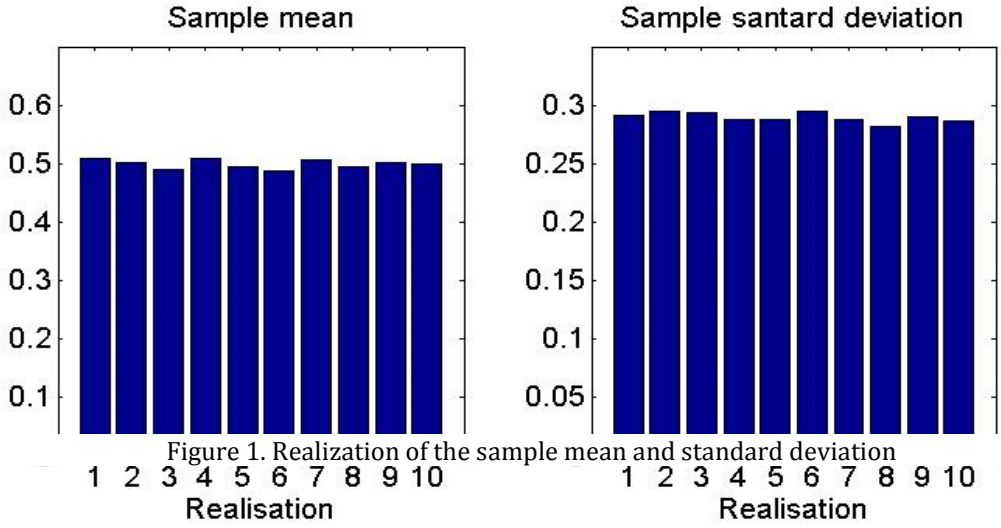

Figure 1. Realization of the sample mean and standard deviation

shown in figure 1 tells that the sample mean and standard deviation are unbiased as the realization values cluster around the true mean (0.5) and standard deviation (0.2887).

1.1.4 As we concluded before, increased number of samples will converge the sample mean to the theoretical value of 0.5. Also, as shown in Figure 2, as

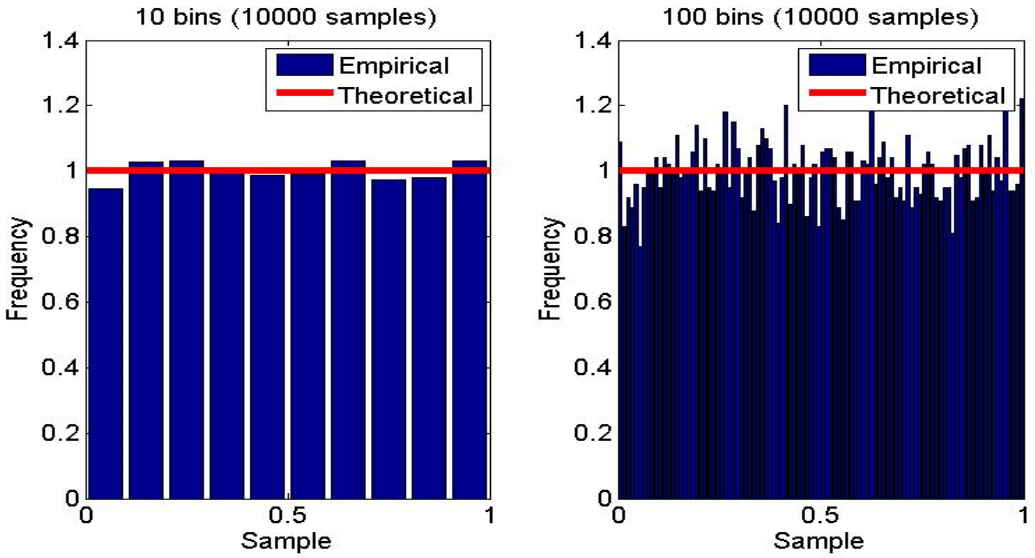

Figure 2. Histogram of the empirical and theoretical value

number of histogram bins increases, the sample mean fluctuates more, leading to an inaccurate analysis of the theoretical mean

1.1.5 The theoretical mean is set to be 0, and standard deviation set to 1. Compared to the mean and standard deviation of the 1000 samples, we can be see that the samples' mean and standard deviation is very close to the theoretical value. As number of sample increases, the sample mean and standard deviation will get closer to the theoretical value.

The result for the two numbers is shown below.

```
The sample mean is: 0.010476
The sample standard deviation is: 1.0131
>>
```

Regenerate ten 1000-samples and show their mean and standard deviation in histograms using similar method as in previous questions, shown in Figure 3 below

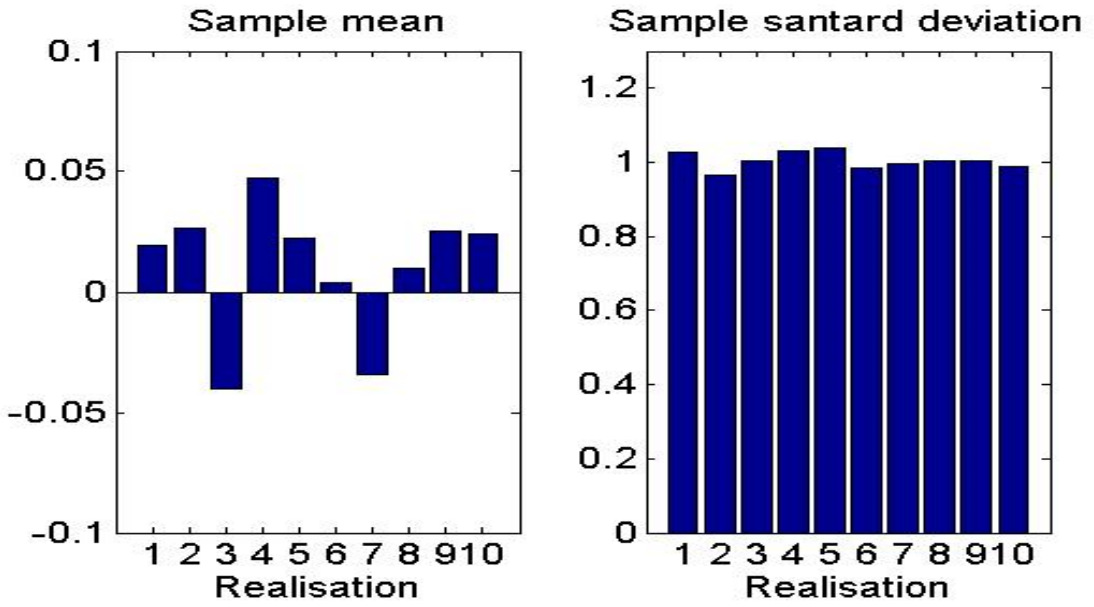

Figure 3. Realization of the sample mean and standard deviation of Guassian distribution

From the realization result, we can see it's similar to the previous question. Sample mean cluster around theoretical value of 0, and sample standard deviation cluster around theoretical value of 1. Again, as number of sample increases, the result can be more accurate.

Again, we analyse the result using probability density function (pdf), and standardize the sample pdf by scaling all bar value by a factor of (maximum of sample/maximum of theoretical). The resulted pdfs are shown in Figure 4. As we concluded previously, increased samples will make the sample result more accurate. Also, for this case, increased bar in the histogram will make the pdf of samples more accurate.

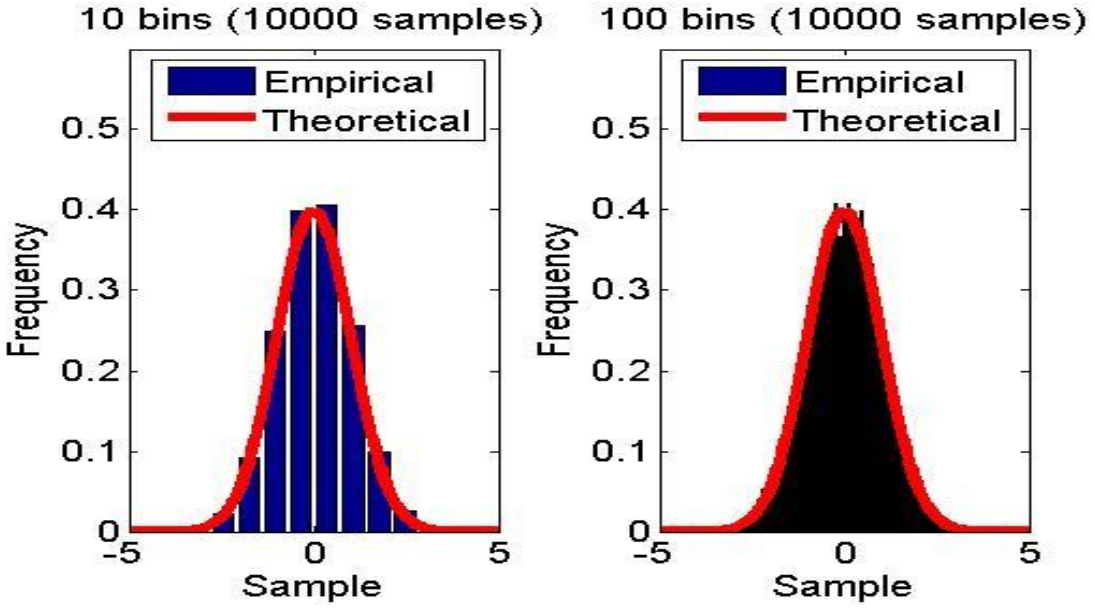

Figure 4. PDF of the empirical and theoretical value

1.2 Stochastic process

1.2.1 Store the 3 random process functions as rp1.m, rp2.m and rp3.m respectively, so we can call them directly from the same directory in the main function. Inputting 100 realizations of length-100 sample, and calculate the ensemble means and standard deviations. Plot the means and standard deviations against time, which is the realization index. As for this particular question, 100 is the number of member in the ensemble, so we put 1 to 100 as the time axis.

The results of the plots for all three processes are shown below in figure 5.

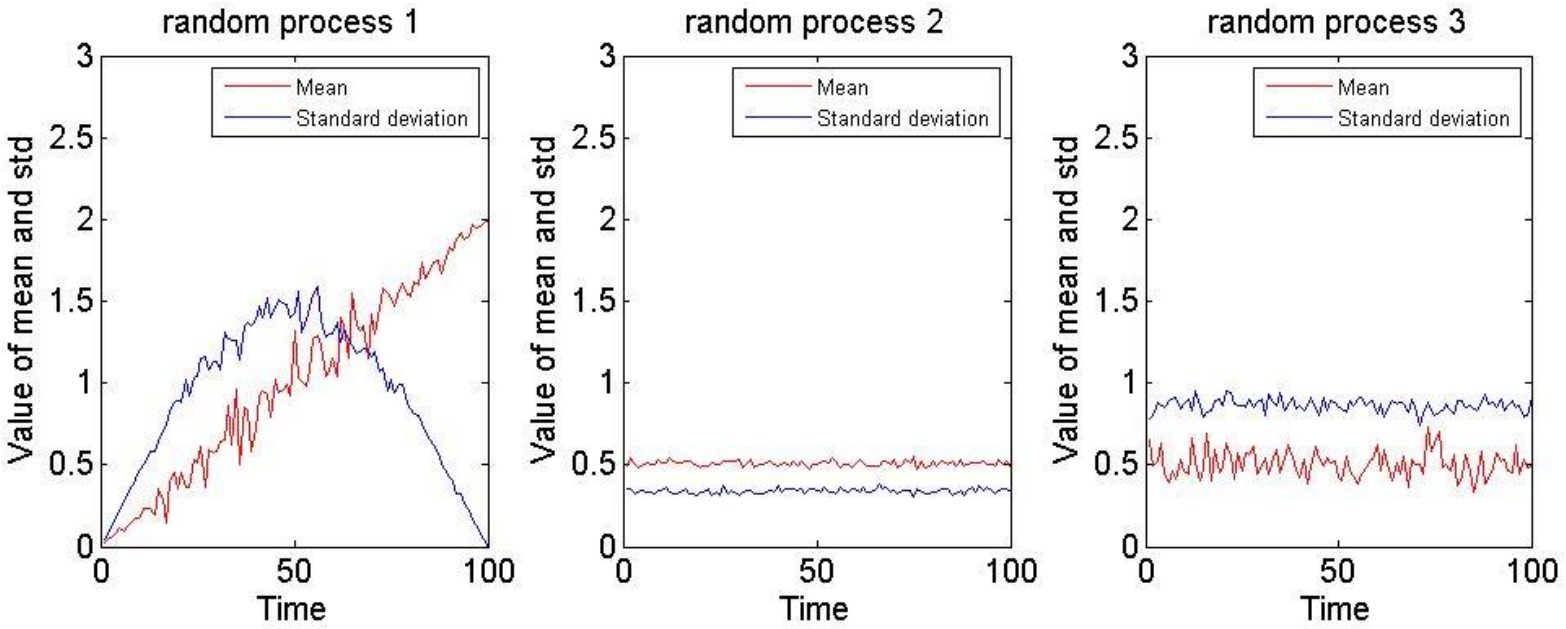

Figure 5. Standard deviation and mean of the 3 random processes

As we can see in the plots, red lines represent the means and blue lines represent the standard deviations. It is easy to tell that process 1 is not stationary as its mean value is increasing, while the other two processes have mean clustering around 0.5. For process 2 and 3, it is also observable that process 2 is more stable than process 3 as process 2 has a lower standard deviation of about 0.304，while process 3 has a higher standard deviation of about 0.8673 (calculated as means of the 100-member ensemble standard deviation of process 2 and 3). Such feature is also demonstrated by the larger fluctuation of the mean of process 3 in the plot.

As for process 1, the standard deviation changes as realization index changes. This is because the realization index N is used in a sine function to generate the multiplier Mc of original randomized samples. Hence, as N approaches the middle point, Mc is at its maximum of 1, and more fluctuation is expected. Such feature is also demonstrated by the larger fluctuation around the 50^{th} realization, and peaking of standard deviation.

1.2.2 Means and standard deviations are calculated for four realizations of all 3 processes. The result shown in the command window are shown below:

```
The means for the four realizations of rp1 are:
10.0334      9.98909       10.046       10.011
The standard deviation for the four realizations of rp1 are:
5.8996      5.8715      5.8849      5.8587
The means for the four realizations of rp2 are:
0.71965      0.2082      0.99977      0.70758
The standard deviation for the four realizations of rp2 are:
0.10771      0.18712      0.10638     0.096944
The means for the four realizations of rp3 are:
0.54963      0.49914      0.50198      0.49617
The standard deviation for the four realizations of rp3 are:
0.8716      0.87584      0.8656      0.86559
```

Ergodicity is decided by whether the process's theoretical mean can be approximated by time averages. The means for the 4 realizations are actually the time average. Comparing to the theoretical mean of all three processes respectively, it is clear that only rp3 (theoretical mean is 0.5) satisfy the condition of ergodicity. Hence only rp3 is an ergodic process.

1.2.3 For rp1, the process has an outcome of:

$Y = (X - 0.5) \times 5 \sin\frac{n\pi}{100} + 0.02n \qquad X \sim U(0,1)$

For rp2, the process has an outcome of:

$Y = (X - 0.5) \times X_1 + X_2 \qquad X, X_1, X_2 \sim U(0,1)$

For rp3, the process has an outcome of:

$Y = (X - 0.5) \times 3 + 0.5 \qquad X \sim U(0,1)$

From the mathematical expressions, we can calculate the theoretical mean and standard deviation for rp1, rp2 and rp3 respectively as theoretical mean (0.5) and standard (0.2887) deviation of X are already known to us in section 1.1

For rp1, theoretical mean is:

$E\{y\} = (E\{x\} - 0.5) \times 5 \sin\frac{n\pi}{100} + 0.02n = 0.02n$ (Since E{x}-0.5=0)

, which correspond to the plot of a linear line with gradient 0.02.

Theoretical standard deviation is:

$\sigma = \sqrt{Var\{y\}} = \sqrt{Var\left\{(x - 0.5) \times 5 \sin\frac{n\pi}{100}\right\}} = \sqrt{25\left(\sin\frac{n\pi}{100}\right)^2 Var\{x - 0.5\}} = 5\sigma_x \sin\frac{n\pi}{100} = 1.4415 \sin\frac{n\pi}{100}$

, which corresponds to the sine distribution of standard deviation of rp1.

It is trivial that standard deviation reaches maximum at n=50, the middle point on the plot, and this feature is reflected by the increased extent of fluctuation around the middle point of the sample mean plot.

For rp2, theoretical mean is:

$E\{y\} = E\{x - 0.5\} \times E\{x_1\} + E\{x_2\} = E\{x_2\} = 0.5$ (Since E{x}-0.5=0)

, which correspond to the plot of constant line clustering around the value of 0.5.

Theoretical standard deviation is:

$\sigma = \sqrt{Var\{y\}} = \sqrt{Var\{(x - 0.5) \times x_1\} + Var\{x_2\}} = \sqrt{E\{(x - 0.5)^2\}E\{{x_1}^2\} - \{E\{x - 0.5\}\}^2\{E\{x_1\}\}^2 + Var\{x_2\}} = \sqrt{E\{(x - 0.5)^2\}E\{{x_1}^2\} + Var\{x_2\}} = \sqrt{\{Var\{x - 0.5\} + E\{x - 0.5\}\}\{Var\{x\} + E\{x\}\} + Var\{x_2\}} = 0.3332$ (Since the three Xs are not correlated)

, which correspond to the plot of constant line clustering around the value of 0.3332.

Since rp2 has smaller standard deviation than rp3, the extent of fluctuation is less than rp3.

For rp3, theoretical mean is:
$E\{y\} = E\{x - 0.5\} \times 3 + 0.5 = 0.5$ (Since E{x}-0.5=0)
, which correspond to the plot of a constant line clustering around the value of 0.5.
Theoretical standard deviation is:
$\sigma = \sqrt{Var\{y\}} = \sqrt{Var\{(x - 0.5) \times 3\}} = 3\sqrt{Var\{x\}} = 0.8661$
, which correspond to the plot of a constant line clustering around the value of 0.8661.
The theoretical standard deviation is also illustrated by the larger extent of fluctuation of sample mean compared to that of rp2.

1.3 Estimation of probability distributions

1.3.1 An m-file named pdf is written to estimate the pdf of samples. Use hist() function to obtain the outcomes and their corresponding probability. Then use bar() to plot probability against outcome. The probability is scaled, so that the area under the estimation pdf graph approximates to 1. Result of testing the code with 100, 10000, 1000000 data length of a Gaussian data array is shown below in Figure 6:

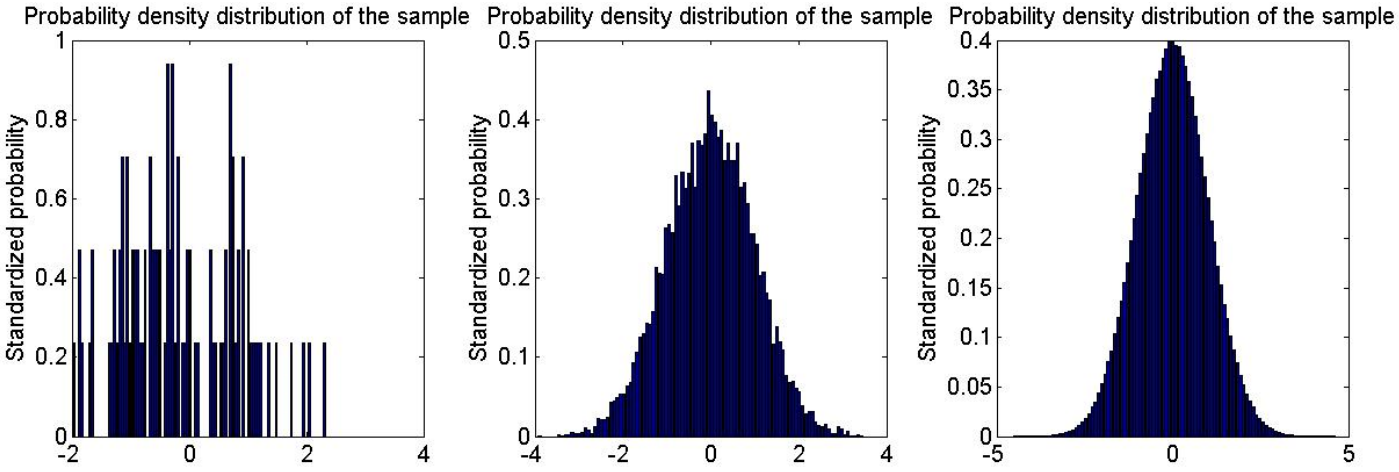

Figure 6. Testing pdf with different-sample-number Gaussian process

As we can see, when number of samples is small, 100 for instance, the pdf is just numbers of spikes with different values, pattern of Gaussian distribution can't be seen at all. This is because the samples coincide at most 4 times, so spikes with 4 discrete levels are seem in the graph.

While number of samples increases to 10000, a shape of Gaussian distribution is almost seem, with discrete levels observed. This is due to the fact that samples start to have the same value, but variation in number of members at each sample value is not as large.

A nice Gaussian pdf distribution is seem as number of sample increases to 1000000, as increased number of sample approximates sample properties to theoretical properties.

1.3.2 The only processes that is stationary and ergodic is rp3. Since the pdf function is already standardized using the method of dividing the number of occurrence at different value with total number of occurrence, the histogram we obtained with 100 bins is a good approximation of the process's pdf. Due to the fact that the area under a pdf is always 1, and rp3 has its samples occurring from -1 to 2 with equal probabilities. We can easily know that the theoretical pdf of rp3 is simply a constant line of 0.3333 from -1 to 2.

The plots of empirical and theoretical pdf of rp3 with number of samples 100, 1000, 10000 are shown below in Figure 7:

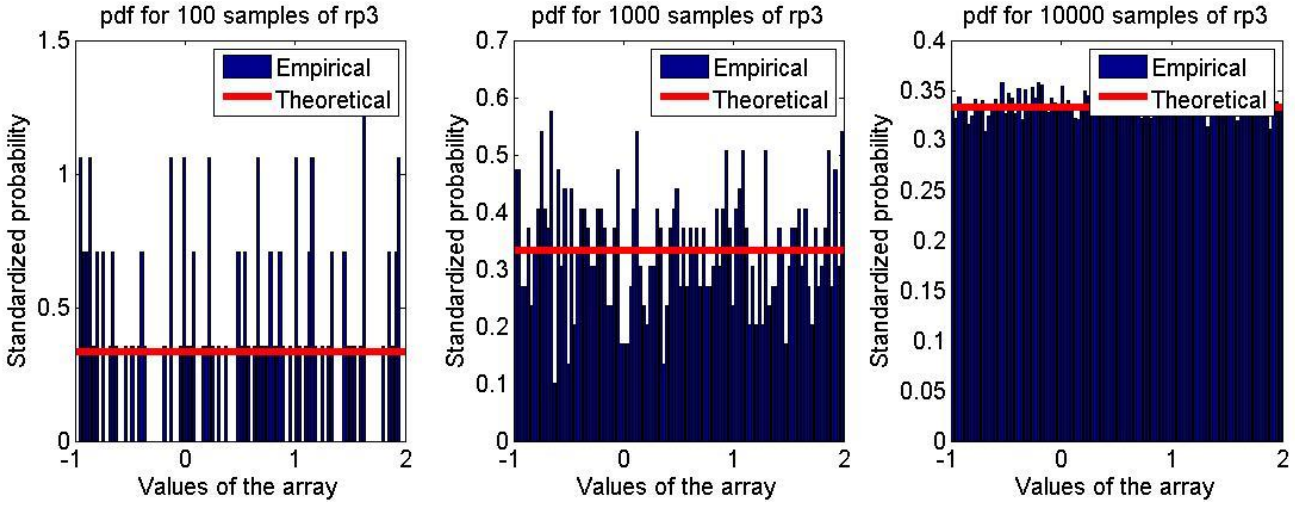

Figure 7. Empirical and theoretical pdf of rp3 with 100, 1000, 10000 samples

It is clear that as number of samples in rp3 increases, the empirical pdf of rp3 gets closer to its theoretical pdf.

1.3.3 The non-stationary process is rp1, and it is not possible to estimate the pdf of rp1 using the pdf function we created. It is because the outcome value of rp1 is not time independent. A second dimension of N is used to simulate the time effect. If a pdf is to be plotted for rp1, a third axis in the z direction should be introduced to reflect the change in N/time, which means the pdf of rp1 varies with time. Hence the pdf of a non-stationary process would have three dimensions, and function pdf can't deal with the third dimension.

Now, we look into an rp1 signal with N=1000 samples. As the mean changes from 0 to 1, N increases from 0 to 500, since theoretical mean of rp1 is 0.02N. However, the standard deviation of the signal increases from 0 to 1.4415 as we calculated in part 1.2, which means the mean of the signal is likely to fluctuate largely, and the sample mean we take can't be a good estimation of theoretical mean.

If we want to compute the pdf of rp1, we can plot the pdf of rp1 at different time/N, so that a set of N pdf plots are obtained. After that, we can stack them together in the z direction, which is a measure of N. In this way, a three dimension pdf plot is obtained, and information about rp1 signal is reflected.