1 Random Signals and Stochastic Processes

1.1 Statistical estimation

A 1000-sample realization, $\mathbf{x}$, of process $X_n \sim \mathcal{U}(0,1) \forall n$, was generated using the `rand` function.

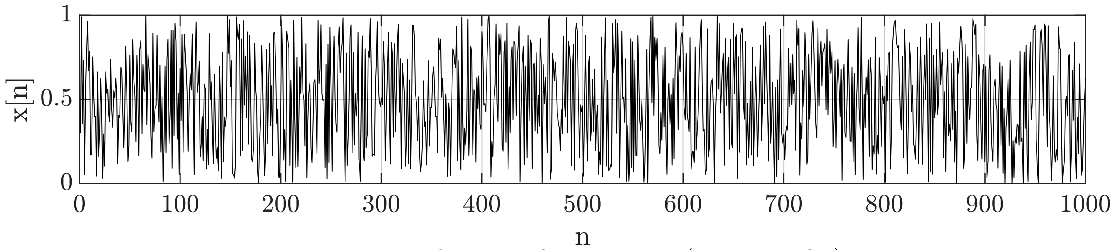

Figure 1.1: Realization of Process X_n (1000 samples).

We will use the sample mean of $\mathbf{x}$, $\widehat{m}(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N} x[n]$, to estimate the theoretical mean, $m = E\{X\}$.

First, we calculate the expected value and standard deviation of $X \sim \mathcal{U}(0,1),$ with $f_X(x) = u(x) - u(x-1)$.[1]

$$m = E\{X\} = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x\ dx = \frac{1}{2} = 0.5 \tag{1}$$

$$\sigma = \sqrt{Var\{X\}} = \sqrt{\int_{-\infty}^{\infty}(x-m)^2 f_X(x)dx} = \sqrt{\int_0^1 (x-0.5)^2 dx} = \sqrt{\frac{1}{3}\Bigg[\ (x-0.5)^3\ \Bigg]_0^1} = \sqrt{\frac{1}{12}} \approx 0.2887 \tag{2}$$

Using the MATLAB `mean` function, the sample mean is: $\widehat{m}(\boldsymbol{x}) \approx 0.4914$. Thus, $B(\widehat{m}(\boldsymbol{x})) = m - \widehat{m}(\boldsymbol{x}) \approx 0.0086$.

We will evaluate the accuracy of estimator $\widehat{m} = \frac{1}{N}\sum_{n=1}^{N} x[n]$ in determining the expected value of a realized variable X with arbitrary pdf, expected value m and standard deviation σ, in terms of three properties.

1. Consistency (Does an infinitely large sample result in perfect estimation of m?)

   By the law of large numbers, $\lim\limits_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} x[n] = E\{X\} \implies \lim\limits_{N\to\infty} \widehat{m} = m \implies \widehat{m}$ is *consistent.*

2. Bias (How big is the expected difference between $\widehat{m}$ and m?)

   $Bias(\widehat{m}) = E\{m - \widehat{m}\} = E\{m - \frac{1}{N}\sum_{n=1}^{N} x[n]\} = m - \frac{1}{N}\sum_{n=1}^{N} m = 0 \implies \widehat{m}$ is an *unbiased* estimator of m.

3. Standard Deviation (How spread out are the values of $\widehat{m}$ from $E\{\widehat{m}\}$?)

   $$\sigma_{\widehat{m}}^2 = Var\{\widehat{m}\} = E\{(\widehat{m} - E\{\widehat{m}\})^2\} = E\left\{\left(\tfrac{1}{N}\sum_{n=1}^{N} x[n] - m\right)^2\right\} = \tfrac{1}{N^2}E\left\{\left[\sum_{n=1}^{N}(x[n]-m)\right]^2\right\} =$$

   $$= \tfrac{1}{N^2}\sum_{n=1}^{N}\sum_{k=1}^{N} E\left\{(x[n]-m)(x[k]-m)\right\} = \tfrac{1}{N^2}\sum_{n=1}^{N}\sum_{k=1}^{N} Cov\{x[n],x[k]\} \tag{3}$$

   Then, since $\{x[1], x[2],\ \dots\ , x[N]\}$ are independent realizations of random variable X:

   $$Cov\{x[n],x[k]\} = \begin{cases} \sigma^2 & ,\ n = k \\ 0 & ,\ otherwise \end{cases} \implies Var\{\widehat{m}\} = \frac{1}{N^2}\sum_{n=1}^{N} Var\{X\} = \frac{\sigma^2}{N} \implies \sigma_{\widehat{m}} = \frac{\sigma}{\sqrt{N}} \tag{4}$$

   The standard deviation of the sample mean is inversely proportional to the square root of the sample size.

   For the vector $\mathbf{x}$ generated above, $N = 1000,\ \sigma \approx 0.2887 \Rightarrow \sigma_{\widehat{m}} \approx \frac{0.2887}{\sqrt{1000}} \approx 0.0091$.

   The sample mean $\widehat{m}(\boldsymbol{x})$ obtained, 0.4914, falls within one standard deviation, $\sigma_{\widehat{m}}$, of $E\{\widehat{m}\}$, 0.5.

We will use the sample standard deviation of $\mathbf{x}$, $\widehat{\sigma}(\boldsymbol{x}) = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x[n] - \widehat{m}(\boldsymbol{x}))^2}$ to estimate σ.

For $X \sim \mathcal{U}(0,1)$ we have already calculated $m = 0.5, \sigma \approx 0.2887$.

The sample standard deviation of $\mathbf{x}$, was obtained using the MATLAB `std` function: $\widehat{\sigma}(\boldsymbol{x}) \approx 0.2832$.

Hence, the estimation bias is: $B(\widehat{\sigma}(\boldsymbol{x})) = \sigma - \widehat{\sigma}(\boldsymbol{x}) \approx 0.0055$.

We will evaluate the accuracy of estimator $\widehat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x[n] - \widehat{m})^2}$ in determining σ for a realized variable X with arbitrary pdf, expected value m and standard deviation σ, in terms of three properties.

1. Consistency (Does an infinitely large sample result in perfect estimation of σ?)

   $$\widehat{\sigma}^2 = \tfrac{1}{N-1}\sum_{n=1}^{N}(x[n] - \widehat{m})^2 = \tfrac{N}{N-1}\left(\tfrac{1}{N}\sum_{n=1}^{N} x[n]^2 - \widehat{m}^2\right) \tag{5}$$

   By the law of large numbers $\lim\limits_{N\to\infty} \widehat{m} = E\{X\} \implies \lim\limits_{N\to\infty} \widehat{m}^2 = E\{X\}^2$ and $\lim\limits_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} x[n]^2 = E\{X^2\}$.

   Thus, $\lim\limits_{N\to\infty} \widehat{\sigma} = \sqrt{\lim\limits_{N\to\infty} \widehat{\sigma}^2} = \sqrt{\lim\limits_{N\to\infty} \frac{N}{N-1}\sigma^2} = \sigma$. Hence, $\widehat{\sigma}$ is *consistent.*

[1] $u(x)$ will be used to denote the unit step function throughout this report

2. Bias (How big is the expected difference between $\widehat{\sigma}$ and σ?)

   Before evaluating the bias of estimator $\widehat{\sigma}$ in determining σ we will evaluate the bias of $\widehat{\sigma}^2$ in determining σ^2:

   $$E\left\{\widehat{\sigma}^2\right\} = E\left\{\frac{1}{N-1}\sum_{n=1}^{N}(x[n]-\widehat{m})^2\right\} = \frac{1}{N-1}E\left\{\sum_{n=1}^{N}x[n]^2 - 2\sum_{n=1}^{N}(\widehat{m}\cdot x[n]) + \sum_{n=1}^{N}\widehat{m}^2\right\} =$$

   $$= \frac{1}{N-1}\left(\sum_{n=1}^{N}E\left\{x[n]^2\right\} - E\left\{2N\widehat{m}^2\right\} + E\left\{N\widehat{m}^2\right\}\right) = \frac{1}{N-1}\left(N(\sigma^2+m^2)\text{-}N(\sigma_{\widehat{m}}^2+m^2)\right) = \sigma^2 \qquad (6)$$

   Hence, $Bias(\widehat{\sigma}^2) = E\{\sigma^2 - \widehat{\sigma}^2\} = \sigma^2 - E\{\widehat{\sigma}^2\} = 0$ and $\widehat{\sigma}^2$ is an *unbiased* estimator of σ^2. However as the square root is not a linear operator, $E\{\widehat{\sigma}^2\} = \sigma^2 \not\Rightarrow E\{\widehat{\sigma}\} = \sigma$ and hence we *cannot* conclude that $\widehat{\sigma}$ is an unbiased estimator. *The bias of the estimator depends on the distribution of X.*

3. Standard Deviation (How spread out are the values of $\widehat{\sigma}$ from $E\{\widehat{\sigma}\}$?)

   As with the bias, *the standard deviation of the estimator depends on the distribution of X.*

To obtain the bias and standard deviation for $X \sim \mathcal{U}(0,1)$, we use MATLAB to generate M, N-sample vectors $\{\boldsymbol{x_1},\ \ldots\ \boldsymbol{x_M}\}$ and obtain the vector of sample standard deviations $\boldsymbol{s_N} = [\widehat{\sigma}_N(\boldsymbol{x_1}),\ \ldots, \widehat{\sigma}_N(\boldsymbol{x_M})]$.Since $\widehat{m}$ and $\widehat{\sigma}$ are consistent estimators, for sufficiently large M: $\widehat{m}(\boldsymbol{s_N}) \to E\{\widehat{\sigma}_N\}$ and $\widehat{\sigma}(\boldsymbol{s_N}) \to \sigma_{\widehat{\sigma}_N}$. By repeating this for a range of N, we produce estimates of the $Bias(\widehat{\sigma}_N)$ vs N, $\frac{Bias(\widehat{\sigma}_N)}{\sigma}$ vs N and $\sigma_{\widehat{\sigma}_N}$ vs N plots.

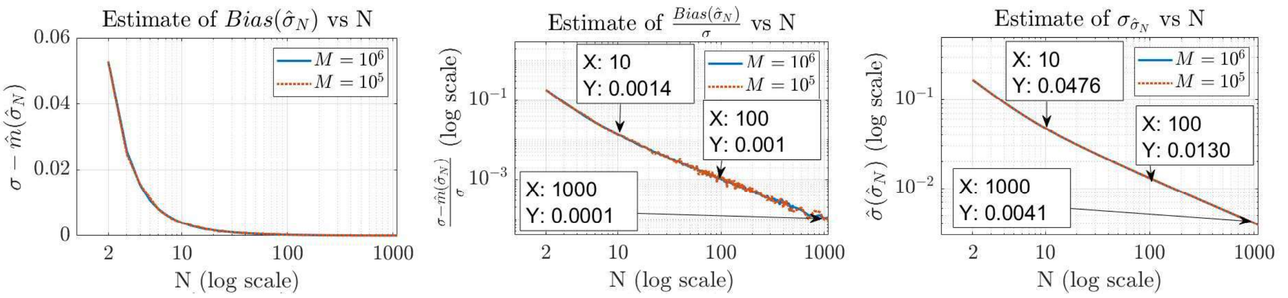

Figure 1.2: Estimate of $Bias(\widehat{\sigma}_N)$ and $\sigma_{\widehat{\sigma}_N}$ for $X \sim \mathcal{U}(0,1)$

Hence, $\widehat{\sigma}$ is a *biased estimator*. However, as N increases $Bias(\widehat{\sigma})$ becomes less significant. In fact $N \geq 10$ results in $Bias(\widehat{\sigma}) < 0.015\sigma$, $N \geq 100$ in $Bias(\widehat{\sigma}) < 0.001\sigma$ and $N \geq 1000$ in $Bias(\widehat{\sigma}) < 0.0001\sigma$. The standard deviation of the sample standard deviation decreases as N increases. For the vector $\mathbf{x}$ generated above, $N = 1000, \sigma \approx 0.2887$ implies that $E\{\widehat{\sigma}\} \approx 0.9999\sigma \approx 0.2887$ and $\sigma_{\widehat{\sigma}} \approx 0.0041$. Thus, $\widehat{\sigma}(\boldsymbol{x})$ obtained, 0.2832, falls within $2\sigma_{\widehat{\sigma}}$, of $E\{\widehat{\sigma}\}$.

We generate ten 1000-sample realizations and use $\{\widehat{m}(\boldsymbol{x_1})..., \widehat{m}(\boldsymbol{x_{10}})\}$ and $\{\widehat{\sigma}(\boldsymbol{x_1})..., \widehat{\sigma}(\boldsymbol{x_{10}})\}$ to estimate m and σ.

Table 1.1: Sample mean $\widehat{m}(\boldsymbol{x_i})$ and corresponding bias $B(\widehat{m}(\boldsymbol{x_i})) = 0.5 - \widehat{m}(\boldsymbol{x_i})$

$\widehat{m}(\boldsymbol{x_i})$	0.5036	0.5005	0.5068	0.4988	0.5118	0.4955	0.5055	0.5060	0.5123	0.4914
$B(\widehat{m}(\boldsymbol{x_i}))$	-0.0036	-0.0005	-0.0068	0.0012	-0.0118	0.0045	-0.0055	-0.0060	-0.0123	0.0086

max$|B(\widehat{\sigma}(\boldsymbol{x_i}))|$=0.0123.8 out of 10 sample means are within one $\sigma_{\widehat{m}}$ of $E\{\widehat{m}\}$. ($E\{\widehat{m}\} = m = 0.5$, $\sigma_{\widehat{m}} \approx 0.0091$)

Table 1.2: Sample standard deviation $\widehat{\sigma}(\boldsymbol{x_i})$ and corresponding bias $B(\widehat{\sigma}(\boldsymbol{x_i})) \approx \sigma - \widehat{\sigma}(\boldsymbol{x_i})$.

$\widehat{\sigma}(\boldsymbol{x_i})$	0.2859	0.2876	0.2848	0.2820	0.2848	0.2933	0.2898	0.2921	0.2895	0.2922
$B(\widehat{\sigma}(\boldsymbol{x_i}))$	0.0028	0.0011	0.0039	0.0067	0.0039	-0.0047	-0.0011	-0.0034	-0.0008	-0.0035

max$|B(\widehat{\sigma}(\boldsymbol{x_i}))|$=0.0067. 8 out of ten 10 values are within one $\sigma_{\widehat{\sigma}}$ of $E\{\widehat{\sigma}\}$. ($E\{\widehat{\sigma}\}$≈0.9999σ≈0.2887, $\sigma_{\widehat{\sigma}}$≈0.0041)

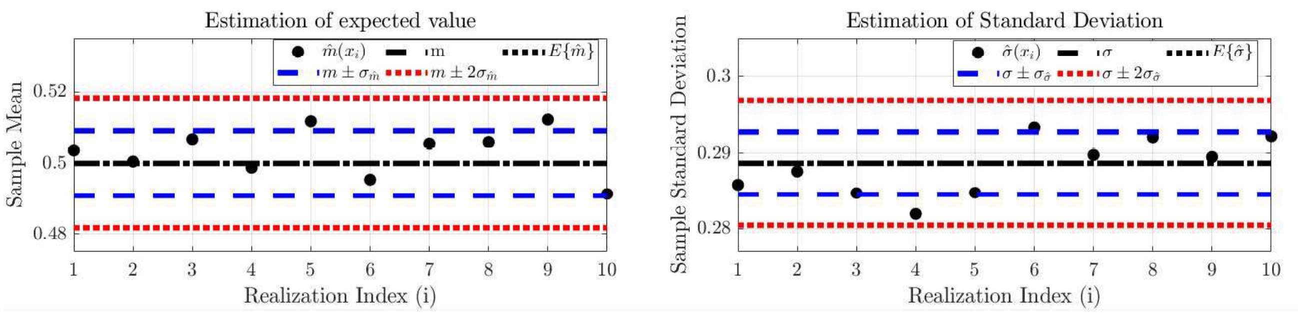

Figure 1.3: Plot of sample mean and sample standard deviation of each vector $\boldsymbol{x_i}$

We will attempt to estimate the pdf of X ($f_X(x)$) using a histogram of $\mathbf{x}$ and normalizing by the total area. The histogram depends on two parameters, the number of samples considered, N, and the number of bins, ν. Hence we represent the pdf estimator as $\widehat{f_{N,\nu}}$. For each N-sample vector $\mathbf{x}$, let us denote the range $r(\boldsymbol{x}) = max(\boldsymbol{x}) - min(\boldsymbol{x})$, the width of each bin $w = \frac{r(\boldsymbol{x})}{\nu}$, the height of each bin $h[i]$ and the beginning of each bin, $b[i] = min(\boldsymbol{x}) + (i-1)w$. Note that by definition $w\sum_{i=1}^{\nu} h[i] = 1$. The pdf estimate obtained from $\mathbf{x}$ can, therefore, be represented by:

$$\widehat{f_{N,\nu}}(x) = \sum_{i=1}^{\nu} \left[\ h[i] \cdot \big(u(x - b[i]) - u(x - b[i] - w)\big)\ \right] \tag{7}$$

To assess estimation accuracy we use Integrated Squared Error: $ISE(\widehat{f_{N,\nu}}) = \int_{-\infty}^{\infty} \left(f_X(x) - \widehat{f_{N,\nu}}(x)\right)^2 dx$

$$ISE(\widehat{f_{N,\nu}}) = \int_{-\infty}^{\infty} \left(\widehat{f_{N,\nu}}(x)^2 - 2\widehat{f_{N,\nu}}(x) f_X(x) + f_X(x)^2\right) dx = w\sum_{i=1}^{\nu} h[i]^2 - 2\sum_{i=1}^{\nu}\left(h[i]\left[F_X(x)\right]_{b[i]}^{b[i]+w}\right) + \int_{-\infty}^{\infty} f_X(x)^2 dx \tag{8}$$

$X \sim \mathcal{U}(0,1), f_X(x) = u(x) - u(x-1) \Rightarrow$

$$ISE(\widehat{f_{N,\nu}}) = w\sum_{i=1}^{\nu} h[i]^2 - 2\sum_{i=1}^{\nu}(h[i] \cdot w) + 1 = w\sum_{i=1}^{\nu} h[i]^2 - 1 \text{ and } ISE(\widehat{f_{N,1}}) = w \cdot h[1]^2 - 1 = \frac{1}{w} - 1 = \frac{1}{r(\boldsymbol{x})} - 1 \tag{9}$$

By the Cauchy-Schwarz inequality: $ISE(\widehat{f_{N,\nu}}) \geq w\left(\sum_{i=1}^{\nu} h[i]\right)^2 - 1 = w\frac{1}{w^2} - 1 = \frac{\nu}{r(\boldsymbol{x})} - 1 \geq ISE(\widehat{f_{N,1}})$ (10)

Hence, the error is minimized for ν=1. This is true when estimating a uniform distribution, as we are effectively just estimating the lower and upper bounds. Generally, the optimal value of ν depends both on N and $f_X(x)$. To demonstrate the effects of ν on estimating the uniform pdf, we plot histograms for N=1000 and $\nu = \{1, 20, 400\}$

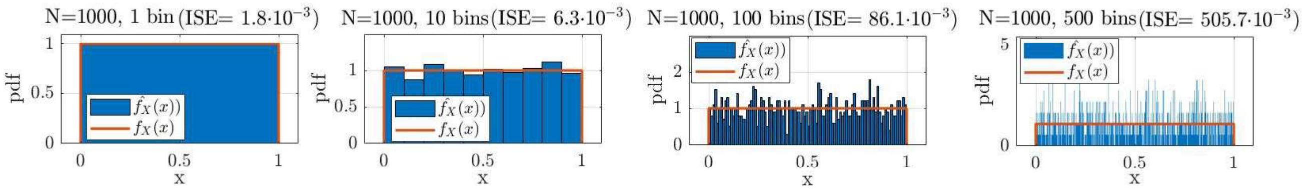

Figure 1.4: Estimation of $f_X(x)$ for $N = 1000$, $\nu = \{1,\ 10, 100, 500\}$. The error increases with increasing ν.

To investigate the effects of N on estimation accuracy, we generate M vectors $\mathbf{x}$ for each size N and calculate their estimation error. For large enough M: $\widehat{m}(ISE(\widehat{f_{N,\nu}})) \to E\{ISE(\widehat{f_{N,\nu}})\}$ and $\widehat{\sigma}()ISE(\widehat{f_{N,\nu}})) \to \sigma_{ISE(\widehat{f_{N,\nu}})}$.

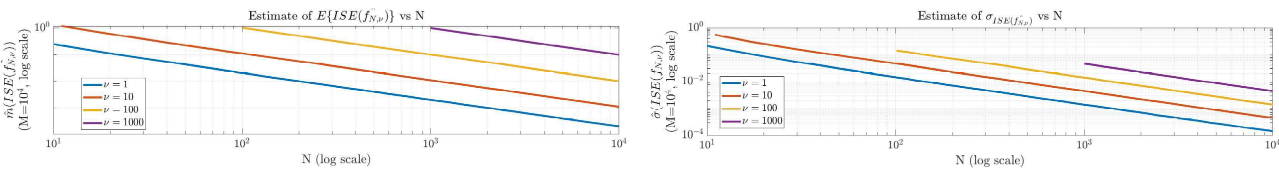

Figure 1.5: Estimation of the error expected value and standard deviation for $\nu = \{1,\ 10, 100,\ 1000\}$ and $N < 10^4$.

The slopes of both the expected value plot and standard deviation plot are approximately -1, for all ν used. Hence, the expected value and standard deviation scale with N^{-1} and for the values of ν examined the estimate converges, i.e. $N \to \infty \Rightarrow \widehat{f_{N,\nu}} \to f_X$. To illustrate this we plot 10-bin histograms for N=$\{100, 1000, 10000\}$. From previous analysis we obtained the expected value and standard deviation of the error to compare with our results.

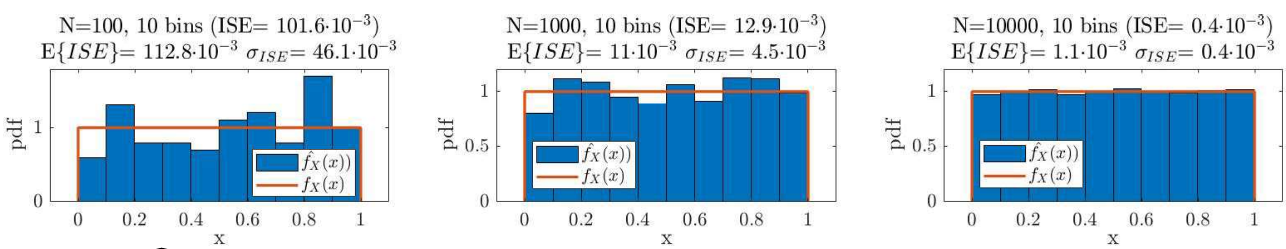

Figure 1.6: $\widehat{f_X(x)}$ for $\nu = 10$, $N = \{10^2,\ 10^3, 10^4\}$. Plot converges as N increases. $|ISE - E\{ISE\}| < 2\sigma_{ISE}$ in all cases.

We generate a 1000-sample realization, $\mathbf{x}$, of $X_n \sim \mathcal{N}(0,1) \forall n$ using the `randn` function.

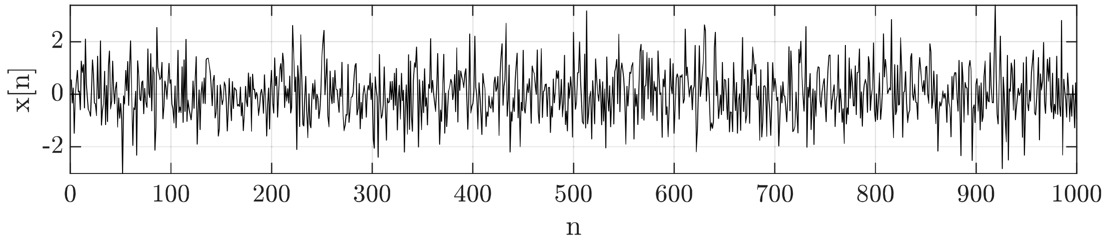

Figure 1.7: Realization of Process X_n (1000 samples).

$m = 0$ and $\sigma = 1$. The sample mean of $\mathbf{x}$, was: $\widehat{m}(\boldsymbol{x}) \approx -0.0039$. Thus, $B(\widehat{m}(\boldsymbol{x})) = m - \widehat{m}(\boldsymbol{x}) \approx 0.0039$.

The properties used to evaluate the accuracy of $\widehat{m}$ are not dependent on the distribution of X, hence we do not need to repeat our analysis. For vector $\mathbf{x}$ generated, with N=1000, $E\{\widehat{m}\}$=m=0 and $\sigma_{\widehat{m}} = \frac{\sigma}{\sqrt{N}} = \frac{1}{\sqrt{1000}} \approx 0.0316$. Thus, the sample mean obtained is within one standard deviation of m.

The sample standard deviation of $\mathbf{x}$, was obtained using the MATLAB `std` function: $\widehat{\sigma}(\boldsymbol{x}) \approx 1.0247$

This means that the sample standard deviation of $\mathbf{x}$ has estimation bias: $B(\widehat{\sigma}(\boldsymbol{x})) = \sigma - \widehat{\sigma}(\boldsymbol{x}) \approx -0.0247$.

The expected value and standard deviation of $\widehat{\sigma}$ depend on the distribution of X. For $X \sim \mathcal{N}(0,1)$, we generate M, N-sample vectors $\{\boldsymbol{x_1 ..., x_M}\}$ and obtain the vector of sample standard deviations $\boldsymbol{s_N} = [\widehat{\sigma}_N(\boldsymbol{x_1}) ..., \widehat{\sigma}_N(\boldsymbol{x_M})]$. For large M: $\widehat{m}(\boldsymbol{s_N}) \rightarrow E\{\widehat{\sigma}_N\}$, $\widehat{\sigma}(\boldsymbol{s_N}) \rightarrow \sigma_{\widehat{\sigma}_N}$. Repeating for a range of N, we estimate $Bias(\widehat{\sigma}_N)$ and $\sigma(\widehat{\sigma}_N)$ vs N.

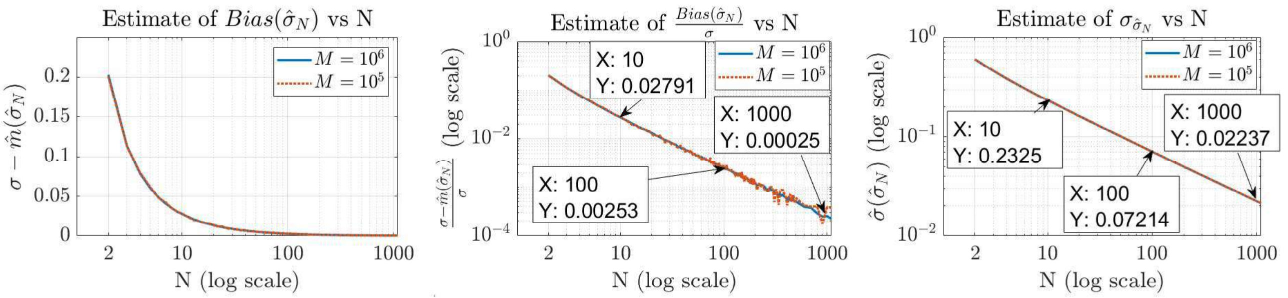

Figure 1.8: Plots produced by estimating $E\{\widehat{\sigma}_N\}$ and $\sigma_{\widehat{\sigma}}$ using MATLAB

$\widehat{\sigma}$ is a *biased estimator*. As N increases, $Bias(\widehat{\sigma})$ decreases. N≥10⇒$Bias(\widehat{\sigma})<0.03\sigma$, N≥100⇒$Bias(\widehat{\sigma})<0.003\sigma$ and N≥1000 ⇒$Bias(\widehat{\sigma})<0.0003\sigma$. The standard deviation of $\widehat{\sigma}$ decreases as N increases. For $\mathbf{x}$, N=1000, σ=1⇒$E\{\widehat{\sigma}\}\approx 0.99975\sigma \approx 1$ and $\sigma_{\widehat{\sigma}} \approx 0.02237$. $\widehat{\sigma}(\boldsymbol{x})$ obtained, 1.0247, within two standard deviations, $\sigma_{\widehat{\sigma}}$, of $E\{\widehat{\sigma}\}$.

We generate ten 1000-sample realizations and use $\{\widehat{m}(\boldsymbol{x_1}) ..., \widehat{m}(\boldsymbol{x_{10}})\}$ and $\{\widehat{\sigma}(\boldsymbol{x_1}) ..., \widehat{\sigma}(\boldsymbol{x_{10}})\}$ to estimate m and σ.

Table 1.3: Sample mean $\widehat{m}(\boldsymbol{x_i})$ and corresponding bias $B(\widehat{m}(\boldsymbol{x_i})) = m - \widehat{m}(\boldsymbol{x_i})$.

$\widehat{m}(\boldsymbol{x_i})$	0.0318	0.0246	0.0107	0.0183	0.0307	-0.0192	-0.0367	0.0175	-0.0287	-0.0280
$B(\widehat{m}(\boldsymbol{x_i}))$	-0.0318	-0.0246	-0.0107	-0.0183	-0.0307	0.0192	0.0367	-0.0175	0.0287	0.0280

max($B(\widehat{m}(\boldsymbol{x_i}))$)=0.0367. 8 out of 10 sample means are within one $\sigma_{\widehat{m}}$ of $E\{\widehat{m}\}$. ($E\{\widehat{m}\} = m = 0$, $\sigma_{\widehat{m}} \approx 0.0316$).

Table 1.4: Sample standard deviation $\widehat{\sigma}(\boldsymbol{x_i})$ and corresponding bias $B(\widehat{\sigma}(\boldsymbol{x_i})) \approx \sigma - \widehat{\sigma}(\boldsymbol{x_i})$

$\widehat{\sigma}(\boldsymbol{x_i})$	1.0602	1.0309	0.9840	0.9926	0.9843	0.9890	0.9663	1.0021	1.0347	1.0335
$B(\widehat{\sigma}(\boldsymbol{x_i}))$	-0.0602	-0.0309	0.0160	0.0074	0.0157	0.0110	0.0337	-0.0021	-0.0347	-0.0335

max($B(\widehat{\sigma}(\boldsymbol{x_i}))$)=0.0602. 5 out of 10 values are within one $\sigma_{\widehat{\sigma}}$ of $E\{\widehat{\sigma}\}$. ($E\{\widehat{\sigma}\} \approx 0.99975\sigma \approx \sigma$, $\sigma_{\widehat{\sigma}} \approx 0.0224$)

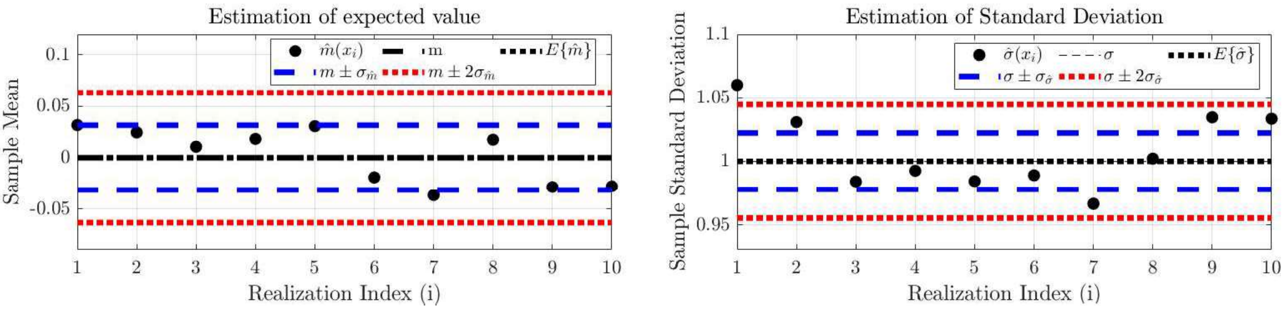

Figure 1.9: Plots produced by estimating $E\{\widehat{\sigma}_N\}$ and $\sigma_{\widehat{\sigma}}$ using MATLAB

We will again estimate the pdf of X (in this case $\phi(x)$) by plotting the histogram of $\mathbf{x}$. Since $X \sim \mathcal{N}(0,1)$:

$$ISE(\widehat{f_{N,\nu}}) = w \sum_{i=1}^{\nu} h[i]^2 - 2\sum_{i=1}^{\nu}\Big[h[i]\Big(\Phi(b[i]+w) - \Phi(b[i])\Big)\Big] + \int_{-\infty}^{\infty}\phi(x)^2 dx \text{ (note that}^2) \quad (11)$$

To determine the relationship between N and ν that minimizes ISE, we generate M, N-sample vectors $\{\boldsymbol{x_1}..., \boldsymbol{x_M}\}$ and create for each of them histograms with different numbers of bins. Then we calculate the value of ν that minimizes ISE for each of them. This results in a vector of optimal values of ν: $\boldsymbol{\nu_{opt_N}} = [\nu_{opt_N}(\boldsymbol{x_1}), ..., \nu_{opt_N}(\boldsymbol{x_{10}})]$. For large M, $\widehat{m}(\boldsymbol{\nu_{opt_N}})$ will be the value of ν that optimizes pdf estimation from a sample of size N. We compare this to the theoretical prediction deduced by using Scott's rule for histogram bin width[3], $w_{opt_N} = \frac{3.5\widehat{\sigma}}{N^{1/3}}$, and the fact that for $\mathbf{x}$ =[x[1]...,x[N]]^T, with $x[n] \sim \mathcal{N}(0,1) \forall n$: $E\{\max(\boldsymbol{x})\text{-}\min(\boldsymbol{x})\} = \int_{-\infty}^{\infty}(1-(1-\Phi(x))^N - \Phi(x)^N)dx$. [4]

Then, $$\nu_{opt_N}(predicted) = \frac{1}{w_{opt}} E\{max(\boldsymbol{x}) - min(\boldsymbol{x})\} = \frac{1}{3.5} N^{\frac{1}{3}} \int_{-\infty}^{\infty}(1-(1-\Phi(x))^N - \Phi(x)^N)dx \quad (12)$$

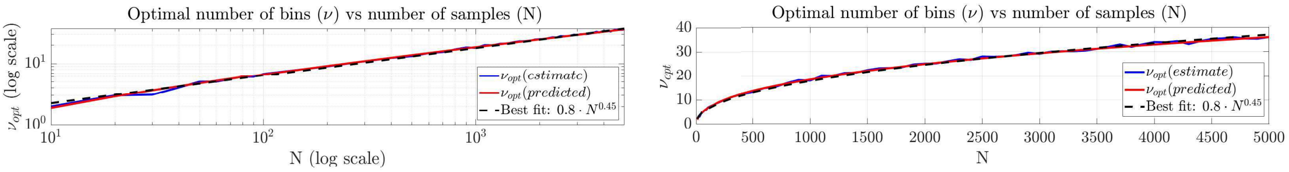

Figure 1.10: Plots of of optimal number of bins vs N. The prediction matches $\widehat{m}(\boldsymbol{\nu_{opt_N}})$.

The above plots show that the optimal number of bins increases with N. Excel uses $\nu \approx \sqrt{N}$ when plotting a histogram. Linear regression on the log-log plot, yields $\nu_{opt_N} \approx 0.8N^{0.45}$, which is consistent with this practice. To demonstrate the effects of ν on estimating $\phi(x)$, we plot estimated pdfs for N=1000, ν=$\{1, 15, 30, 100\}$.

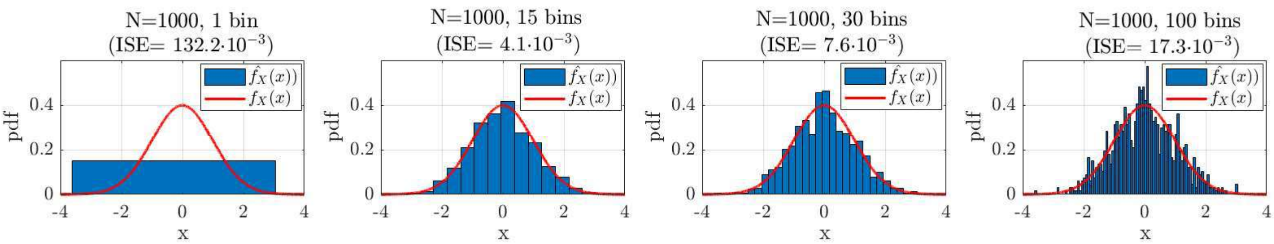

Figure 1.11: Estimation of $\phi(x)$ for N=100, $\nu = \{1, 15, 30, 100\}$. The error is minimized close to $0.8 \cdot 1000^{0.45} \approx 18$

To investigate the effects of N on estimation accuracy, we generate M vectors $\mathbf{x}$ for each size N and calculate their estimation error. For large enough M: $\widehat{m}(ISE(\widehat{f_{N,\nu}})) \rightarrow E\{ISE(\widehat{f_{N,\nu}})\}$ and $\widehat{\sigma}()ISE(\widehat{f_{N,\nu}})) \rightarrow \sigma_{ISE(\widehat{f_{N,\nu}})}$.

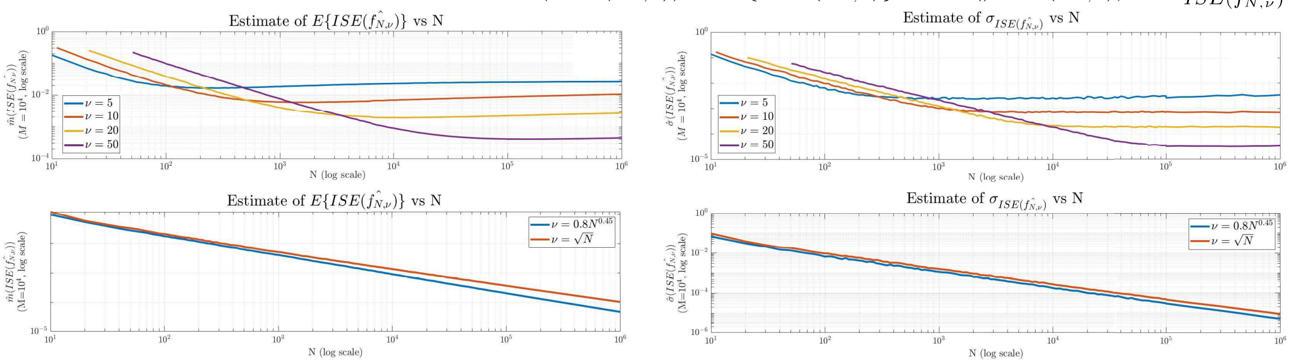

Figure 1.12: Estimate of the error expected value and standard deviation for $\nu = \{5, 10, 20, 50, 0.8N^{0.45}, \sqrt{N}\}$ and N<10^6.

Thus, for constant ν, $E\{ISE\}$ and σ_{ISE} do not tend to 0 as $N \rightarrow \infty$. For $\nu = 0.8N^{0.45}$ or $\nu = \sqrt{N}$, the slope of the log-log plot for both $E\{ISE\}$ and σ_{ISE} is constant and negative. Hence for ν=0.8N$^{0.45}$ or ν=$\sqrt{N}$, $\lim_{N\rightarrow\infty} E\{ISE\}$=0 and $\lim_{N\rightarrow\infty} \sigma_{ISE}$=0 and the pdf estimate converges, i.e. $\lim_{N\rightarrow\infty} f_{N,\nu}(x) = \phi(x)$.

We plot 10-bin normalized histograms for N=$\{100, 1000, 10000\}$ to demonstrate that for a constant value of ν, the pdf does not converge with increasing N. Then we plot normalized histograms for N=$\{100, 1000, 10000\}$ with optimal bin size, ν = round$(0.8N^{0.45})$, and demonstrate that the pdf will in this case converge. From previous analysis we have obtained the expected value and standard deviation of the error to compare with our results.

[2] $\int_{-\infty}^{\infty}\phi(x)^2 dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{x^2}dx = \frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}\phi(x)dx = \frac{1}{2\sqrt{\pi}}$

[3] Scott, D. W. (1979). On Optimal and Data-Based Histograms. www.jstor.org/stable/2335182

[4] Tippett, L. (1925). www.jstor.org/stable/2332087

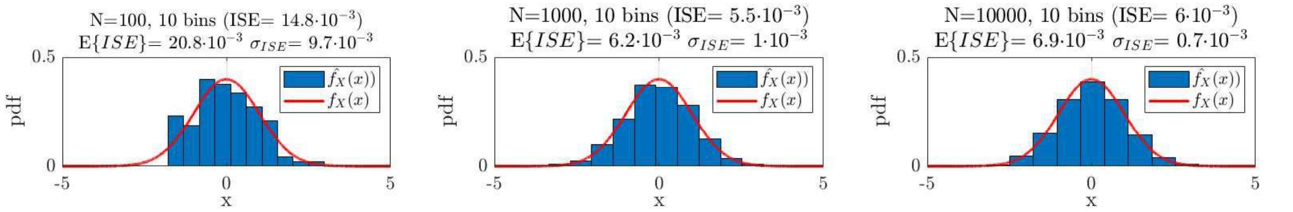

Figure 1.13: $\widehat{f_{N,\nu}(x)}$ for N=$\{10^2, 10^3, 10^4\}$, ν=10. No convergence as N increases.$|ISE - E\{ISE\}| < 2\sigma_{ISE}$ in all cases.

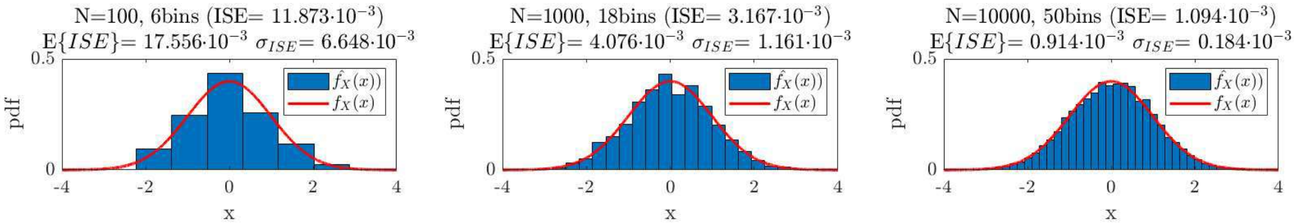

Figure 1.14:$\widehat{f_{N,\nu}(x)}$ for N=$\{10^2$,10^3,$10^4\}$,$\nu\approx$round($0.8N^{0.45}$).Convergence as N increases.$|ISE$-$E\{ISE\}|<\sigma_{ISE}$ in all cases.

1.2 Stochastic processes

We generate M N-sample realizations of the processes below and examine their stationarity and ergodicity.

rp1: $v[k,n] = (x[k,n] - 0.5) \cdot bsin(\frac{n\pi}{N}) + a \cdot n$, with $a = 0.02,\ b = 5$ and $x[k,n] \sim \mathcal{U}(0,1)\forall k,n$.[5]

rp2: $v[k,n]$=$(x[k,n]$-0.5)$M_r[k]$+$A_r[k]$, with $x[k,n]$~$\mathcal{U}(0,1)$, $M_r[k,n]$=$M_r[k]$~$\mathcal{U}(0,1)$, $A_r[k,n]$=$A_r[k]$~$\mathcal{U}(0,1)\forall k,n$

rp3: $v[k,n] = (x[k,n] - 0.5) \cdot 3 + 0.5$, with $x[k,n] \sim \mathcal{U}(0,1)\forall k,n$.

To study these properties, we will calculate the ensemble mean and standard deviation at each time n:

$$\widehat{m}_{ens}[n] = \frac{1}{M}\sum_{k=1}^{M} v[k,n] \text{ and } \widehat{\sigma}_{ens}[n] = \sqrt{\frac{1}{M-1}\sum_{k=1}^{M}(v[k,n] - \widehat{m}_{ens}[n])^2} \tag{13}$$

as well as the sample mean and sample standard deviation of each realization:

$$\widehat{m}_{samp}[k] = \frac{1}{N}\sum_{n=1}^{N} v[k,n] \text{ and } \widehat{\sigma}_{samp}[k] = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(v[k,n] - \widehat{m}_{samp}[n])^2} \tag{14}$$

We generate 100 100-sample realizations of all processes and plot $\widehat{m}_{ens}[n]$ and $\widehat{\sigma}_{ens}[n]$ as a function of time n.

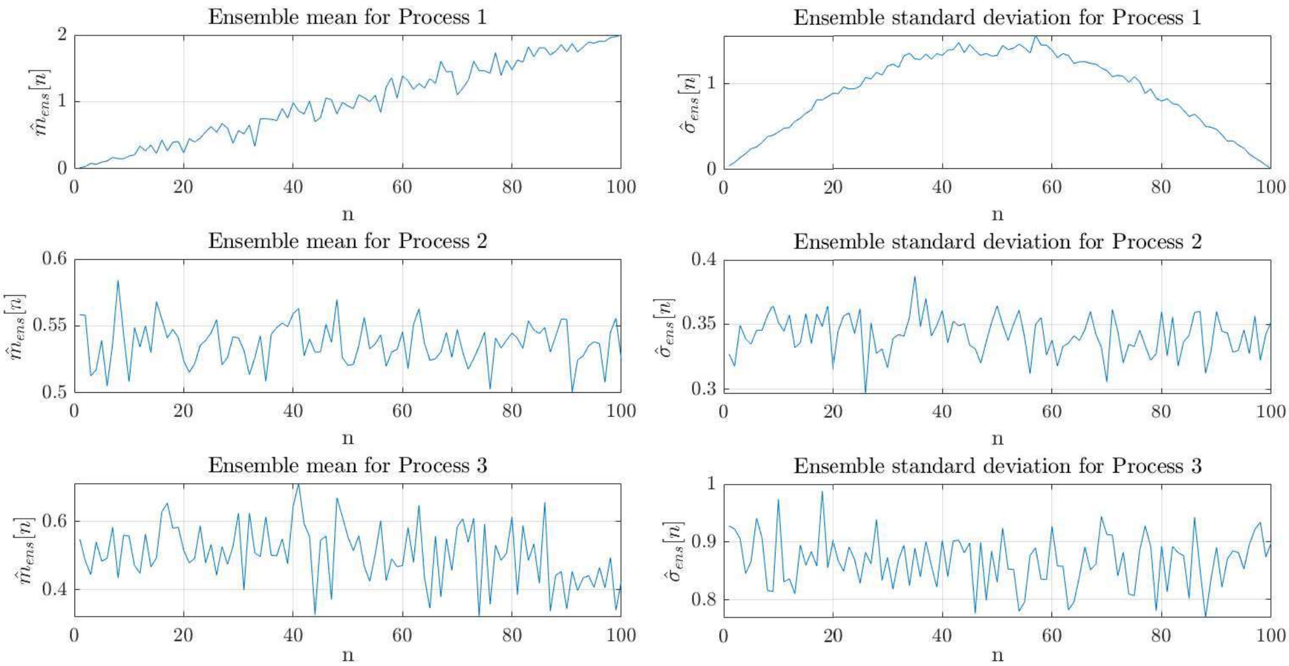

Figure 1.15: Ensemble mean and ensemble standard deviation for processes 1,2,3.

A random process is *strict sense stationary of order N* if its joint probability density functions up to order N do not depend on time shifts. A weaker condition is that a random process is *wide sense stationary* if its mean does not vary in time and the autocorrelation depends only on time lag. Thus, to show that a process is non-stationary it is sufficient to show that the mean or standard deviation vary with time.

[5]Index k represents the kth realization (from a total of M). The notation "m" is reserved for the expected value.

For rp1, both $\widehat{m}_{ens}$ and $\widehat{\sigma}_{ens}$ show a trend that varies in time. Thus rp1 is *not stationary.* For rp2 and rp3, $\widehat{m}_{ens}$ and $\widehat{\sigma}_{ens}$ vary due to noise and are otherwise constant in time. This is an indication, *not a proof*, that they may be stationary. Showing that they are WSS requires checking that their autocorrelation only depends on time lag.

We generate 4 1000-sample realizations of all processes and calculate $\widehat{m}_{samp}$ and $\widehat{\sigma}_{samp}$ for each realization.

Table 1.5: $\widehat{m}_{samp}$ and $\widehat{\sigma}_{samp}$ for each realization of each process.

	rp1		rp2		rp3	
k	$\widehat{m}_{samp}$	$\widehat{\sigma}_{samp}$	$\widehat{m}_{samp}$	$\widehat{\sigma}_{samp}$	$\widehat{m}_{samp}$	$\widehat{\sigma}_{samp}$
1	9.9317	5.8888	0.0399	0.0453	0.5288	0.8595
2	10.0175	5.8609	0.9258	0.1998	0.4834	0.8551
3	9.9901	5.8710	0.9282	0.2609	0.4386	0.8521
4	9.9944	5.8575	0.1481	0.2620	0.5412	0.8899

Table 1.6: Statistics to identify ergodicity.

	rp1	rp2	rp3
$\widehat{m}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}})$	9.9834	0.5105	0.4980
$\widehat{\sigma}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}})$	0.0365	0.4830	0.0467
$\widehat{m}(\widehat{\boldsymbol{\sigma}}_{\boldsymbol{samp}})$	5.8696	0.1920	0.8641
$\widehat{\sigma}(\widehat{\boldsymbol{\sigma}}_{\boldsymbol{samp}})$	0.0141	0.1020	0.0174

A process, X, is ergodic for the mean if, for any single realization of the process: $\lim_{N\to\infty} \widehat{m}_{samp} = E\{x[n]\} = m \;\forall k, n$. Firstly, we must note that the definition of ergodicity of the mean requires that the mean does not vary in time. Secondly, since we only have access to finite samples we will be using estimators to conduct our assessment.

- rp1: sample mean does not vary significantly between realizations ($\widehat{m}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}}) \approx 9.983$, $\widehat{\sigma}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}}) \approx 0.037$). However, we have already shown that the mean of rp1 varies with time. Hence, the process cannot be ergodic.
- rp2: sample mean varies significantly ($\widehat{m}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}})\approx 0.511$, $\widehat{\sigma}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}})\approx 0.483$). Thus, the process is not ergodic.
- rp3: We have shown that rp3 has time invariant mean and the sample mean does not show significant variation ($\widehat{m}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}}) \approx 0.498$, $\widehat{\sigma}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}}) \approx 0.047$). Thus we can say that it is mean-ergodic. To conclude that it is wide sense ergodic, we would also have to show it is WSS and that the temporal autocorrelation matches the ensemble autocorrelation. Thus, the fact that the sample standard deviation does not vary significantly ($\widehat{m}(\widehat{\boldsymbol{\sigma}}_{\boldsymbol{samp}}) \approx 0.864$, $\widehat{\sigma}(\widehat{\boldsymbol{\sigma}}_{\boldsymbol{samp}}) \approx 0.017$), is a good indication but not sufficient.

To summarize the discussion on stationarity and ergodicity:

1. rp1 is not stationary and hence not ergodic.
2. rp2 has stationary mean and standard deviation and analysis needs to be performed to assess whether it is wide sense stationary. The temporal averages vary significantly and hence it is not ergodic.
3. rp3 has stationary mean and standard deviation and is mean-ergodic. Further analysis needs to be performed to assess whether it is wide sense stationary and wide sense ergodic.

We will now obtain theoretical values for the mean and standard deviation of the processes examined.

rp1: $v[k,n] = (x[k,n] - 0.5) \cdot bsin(\frac{n\pi}{N}) + a \cdot n$, with $a = 0.02,\; b = 5$ and $x[k,n] \sim \mathcal{U}(0,1)\forall k, n$.

$$E\{v[k,n]\} = E\{(x[k,n] - 0.5)\} \cdot bsin(\tfrac{n\pi}{N}) + a \cdot n = a \cdot n = 0.02n \tag{15}$$

$$\begin{aligned} Var\{v[k,n]\} &= E\{v[k,n]^2\} - E\{v[k,n]\}^2 = E\{((x[k,n] - 0.5) \cdot bsin(\tfrac{n\pi}{N}) + a \cdot n)^2\} - (an)^2 = \\ &= E\{x[k,n]^2 - x[k,n] + 0.25\} \cdot b^2 sin(\tfrac{n\pi}{N})^2 = (E\{x[k,n]^2\} - 0.25) \cdot b^2 sin(\tfrac{n\pi}{N})^2 = \\ &= b^2 sin(\tfrac{n\pi}{N})^2 (\textstyle\int_0^1 x^2 dx - 0.25) = \tfrac{25}{12} sin(\tfrac{n\pi}{N})^2 \Rightarrow \sigma_{v[k,n]} = \tfrac{5}{\sqrt{12}} sin(\tfrac{n\pi}{N}) \end{aligned} \tag{16}$$

These results are consistent with $\widehat{m}_{ens}$ and $\widehat{\sigma}_{ens}$ plotted above. However, as expected from the fact that rp1 is not ergodic, the temporal statistics $\widehat{m}_{samp}$ and $\widehat{\sigma}_{samp}$ do not reflect the theoretical values.

rp2: $v[k,n]$=$(x[k,n]$-$0.5)M_r[k]$+$A_r[k]$, with $x[k,n]\sim\mathcal{U}(0,1)$, $M_r[k,n]$=$M_r[k]\sim\mathcal{U}(0,1)$, $A_r[k,n]$=$A_r[k]\sim\mathcal{U}(0,1)\forall k, n$

$$E\{v[k,n]\} = E\{(x[k,n] - 0.5)M_r[k]\} + E\{A_r[k]\} = E\{x[k,n]M_r[k]\} - 0.5E\{M_r[k]\} + E\{A_r[k]\} = 0.5 \tag{17}$$

$$\begin{aligned} Var\{v[k,n]\} &= E\{v[k,n]^2\} - E\{v[k,n]\}^2 = E\{((x[k,n] - 0.5)M_r[k] + A_r[k])^2\} - 0.25 = \\ &= E\{(x[k,n] - 0.5)^2\}E\{M_r^2\} - 2E\{(x[k,n]\text{-}0.5)\}E\{M_r[k]\}E\{A_r[k]\} + E\{A_r[k]^2\} - 0.25 = \\ &= \tfrac{1}{3}\textstyle\int_0^1 (x - 0.5)^2 dx - 0 + \tfrac{1}{3} - 0.25 = \tfrac{1}{3}\tfrac{1}{12} + \tfrac{1}{12} = \tfrac{1}{9} \Rightarrow \sigma_{v[k,n]} = \tfrac{1}{3} \end{aligned} \tag{18}$$

These results are consistent with $\widehat{m}_{ens}$ and $\widehat{\sigma}_{ens}$ plotted above. However, as expected from the fact that rp2 is not ergodic, the temporal statistics $\widehat{m}_{samp}$ and $\widehat{\sigma}_{samp}$ do not reflect the theoretical values.

rp3: $v[k,n] = (x[k,n] - 0.5)3 + 0.5 = 3x[k,n] - 1$, with $x[k,n]\sim\mathcal{U}(0,1)\forall k, n$.

$$E\{v[k,n]\} = E\{(3x[k,n] - 1\} = 3E\{x[k,n]\} - 1 = 0.5 \tag{19}$$

$$Var\{v[k,n]\}=E\{v[k,n]^2\}+E\{v[k,n]\}^2=E\{9x[k,n]^2\text{-}6x[k,n]+1\}-\frac{1}{4}=3-3+1-\frac{1}{4}=\frac{3}{4}\Rightarrow\sigma_{v[k,n]}=\frac{\sqrt{3}}{2} \quad (20)$$

Results are consistent with $\widehat{m}_{ens}$ and $\widehat{\sigma}_{ens}$ plotted. The sample mean and sample standard deviation reflect the theoretical values ($\widehat{m}(\widehat{\boldsymbol{m}}_{\boldsymbol{samp}})\approx0.498\approx0.5$, $\widehat{m}(\widehat{\boldsymbol{\sigma}}_{\boldsymbol{samp}})\approx0.864\approx\frac{\sqrt{3}}{2}$), due to the process being ergodic.

1.3 Estimation of probability distributions

We will write an m-file that gives an estimate of the pdf of a collection of N samples. The pdf estimator will be a histogram implemented using the MATLAB `hist` function and normalized by the total area. The bin number will be `round`$(0.8N^{0.45})$. Setting the number of bins to increase with N allows for convergence, as $N\rightarrow\infty$, for non-uniform distributions. This specific relation was found to be optimal in estimating a standard normal distribution.

```
%varargout used to allow the user to access the pdf plot coordinates
%eg: [xn,yn]=pdf(x) will  plot the estimate and also
%give the height and center of each bin
function varargout=pdf(v)
    N=length(v);
    bins=round(0.8*(N^0.45));%found to be optimal for standard normal
    [counts,xn] = hist(v,bins);
    width=(max(v)-min(v))/bins;
    pdf_at_xn=counts/(N*width);
    bar(xn,pdf_at_xn,1)
    xlabel('x')
    ylabel('pdf.estimate(x)')
    title(strcat('PDF Estimation (',num2str(bins),' bins)'))
    if nargout==2
        varargout(1)={xn};
        varargout(2)={pdf_at_xn};
    end
end
```

We test our code with N-sample stationary Gaussian processes, $v[n]\sim\mathcal{N}(0,1)\forall n$, for $N=\{100,1000,10000\}$.

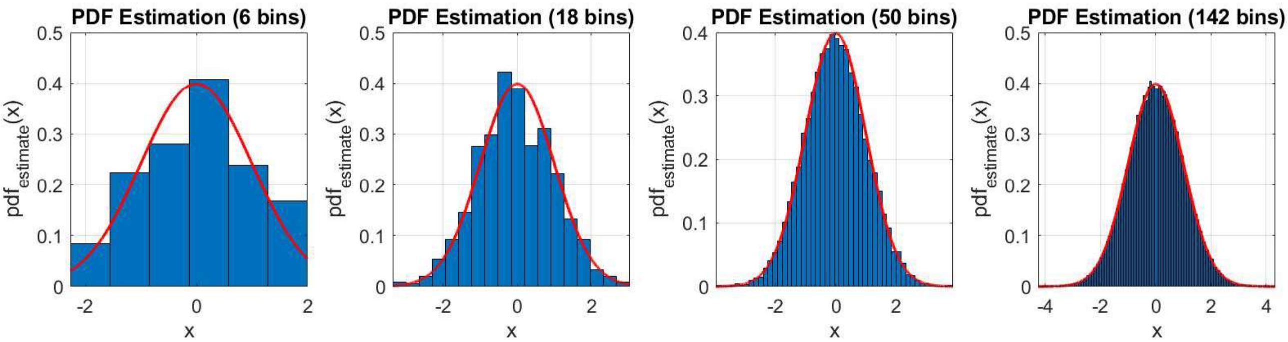

Figure 1.16: Estimate of Gaussian process, N=$\{10^2,10^3,10^4,10^5\}$. Estimate converges as N increases. $\phi(x)$ in red.

We will now use our code to estimate rp3, the only process that was shown to have stationary mean and be mean-ergodic. Before estimating, we will calculate the theoretical pdf and use it to compare our results.

$$v[k,n]=(x[k,n]-0.5)3+0.5=3x[k,n]-1\Rightarrow F_{v[k,n]}(v)=P(v[k,n]\leq v)=P(x[k,n]\leq\tfrac{v+1}{3})=F_{x[k,n]}(\tfrac{v+1}{3}).$$

$$\Rightarrow f_{v[k,n]}(v)=\frac{dF_{v[k,n]}(v)}{dv}=\frac{dF_{x[k,n]}(\frac{v+1}{3})}{d(\frac{v+1}{3})}\cdot\frac{d(\frac{v+1}{3})}{dv}=\frac{1}{3}f_{x[k,n]}(\tfrac{v+1}{3})=\frac{u(v+1)-u(v-2)}{3} \quad (21)$$

The autocorrelation function is given by:

$$E\{v[k,n_1]v[k,n_2]\}=9E\{x[k,n_1]x[k,n_2]\}\text{-}3E\{x[k,n_1]\text{-}x[k,n_2]\}+1=9E\{x[k,n_1]x[k,n_2]\}\text{-}2=\begin{cases}1\text{ , for } n_1=n_2\\ \frac{1}{4}\text{ , } otherwise\end{cases} \quad (22)$$

Hence, the autocorrelation depends only on time difference and rp3 is WSS.

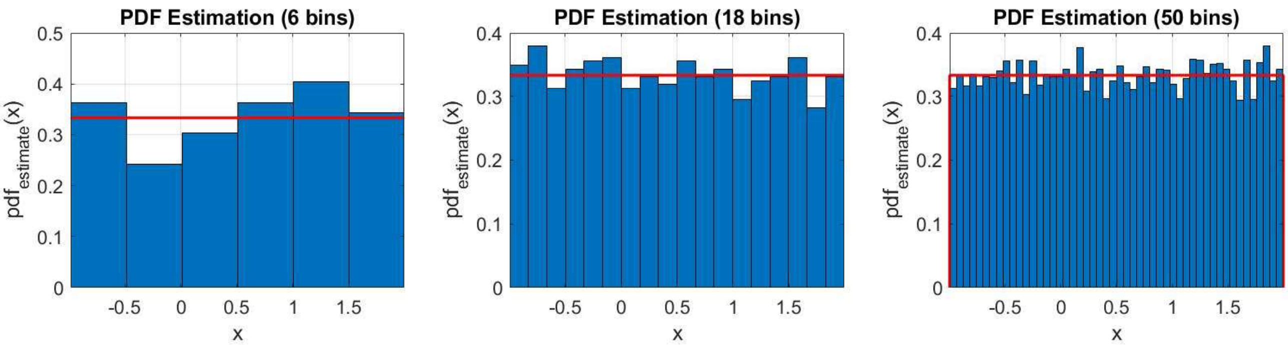

Figure 1.17: Estimate of rp3, N=$\{10^2, 10^3, 10^4\}$. The estimate converges as N increases. Theoretical pdf in red.

The function created cannot estimate non-stationary processes as it inherently assumes that the pdf is time invariant. It is, also, only applicable to ergodic processes as it assumes that the distribution of values in a single realization is equivalent to the pdf. In cases where non-stationarity occurs only due to variation of the deterministic component of the signal, we can use a filter to reveal the deterministic change and account for it. Consider example: $x[n]$=$u[n\text{-}t]$+$w[n]$, where $w[n]$~$\mathcal{N}(0,1)$, t=500 and n≤1000⇒$f_{x[n]}(x)$=$\phi(x-u(n-500))$. Estimating $f_{x[n]}(x)$, we separate the deterministic $x_d[n]$, from the zero-mean stochastic part,$x_{st}[n]$=$x[n]-x_d[n]$. Then, $f_{x[n]}(x)$=$f_{x_{st}[n]}(x-x_d[n])$ and we can use our function to estimate $f_{x_{st}[n]}$. Our estimate will be $\widehat{f}_{x[n]}(x)$=$\widehat{f}_{x_{st}[n]}(x-\widehat{x_d}[n])$. We generate the process and estimate $x_d[n]$ using a moving average filter.

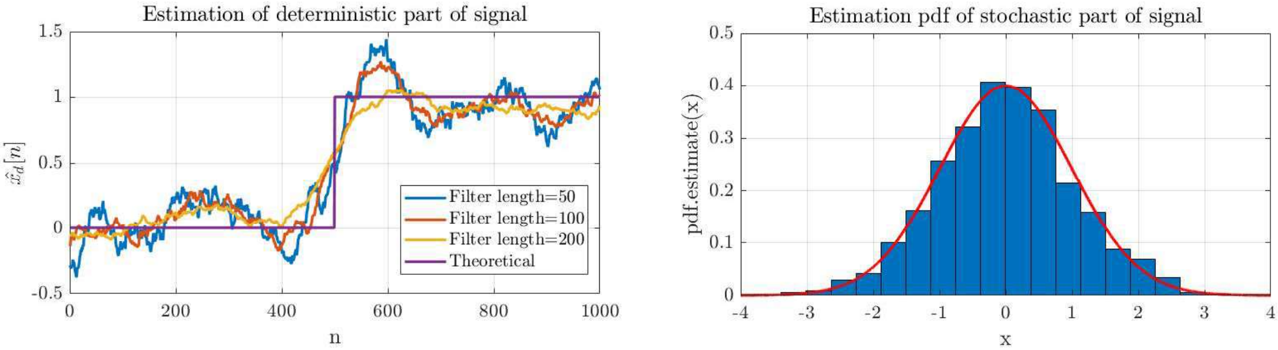

Figure 1.18: Estimate of $x_d[n]$ using MA filter of length 50,100,200 (theoretical in purple, value used in estimations: 100) and estimate of pdf of zero mean stochastic part of process (theoretical in red).

Then we can plot $\widehat{f_{x[n]}}(x)$ for different n and see how they align with the theoretical pdf: $f_{x[n]}(x)$=$\phi(x\text{-}u(n\text{-}500))$.

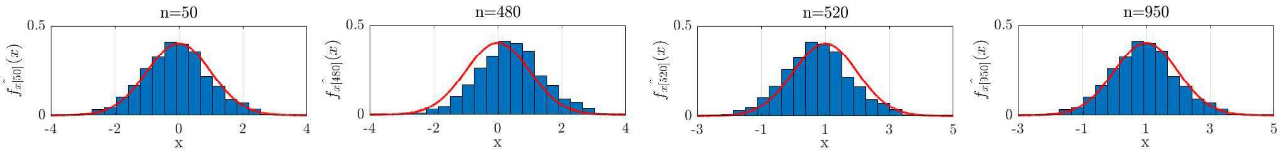

Figure 1.19: $\widehat{f_{x[n]}}(x)$ for n=$\{50, 480, 520, 950\}$ (theoretical in red).

Most of the error in the filter comes from the step response of the MA filter and becomes more significant closer to 500. Additionally, this method is very narrow in application as it tackles only deterministic changes in mean. As such, it cannot be applied to processes where the variance changes in time (such as rp1).

Finally, we use our method to tackle the process: $x[n]$=$sin(\frac{2n\pi}{N})+w[n]$, where $w[n]$~$\mathcal{U}(0,1)$ and n≤N=1000. Then, $x[n]\sim\mathcal{U}(sin(\frac{2n\pi}{N}), sin(\frac{2n\pi}{N})+1)$. Note that in this case, the theoretical results are: $x_d[n]$=$sin(\frac{2n\pi}{N})$+0.5 and $x_{st}[n]\sim\mathcal{U}(-0.5, 0.5)$ with the transformation done to ensure that the stochastic part is zero-mean.

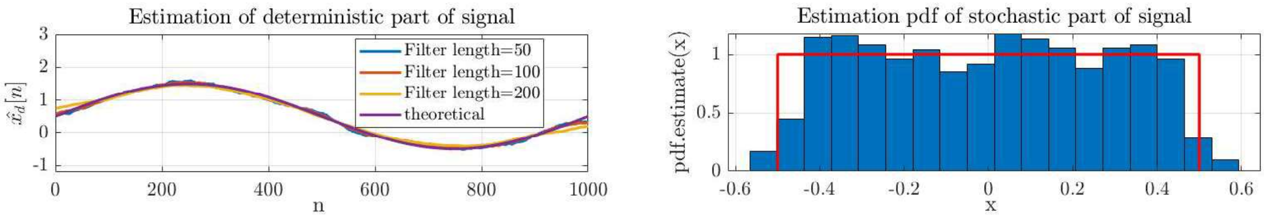

Figure 1.20: Estimate of $x_d[n]$ using MA filter of length 50, 100, 200 (theoretical in purple, value in estimations: 100) and estimate of pdf of zero mean stochastic part of process (theoretical in red).

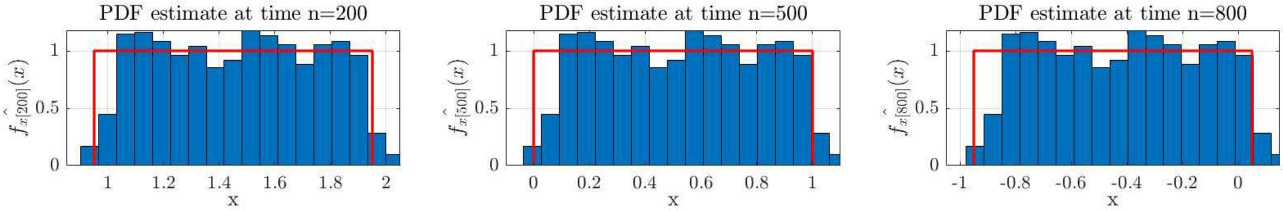

Figure 1.21: $\widehat{f_{x[n]}}(x)$ for n=$\{200, 500, 800\}$ (theoretical in red).