S8 Time Series Analysis
Coursework 2

Handout: Friday 30 November 2018.
Deadline: Thursday 13 December 2018, 4pm.

You must type up your solutions in LaTeX or Word, including all the code you used in the body of your report. You need to submit a hard copy to the Undergraduate Office and upload the Word or pdf document to the TurnItIn Assignment on Blackboard. Both submissions must be done by the deadline. Blackboard is not optimised to Safari and can hang when uploading. I suggest using Firefox instead.

These questions are about spectral estimation. Make use of the `fft` function, available in MATLAB, R and via `numpy.fft` in Python. Carefully read the definition of the function in the software to get the indexing correct (part of the question!).

For all calculations assume $\Delta_t = 1$ and $\sigma_\epsilon^2 = 1$.

1. For a portion $X_1, ..., X_N$ of a zero mean stationary random process $\{X_t\}$, the periodogram is defined as

$$\hat{S}^{(p)}(f) = \frac{1}{N}\left|\sum_{t=1}^{N} X_t \mathrm{e}^{-\mathrm{i}2\pi f t}\right|^2.$$

Asymptotically as $N \to \infty$

$$E\{\hat{S}^{(p)}(f)\} = S(f)$$
$$\mathrm{var}\,\{\hat{S}^{(p)}(f)\} = S^2(f), \qquad 0 < f < \frac{1}{2},$$

where $S(f)$ is the spectral density function of $\{X_t\}$. If we restrict ourselves just to the Fourier frequencies $f_k = k/N$, we also find that the $(N/2)+1$ random variables (N even), $\hat{S}^{(p)}(f_0), \hat{S}^{(p)}(f_1), \ldots, \hat{S}^{(p)}(f_{N/2})$, are all approximately pairwise uncorrelated for N large enough; i.e.,

$$\mathrm{corr}\,\{\hat{S}^{(p)}(f_j), \hat{S}^{(p)}(f_k)\} \approx 0, \;\; j \neq k \text{ and } 0 \leq j, k \leq N/2.$$

(A) Use a Gaussian random number generator to create a realization of a portion $\epsilon_1, \epsilon_2, \ldots, \epsilon_{128}$ of a Gaussian white noise process with zero mean and unit variance. Generate a corresponding realization of a portion X_1, X_2, $\ldots$, X_{128} of the zero mean AR(2) process $X_t = \frac{3}{4}X_{t-1} - \frac{1}{2}X_{t-2} + \epsilon_t$. To do this generate:

$$\begin{aligned}
X_1 &= \tfrac{4}{3}\epsilon_1 \\
X_2 &= \tfrac{1}{2}X_1 + \tfrac{2}{\sqrt{3}}\epsilon_2 \\
X_3 &= \tfrac{3}{4}X_2 - \tfrac{1}{2}X_1 + \epsilon_3 \\
&\;\;\vdots \\
X_{128} &= \tfrac{3}{4}X_{127} - \tfrac{1}{2}X_{126} + \epsilon_{128}
\end{aligned}$$

(The first two values are structurally different to the rest in order to avoid "start-up" transients which would occur if we just set $X_1 = X_2 = 0$. You explored start-up transients in Coursework 1).

(B) Compute the periodogram for $\{X_t\}$ at three adjacent Fourier frequencies, $\hat{S}_0^{(p)}(f_{12})$, $\hat{S}_0^{(p)}(f_{13})$ and $\hat{S}_0^{(p)}(f_{14})$, where $f_{12} = 12/128$, $f_{13} = 13/128$ and $f_{14} = 14/128$.

Repeat steps (A) and (B) $N_r = 10,000$ times (using a different realization of $\{\epsilon_t\}$ each time) to obtain the sequences $\{\hat{S}_j^{(p)}(f_{12}) : j = 0, \ldots, N_r - 1\}$, $\{\hat{S}_j^{(p)}(f_{13}) : j = 0, \ldots, N_r - 1\}$ and $\{\hat{S}_j^{(p)}(f_{14}) : j = 0, \ldots, N_r - 1\}$.

(a) Compute the sample mean and sample variance for the three sequences. Compare your results with the large sample results given above and give a brief comment.

[3 marks]

(b) For the three sequences, also compute the sample correlation coefficient $\hat{\rho}$ (the Pearson product moment correlation coefficient) between $\{\hat{S}_j^{(p)}(f_{12})\}$ and $\{\hat{S}_j^{(p)}(f_{13})\}$, between $\{\hat{S}_j^{(p)}(f_{12})\}$ and $\{\hat{S}_j^{(p)}(f_{14})\}$ and between $\{\hat{S}_j^{(p)}(f_{13})\}$ and $\{\hat{S}_j^{(p)}(f_{14})\}$. Compare your results with the large sample results given above and give a brief comment.

[2 marks]

(c) Compute and present histograms for $\{\hat{S}_j^{(p)}(f_{12})\}$, $\{\hat{S}_j^{(p)}(f_{13})\}$, $\{\hat{S}_j^{(p)}(f_{14})\}$, and compare them (graphically) to the pdfs suggested by the fact that, as $N \to \infty$, $\hat{S}^{(p)}(f) \stackrel{\mathrm{d}}{=} S(f)\chi_2^2/2,\ 0 < f < \frac{1}{2}$. Give a brief comment.

[3 marks]

2. The direct spectral estimator is defined as $\hat{S}^{(d)}(f) = \left|\sum_{t=1}^{N} h_t X_t \mathrm{e}^{-\mathrm{i}2\pi f t}\right|^2$, where $\{h_t\}$ is a taper normalized so that $\sum_{t=1}^{N} h_t^2 = 1$. In this question you will use the Hanning taper, defined here as

$$h_t = \frac{1}{2}\left[\frac{8}{3(N+1)}\right]^{1/2}\left[1 - \cos\left(\frac{2\pi t}{N+1}\right)\right], \quad t = 1, \ldots, N.$$

The Yule-Walker estimator for the parameters of an AR(p) process is given in the notes. You will need to write **your own** function for implementing this. You can assume you know the process is zero mean and make use of the acvs estimator you will have coded up in Coursework 1.

(a) For the AR(2) process defined in Question 1, generate three independent realisations. One of length $N = 64$, one of length $N = 256$ and one of length $N = 1024$. Create a 4 x 3 array of plots on a single page (MATLAB's `subplot` routine is good for this). Each column of plots will correspond to a value of N. On each of the 12 set of axes, plot the true spectral density function (sdf) evaluated at the Fourier frequencies along with:

- Row 1: the periodogram estimate of the sdf.
- Row 2: the direct spectral estimate of the sdf using the Hanning taper.
- Row 3: the Yule-Walker estimate of the sdf assuming an AR(2) model without using tapering.
- Row 4: the Yule-Walker estimate of the sdf assuming an AR(2) model using the Hanning taper.

In less than 100 words, comment on your results.

(b) In this question you will create 3 plots.

For the AR(2) process defined in Question 1, using 10000 independent realisations of length $N = 64$, plot on the same set of axes

- The true sdf.
- The sample mean periodogram.
- The sample mean direct spectral estimate of the sdf using the Hanning taper.
- The sample mean Yule-Walker estimate of the sdf assuming an AR(2) model without using tapering.
- The sample mean Yule-Walker estimate of the sdf assuming an AR(2) model using the Hanning taper.

Repeat for $N = 256$ and $N = 1024$.

In less than 100 words, comment on your results.

(c) For the zero mean MA(3) model

$$X_t = \epsilon_t + \frac{1}{2}\epsilon_{t-1} - \frac{1}{4}\epsilon_{t-2} + \frac{1}{2}\epsilon_{t-3},$$

where $\{\epsilon_t\}$ is a Gaussian white noise process with zero mean and unit variance, generate three independent realisations; one of length $N = 64$, one of length $N = 256$ and one of length $N = 1024$. Create a 4 x 3 array of plots on a single page. Each column of plots will correspond to a value of N. On each of the 12 set of axes, plot the true spectral density function (sdf) evaluated at the Fourier frequencies along with:

- Row 1: the periodogram estimate of the sdf.
- Row 2: the direct spectral estimate of the sdf using the Hanning taper.
- Row 3: the Yule-Walker estimate of the sdf assuming an AR(2) model without using tapering.
- Row 4: the Yule-Walker estimate of the sdf assuming an AR(2) model using the Hanning taper.

In less than 100 words, comment on your results.

(d) In this question you will create 3 plots.

For the MA(3) model in 2(c), using 10000 independent realisations of length $N = 64$, plot on the same set of axes

- The true sdf.
- The sample mean periodogram.
- The sample mean direct spectral estimate of the sdf using the Hanning taper.
- The sample mean Yule-Walker estimate of the sdf assuming an AR(2) model without using tapering.
- The sample mean Yule-Walker estimate of the sdf assuming an AR(2) model using the Hanning taper.

Repeat for $N = 256$ and $N = 1024$.

In less than 100 words, comment on your results.

[12 marks]