Teaching Office Correction:
Please note:
Crib Q1 = Exam Q1
Crib Q2 = Exam Q4
Crib Q3 = Exam Q2
Crib Q4 = Exam Q3

Q1a-i. The transition probability matrix P of the Markov chain with outputs $\{A, B\}$ is

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{bmatrix}.$$

Check that matrix P has stationary probability mass function $\pi = [p(A), p(B)] = [0.25, 0.75]$ by checking that $\pi P = \pi$. Note that πP is finding the values of $p(i_{n+1})$ using the equation

$$p(i_{n+1}) = p(i_{n+1}|A)p(A) + p(i_{n+1}|B)p(B).$$

Q1a-ii. By the Markov property, $p(i_{2n}|i_{2(n-1)}, i_{2(n-2)}, \ldots, i_0)$ is $p(i_{2n}|i_{2(n-1)})$ since the future behaviour of the process once it has arrived at state $i_{2(n-1)}$ at time $2n-2$ does not depend on the precise trajectory it took to arrive at state $i_{2(n-1)}$.

We have to compute

$$\begin{aligned} p(i_n|i_{n-2}) &= \frac{p(i_{n-2}, i_n)}{p(i_{n-2})} \\ &= \frac{\sum_{i_{n-1}\in\{A,B\}} p(i_{n-2}, i_{n-1}, i_n)}{p(i_{n-2})} \\ &= \sum_{i_{n-1}\in\{A,B\}} p(i_{n-1}, i_n|i_{n-2}) \\ &= \sum_{i_{n-1}\in\{A,B\}} p(i_{n-1}|i_{n-2})p(i_n|i_{n-1}) \quad \text{(Markov property)} \end{aligned}$$

for $i_{n-2} = A, i_n = A$ and $i_{n-2} = A, i_n = B$. Other values can be recovered since $p(i_n|i_{n-2})$ is a probability mass function for each i_{n-2}. This calculation is equivalent to computing P^2 since row i_{n-2} and column i_n of the matrix P^2 is

$$[p(A|i_{n-2}), p(B|i_{n-2})] \begin{bmatrix} p(i_n|A) \\ p(i_n|B) \end{bmatrix}$$

which is precisely $\sum_{i_{n-1}\in\{A,B\}} p(i_{n-1}|i_{n-2})p(i_n|i_{n-1})$

$$\begin{bmatrix} 0.52 & 0.48 \\ 0.16 & 0.84 \end{bmatrix}.$$

Q1b-i.

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{ixt} f_X(x)dx.$$

Let $F_X(f)$ denote the Fourier transform of f_X. So $F_X(f) = \varphi_X(-f)$.

$$\varphi_X(t) = F_X(-t) = ab\text{sinc}^2(-tb/2) = 4ab\sin^2(-tb/2)/t^2b^2.$$

Q1b-ii. Using $2\sin^2 x = 1 - \cos(2x)$,

$$\begin{aligned}
\varphi_X(t) &= \frac{2a}{t^2 b}\left(1 - \cos(tb)\right) \\
&= \frac{2a}{t^2 b}\left(1 - \cos(tb)\right) \\
&= \frac{2a}{t^2 b}\left(\frac{t^2 b^2}{2!} - \frac{t^4 b^4}{4!} + \frac{t^6 b^6}{6!} + \dots\right) \\
&= ab\left(1 - \frac{t^2 b^2}{12} + \frac{t^4 b^4}{360} + \dots\right)
\end{aligned}$$

Relate characteristic function derivatives to moments: $i^n E(X^n) = \left(\frac{d^n}{dt^n}\varphi_X(t)\right)_{t=0}$.

So $E(X^0) = ab$ which should be 1. $E(X^2) = ab^3/6 = b^2/6$. $E(X^4) = ab^5/15 = b^4/15$.

Q2a. Your answer should state the definition of strict sense and wide sense stationary.

Strictly stationary requires that $p(x_n, \ldots, x_m) = p(x_{n+k}, \ldots, x_{m+k})$ for all n, $m \geq n$, and k. This implies any two sections of the process is identically distributed. In the lectures we have seen how this is an unrealistic assumption if X_n describes the average temperature of a particular day n of the year because strict stationary implies the average daily temperature in summer and winter is the same, which is clearly inappropriate.

Wide sense stationary does not require equality of the probability distributions of the any two sections of the process but just that process' first and second oder statistics do not change with time. To see why this a more appropriate assumption, in the lectures we saw how the process $X_n = \sin(f_0 n + \phi)$ where ϕ is a randomly distributed phase, drawn uniformly from $(0, 2\pi)$, is WSS but not strictly stationary.

Q2b. We may write

$$Y_n = \sum_{i=-\infty}^{\infty} h_i X_{n-i}.$$

We see that $E(Y_n) = 0$ since input process has zero mean.

$$Y_n Y_{n+k} = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i X_{n-i} h_j X_{n+k-j}$$

and

$$\begin{aligned} E\{Y_n Y_{n+k}\} &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j \mathbb{E}\{X_{n-i} X_{n+k-j}\} \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j c\delta(k-j+i) \\ &= c \sum_{i=-\infty}^{\infty} h_i h_{k+i}. \end{aligned}$$

Here $\delta(l) = 0$ for $l \neq 0$ and $\delta(0) = 1$. We see that $E\{Y_n Y_{n+k}\}$ does not depend on n but depends on k only. Let $R_Y(k) = E\{Y_n Y_{n+k}\}$.

Q2c. Using $h_i = a\exp(-ib)$ for non-negative i, for $k \geq 0$

$$R_Y(k) = c \sum_{i=-\infty}^{\infty} h_i h_{k+i} = ca^2 \exp(-kb) \sum_{i=0}^{\infty} \exp(-i2b) = \frac{ca^2 \exp(-kb)}{1 - \exp(-2b)}.$$

Also $R_Y(-k) = R_Y(k)$.

Q2d.

$$\sum_{k=-\infty}^{\infty} e^{-|k|b} e^{-j2\pi f k} = \frac{1}{1 - e^{-b} e^{j2\pi f}} + \frac{1}{1 - e^{-b} e^{-j2\pi f}} - 1$$

by splitting the sum into two geometrical series. Further simplification possible. Then multiply the expression on the RHS with $\frac{ca^2}{1-\exp(-2b)}$ to get the PSD $S_X(f)$.

Version SJG/1

EGT2
ENGINEERING TRIPOS PART IIA

Wednesday 26 April 2017   9.30 to 11

Module 3F3

SIGNAL AND PATTERN PROCESSING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) A source generates a stream of symbols S_n, $n = 0, 1, \ldots$ and each symbol takes one of two possible values, either A or B. The probability of symbol S_n depends only upon the value of symbol S_{n-1}. Let $p(i_n|i_{n-1})$ denote the probability that $S_n = i_n$ given $S_{n-1} = i_{n-1}$. These probabilities are given in the following table.

	$p(i_n\|i_{n-1})$	
	$S_{n-1} = A$	$S_{n-1} = B$
$S_n = A$	0.7	0.1
$S_n = B$	0.3	0.9

(i) Let $p(i_0)$ denote the probability that $S_0 = i_0$. Explain how $p(i_n)$, the probability that $S_n = i_n$, may be calculated. [10%]

(ii) Show that the stationary probability distribution, for which $p(i_n = A) = p(i_0 = A)$, is given by [10%]

	$p(i_n)$
$i_n = A$	0.25
$i_n = B$	0.75

(iii) Show that the random process $S_0, S_2, S_4, \ldots$, generated by the same source but retaining only source symbols with even time indices, is a Markov chain, and determine its transition probability matrix. [30%]

(b) The characteristic function of a random variable X is defined using the mathematical expectation $\mathbb{E}$ as $\varphi_X(t) = \mathbb{E}[\exp(iXt)]$ where t is a real number.

(i) Let X have probability density function $f_X(x)$. Determine the relationship between $\varphi_X(t)$ and the Fourier transform of $f_X(x)$. [10%]

(ii) Let $f_X(x)$ be the following triangular shaped function

$$f_X(x) = 1/b\left(1 - \frac{|x|}{b}\right) \qquad \text{for } |x| \le b$$

and $f_X(x) = 0$ for $|x| > b$. Determine $\varphi_X(t)$ (using the Data book). [10%]

(iii) Express $\varphi_X(t)$ as a power series in t and hence find $\mathbb{E}[X^0]$, $\mathbb{E}[X^2]$ and $\mathbb{E}[X^4]$. [30%]

2 (a) For a random process X_n, $n = 0, 1, \ldots$ explain the difference between *strict* and *wide-sense stationary* (WSS). Why might WSS be the more practical assumption for modelling of a real-world physical process? [20%]

(b) A zero-mean random process X_n has autocorrelation function $R_X(k) = c$ for $k = 0$ and $R_X(k) = 0$ for $|k| > 0$, where c is a constant. It is passed through a linear system with infinite impulse response $\{h_n\}_{n=-\infty}^{\infty}$. If Y_n denotes the output process, find an expression for the autocorrelation function $R_Y(k)$ of the system output in terms of the impulse response and c. [30%]

(c) Find $R_Y(k)$ when $h_n = 0$ for $n < 0$ and

$$h_n = a \exp(-nb)$$

for $n \geq 0$, where a, b are positive constants. [25%]

(d) Determine the power spectral density of the random process Y_n. [25%]

make one of these a show that?

3 A stationary, ergodic random process $\{X_n\}$ is measured over a time interval $n = 0, 1, ..., N-1$, leading to a measured vector of time samples:

$$\mathbf{x} = [x_0\, x_1 \ldots x_{N-1}]^T\,.$$

It is required to estimate an unknown quantity θ relating to $\{X_n\}$ using an estimator $\hat{\theta}(\mathbf{x})$ which is a function of the measured data.

(a) Define the terms *bias* and *variance* for the estimator $\hat{\theta}$. [15%]

Solution:

Bias is $E[\hat{\theta}] - \theta$

Variance is $E[(\hat{\theta}] - E[\hat{\theta}])^2]$, , where expectations are taken with respect to $p(\mathbf{x}|\theta)$.

(b) The mean and autocorrelation function for the process are to be estimated according to the formulae:

$$\hat{\mu} = \frac{1}{N}\sum_{n=0}^{N-1} x_n$$

and

$$\hat{R}_X[k] = \frac{1}{N-k}\sum_{n=0}^{N-1-k} x_n x_{n+k},\ (k \geq 0)$$

Explain why these estimation formulae are valid, given the stated assumptions about the process. [10%]

Solution:

The process is ergodic and stationary, hence it is appropriate to estimate mean and autocorrelation functions directly from one vector of measurements $\mathbf{x}$.

(c) Show whether each estimator is unbiased or not. [30%]

Solution:

$$E[\hat{\mu}] = E\frac{1}{N}\sum_{n=0}^{N-1} x_n = \frac{1}{N}\sum_{n=0}^{N-1} Ex_n = \frac{1}{N}\sum_{n=0}^{N-1} x_n = \frac{1}{N}\sum_{n=0}^{N-1} \mu = \mu$$

hence unbiased.

$$E\hat{R}_X[k] = E\frac{1}{N-k}\sum_{n=0}^{N-1-k} x_n x_{n+k} = \frac{1}{N-k}\sum_{n=0}^{N-1-k} Ex_n x_{n+k} = \frac{1}{N-k}\sum_{n=0}^{N-1-k} R_X[k] = R_x[k]$$

hence also unbiased.

(d) The mean value of the process is now assumed to be zero. Some autocorrelation function values are now estimated according to the above estimation formula, leading to:

$$\hat{R}_X[0] = 10.5, \ \hat{R}_X[1] = -9.1, \ \hat{R}_X[2] = 7 .$$

It is required to predict the next value of the signal based upon previous values using a linear filter:

$$\hat{x}_{n+1} = h_0 x_n + h_1 x_{n-1}$$

Assuming that the estimated autocorrelation values are accurate, determine the coefficients h_0 and h_1 such that the mean-squared prediction error $\mathbb{E}[(\hat{x}_{n+1} - x_{n+1})^2]$ is minimised. [30%]

Solution:

Let

$$\varepsilon = (\hat{x}_{n+1} - x_{n+1}) = (h_0 x_n + h_1 x_{n-1} - x_{n+1})$$

Then

$$E = \mathbb{E}[(\hat{x}_{n+1} - x_{n+1})^2] = \mathbb{E}[\varepsilon^2]$$

Then,

$$\frac{\partial E}{\partial h_i} = \mathbb{E}[2\varepsilon \frac{\partial \varepsilon}{\partial h_i}]$$

for $i = 0, 1$.

But,

$$\frac{\partial \varepsilon}{\partial h_0} = x_n, \ \frac{\partial \varepsilon}{h_1} = x_{n-1},$$

So

$$\frac{\partial E}{\partial h_0} = \mathbb{E}[2\varepsilon x_n] = \mathbb{E}[2(h_0 x_n + h_1 x_{n-1} - x_{n+1})x_n] = h_0 R_X[0] + h_1 R_X[1] - R_X[1]$$

$$\frac{\partial E}{\partial h_1} = \mathbb{E}[2\varepsilon x_{n-1}] = \mathbb{E}[2(h_0 x_n + h_1 x_{n-1} - x_{n+1})x_{n-1}] = h_0 R_X[1] + h_1 R_X[0] - R_X[2]$$

where we have used the result $R_X[k] = R_X[-k]$.

Setting both partial derivatives to zero we get:

$$h_0 R_X[0] + h_1 R_X[1] = R_X[1]$$

$$h_0 R_X[1] + h_1 R_X[0] = R_X[2]$$

or,

$$\begin{bmatrix} R_X[0] & R_X[1] \\ R_X[1] & R_X[0] \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} R_X[1] \\ R_X[2] \end{bmatrix}$$

$$\mathbf{Rh} = \mathbf{r}$$

Hence, plugging in the given estimates of autocorrelation,

$$\begin{bmatrix} 10.5 & -9.1 \\ -9.1 & 10.5 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} -9.1 \\ 7 \end{bmatrix}$$

i.e.

$$h_0 = -1.16,\ h_1 = -0.34$$

(e) Determine the mean-squared prediction error of this optimal filter and compare it with a filter which takes the previous value of the process as the prediction, i.e. $\hat{x}_{n+1} = x_n$, commenting on why this simpler estimator would not be expected to perform well. [15%]

Solution:

Mean-squared prediction error for optimal filter is, using formula from notes:

$$\mathbb{E}[\varepsilon^2] = R_X[0] - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} = 2.31$$

For the simple filter however, $\varepsilon = x_n - x_{n+1}$, and

$$\mathbb{E}[\varepsilon^2] = \mathbb{E}[(x_n - x_{n+1})^2] = 2R_X[0] - 2R_X[1] = 39.2$$

This is a very poor choice of estimator since the sequence is negatively correlated, so x_{n+1} is nowhere near x_n on average.

4 (a) A pilot tone in an RF communications channel is measured at the receiver in the following form:

$$X_n = A + B\sin(\omega n) + V_n$$

where $\{V_n\}$ is a white Gaussian noise process with zero mean and variance σ_V^2, $\omega < \pi$ is a known frequency of transmission, A is an unknown DC offset and B is an unknown received signal amplitude. It is required to estimate A and B from a measured vector of samples from the process $\{X_n\}$,

$$\mathbf{x} = [x_0\, x_1 \ldots x_{N-1}]^T.$$

For a particular set of parameter values $A = a$ and $B = b$, show that the total squared error term $\varepsilon = \sum_{n=0}^{N-1}(x_n - a - b\sin(\omega n))^2$ can be expressed in terms of the unknown parameter vector $\theta = [a\,b]^T$ as

$$\varepsilon = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{G}\theta + \theta^T\mathbf{G}^T\mathbf{G}\theta$$

where $\mathbf{G}$ should be carefully defined. [20%]

Solution:

We can write, as in lectures:

$$\varepsilon = |\mathbf{x} - \mathbf{G}\theta|^2$$

where

$$\begin{bmatrix} 1 & 0 \\ 1 & \sin(\omega) \\ 1 & \sin(2\omega) \\ 1 & \sin(3\omega) \\ \dots & \dots \\ 1 & \sin((N-1)\omega) \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_2 \end{bmatrix}$$

where the columns will be referred to later.

Expanding:

$$\varepsilon = \mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{G}\theta - \theta^T\mathbf{G}^T\mathbf{x} + \theta^T\mathbf{G}^T\mathbf{G}\theta = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{G}\theta + \theta^T\mathbf{G}^T\mathbf{G}\theta$$

as required, since $\mathbf{x}^T\mathbf{G}\theta = \theta^T\mathbf{G}^T\mathbf{x}$ (they are both scalar quantities and one is the 'transpose' of the other).

(b) Show that the Maximum Likelihood estimator for the parameter vector is found by maximising the following expression:

$$-0.5N\log(2\pi\sigma_V^2) - \frac{1}{2\sigma_V^2}\varepsilon$$

and hence that the ML estimator is

$$\theta^{\text{ML}} = \mathbf{M}^{-1}\mathbf{b}$$

where

$$\mathbf{b} = \begin{bmatrix} \sum_{n=0}^{N-1} x_n & \sum_{n=0}^{N-1} \sin(\omega n)x_n \end{bmatrix}^T$$

$$\mathbf{M} = \begin{bmatrix} N & \frac{\sin(N\omega/2)}{\sin(\omega/2)}\sin(\omega(N-1)/2) \\ \frac{\sin(N\omega/2)}{\sin(\omega/2)}\sin(\omega(N-1)/2) & N/2 - \frac{\sin(N\omega)}{2\sin(\omega)}\cos(\omega(N-1)) \end{bmatrix}$$

[40%]

You may find the following result helpful:

$$\sum_{n=0}^{N-1} \exp(inb) = \exp(i(N-1)b/2)\frac{\sin(Nb/2)}{\sin(b/2)}$$

Solution:

The error term $v_n = x_n - A - B\sin(n\omega)$ is Gaussian zero mean white noise, with variance σ_V^2. Hence its probability is:

$$p(\mathbf{v}) = \prod_{n=0}^{N-1} \mathcal{N}(v_n|0,\sigma_V^2) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp(-\frac{1}{2\sigma_V^2}v_n^2) = \frac{1}{\sqrt{2\pi\sigma_V^2}^N} \exp(-\frac{1}{2\sigma_V^2}\sum_{n=0}^{N-1} v_n^2)$$

But the change of variables $\mathbf{x} = \mathbf{G}\theta + \mathbf{v}$ has unity Jacobian (θ and $\mathbf{G}$ considered fixed), so the likelihood is:

$$p(\mathbf{x}|\theta) = \frac{1}{\sqrt{2\pi\sigma_V^2}^N} \exp(-\frac{1}{2\sigma_V^2}\sum_{n=0}^{N-1}(x_n - a - b\sin(\omega n))^2 = \frac{1}{\sqrt{2\pi\sigma_V^2}^N} \exp(-\frac{1}{2\sigma_V^2}\varepsilon)$$

and the log-likelihood is:

$$-N/2\log(2\pi\sigma_V^2) - \frac{1}{2\sigma_V^2}\varepsilon$$

which we must maximise to find the ML solution.

This is equivalent to maximising ε, so differentiate and set to zero:

$$\frac{d\varepsilon}{d\theta} = 2\mathbf{G}^T\mathbf{G}\theta - 2\mathbf{G}^T\mathbf{x} = \mathbf{0}$$

or,

$$\theta^{ML} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{x}$$

(the familiar OLS solution)

and then the the forms of $\mathbf{M}$ and $\mathbf{b}$ correspond to calculating the specific form of $\mathbf{G}^T\mathbf{G}$ and $\mathbf{G}^T\mathbf{x}$ for this model. Working this through:

Let $\mathbf{G} = \begin{bmatrix}\mathbf{g}_1 & \mathbf{g}_2\end{bmatrix}$ where $\mathbf{g}_1$ and $\mathbf{g}_2$ are the columns of $\mathbf{G}$. Then,

$$\mathbf{M} = \mathbf{G}^T\mathbf{G} = \begin{bmatrix}\mathbf{g}_1^T\mathbf{g}_1 & \mathbf{g}_1^T\mathbf{g}_2 \\ \mathbf{g}_2^T\mathbf{g}_1 & \mathbf{g}_2^T\mathbf{g}_2\end{bmatrix}$$

Now,

$$\mathbf{g}_1^T\mathbf{g}_1 = N$$

$$\mathbf{g}_1^T\mathbf{g}_2 = \mathbf{g}_2^T\mathbf{g}_1 = \sum_{n=0}^{N-1} \sin(n\omega) = \mathscr{I}\sum_{n=0}^{N-1} \exp(j\omega) = \sin((N-1)\omega/2)\frac{\sin(N\omega/2)}{\sin(\omega/2)}$$

and

$$\mathbf{g}_2^T\mathbf{g}_2 = \sum_{n=0}^{N-1} \sin^2(n\omega) = 0.5\sum_{n=0}^{N-1}(1-\cos(2n\omega)) = N/2 - 0.5\sum_{n=0}^{N-1}\cos(2n\omega)$$

$$= N/2 - 0.5\mathscr{R}\sum_{n=0}^{N-1}\exp(2j\omega) = N/2 - \cos((N-1)\omega)\frac{\sin(N\omega)}{2\sin(\omega)}$$

where the last two results use the real and imaginary parts of the suggested summation formula.

b follows fairly straightforwardly as $\mathbf{G}^T\mathbf{x}$.

(c) With $\omega = \pi/5$ and $N = 1000$, show that the solution simplifies to:

$$\hat{a} = \frac{1}{N}\sum_{n=0}^{N-1} x_n, \quad \hat{b} = \frac{2}{N}\sum_{n=0}^{N-1}\sin(\omega n)x_n$$

[20%]

Solution:

In this case the columns of **G** are orthogonal and hence $\mathbf{G}^T\mathbf{G}$ is diagonal, which decouples the solution into the two simple equations given.

(d) Explain why this simplification occurs. How should the data length be chosen in general, relative to ω, to ensure that this is the case? [20%]

Solution:

This is because the sine basis function $\mathbf{g}_2$ in this case is orthgonal to the dc component $\mathbf{g}_1$. The dc component is obtained in this case as the mean of $\mathbf{x}$ without reference to the sine term and b is the normalised projection of the sin-wave onto the data.

This can be ensured provided the data length is an integer multiple of the period of the sin-wave, $2\pi/\omega$.

END OF PAPER

THIS PAGE IS BLANK