Engineering Tripos Part IIA, 2019: Solutions to module 3F3, Statistical Signal Processing

May 21, 2019

Question 1

Part (a)

For $0 < i < M$

$$
\begin{aligned}
&\Pr(X_{n+1} = i) \\
&= \sum_{j=0}^{M} \Pr(X_{n+1} = i, X_n = j) \\
&= \sum_{j=0}^{M} \Pr(X_{n+1} = i | X_n = j) \Pr(X_n = j) \\
&= \Pr(X_{n+1} = i | X_n = i-1) \Pr(X_n = i-1) + \Pr(X_{n+1} = i | X_n = i+1) \Pr(X_n = i+1) \\
&= \alpha \Pr(X_n = i-1) + (1-\alpha) \Pr(X_n = i+1).
\end{aligned}
$$

For $i = M$ the same procedure gives

$$
\begin{aligned}
&\Pr(X_{n+1} = M) \\
&= \Pr(X_{n+1} = M | X_n = M-1) \Pr(X_n = M-1) + \Pr(X_{n+1} = M | X_n = M) \Pr(X_n = M) \\
&= \alpha \Pr(X_n = M-1) + \alpha \Pr(X_n = M).
\end{aligned}
$$

For $i = 0$

$$
\begin{aligned}
&\Pr(X_{n+1} = 0) \\
&= (1-\alpha) \Pr(X_n = 0) + (1-\alpha) \Pr(X_n = 1).
\end{aligned}
$$

[30%]

Part (b)

The Markov chain is irreducible since the chain can visit any state from any starting point.

Let P be the $(M+1)\times(M+1)$ transition probability matrix where row i contains the transition probabilities out of state $i-1$. Let $\pi_n = (\Pr(X_n = 0), \ldots, \Pr(X_n = M))$. Then

$$\pi_{n+1} = \pi_n P.$$

The stationary probability satisfies

$$\pi = \pi P$$

for a row vector π with non-negative elements that add to 1.

When $\alpha = 0.5$, by inspection we see that $\pi = 1/(M+1)\,(1, \ldots, 1)$ will satisfy $\pi = \pi P$.

[20%]

Part (c)

$$\begin{aligned}
&\log \Pr(X_1 = x_1, \ldots X_T = x_T | X_0 = x_0) \\
&= \sum_{i=1}^{T} \log \Pr(X_i = x_i | X_{i-1} = x_{i-1}) \\
&= (s + r) \log(\alpha) + (T - s - r) \log(1 - \alpha)
\end{aligned}$$

Differentiating with respect to α and setting to 0 gives

$$\frac{s+r}{\alpha} - \frac{T-s-r}{1-\alpha} = 0$$

and hence

$$\alpha = \frac{s+r}{T}.$$

[20%]

Part (d)-i

$$\begin{aligned}
\sum_{i=1}^{\infty} i \ \Pr(X_n - 1 = i) &= \Pr(X_n - 1 = -1) + \sum_{i=-1}^{\infty} i \ \Pr(X_n - 1 = i) \\
&\geq \mathbf{E}(X_n - 1) \\
&= \mathbf{E}(X_n) - 1.
\end{aligned}$$

[10%]

Part (d)-ii

Equation (1) that was derived still holds for $i > 0$.

$$\begin{aligned}\mathbf{E}(X_{n+1}) &= \sum_{i=0}^{\infty} i \ \Pr(X_{n+1} = i) \\ &= \sum_{i=1}^{\infty} i \ \Pr(X_{n+1} = i) \\ &= \alpha \sum_{i=1}^{\infty} i \ \Pr(X_n = i-1) + (1-\alpha) \sum_{i=1}^{\infty} i \ \Pr(X_n = i+1).\end{aligned}$$

The first sum

$$\sum_{i=1}^{\infty} i \ \Pr(X_n = i-1) = \sum_{i=1}^{\infty} i \ \Pr(X_n + 1 = i) = \mathbf{E}(X_n + 1) = \mathbf{E}(X_n) + 1.$$

The second sum was shown to be:

$$\begin{aligned}\sum_{i=1}^{\infty} i \ \Pr(X_n = i+1) &= \sum_{i=1}^{\infty} i \ \Pr(X_n - 1 = i) \\ &\geq \mathbf{E}(X_n) - 1.\end{aligned}$$

Combining gives

$$\mathbf{E}(X_{n+1}) \geq \alpha\,(\mathbf{E}(X_n) + 1) + (1-\alpha)\,(\mathbf{E}(X_n) - 1) = 2\alpha - 1 + \mathbf{E}(X_n).$$

When $\alpha > 1$ we see that the expected value of the queue length is strictly increasing in time. Thus the queue must be growing in length.

[20%]

Examiner's comments: Attempted by 60% of candidates. Parts (a) and (b) were easy point earners. Part (c) was poorly answered by a significant number although the question helps in setting up the maximum likelihood problem. Part (d) was very poorly answered with only a handful of candidates getting (d)-ii correct even though the previous parts make clear which results to call on. In part (d)-i, many students failed to realise that the range of $X_n - 1$ is all the integers commencing from -1; this was the primary reason many were unable to show the required result.

Question 2

Part (a)

Formulate the problem in vector notation:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + W = G\theta + W$$

Minimise

$$J(\theta) = (X - G\theta)^T (X - G\theta) = \theta^T G^T G\theta - 2\theta^T G^T X + X^T X$$

with respect to θ. Differentiate and set to zero,

$$\begin{aligned} \begin{bmatrix} \frac{\partial}{\partial a} J \\ \frac{\partial}{\partial b} J \end{bmatrix} &= 0 \\ 2G^T G\theta - 2G^T X &= 0 \\ \hat{\theta} &= \left(G^T G\right)^{-1} G^T X. \end{aligned}$$

Compute the expected value to get

$$\begin{aligned} \mathbf{E}(\hat{\theta}) &= \mathbf{E}\left(\left(G^T G\right)^{-1} G^T X\right) \\ &= \left(G^T G\right)^{-1} G^T \left(\mathbf{E}\left(X\right)\right) \\ &= \left(G^T G\right)^{-1} G^T G\theta \\ &= \theta. \end{aligned}$$

Thus the estimate is unbiased. [30%]

Part (b)

Need to compute

$$\mathbf{E}\left[\left(\hat{\theta} - \theta\right)\left(\hat{\theta} - \theta\right)^T\right] = \mathbf{E}\left[\hat{\theta}\hat{\theta}^T\right] - \theta\theta^T.$$

The first expectation is

$$\begin{aligned} \mathbf{E}\left[\hat{\theta}\hat{\theta}^T\right] &= \mathbf{E}\left[\left(G^T G\right)^{-1} G^T X \left(\left(G^T G\right)^{-1} G^T X\right)^T\right] \\ &= \left(G^T G\right)^{-1} G^T \mathbf{E}\left[X X^T\right] G \left(\left(G^T G\right)^{-1}\right)^T \\ &= H\mathbf{E}\left[X X^T\right] H^T. \end{aligned}$$

$$
\begin{aligned}
\mathbf{E}\left[XX^T\right] &= G\theta\theta^T G^T + E\left(WW^T\right) + E\left(G\theta W^T + W(G\theta)^T\right) \\
&= G\theta\theta^T G^T + E\left(WW^T\right) \\
&= G\theta\theta^T G^T + R
\end{aligned}
$$

where the matrix $R = E\left(WW^T\right)$. Then

$$
\begin{aligned}
H\mathbf{E}\left[XX^T\right] H^T &= H\left[G\theta\theta^T G^T + R\right] H^T \\
H\left[G\theta\theta^T G^T\right] H^T &= \left(G^T G\right)^{-1} G^T G \left[\theta\theta^T\right] G^T G \left(\left(G^T G\right)^{-1}\right)^T \\
&= \theta\theta^T.
\end{aligned}
$$

Thus the variance is HRH^T. [20%]

Part (c)

$$
HRH^T = cHH^T = c\left(G^T G\right)^{-1} G^T G \left(\left(G^T G\right)^{-1}\right)^T = c\left(G^T G\right)^{-1}
$$

where $c = \mathbf{E}(W_i^2)$ if we assume W_i is white noise, i.e. $R = cI$ with I being the identity matrix.

The matrix

$$
G^T G = \begin{bmatrix} N & \sum_{i=1}^{N-1} i \\ \sum_{i=1}^{N-1} i & \sum_{i=1}^{N-1} i^2 \end{bmatrix}
$$

The inverse of this matrix tends to the zero matrix. Use the data book to obtain

$$
G^T G = \begin{bmatrix} N & N(N-1)/2 \\ N(N-1)/2 & (N-1)N(2N-1)/6 \end{bmatrix}
$$

The determinant is $(N-1)N^2(2N-1)/6 - N^2(N-1)^2/4$ and the highest order term is $N^4/3 - N^4/4$. This term causes all terms in $(G^T G)^{-1}$ to tend to zero. [10%]

Part (d)-i

When $W_n = c\sin(\omega n + \phi)$,

$$
\mathbf{E}(W_n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} c\sin(\omega n + \phi)d\phi = 0 \tag{1}
$$

since periodic function is integrated over a complete period.

$$
\begin{aligned}
\mathbf{E}(W_n W_m) &= c^2 \mathbf{E}\left(\sin(\omega n + \phi)\sin(\omega m + \phi)\right) \\
&= \frac{c^2}{2}\mathbf{E}\left(\cos(\omega(n-m)) - \cos(\omega(m+n) + 2\phi\right) \\
&= \frac{c^2}{2}\cos(\omega(n-m))
\end{aligned}
$$

since $\mathbf{E}\left(\cos(\omega(m+n) + 2\phi\right) = 0$ as it is being integrated over 2 periods when it is expressed as in (1). Thus W_n is wide sense stationary and its autocorrelation function is

$$R_W(k) = c^2 \cos(\omega k)/2.$$

[20%]

Part (d)-ii

The PSD of $R_W(k) = c^2 \cos(\omega k)/2$ is

$$S_W(f) = \sum_{k=-\infty}^{\infty} \frac{c^2}{2}\cos(\omega_0 k) e^{i2\pi f k}$$

Form the data book this gives a spike when $f = \omega_0/2\pi$.

[10%]

Part (d)-iii

Call the residual data

$$e_n = x_n - \hat{a} - \hat{b}n = w_n + (a - \hat{a}) + (bn - \hat{b}n).$$

We can estimate the autocorrelation function of the residuals and then compute its PSD. Find the peak in power spectrum. This gives an estimate of ω_0. This should be a reasonable estimator if $\hat{a}$ and $\hat{b}$ estimates the trend well. Which it is as the variance of least squares tends to 0 for large amounts of data.

[10%]

Examiner's comments: Attempted by 75% of candidates. Part (a) should have been answered by setting up the optimisation problem in matrix notation. Candidate who did not do this had difficult algebra to contend with. Part (b) was mainly answered well. Part (c) was poorly answered and results in the data book should have been used to find the inverse. Part (d)-(i) and (ii) were easy point earners. Almost all did not answer part (d)-iii correctly by failing to subtract the linear trend from the data x_n.

Question 3

Part (a)

The joint pdf can be written as the product of a conditional and marginal pdf as follows,

$$\begin{aligned} p(x, y_1, \ldots, y_n) &= p(x)p(y_1, \ldots, y_n|x) \\ &= p(x)p(y_1|x) \cdots p(y_n|x_1) \end{aligned}$$

where the second equality follows since given $X = x$, the observations $Y_1, \ldots, Y_n$ are independent. Furthermore,

$$p(y_i|x) = (2\pi\sigma_w^2)^{-0.5} \exp\left(-0.5(y_i - x)^2/\sigma_w^2\right).$$

[10%]

Part (b)

The conditional density is

$$p(x|y_1) = \frac{p(x, y_1)}{p(y_1)} = \frac{p(x)p(y_1|x)}{p(y_1)}.$$

We can write the density $q(x)$ of a Gaussian random variable with mean m and variance s^2 as

$$\log q(x) = -\frac{1}{2s^2}x^2 + \frac{m}{s^2}x + C$$

where C denotes the terms not depending on x.

$$\begin{aligned} \log p(x|y_1) &= -\frac{1}{2\sigma^2}(x^2 - 2\mu x) - \frac{1}{2\sigma_w^2}(x^2 - 2y_1 x) + C \\ &= -\frac{1}{2}\left(x^2\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_w^2}\right) - 2x\left(\frac{\mu}{\sigma^2} + \frac{y_1}{\sigma_w^2}\right)\right) + C \\ &= -\frac{1}{2}\left(x^2\left(\frac{\sigma_w^2 + \sigma^2}{\sigma^2\sigma_w^2}\right) - 2x\left(\frac{\sigma_w^2\mu + \sigma^2 y_1}{\sigma^2\sigma_w^2}\right)\right) + C \\ &= -\frac{1}{2}\left(\frac{\sigma_w^2 + \sigma^2}{\sigma^2\sigma_w^2}\right)\left(x^2 - 2x\left(\frac{\sigma_w^2\mu + \sigma^2 y_1}{\sigma^2 + \sigma_w^2}\right)\right) + C \end{aligned}$$

which implies $p(x|y_1)$ is Gaussian with mean

$$\mu_1 = \frac{\sigma_w^2\mu + \sigma^2 y_1}{\sigma^2 + \sigma_w^2}$$

and variance

$$\sigma_1^2 = \frac{\sigma_w^2\sigma^2}{\sigma^2 + \sigma_w^2}.$$

[30%]

Part (c)

The density $p(x|y_1, y_2)$ is

$$\begin{aligned} p(x|y_1, y_2) &= \frac{p(x)p(y_1|x)p(y_2|x)}{p(y_1, y_2)} \\ &= \frac{p(x)p(y_1|x)}{p(y_1)} \frac{p(y_2|x)}{p(y_2|y_1)} \\ &= p(x|y_1) \frac{p(y_2|x)}{p(y_2|y_1)}. \end{aligned}$$

Using the result from the previous part, as $p(x|y_1)$ is $\mathcal{N}(\mu_1, \sigma_1^2)$, it follows that $p(x|y_1, y_2)$ is $\mathcal{N}(\mu_2, \sigma_2^2)$ where

$$\mu_2 = \frac{\sigma_w^2 \mu_1 + \sigma_1^2 y_2}{\sigma_1^2 + \sigma_w^2}$$

and

$$\sigma_2^2 = \frac{\sigma_w^2 \sigma_1^2}{\sigma_1^2 + \sigma_w^2}.$$

The same equations apply for other values of k:

$$\sigma_k^2 = \frac{\sigma_w^2 \sigma_{k-1}^2}{\sigma_{k-1}^2 + \sigma_w^2}$$

and

$$\mu_k = \frac{\sigma_w^2 \mu_{k-1} + \sigma_{k-1}^2 y_k}{\sigma_{k-1}^2 + \sigma_w^2}.$$

[20%]

Part (d)

We see that

$$V = \frac{1}{n} \left(W_1 + \ldots + W_n \right).$$

Any linear transformation of a vector of independent Gaussian random variables is still Gaussian. (The fact that V is Gaussian can be verified through its characteristic function. This is bookwork.) The variance of V is σ_w^2/n while its mean is zero.

Using the solution to the earlier part $p(x|z)$ is $\mathcal{N}(m, s^2)$ where

$$m = \frac{\mu \sigma_w^2/n + \sigma^2 z}{\sigma^2 + \sigma_w^2/n}$$

and

$$s^2 = \frac{\sigma^2 \ \sigma_w^2/n}{\sigma^2 + \sigma_w^2/n}.$$

[20%]

Part (e)

There is no difference between $p(x|y_1,\ldots,y_n)$ and $p(x|z)$. Let $nz = y_1+\ldots+y_n$.

$$\begin{aligned}\log p(x|y_1,\ldots,y_n) &= -\frac{1}{2\sigma^2}(x^2-2\mu x) - \frac{1}{2\sigma_w^2}(nx^2-2nzx)+C \\ &= -\frac{1}{2\sigma^2}(x^2-2\mu x) - \frac{1}{2\sigma_w^2/n}(x^2-2zx)+C\end{aligned}$$

In an earlier part we have already shown this to correspond to a Gaussian density with mean m and variance s^2. The second scheme though is more memory efficient since only the average of the observations are stored. This is important if n is very large.

[20%]

Examiner's comments: Attempted by 67% of candidates. Parts (a) and (b) were well answered by most though those that found part (a) difficult were ill equipped to answer the whole question properly. The proof in part (c) was circular in some cases though the arguments to be used were trivial. It was disappointing to see part (d) answered poorly by a significant majority even though part (b) equipped candidates with the necessary tool. Part (e) was poorly done by the vast majority and many failed to realise both techniques yield the same estimate for X.

Question 4

Part (a)

To find lag 0, square both sides and take the expectation

$$\mathbf{E}(X_n^2) = \alpha^2 \mathbf{E}(X_{n-1}^2) + \mathbf{E}(W_n^2) + 2\alpha \mathbf{E}(X_{n-1}W_n)$$

Thus $R_X(0) = \alpha^2 R_X(0) + \sigma^2$. To find $R_X(k)$ for $k > 0$, multiply both sides with X_{n-k} and take the expectation

$$\begin{aligned}\mathbf{E}\left(X_n X_{n-k}\right) &= \alpha \mathbf{E}\left(X_{n-1} X_{n-k}\right) + \mathbf{E}\left(W_n X_{n-k}\right)\\ R_X(k) &= \alpha R_X(k-1).\end{aligned}$$

So $R_X(k) = \alpha^k R_X(0)$ where $R_X(0) = \sigma^2/(1-\alpha^2)$. Also $R_X(-k) = R_X(k)$.

[20%]

Part (b)

The autocorrelation of U_n is

$$\begin{aligned}\mathbf{E}\left(U_n U_{n+k}\right) =& \mathbf{E}\left((X_n + V_n)(X_{n+k} + V_{n+k})\right)\\ &= \mathbf{E}\left(X_n X_{n+k}\right) + \mathbf{E}\left(V_n V_{n+k}\right) + \mathbf{E}\left(X_n V_{n+k}\right) + \mathbf{E}\left(X_{n+k} V_n\right)\\ &= R_X(k) + R_V(k) + 0 + 0.\end{aligned}$$

Since X_n is a weighted sum of all previous values of $W_n, W_{n-1}, \ldots$ and $\mathbf{E}\left(W_m V_{n+k}\right) = 0$, it follows that $\mathbf{E}\left(X_n V_{n+k}\right) = 0$. Similarly for the term $\mathbf{E}\left(X_{n+k} V_n\right) = 0$.

Finally, $R_V(0) = \sigma_v^2$ and $R_V(k) = 0$ for other values of k thus

$$R_U(k) = R_X(k) + \sigma_v^2 \mathbb{I}_{[k=0]}.$$

[15%]

Part (c)

Differentiate the cost function wrt to h_1 to get

$$\mathbf{E}\left(-2\left(X_n - \hat{X}_n\right) U_{n-1}\right) = \mathbf{E}\left(-2\left(X_n - h_1 U_{n-1}\right) U_{n-1}\right)$$

Set the derivative to zero to get

$$\begin{aligned}h_1 &= \frac{\mathbf{E}\left(X_n U_{n-1}\right)}{\mathbf{E}\left(U_{n-1}^2\right)}\\ &= \frac{\mathbf{E}\left(X_n X_{n-1}\right) + \mathbf{E}\left(X_n V_{n-1}\right)}{\mathbf{E}\left(X_{n-1}^2\right) + \mathbf{E}\left(V_{n-1}^2\right) + 0}\\ &= \frac{\mathbf{E}\left(X_n X_{n-1}\right)}{\mathbf{E}\left(X_{n-1}^2\right) + \mathbf{E}\left(V_{n-1}^2\right)}\\ &= \frac{\alpha R_X(0)}{R_X(0) + \sigma_v^2}\end{aligned}$$

and the solution is not α because $\sigma_v^2 > 0$. [15%]

Part (d)

We need to predict X_n using noisy measurements of its p previous values. Differentiate the cost function wrt to each h_i and set to zero. A vector formulation is more convenient.

Let

$$\hat{X}_n = h^T \left[U_{n-1}, \ldots, U_{n-p}\right]^T = h^T \overline{U}_{n-1} = h^T \overline{X}_{n-1} + h^T \overline{V}_{n-1}$$

where

$$\overline{X}_{n-1} = \left[X_{n-1}, \ldots, X_{n-p}\right]^T, \qquad \overline{V}_{n-1} = \left[V_{n-1}, \ldots, V_{n-p}\right]^T$$

The cost function is

$$\mathbf{E}\left((X_n - h^T \overline{U}_{n-1})^2\right).$$

Differentiating wrt to the vector h and set to 0

$$-2\mathbf{E}\left((X_n - h^T \overline{U}_{n-1})\overline{U}_{n-1}\right) = 0$$

or

$$\mathbf{E}\left(X_n \overline{U}_{n-1}\right) = \mathbf{E}\left(\overline{U}_{n-1}\overline{U}_{n-1}^T\right) h$$

or

$$\mathbf{E}\left(\overline{U}_{n-1}\overline{U}_{n-1}^T\right)^{-1} \mathbf{E}\left(X_n \overline{U}_{n-1}\right) = h.$$

[30%]

Part (e)

When $\sigma_v^2 = 0$,

$$\begin{aligned}
\mathbf{E}\left(\overline{U}_{n-1}\overline{U}_{n-1}^T\right) &= \mathbf{E}\left(\overline{X}_{n-1} X_{n-1}^T\right) \\
&= R_X(0) \begin{bmatrix} 1 & \alpha & \cdots & \alpha^{p-1} \\ \alpha & 1 & & \\ \vdots & & \ddots & \\ \alpha^{p-1} & & & 1 \end{bmatrix} \\
\mathbf{E}\left(X_n \overline{U}_{n-1}\right) &= R_X(0) \begin{bmatrix} \alpha \\ \vdots \\ \alpha^p \end{bmatrix}
\end{aligned}$$

Clearly if $h = [\alpha, 0, \ldots, 0]^T$it picks the first column of $\mathbf{E}\left(\overline{U}_{n-1}\overline{U}_{n-1}^T\right)$ which is equal to $\mathbf{E}\left(X_n \overline{U}_{n-1}\right)$. Thus in the absence of noise in the observation $p = 1$ is sufficient.

[20%]

Examiner's comments: Attempted by 97% of candidates and with the exception of part (e) done well by the majority. The optimal value for p in part (e) should have been found by using the solution for part (d) when the noise has zero variance.

S.S.Singh

May 2019