Module 4F7, Statistical Signal Analysis, Easter 2022

S.S.Singh

January 31, 2022

Question 1

Part a

$$\Sigma_x = \mathbf{E}(XX^T) = \begin{bmatrix} \mathbf{E}(X_1^2) & \mathbf{E}(X_1X_2) & \mathbf{E}(X_1X_3) \\ \mathbf{E}(X_2X_1) & \mathbf{E}(X_2^2) & \mathbf{E}(X_2X_3) \\ \mathbf{E}(X_3X_1) & \mathbf{E}(X_3X_2) & \mathbf{E}(X_3^2) \end{bmatrix}$$

so we need lag 0, lag 1 and lag 2 terms of the autocorrelation, which are $r_0 = \sigma^2(1-\alpha^2)$, $r_1 = \alpha r_0$and $r_3 = \alpha r_2$ respectively.

$$YY^T = XX^T + VV^T + XV^T + VX^T$$

and thus $\Sigma_y = \Sigma_x + I$ as cross terms have zero expectation. Similarly,

$$YX^T = XX^T + VX^T$$

and $\Sigma_{yx} = \Sigma_x$.

[Total marks: 30%]

Part b-(i)

To find b_i set $\mathbf{E}(\hat{X}_i) = \mathbf{E}(X_i)$ or $b_i + B_i\mathbf{E}(Y) = \mathbf{E}(X_i)$. Firstly, $\mathbf{E}(X_i) = \alpha\mathbf{E}(X_{i-1}) = 0$. Secondly, $\mathbf{E}(Y)$ is the zero vector. This gives $b_i = 0$.

To find B_i differentiate as follows:

$$\begin{aligned} e_i^2 &= X_i^2 + (\hat{X}_i)^2 - 2X_i\hat{X}_i \\ e_i^2 &= X_i^2 + (B_iY)^2 - 2B_iYX_i \\ \nabla e_i^2 &= 2(B_iY)Y - 2YX_i \\ &= 2Y(Y^TB_i^T) - 2YX_i \\ \mathbf{E}(\nabla e_i^2) &= 2\Sigma_yB_i^T - 2(\Sigma_{yx_i}) \end{aligned}$$

setting to zero gives $B_i^T = \Sigma_y^{-1}\Sigma_{yx_i}$ or $B_i = \Sigma_{yx_i}^T\Sigma_y^{-1}$ or

$$\hat{X} = \Sigma_{yx}^T\Sigma_y^{-1}Y$$

[Total marks: 30%]

Part b-(ii)

From the previous part

$$(X - \hat{X})(X - \hat{X})^T = XX^T + \hat{X}\hat{X}^T - \hat{X}X^T - X\hat{X}^T$$
$$\hat{X}X^T = \Sigma_{yx}^T \Sigma_y^{-1} Y X^T$$
$$X\hat{X}^T = XY^T \Sigma_y^{-1} \Sigma_{yx}$$
$$\hat{X}\hat{X}^T = \Sigma_{yx}^T \Sigma_y^{-1} Y Y^T \Sigma_y^{-1} \Sigma_{yx}$$

Taking the expected value gives

$$\mathbf{E}(\hat{X}X^T) = \Sigma_{yx}^T \Sigma_y^{-1} \Sigma_{yx}$$
$$\mathbf{E}(X\hat{X}^T) = \Sigma_{yx}^T \Sigma_y^{-1} \Sigma_{yx}$$
$$\mathbf{E}(\hat{X}\hat{X}^T) = \Sigma_{yx}^T \Sigma_y^{-1} \Sigma_{yx}$$

Adding all terms gives

$$\mathbf{E}\left[(X - \hat{X})(X - \hat{X})^T\right] = \Sigma_x - \Sigma_{yx}^T \Sigma_y^{-1} \Sigma_{yx}$$

[Total marks: 40%]

End of Question 1.

Question 2

Part a

$$p_\theta(x_1, y_1, \ldots, x_T, y_T) = p_\theta(x_1)p(y_1|x_1)\cdots p_\theta(x_T)p(y_T|x_T)$$

where

$$\begin{aligned} p_\theta(x_i) &= \frac{1}{\sqrt{2\pi}\theta}\exp\left(-\frac{x_i^2}{2\theta^2}\right) \\ p(y_i|x_i) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i - x_i)^2}{2\sigma^2}\right) \end{aligned}$$

[Total marks: 20%]

Part b

We can solve for

$$\begin{aligned} p_\theta(x_t, y_t) &= \frac{1}{2\pi\theta\sigma}\exp\left(-\frac{x_t^2}{2\theta^2} - \frac{(y_t - x_t)^2}{2\sigma^2}\right) \\ &= D\exp\left(x_t^2(-\frac{1}{2\theta^2} - \frac{1}{2\sigma^2}) + \frac{x_t y_t}{\sigma^2}\right) \\ p_\theta(x_t|y_t) &= C\exp\left(-\frac{x_t^2}{2}(\frac{1}{\theta^2} + \frac{1}{\sigma^2}) + \frac{x_t y_t}{\sigma^2}\right) \end{aligned}$$

where constant D contains non-x_t terms and constant C is the normalising constant so that $p_\theta(x_t|y_t)$ integrates to 1. Find answer to be a Gaussian with variance

$$(\frac{1}{\theta^2} + \frac{1}{\sigma^2})^{-1}$$

and mean

$$\frac{y_t}{\sigma^2}(\frac{1}{\theta^2} + \frac{1}{\sigma^2})^{-1} = y_t\frac{\theta^2}{\theta^2 + \sigma^2}$$

[Total marks: 20%]

Part c

$$\text{log} p_{\tilde{\theta}}(x_1)\cdots p_{\tilde{\theta}}(x_T) = -\frac{T}{2}\log(2\pi) - T\log(\tilde{\theta}) - \sum_{i=1}^{T}\frac{x_i^2}{2\tilde{\theta}^2}$$

Taking the expectation gives

$$\begin{aligned} \int x_t^2 p_{\hat{\theta}}(x_t|y_t)dx_t &= \text{mean}^2 + \text{variance} \\ &= y_t^2\left(\frac{\hat{\theta}^2}{\hat{\theta}^2 + \sigma^2}\right)^2 + (\frac{1}{\hat{\theta}^2} + \frac{1}{\sigma^2})^{-1} \\ &= s_t^2 \end{aligned}$$

$$Q(\hat{\theta}, \tilde{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\tilde{\theta}^2) - \sum_{t=1}^{T}\frac{s_t^2}{2\tilde{\theta}^2}$$

[Total marks: 20%]

Part d

Differentiating with respect to $\tilde{\theta}^2$ and set to 0 to get

$$-\frac{T}{2\tilde{\theta}^2} + \sum_{t=1}^{T} \frac{s_t^2}{2\tilde{\theta}^4} = 0$$

$$\tilde{\theta}^2 = \frac{1}{T} \sum_{t=1}^{T} s_t^2$$

When $\sigma = 0$, $s_t^2 = y_t^2 + 0 = y_t^2$.

[Total marks: 20%]

Part e

By differentiating with respect to $\tilde{\theta}^2 = \frac{1}{T} \sum_{t=1}^{T} x_t^2$ which does coincide the answer in part-(d) when $\sigma = 0$ since, for a noiseless observation, it must be that $y_t = x_t$.

[Total marks: 20%]

End of Question 2.

Question 3

Part a

When $p(y_n|x_n = i) = \frac{1}{\sqrt{2\pi}\theta_i} \exp\left(-\frac{1}{2\theta_i^2} y_n^2\right)$.

[Total marks: 5%]

Part b

If $S_n > S_{n-1}$ then $Y_n > 0$ and if $S_n < S_{n-1}$ then $Y_n < 0$. Thus Y_n can take positive and negative values as it is assumed by the model $p(y_n|x_n)$. In this case the instantaneous value X_n of the Markov sequence $X_1, X_2, \dots$ determines the variance of Y_n and the variance of the observation sequence assumes one of two values only, θ_1^2 or θ_2^2. Although this may be a simplification, it is still consistent with the data since $p(Y_1, \dots, Y_n) > 0$ for all possible values of $Y_1, \dots, Y_n$.

[Total marks: 5%]

Part c

$$\begin{aligned}
p_{n+1} &= \Pr(X_{n+1} = 1|y_{1:n}) \\
&= \Pr(X_{n+1} = 1, X_n = 1|y_{1:n}) + \Pr(X_{n+1} = 1, X_n = 2|y_{1:n}) \\
&= \Pr(X_{n+1} = 1|X_n = 1, y_{1:n}) \Pr(X_n = 1|y_{1:n}) \\
&\quad + \Pr(X_{n+1} = 1|X_n = 2, y_{1:n}) \Pr(X_n = 2|y_{1:n}) \\
&= (1 - q) \Pr(X_n = 1|y_{1:n}) \\
&\quad + q(1 - \Pr(X_n = 1|y_{1:n}))
\end{aligned}$$

Thus $\Pr(X_n = 1|y_{1:n})$ is needed which is

$$\frac{p_n \frac{1}{\sqrt{2\pi}\theta_1} \exp\left(-\frac{1}{2\theta_1^2} y_n^2\right)}{p_n \frac{1}{\sqrt{2\pi}\theta_1} \exp\left(-\frac{1}{2\theta_1^2} y_n^2\right) + (1 - p_n) \frac{1}{\sqrt{2\pi}\theta_2} \exp\left(-\frac{1}{2\theta_2^2} y_n^2\right)}$$

[Total marks: 20%]

Part d

$$\begin{aligned}
a_n &= p(y_{n:T}|X_n = 1) \\
&= p(y_n|X_n = 1)p(y_{n+1:T}|X_n = 1) \\
&= p(y_n|X_n = 1)p(y_{n+1:T}, X_{n+1} = 1|X_n = 1) \\
&+ p(y_n|X_n = 1)p(y_{n+1:T}, X_{n+1} = 2|X_n = 1) \\
&= p(y_n|X_n = 1)a_{n+1}(1 - q) + p(y_n|X_n = 1)b_{n+1}q
\end{aligned}$$

Similarly

$$\begin{aligned}
b_n &= p(y_{n:T}|X_n = 2) \\
&= p(y_n|X_n = 2)p(y_{n+1:T}|X_n = 2) \\
&= p(y_n|X_n = 2)p(y_{n+1:T}, X_{n+1} = 1|X_n = 2) \\
&+ p(y_n|X_n = 2)p(y_{n+1:T}, X_{n+1} = 2|X_n = 2) \\
&= p(y_n|X_n = 2)a_{n+1}q + p(y_n|X_n = 2)b_{n+1}(1 - q)
\end{aligned}$$

[Total marks: 20%]

Part e

$$\begin{aligned}\Pr(X_n = 1|y_{1:T}) &= \frac{\Pr(X_n = 1|y_{1:n-1})p(y_{n:T}|X_n = 1)}{\Pr(X_n = 1|y_{1:n-1})p(y_{n:T}|X_n = 1) + \Pr(X_n = 2|y_{1:n-1})p(y_{n:T}|X_n = 2)} \\ &= \frac{p_n a_n}{p_n a_n + (1 - p_n)b_n}\end{aligned}$$

[Total marks: 10%]

Part f

$$\frac{d}{d\theta_1} \log p(y_n|1) = -\frac{1}{\theta_1} + \frac{1}{\theta_1^3} y_n^2$$

Thus answer sought is

$$\left(-\frac{1}{\theta_1} + \frac{1}{\theta_1^3} y_n^2\right) \pi_n$$

and

$$\left(-\frac{1}{\theta_2} + \frac{1}{\theta_2^3} y_n^2\right) (1 - \pi_n)$$

[Total marks: 20%]

Part g

$$\begin{aligned}\frac{d}{d\theta_i} \log p(y_{1:T}) &= \frac{1}{p(y_{1:T})} \frac{d}{d\theta_i} \sum_{x_{1:T}} p(y_{1:T}, x_{1:T}) \\ &= \frac{1}{p(y_{1:T})} \frac{d}{d\theta_i} \sum_{x_{1:T}} p(y_{1:T} x_{1:T}) \\ &= \frac{1}{p(y_{1:T})} \sum_{x_{1:T}} \frac{d}{d\theta_i} \log p(y_{1:T} x_{1:T}) p(x_{1:T}, y_{1:T}) \\ &= \sum_{x_{1:T}} \frac{d}{d\theta_i} \log p(y_{1:T} x_{1:T}) p(x_{1:T}|y_{1:T}) \\ &= \sum_{x_{1:T}} \frac{d}{d\theta_i} \sum_{t=1}^{T} \log p(y_t|x_t) p(x_{1:T}|y_{1:T}) \\ &= \sum_{t=1}^{T} \sum_{x_t} \frac{d}{d\theta_i} \log p(y_t|x_t) p(x_t|y_{1:T}) \\ &= \sum_{t=1}^{T} \frac{d}{d\theta_i} \log p(y_t|x_t = i) p(x_t = i|y_{1:T})\end{aligned}$$

The answer is

$$\sum_{t=1}^{T} \left(-\frac{1}{\theta_1} + \frac{1}{\theta_1^3} y_t^2\right) \pi_t$$

and

$$\sum_{t=1}^{T} \left(-\frac{1}{\theta_2} + \frac{1}{\theta_2^3} y_n^2\right) (1 - \pi_t)$$

[Total marks: 20%]

End of Question 3.

Question 4

Part a

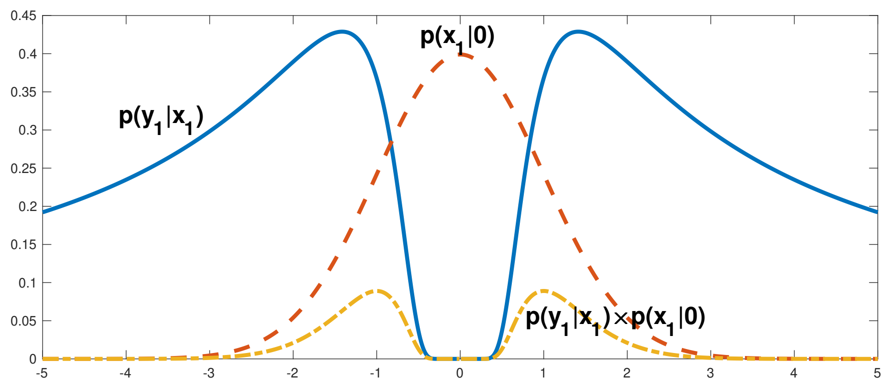
$$p(y_1|x_1) = \frac{1}{\sqrt{2\pi x_1^2}} \exp\left(-\frac{1}{2x_1^2}(y_1-\theta)^2\right)$$

Take the log and differentiate to find this mode:

$$\frac{d}{dx_1^2} \log p(y_1|x_1) = -\frac{1}{2x_1^2} + \frac{1}{2x_1^4}(y_1-\theta)^2$$

or $x_1^2 = (y_1-\theta)^2$. So there are two distinct modes for x_1. The function $p(y_1|x_1)$ is bimodal too, close to zero for large x_1^2 and similarly becoming zero as $x_1^2 \to 0$ since the variance is 0 and the gaussian density will be 0 unless $y_1 = \theta$.

$p(x_1|x_0)$ is a gaussian centered at $\alpha x_0 = 0$. So $p(x_1|x_0)$ is symmetric around 0. The posterior will be bi-modal. The exact plot is - only a sketch showing unimodality and bimodality is needed.



[Total marks: 30%]

Part b

Bayes theorem can be expressed as a prediction step and an update step.

The prediction step is

$$p(x_{t+1}|y_{1:t}) = \int p(x_{t+1}|x_t)p(x_t|y_{1:t})dx_t.$$

The update step is

$$p(x_{t+1}|y_{1:t+1}) = \frac{p(y_{t+1}|x_t)p(x_{t+1}|y_{1:t})}{\int p(y_{t+1}|x_t)p(x_{t+1}|y_{1:t})dx_{t+1}}$$

[Total marks: 20%]

Part c

$$\begin{aligned}\frac{d}{d\theta}\log p(y_{1:T}, x_{1:T}) &= \frac{d}{d\theta}\log\left(p(x_1|x_0)p(y_1|x_1)\cdots p(x_T|x_{T-1})p(y_T|x_T)\right)\\ &= \sum_{t=1}^{T}\frac{d}{d\theta}\log p(y_t|x_t)\\ &= \sum_{t=1}^{T}\frac{y_t-\theta}{x_t^2}\end{aligned}$$

[Total marks: 20%]

Part d

$$
\begin{aligned}
\frac{d}{d\theta}\log p(y_{1:T}) &= \frac{1}{p(y_{1:T})}\frac{d}{d\theta}p(y_{1:T}) \\
\frac{d}{d\theta}p(y_{1:T}) &= \frac{d}{d\theta}\int p(y_{1:T}, x_{1:T})dx_{1:T} \\
&= \int \frac{d}{d\theta}p(y_{1:T}, x_{1:T})dx_{1:T} \\
&= \int \left(\frac{d}{d\theta}\log p(y_{1:T}, x_{1:T})\right) p(y_{1:T}, x_{1:T})dx_{1:T} \\
\frac{d}{d\theta}\log p(y_{1:T}) &= \int \left(\frac{d}{d\theta}\log p(y_{1:T}, x_{1:T})\right) p(x_{1:T}|y_{1:T})dx_{1:T} \\
&= \sum_{t=1}^{T}\int \frac{d}{d\theta}\log p(y_t|x_t)p(x_{1:T}|y_{1:T})dx_{1:T} \\
&= \sum_{t=1}^{T}\int \frac{d}{d\theta}\log p(y_t|x_t)p(x_t|y_{1:T})dx_{1:T} \\
&= \sum_{t=1}^{T}(y_t-\theta)s_t
\end{aligned}
$$

[Total marks: 20%]

Part e

Convert the samples into importance samples from $p(x_{1:T+1}|y_{1:T+1})$

Prediction step:

Sample $X^i_{T+1} \sim p(x_{T+1}|X^i_T)$ and extend the particle to get $X^i_{1:T+1}$. Do this for $i = 1, \dots, N$.

Update step:

Assign the particle $X^i_{1:T+1}$ the weight $W^i_{T+1} = p(y_{T+1}|X^i_{T+1})$. Do this for $i = 1, \dots, N$. From the previous part, the desired estimate of $\frac{d}{d\theta}\log p(y_{1:T+1})$ is

$$\sum_{t=1}^{T+1}(y_t-\theta)c_t$$

where

$$c_t = \frac{\sum_{i=1}^{N}\frac{1}{X^i_t}\frac{1}{X^i_t}W^i_{T+1}}{\sum_{i=1}^{N}W^i_{T+1}}.$$

[Total marks: 10%]

End of Question 4.