4F7 worked solutions May 2023

May 22, 2023

Question 1

Examiner's comments:

A popular and well answered question. Parts a) and b) well done, although details in b) were often inaccurate and few candidates managed the full solution. c) was surprisingly poor given that it could technically be done by 3F3/3F8 students and the simplification result is in the data book.

Part a

Solve for h_0: let $\bar{Y}_i = Y_i - m_i$ then

$$\frac{d}{dh_0}\mathbf{E}\left\{\left(X - h_0 - h_1\bar{Y}_1 - \cdots - h_n\bar{Y}_n\right)^2\right\} = 0$$

which gives $h_0 = \mathbf{E}(X) = x$.

Solving similarly, the solution for $\mathbf{h} = (h_1, \ldots, h_n)^T$ solves

$$\Sigma\mathbf{h} = \mathbf{p}$$

where $(\Sigma)_{i,j} = \mathrm{cov}(Y_i, Y_j)$ and $(\mathbf{p})_i = \mathrm{cov}(X, Y_i)$.

So

$$A = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & \Sigma & \\ 0 & & & \end{bmatrix}, \qquad b = \begin{bmatrix} x \\ \mathbf{p} \end{bmatrix}$$

[30%]

Part b

We obtain immediately $h_0 = \mathbf{E}(X) = 0$.

$$\begin{aligned} \mathrm{cov}(Y_i, Y_j) &= \mathbf{E}(Y_iY_j) - \mathbf{E}(Y_i)\mathbf{E}(Y_j) \\ &= \mathbf{E}(X^2) + \mathbf{E}(W_iW_j) - \mathbf{E}(|X|)\mathbf{E}(|X|) \end{aligned}$$

and $\mathbf{E}(W_iW_j) = 0$ when $i \neq j$.

$$\begin{aligned}\mathrm{cov}(Y_i, X) &= \mathbf{E}(Y_i X) - \mathbf{E}(Y_i)\mathbf{E}(X) \\ &= \mathbf{E}(|X|X)\end{aligned}$$

Thus

$$\Sigma = diag(\sigma^2) + (\mathbf{E}(X^2) - \mathbf{E}(|X|)\mathbf{E}(|X|))\mathbf{1}^{n\times n}$$

and

$$\mathbf{p} = \mathbf{E}(|X|X)\mathbf{1}^{n\times 1}$$

where $\mathbf{1}^{a\times b}$ is a $a\times b$ matrix containing all ones and $diag(a)$ is a diagonal matrix with the value a down its leading diagonal and zeroes elsewhere.

Subtracting row i from row j on left and right leads to $h_i = h_j = h$, say. Then substituting $h_i = h$ into any row gives:

$$(n(\mathbf{E}(X^2) - \mathbf{E}(|X|)\mathbf{E}(|X|)) + \sigma^2)h = \mathbf{E}(|X|X)$$

from which h is obtained as:

$$h = \frac{\mathbf{E}(|X|X)}{(n(\mathbf{E}(X^2) - \mathbf{E}(|X|)\mathbf{E}(|X|)) + \sigma^2)}$$

[30%]

Part c.i

$$\int_{-\infty}^{\infty} xp(x, y_1)dx = \int_{-\infty}^{\infty} xp(x)p(y_1||x|) = 0$$

since $xp(x)p(y_1||x|)$ is an odd function of x. Thus $\mathbf{E}(X|Y_1 = y_1) = 0$. [30%]

Part c.ii

Since $\mathrm{cov}(Y_i, X) = \mathbf{E}(X|X|) = 0$ in this case, because:

$$\mathbf{E}(X|X|) = \mathbf{E}(X^2 sign(X)) = \int X^2 sign(X)p(X)dX = 0$$

since $X^2 sign(X)p(X)$ is an odd function for any symmetric $p(X)$, e.g. zero mean Gaussian as in this case, we have $h = 0$ and thus $\hat{X} = 0 = \mathbf{E}(X|Y_1 = y_1)$. [20%]

Question 2

Examiner's comments:

A much less popular question, though well handled by most. No parts caused particular problems, candidates were generally well on top of this material.

Part a

With $c_k = 0$ we have

$$
\begin{aligned}
&p\left(X_{k+1} = (c_{k+1}, d_{k+1}) \,|\, X_k = (c_k = 0, d_k)\right) \\
&= \begin{cases} \theta_3 d_k, & \text{if } (c_{k+1}, d_{k+1}) = (c_k, d_k - 1) \\ 1 - \theta_3 d_k & \text{if } (c_{k+1}, d_{k+1}) = (c_k, d_k) \\ 0 & \textit{Otherwise} \end{cases}
\end{aligned}
$$

and since prey count is zero neither prey nor predator numbers can increase. Candidates gave solutions both using the transition probability matrix and state transition diagram, both of which received credit. The transition matrix approach is:

$$
\begin{bmatrix}
1 & 0 & \cdots & & & \\
\theta_3 & 1-\theta_3 & 0 & \cdots & & \\
 & 2\theta_3 & 1-2\theta_3 & 0 & \cdots & \\
 & & & & & \\
0 & \cdots & & 0 & M\theta_3 & 1-M\theta_3
\end{bmatrix}
$$

Here row i is $X_k = (0, d_k)$ and column j corresponds to $X_{k+1} = (0, d_{k+1})$.[15%]

Part b

$$
p(d_0, \ldots, d_n | c_0, \ldots, c_n) = \frac{p(d_0, \ldots, d_n, c_0, \ldots, c_n)}{p(c_0, \ldots, c_n)} \frac{p(c_0, d_0, \ldots, c_n, d_n)}{\sum_{d_0=1}^{M_d} \cdots \sum_{d_n=1}^{M_d} p(c_0, d_0, \ldots, c_n, d_n)}
$$

Use the Markov property to further express the numerator and denominator as

$$
p(c_0, d_0, \ldots, c_n, d_n) = p(c_0, d_0) p(c_1, d_1 | c_0, d_0) \cdots p(c_n, d_n | c_{n-1}, d_{n-1})
$$

[15%]

Part c

The prediction step is:

$$
p(d_{n-1}, c_n, d_n | c_{0:n-1}) = p(c_n, d_n | c_{n-1}, d_{n-1}) \times p(d_{n-1} | c_{0:n-1})
$$

So the marginal of interest is

$$p(c_n, d_n | c_{0:n-1}) = \sum_{d_{n-1}=0}^{M_d} p(c_n, d_n | c_{n-1}, d_{n-1}) \times p(d_{n-1} | c_{0:n-1})$$

The update step is:

$$p(d_n | c_{0:n}) = \frac{p(c_n, d_n | c_{0:n-1})}{\sum_{d_n=1}^{M_d} p(c_n, d_n | c_{0:n-1})}$$

[30%]

Part d

[Changing T for n]:

$$\log p(c_0, \ldots, c_n) = \log \left(\sum_{d_0} \cdots \sum_{d_n} p(c_0, d_0, \ldots, c_n, d_n) \right)$$

Differentiate with respect to θ_i, using the result $\frac{d \log(f)}{dx} = 1/f \frac{df}{dx}$, to get

$$\begin{aligned}
\frac{d}{d\theta_i} \log p(c_0, \ldots, c_n) &= \frac{\sum_{d_0} \cdots \sum_{d_n} \frac{d}{d\theta_i} p(c_0, d_0, \ldots, c_n, d_n)}{\sum_{d_0} \cdots \sum_{d_n} p(c_0, d_0, \ldots, c_n, d_n)} \\
&= \frac{\sum_{d_0} \cdots \sum_{d_n} p(c_0, d_0, \ldots, c_n, d_n) \frac{d}{d\theta_i} \log p(c_0, d_0, \ldots, c_n, d_n)}{p(c_0, \ldots, c_n)} \\
&= \frac{\sum_{d_0} \cdots \sum_{d_n} p(d_0, \ldots, d_n | c_0, \ldots, c_n) \frac{d}{d\theta_i} \log p(c_0, d_0, \ldots, c_n, d_n)}{p(c_0, \ldots, c_n)}
\end{aligned}$$

[20%]

Part e

Transition probabilities have been found in Part a. Differentiating gives

$$\begin{aligned}
\frac{d}{d\theta_3} \log p(0, d_k | 0, d_{k-1}) &= \frac{\frac{d}{d\theta_3} p(0, d_k | 0, d_{k-1})}{p(0, d_k | 0, d_{k-1})} \\
&= \begin{cases} \frac{d_{k-1}}{\theta_3 d_{k-1}} & \text{if } d_k = d_{k-1} - 1 \\ \frac{-d_{k-1}}{1 - \theta_3 d_{k-1}} & \text{if } d_k = d_{k-1} \end{cases}
\end{aligned}$$

Finally,

$$\frac{d}{d\theta_1} \log p(0, d_k | 0, d_{k-1}) = \frac{d}{d\theta_2} \log p(0, d_k | 0, d_{k-1}) = 0$$

[20%]

Question 3

Examiner's comments:

A complex and (with hindsight) probably over long question. Few candidates made a full correct solution to b) because of the subtlety of the problem, and 'reasonable attempts were given quite a lot of credit. Few candidates really understood what was being asked in c).

Part a.i

The estimate is

$$\frac{1}{N}\sum_{i=1}^{N} h(\theta^i)\frac{p(\theta^i|y)}{q(\theta^i)}$$

[no self-normalisation required since $p(\theta|y)$ can be calculated explicitly.]

The mean (first moment) of the estimate is

$$m = \mathbf{E}_q(1/N\sum h(\theta^i)p(\theta^i|y) = N/N\int h(\theta)p(\theta|y)q(\theta)/q(\theta)d\theta = \int h(\theta)p(\theta|y)d\theta$$

$$= \frac{1}{2\alpha}\int_{y-\alpha}^{y+\alpha} h(\theta)p(\theta)/p(y)d\theta$$

because $p(\theta|y) = p(y|\theta)p(\theta)/p(y) = \frac{1}{2\alpha}p(\theta)/p(y)$ for $\theta \in [y-\alpha, y+\alpha)$.

The variance is

$$\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}_q\left(\left(h(\theta^i)\frac{p(\theta^i|y)}{q(\theta^i)} - m\right)^2\right) = \frac{1}{N}\mathbf{E}_q\left(\left(h(\theta^1)\frac{p(\theta^1|y)}{q(\theta^1)}\right)^2\right) - \frac{1}{N}m^2$$

$$= \frac{1}{N}\left(\int h(\theta)^2\frac{p(\theta|y)^2}{q(\theta)}d\theta - m^2\right)$$

$$= \frac{1}{N}\left(\frac{1}{4\alpha^2}\int_{y-\alpha}^{y+\alpha} h(\theta)^2\frac{(p(\theta)/p(y))^2}{q(\theta)}d\theta - m^2\right)$$

[20%]

Part a.ii

Now use $q(\theta) = p(\theta)$ and

$$p(y|\theta) = 1/(2\alpha), \ \ \theta \in [y-\alpha, y+\alpha)$$

to get

$$\frac{1}{N}\left(\frac{1}{4\alpha^2}\int_{y-\alpha}^{y+\alpha} h(\theta)^2\frac{(p(\theta)/p(y))^2}{q(\theta)}d\theta - m^2\right) = \frac{1}{N}\left(\frac{1}{4\alpha^2}\int_{y-\alpha}^{y+\alpha} h(\theta)^2\frac{p(\theta)}{p(y)^2}d\theta - m^2\right)$$

[10%]

Part b.i

$$\pi_k = \begin{cases} \frac{p(\theta)\prod_{i=1}^k p(y_i|\theta)}{\int p(\theta)\prod_{i=1}^k p(y_i|\theta)d\theta} = \frac{p(\theta)\prod_i 1/(2\alpha_i)}{\prod_i 1/(2\alpha_i)\int_{y-min(\alpha_i)}^{y+min(\alpha_i)} p(\theta)d\theta} = \frac{p(\theta)}{\int_{y-min(\alpha^i)}^{y+min(\alpha_i)} p(\theta)d\theta}, & \theta \in [y - min(\alpha_i), y + min(\alpha_i)) \\ 0, & otherwise \end{cases}$$

where the minumum is over $i \in 1, 2, ..., k$.

There is no prediction step since the parameter θ does not change over time. The update step is

$$\begin{aligned} \pi_{k+1}(\theta) &= \frac{p(y_{k+1}|\theta)\pi_k(\theta)}{\int p(y_{k+1}|\theta)\pi_k(\theta)\, d\theta} \\ &= \begin{cases} \frac{\pi_k(\theta)}{\int_{y_{k+1}-a_{k+1}}^{y_{k+1}+a_{k+1}} \pi_k(\theta)\, d\theta} & \text{if } |\theta - y_{k+1}| \le a_{k+1} \\ 0 & \text{if } |\theta - y_{k+1}| > a_{k+1} \end{cases} \end{aligned}$$

[10%]

Part b.ii

$$\begin{aligned} p(y_{1:k}) &= \int p(\theta)p(y_1|\theta)\cdots p(y_k|\theta)d\theta \\ \frac{d}{da_1}p(y_{1:k}) &= \int p(\theta)\left(\frac{d}{da_1}p(y_1|\theta)\right)p(y_2|\theta)\cdots p(y_k|\theta)d\theta \\ &= \int p(\theta, y_{1:k})\left(\frac{d}{da_1}\log p(y_1|\theta)\right)d\theta \end{aligned}$$

where we have used the result $\frac{d\log(f)}{dx} = 1/f\,\frac{df}{dx}$.

Thus

$$\frac{d}{da_1}\log p(y_{1:k}) = \int p(\theta|y_{1:k})\left(\frac{d}{da_1}\log p(y_1|\theta)\right)d\theta$$

The importance sampling estimate is

$$\left(\frac{1}{\sum_{j=1}^N w_k^j}\right)\sum_{i=1}^N\left(\frac{d}{da_1}\log p(y_1|\theta_k^i)\right)w_k^i$$

[20%]

Part b.iii

Resample the particles to get new samples $\theta_k^{J_1}, \ldots, \theta_k^{J_N}$ and assign all these particles the same weight, namely

$$\bar{w}_k = \frac{1}{N}\sum_{i=1}^N w_k^i.$$

Now approximate $\pi_{k+1}(\theta)$ by approximating the update step with the resampled and reweighted particles:

..

$$\begin{aligned}
\int h(\theta)\pi_{k+1}(\theta)d\theta &= \frac{\int h(\theta)p(y_{k+1}|\theta)\pi_k(\theta)\ d\theta}{\int p(y_{k+1}|\theta)\pi_k(\theta)\ d\theta} \\
&= \frac{\int h(\theta)p(y_{k+1}|\theta)p(\theta, y_{1:k})\ d\theta}{\int p(y_{k+1}|\theta)p(\theta, y_{1:k})\ d\theta} \\
&\approx \frac{h(\theta_k^{J_1})p(y_{k+1}|\theta_k^{J_1})\bar{w}_k + \cdots + h(\theta_k^{J_N})p(y_{k+1}|\theta_k^{J_N})\bar{w}_k}{p(y_{k+1}|\theta_k^{J_1})\bar{w}_k + \cdots + p(y_{k+1}|\theta_k^{J_N})\bar{w}_k}
\end{aligned}$$

Note that the term $\bar{w}_k$ cancels out.

[20%]

Part c.i

You can choose any positive non-increasing sequence $a_1 \geq a_2 \geq \cdots \geq a_T = \alpha$, since

$$p(Y_1 = y|\theta)p(Y_2 = y|\theta)\cdots p(Y_T = y|\theta) = p(Y = y|\theta) \times \text{constant}$$

where term ‘constant’ does not depend on θ.

[10%]

Part c.ii

No benefit since the original set of samples from time 1, namely $\theta_1^1, \ldots, \theta_1^N$ are resampled at every time step which will result in the number of unique samples diminishing. Also, the estimate of Part (a) does not have to estimate the denominator $p(y)$.

[10%]

Question 4

Examiner's comments:

A well answered and popular question. Most made a very good attempt at a), although there was confusion about how to deal with h0. b) required some insight about the integrals and most candidates did not spot the symmetry in the integrals which leads to the answer of 0

Part a.i

Set $\hat{X}_0 = m$ and verify $\mathbf{E}(\hat{X}_1) = m$. Now do the same verification to show $\mathbf{E}(\hat{X}_n) = m$ when $\mathbf{E}(\hat{X}_{n-1}) = m$.

[10%]

Part a.ii

Square the expression for $\hat{X}_n - X$:

$$\begin{aligned}
\hat{X}_n - X &= G_n(V_n + X - \hat{X}_{n-1}) + \hat{X}_{n-1} - X \\
&= G_n V_n + (1 - G_n)(\hat{X}_{n-1} - X) \\
(\hat{X}_n - X)^2 &= G_n^2(V^2 + (1 - G_n)^2(X - \hat{X}_{n-1})^2 + 2G_n V_n(1 - G_n)(\hat{X}_{n-1} - X) \\
\mathbf{E}(\hat{X}_n - X)^2 &= G_n^2\sigma_v^2 + (1 - G_n)^2 s_{n-1}^2 + 0
\end{aligned}$$

where $s_{n-1}^2 = \mathbf{E}((X - \hat{X}_{n-1})^2)$.

Differentiate with respect to G_n to find minimising G_n which is $G_n = s_{n-1}^2/(s_{n-1}^2 + \sigma_v^2)$.

[30%]

Part b.i

(x, y_1) will be bivariate Gaussian as it is a product of 2 Gaussians.

Covariance matrix for (x, y_1) is

$$S = \begin{bmatrix} \mathbf{E}X^2 & \mathbf{E}XY_1 \\ \mathbf{E}XY_1 & \mathbf{E}Y_1^2 \end{bmatrix} - \begin{bmatrix} \mathbf{E}X \\ \mathbf{E}Y_1 \end{bmatrix} \begin{bmatrix} \mathbf{E}X \\ \mathbf{E}Y_1 \end{bmatrix}^T = \begin{bmatrix} \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 + \sigma_v^2 \end{bmatrix}$$

$p(x, y_1)$ is Gaussian with mean $(m, m)^T$ and covariance S. From the information engineering data book, or from first principles, $p(x|y_1)$ is Gaussian with mean $\mu_1 = m + (y_1 - m)\sigma_0^2/(\sigma_0^2 + \sigma_v^2) = (m\sigma_v^2 + \sigma_0^2 y_1)/(\sigma_0^2 + \sigma_v^2)$ and variance $\sigma_1^2 = \sigma_0^2 - \sigma_0^4/(\sigma_0^2 + \sigma_v^2) = \sigma_0^2\sigma_v^2/(\sigma_0^2 + \sigma_v^2)$.

This part can also be done from first principles from the Gaussian formulae and not using data book - many candidates did this successfully.

Part b.ii

$$p(x|y_1,y_2) = p(y_2|x)\frac{p(y_1|x)p(x)}{p(y_1,y_2)} = p(y_2|x)\frac{p(y_1|x)p(x)}{p(y_1)}\frac{1}{p(y_2|y_1)}$$

which is the Bayes update formula for a prior $p(x|y_1)$ and likelihood $p(y_2|x)$. We found this to be Gaussian with mean $\mu_2 = \mu_1 + (y_2 - \mu_1)\sigma_1^2/(\sigma_1^2 + \sigma_v^2)$ and variance $\sigma_2^2 = \sigma_1^2\sigma_v^2/(\sigma_1^2 + \sigma_v^2)$. Iterating gives the desired mean and variance solution. Can also be done from first principles (multiplying out the Gaussians and completing the squares of the exponent), resulting in: $\mu_n = (m\sigma_v^2 + \sigma_0^2 \sum_{i=1}^n y_i)/(n\sigma_0^2 + \sigma_v^2)$ and variance $\sigma_n^2 = \sigma_0^2\sigma_v^2/(n\sigma_0^2 + \sigma_v^2)$

Part b.iii

Use $\mathbf{E}(X|y_{1:n}) = \mu_n$. This is unbiased because

$$\int \mathbf{E}(X|y_{1:n})p(y_{1:n})dy_{1:n} = \int\int xp(x|y_{1:n})p(y_{1:n})dxdy_{1:n} = \int xp(x)dx = \mathbf{E}(X)$$