4F7- Statistical Signal Analysis, 2019. Crib and examiner's report.

May 20, 2019

Question 1

Part (a)

Differentiate wrt to each h_i

$$\frac{\partial}{\partial h_i}\mathbf{E}\left\{(X-\hat{X}_n)^2\right\} = \mathbf{E}\left\{-2(X-\hat{X}_n)(Y_i - \mathbf{E}\{Y_i\})\right\}$$

for $i = 1, \ldots, n$. Let

$$\bar{Y} = \begin{bmatrix} Y_1 - \mathbf{E}\{Y_1\} \\ \vdots \\ Y_n - \mathbf{E}\{Y_n\} \end{bmatrix}, \qquad \bar{h} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}.$$

Note that

$$\hat{X}_n = \bar{h}^T\bar{Y} = \bar{Y}^T\bar{h}.$$

Setting the derivative to zero, in vector form, gives

$$\begin{aligned} \mathbf{E}\left\{(X-\hat{X}_n)\bar{Y}\right\} &= 0 \\ \mathbf{E}\left\{X\bar{Y}\right\} &= \mathbf{E}(\bar{Y}\bar{Y}^T)\bar{h} \\ b &= A\bar{h}. \end{aligned}$$

[30%]

Part (b)-i

This can be seen from $A\bar{h} = b$ directly. Alternatively, differentiate the cost function wrt h_n and set the derivative to zero, that is $\frac{\partial}{\partial h_n}\mathbf{E}\left\{(X-\hat{X}_n)^2\right\} = 0$

yields

$$\begin{aligned}
\mathbf{E}\left\{(X-\hat{X}_n)(Y_n-\mathbf{E}\{Y_n\})\right\} &= 0 \\
\mathbf{E}\{X(Y_n-\mathbf{E}\{Y_n\})\} &= \sum_{i=1}^{n} h_i \mathbf{E}\{(Y_i-\mathbf{E}\{Y_i\})(Y_n-\mathbf{E}\{Y_n\})\} \\
&= \sum_{i=1}^{n} h_i \left(\mathbf{E}\{Y_iY_n\}-\mathbf{E}\{Y_i\}\mathbf{E}\{Y_n\}\right) \\
&= h_n\left(\mathbf{E}\left\{Y_n^2\right\}-\mathbf{E}\{Y_n\}^2\right).
\end{aligned}$$

Note that $\mathbf{E}\{X(Y_n-\mathbf{E}\{Y_n\})\}=\mathbf{E}\{XY_n\}$ as $\mathbf{E}\{X\}=0$. Thus

$$h_n = \frac{\mathbf{E}\{XY_n\}}{\mathbf{E}\{Y_n^2\}-\mathbf{E}\{Y_n\}^2}.$$

[10%]

Part (b)-ii

The last row of the matrix A is

$$[0,\ldots,0,\mathbf{E}\{(Y_n-\mathbf{E}\{Y_n\})(Y_n-\mathbf{E}\{Y_n\})\}]$$

since $\mathbf{E}\{(Y_i-\mathbf{E}\{Y_i\})(Y_n-\mathbf{E}\{Y_n\})\}=0$ for $i\neq n$. Since A is symmetric, the last column is the transpose of the last row. Thus

$$A = \begin{bmatrix} & & 0 \\ & B & \vdots \\ & & 0 \\ 0,\ldots,0 & & \mathbf{E}\left\{Y_n^2\right\}-\mathbf{E}\{Y_n\}^2 \end{bmatrix}$$

From part (a), we can see that the solution to the best linear estimate $\hat{X}_{n-1}$ is given by reduced problem

$$B\begin{bmatrix} h_1 \\ \vdots \\ h_{n-1}\end{bmatrix} = \begin{bmatrix}\mathbf{E}\{XY_1\} \\ \vdots \\ \mathbf{E}\{XY_{n-1}\}\end{bmatrix} \tag{1}$$

and thus we can express $\hat{X}_n$ as

$$\begin{aligned}
\hat{X}_n &= \hat{X}_{n-1} + h_n(Y_n-\mathbf{E}\{Y_n\}). \\
&= \hat{X}_{n-1} + \frac{\mathbf{E}\{XY_n\}}{\mathbf{E}\{Y_n^2\}-\mathbf{E}\{Y_n\}^2}(Y_n-\mathbf{E}\{Y_n\}).
\end{aligned}$$

[20%]

Part (c)-i

Note $\mathbf{E}\{Y_i\} = 0$ for $i = 1, \ldots, n-1$ since $\text{sign}(X)$ is either 1 or -1 with equal probabilities. Thus

$$\hat{X}_{n-1} = h_1 Y_1 + \ldots + h_{n-1} Y_{n-1}$$

given by the solution of (1). Also

$$\mathbf{E}\{Y_i Y_j\} = 1$$

for $i \neq j$, $\mathbf{E}\left\{Y_i^2\right\} = 1 + \sigma^2$ and also

$$\mathbf{E}\{Y_i X\} = \mathbf{E}\{\text{sign}(X)X\} + \mathbf{E}\{W_i X\} = \mathbf{E}\{|X|\}$$

and thus

$$\begin{bmatrix} 1+\sigma^2 & 1 & \cdots & 1 \\ 1 & 1+\sigma^2 & \vdots & \vdots \\ \vdots & & \ddots & 1 \\ 1 & 1 & \cdots & 1+\sigma^2 \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_{n-1} \end{bmatrix} = \mathbf{E}\{|X|\} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

[20%]

Part (c)-ii

For the measurement Y_n,

$$\begin{aligned} \mathbf{E}\{Y_i Y_n\} &= \mathbf{E}\{(\text{sign}(X) + W_i)(|X| + W_n)\} \\ &= \mathbf{E}\{\text{sign}(X)|X|\} + \mathbf{E}\{W_i|X|\} + \mathbf{E}\{\text{sign}(X)W_n\} + \mathbf{E}\{W_i W_n\} \\ &= \mathbf{E}\{\text{sign}(X)|X|\} \\ &= \mathbf{E}\{X\} \\ &= 0 \\ &= \mathbf{E}\{Y_i\}\,\mathbf{E}\{Y_n\}. \end{aligned}$$

So use the solution to part (b)-ii where it was shown

$$\hat{X}_n = \hat{X}_{n-1} + h_n(Y_n - \mathbf{E}\{Y_n\}).$$

But $h_n = 0$ since

$$\begin{aligned} \mathbf{E}\{XY_n\} &= \mathbf{E}\{X|X|\} + \mathbf{E}\{XW_n\} \\ &= \mathbf{E}\{\text{sign}(X)|X|\,|X\} + \mathbf{E}\{XW_n\} \\ &= \mathbf{E}\left\{\text{sign}(X)|X|^2\right\} + \mathbf{E}\{X\}\,\mathbf{E}\{W_n\} \\ &= 0. \end{aligned}$$

[20%]

Examiner's comments: The most popular (answered by 90% of candidates) and straightforward question, well-answered by most candidates. The solution for part (b)-ii should have been a simple corollary of the previous parts but some candidates did not make the connection and some even tried to employ a state-space formulation to arrive at the answer which is irrelevant here. Part (c)-i was surprisingly difficult for many. Many candidates did not manage to successfully apply the results of the previous parts to this applied problem, similarly for part (c)-ii. Calculations involving sign(X) and |X| was difficult for a significant number of candidates.

Question 2

Part (a)

Do the update step first:

$$p\left(x_n|y_{1:n-1}, y_n\right) = \frac{p\left(x_n|y_{1:n-1}\right)p(y_n|x_n)}{p(y_n|y_{1:n-1})}$$

$$p(y_n|y_{1:n-1}) = p(y_n|1)\Pr\left(X_n = 1|y_{1:n-1}\right) + p(y_n|-1)\Pr\left(X_n = -1|y_{1:n-1}\right)$$

Do the prediction step next:

$$\Pr\left(X_{n+1} = 1|y_{1:n}\right) = \Pr\left(X_n = 1|y_{1:n}\right)(1-\alpha) + \Pr\left(X_n = -1|y_{1:n}\right)\alpha$$

[20%]

Part (b)

$$\begin{aligned} p(y_{n+1:T}|x_n) &= \sum_{x_{n+1}} p(y_{n+1:T}, x_{n+1}|x_n) \\ &= \sum_{x_{n+1}} p(y_{n+1:T}|x_{n+1}, x_n)p(x_{n+1}|x_n) \\ &= \sum_{x_{n+1}} p(y_{n+1:T}|x_{n+1})p(x_{n+1}|x_n) \end{aligned}$$

Then calculate

$$\begin{aligned} p(y_{n:T}|x_n) &= p(y_{n+1:T}|y_n, x_n)p(y_n|x_n) \\ &= p(y_{n+1:T}|x_n)p(y_n|x_n) \end{aligned}$$

[20%]

Part (c)

To compute $p(x_n|y_{1:T})$, use computed quantities from previous parts:

$$\begin{aligned} p(x_n|y_{1:T}) &= \frac{p(x_n, y_{1:T})}{\sum_{x_n} p(x_n, y_{1:T})} \\ p(x_n, y_{1:T}) &= p(y_{n:T}|x_n, y_{1:n-1})p(x_n, y_{1:n-1}) \\ &= p(y_{n:T}|x_n)p(x_n, y_{1:n-1}) \\ &= p(y_{n:T}|x_n)p(x_n|y_{1:n-1})p(y_{1:n-1}). \end{aligned}$$

Now normalise to get

$$p(x_n|y_{1:T}) = \frac{p(y_{n:T}|x_n)p(x_n|y_{1:n-1})}{\sum_{x_n} p(y_{n:T}|x_n)p(x_n|y_{1:n-1})}.$$

[10%]

Part (d)

Let s denote the number of instances $y_i = x_i$. The

$$\begin{aligned}\log p(y_1, \ldots, y_T | x_1, \ldots, x_T) &= \sum_{i=1}^{T} \log p(y_i | x_i) \\ &= s \log(1-\beta) + (T-s)\log\beta.\end{aligned}$$

The maximiser (via calculus) is

$$\beta = \frac{T-s}{T}.$$

[20%]

Part (e)

The EM algorithm is comprised of the E-step and the M-step.

Let $\hat{\beta}$ be the current best estimate of β. Compute $p(x_i|y_{1:T})$, for $i = 1, \ldots, T$ with the current estimate $\hat{\beta}$ and then compute the Q-function

$$\begin{aligned}Q(\beta) &= \sum_{i=1}^{T} \sum_{x_i} p(x_i|y_{1:T}) \log p(y_i|x_i) \\ &= \sum_{i=1}^{T} \left(\Pr(X_i = y_i|y_{1:T}) \log(1-\beta) + \Pr(X_i \neq y_i|y_{1:T}) \log\beta\right) \\ &= s\log(1-\beta) + (T-s)\log\beta\end{aligned}$$

where

$$s = \sum_{i=1}^{T} \Pr(X_i = y_i|y_{1:T}).$$

The the M-step then maximises $Q(\beta)$ to get the new best estimate of β. This E-M steps are repeated until convergence of the estimate is observed.

[20%]

Part (f)

The hidden state represents the trend in the time series, that is if $X_n = 1$ then we expect a price increase or $S_n \geq S_{n-1}$. If $X_n = -1$ then we expect a price decrease or $S_n < S_{n-1}$. When $\beta = 0$ or $\beta = 1$ the hidden state sequence is a deterministic function of Y_n. If $1 > \beta > 0$ then the model allows the observed price trend Y_n to momentarily depart from the price trend X_n given by the hidden state; the smoother in part (c) will denoise the estimate of X_n. If $\beta = 0.5$ then not possible to estimate X_n.

[10%]

Examiner's comments: A popular question answered by 76% of candidates. Parts (a) and (b) were easy point earners for many. Some had trouble solving part (c) by drawing on the results of the previous two parts. Part (e) was disappointing and largely answered incompletely although part (d) did build up to it. The application in part (f) was poorly answered by the majority.

Question 3

Part (a)-i

$$
\begin{aligned}
&p(x_{n+1}|y_{0:n}) \\
&= \int p(x_n|y_{0:n})p(x_{n+1}|x_n)dx_n \\
&= \int p(x_n|y_{0:n})\frac{1}{\sqrt{2\pi\sigma_w}}\exp\left(-\frac{1}{2\sigma_w}(x_{n+1}-x_n)^2\right)dx_n.
\end{aligned}
$$

Now use the hint. $p(x_n|y_{0:n})$ is $\mathcal{N}(\mu_n, \sigma_n)$. Using the hint, the above integral is the pdf of the sum of 2 independent Gaussian random variables. The first is $\mathcal{N}(\mu_n, \sigma_n)$ and second is $\mathcal{N}(0, \sigma_w)$. Thus $p(x_{n+1}|y_{0:n})$ is $\mathcal{N}(\mu_n, \sigma_n + \sigma_w)$.

[20%]

Part (a)-ii

The update equation

$$
\begin{aligned}
&p(x_{n+1}|y_{0:n+1}) \\
&= \frac{p(x_{n+1}|y_{0:n})p(y_{n+1}|x_{n+1})}{p(y_{n+1}|y_{0:n})} \\
&= \frac{1}{p(y_{n+1}|y_{0:n})}\frac{1}{\sqrt{2\pi\sigma_v}}\exp\left(-\frac{1}{2\sigma_v}(y_{n+1}-x_{n+1})^2\right)p(x_{n+1}|y_{0:n})
\end{aligned}
$$

Using the hint, $p(x_{n+1}|y_{0:n+1})$ is the conditional density of U_1 given $U_1 + U_2 = y_{n+1}$ where U_1 is $\mathcal{N}(\mu_n, \sigma_n + \sigma_w)$ and U_2 is $\mathcal{N}(0, \sigma_v)$. Thus the variance is

$$
\sigma_{n+1} = \frac{(\sigma_n + \sigma_w)\,\sigma_v}{\sigma_n + \sigma_w + \sigma_v}
$$

and the mean is

$$
\mu_{n+1} = \frac{(\sigma_n + \sigma_w)\,y_{n+1} + \sigma_v\mu_n}{(\sigma_n + \sigma_w) + \sigma_v}.
$$

[20%]

Part (b)

We can find the fixed point $\sigma_{n+1} = \sigma_n$.

$$
\begin{aligned}
\frac{(\sigma_n + \sigma_w)\,\sigma_v}{\sigma_n + \sigma_w + \sigma_v} &= \sigma_n \\
\sigma_n\sigma_v + \sigma_w\sigma_v &= \sigma_n^2 + \sigma_w\sigma_n + \sigma_v\sigma_n \\
\sigma_w\sigma_v - \sigma_w\sigma_n &= \sigma_n^2
\end{aligned}
$$

Clearly, by sketching the l.h.s. and r.h.s. as a function of σ_n, the solution lies in the range $0 \leq \sigma_n^2 \leq \sigma_w \sigma_v$. (Another way to see this is to note that $\sigma_w \sigma_n$ is positive.)

After sometime, we will have $\sigma_{n+1} = \sigma_n$, that is the variance of $p(x_n|y_0, \ldots, y_n)$ will stop changing. Without needing to solve for the fixed point, the crude bound $\sigma_w \sigma_v$ can be used to decide how accurate the sensor has to be to obtain a given precision in the variance σ_n of the estimate of the hidden state. [30%]

Part (c)

From the solution for μ_{n+1} we see that $\mu_{n+1} = s_n y_{n+1} + r_n \mu_n$ which is linear combination of the new observation and previous mean. Extrapolating back to time 0 we see that μ_n is indeed a linear combination of μ_0 and $y_0, \ldots, y_n$.

[10%]

Part (d)

Consider some $\hat{x}_n = \mu_n + \delta$. The error is

$$\begin{aligned}
&\int (\mu_n + \delta - x_n)^2 \, p(x_n|y_{0:n}) dx_n \\
&= \int (\mu_n - x_n)^2 \, p(x_n|y_{0:n}) dx_n + \delta^2 \\
&+ 2\delta \int (\mu_n - x_n) \, p(x_n|y_{0:n}) dx_n \\
&= \int (\mu_n - x_n)^2 \, p(x_n|y_{0:n}) dx_n + \delta^2.
\end{aligned}$$

So the error is minimised by setting $\delta = 0$. [20%]

Examiner's comments: Least popular and attempted by 57% of candidates; perhaps many were not comfortable with Bayesian calculations involving univariate Gaussian models. Many had an unnecessary amount of trouble with the two components of part (a) although this was bookwork; the hint was not effectively employed. This had a knock-on effect on part (c). Part (d) again was not well done as many failed to realise that μ_n was the conditional mean of the pdf $p(x_n|y_1, \ldots, y_n)$ and thus does indeed achieve the least possible mean square error.

Question 4

Part (a)

Let P be the transition probability matrix of the Markov chain,

$$P_{i,j} = \Pr\left(X_{n+1} = j | X_n = i\right).$$

The prediction and update steps are: for $x_{n+1} > 0$

$$\begin{aligned} p(x_{n+1}|y_{0:n}) &= \sum_{x_n=0}^{\infty} p(x_{n+1}|x_n)p(x_n|y_{0:n}) \\ &= P_{x_{n+1}-1,x_{n+1}}p(x_{n+1}-1|y_{0:n}) + P_{x_{n+1}+1,x_{n+1}}p(x_n+1|y_{0:n}) \\ &= \alpha p(x_{n+1}-1|y_{0:n}) + (1-\alpha)p(x_n+1|y_{0:n}) \end{aligned}$$

For $x_{n+1} = 0$,

$$p(x_{n+1} = 0|y_{0:n}) = (1-\alpha)p(x_n = 1|y_{0:n}) + (1-\alpha)p(x_n = 0|y_{0:n}).$$

The update step is

$$p(x_{n+1}|y_{0:n+1}) = \frac{p(y_{n+1}|x_{n+1})p(x_{n+1}|y_{0:n})}{\sum_{x_{n+1}=0}^{\infty} p(y_{n+1}|x_{n+1})p(x_{n+1}|y_{0:n})}$$

where

$$p(y_{n+1}|x_{n+1}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma}\left[y_{n+1} - x_{n+1}\right]^2\right).$$

[20%]

Part (b)-i

The pdf of the data is

$$\begin{aligned} p(y_{0:n}) &= \sum_{x_0=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} p(y_{0:n}|x_{0:n})p(x_{0:n}) \\ &= \sum_{x_0=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} p(y_0|x_0)\cdots p(y_n|x_n)p(x_{0:n}) \end{aligned}$$

The unbiased estimate of $p(y_0, \ldots, y_n)$ is

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} \frac{p(y_0|X_0^i)\cdots p(y_n|X_n^i)p(X_0^i, \ldots, X_n^i)}{p(X_0^i, \ldots, X_n^i)} \\ &= \frac{1}{N}\sum_{i=1}^{N} p(y_0|X_0^i)\cdots p(y_n|X_n^i) \\ &= \frac{1}{N}\sum_{i=1}^{N} w_n^i \end{aligned}$$

since the samples are drawn from the pmf of the hidden states. The final equation introduces convenient notation for the subsequent parts.

[10%]

Part (b)-ii

For any function $h(x_0, \ldots, x_n)$ the integral

$$\begin{aligned}&\sum_{x_0=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} h(x_{0:n}) p(x_{0:n}|y_{0:n}) \\ &= \sum_{x_0=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} h(x_{0:n}) \frac{p(x_{0:n}, y_{0:n})}{p(y_{0:n})}\end{aligned}$$

can be approximated by a the ratio of importance sampling estimates of the numerator and denominator which gives

$$\frac{\sum_{i=1}^{N} h(X_{0:n}^i) w_n^i}{\sum_{i=1}^{N} w_n^i} = \frac{1}{W_n} \sum_{i=1}^{N} h(X_{0:n}^i) w_n^i$$

where $W_n = \sum_{i=1}^{N} w_n^i$. [10%]

Part (b)-iii

The importance sampling estimate of the derivative is

$$\frac{1}{W_n} \sum_{i=1}^{N} w_n^i \frac{d}{d\alpha} \log p(X_1^i, \ldots, X_n^i | X_0^i).$$

$$\begin{aligned}\log p(X_1^i, \ldots, X_n^i | X_0^i) &= \sum_{k=1}^{n} \log p(X_k^i | X_{k-1}^i) \\ &= s_n^i \log \alpha + \left(n - s_n^i\right) \log(1 - \alpha)\end{aligned}$$

where s_n^i is the number of instances in the sequence $X_0^i, \ldots, X_n^i$ that $X_k^i > X_{k-1}^i$.Thus

$$\begin{aligned}
&\frac{1}{W_n}\sum_{i=1}^{N} w_n^i \frac{d}{d\alpha}\log p(X_1^i, \ldots, X_n^i | X_0^i) \\
&= \frac{1}{W_n}\sum_{i=1}^{N} w_n^i \frac{d}{d\alpha}\left(s_n^i \log\alpha + \left(n - s_n^i\right)\log(1-\alpha)\right) \\
&= \frac{1}{W_n}\sum_{i=1}^{N} w_n^i \left(\frac{s_n^i}{\alpha} - \frac{n - s_n^i}{1-\alpha}\right) \\
&= \frac{1}{W_n}\sum_{i=1}^{N} w_n^i \left(\frac{s_n^i}{\alpha} + \frac{s_n^i}{1-\alpha} - \frac{n}{1-\alpha}\right) \\
&= \frac{-n}{1-\alpha} + \frac{1}{W_n}\sum_{i=1}^{N} w_n^i s_n^i \left(\frac{1}{\alpha(1-\alpha)}\right).
\end{aligned}$$

Set the derivative to zero to find α,

$$\frac{1}{W_n}\sum_{i=1}^{N} w_n^i s_n^i \left(\frac{1}{\alpha(1-\alpha)}\right) = \frac{n}{1-\alpha}$$

$$\frac{1}{n}\frac{1}{W_n}\sum_{i=1}^{N} w_n^i s_n^i = \alpha.$$

[30%]

Part (b)-iv

Perform the resampling operation as follows: sample $J_1, \ldots, J_N$ independently such that

$$\Pr(J_i = k) = \frac{w_n^k}{W_n}, \qquad k = 1, \ldots, N.$$

Then for each $i = 1, \ldots, N$, sample X_{n+1}^i from the transition probability matrix $p(x_{n+1}|X_n^{J_i})$ and let

$$X_{0:n+1}^i = (X_{0:n}^{J_i}, X_{n+1}^i), \qquad w_{n+1}^i = \frac{W_n}{N}.$$

For any function $h(x_0, \ldots, x_{n+1})$ the importance sampling estimate of

$$\sum_{x_{0:n+1}} h(x_{0:n+1}) p(x_{0:n+1}|y_{0:n})$$

is

$$\frac{\sum_{i=1}^{N} w_{n+1}^i h(X_{0:n+1}^i)}{\sum_{i=1}^{N} w_{n+1}^i} = \frac{1}{N}\sum_{i=1}^{N} h(X_{0:n+1}^i).$$

[10%]

Part (b)-v

The importance sampling estimate of $p(y_{n+1}|y_{0:n})$ is

$$\hat{p}(y_{n+1}|y_{0:n}) = \frac{1}{N}\sum_{i=1}^{N} p(y_{n+1}|X_{n+1}^{i}).$$

If we take the expected value of $p(y_{n+1}|X_{n+1}^{i})$ with respect to the law of (J_i, X_{n+1}^{i}) we get

$$\begin{aligned}
\mathbf{E}\left\{p(y_{n+1}|X_{n+1}^{i})\right\} &= \sum_{k=1}^{N} \Pr(J_i = k) \int p(y_{n+1}|x_{n+1}) p(x_{n+1}|X_n^k) dx_{n+1} \\
&= \sum_{k=1}^{N} \Pr(J_i = k) p(y_{n+1}|X_n^k) \\
&= \sum_{k=1}^{N} \frac{w_n^k}{W_n} p(y_{n+1}|X_n^k).
\end{aligned}$$

So

$$\mathbf{E}\left\{\hat{p}(y_{n+1}|y_{0:n})\right\} = \sum_{k=1}^{N} \frac{w_n^k}{W_n} p(y_{n+1}|X_n^k)$$

Finally

$$\begin{aligned}
\mathbf{E}\left\{\hat{p}(y_{0:n})\hat{p}(y_{n+1}|y_{0:n})\right\} &= \mathbf{E}\left\{\frac{W_n}{N}\sum_{k=1}^{N} \frac{w_n^k}{W_n} p(y_{n+1}|X_n^k)\right\} \\
&= \mathbf{E}\left\{\frac{1}{N}\sum_{k=1}^{N} w_n^k p(y_{n+1}|X_n^k)\right\} \\
&= p(y_{0:n+1}).
\end{aligned}$$

[20%]

Examiner's comments: Attempted by 75% of candidates. Part (b)-iii proved difficult for many and only partially complete answers were provided by the majority. Some even failed to realise that the solution to part (b)-ii was needed to compute the integral in the question. Quite a few candidates did not use resampling in part (b)-iv although being explicitly asked to. Proving unbiasedness in part (b)-v, although bookwork, was not well done by the majority; some even ignored the effect of resampling in their proof.