4F7: STATISTICAL SIGNAL ANALYSIS

S.S. Singh, Easter 2018

Question 1. (a.i) Finding the limit of $\mathbf{h}(n)$:

$$\mathbf{h}^* = \mathbf{h}^* + \mu\mathbf{R}^{-1}(\mathbf{p} - \mathbf{R}\mathbf{h}^*)$$
$$\mathbf{h}^* = \mathbf{R}^{-1}\mathbf{p}$$

Conditions on μ:

(a.ii) Let $\mathbf{v}(n) = \mathbf{h}(n) - \mathbf{h}^*$. Thus

$$\begin{aligned}
\mathbf{h}(n) - \mathbf{h}^* &= \mathbf{h}(n-1) - \mathbf{h}^* + \mu(\mathbf{R}\mathbf{h}^* - \mathbf{R}\mathbf{h}(n-1)) \\
\mathbf{v}(n) &= \mathbf{v}(n-1) - \mu\mathbf{R}\mathbf{v}(n-1) \\
&= (\mathbf{I} - \mu\mathbf{R})\mathbf{v}(n-1) \\
\mathbf{v}(n) &= (\mathbf{I} - \mu\mathbf{R})^n\mathbf{v}(0)
\end{aligned}$$

From the hint, $(\mathbf{I} - \mu\mathbf{R})^n \to 0$ when $|1 - \mu\lambda_i| < 1$ for all eigenvalues λ_i of $\mathbf{R}$.

(b) Repeat the same anlysis before with new step-size $\mu\mathbf{R}$.

guarantees convergence. Optimal $\mu = 1$ as it would result in convergence in one step.

Let $\mathbf{v}(n) = \mathbf{h}(n) - \mathbf{h}^*$. Thus

$$\begin{aligned}
\mathbf{h}(n) - \mathbf{h}^* &= \mathbf{h}(n-1) - \mathbf{h}^* + \mu\mathbf{S}(\mathbf{R}\mathbf{h}^* - \mathbf{R}\mathbf{h}(n-1)) \\
\mathbf{v}(n) &= \mathbf{v}(n-1) - \mu\mathbf{S}\mathbf{R}\mathbf{v}(n-1) \\
&= (\mathbf{I} - \mu\mathbf{S}\mathbf{R})\mathbf{v}(n-1)
\end{aligned}$$

If we set $\mathbf{S} = \mathbf{R}^{-1}$ then convergence assured when $|1 - \mu| < 1$. Optimal $\mu = 1$ as it would result in convergence in one step.

(c.i) The desired response $d(n) = u(n)$ which is the noisy measurement for $x(n)$ as direct measurements of $x(n)$ are unavailable. The input to the LMS is $\mathbf{u}(n-1) = [u(n-1), \ldots, u(n-M)]^{\mathrm{T}}$. The cost function is

$$J(\mathbf{h}) = E\left\{\left(d(n) - \mathbf{h}^T\mathbf{u}(n-1)\right)^2\right\}$$

where $\mathbf{h} = [h_1, \ldots, h_M]^{\mathrm{T}}$. The gradient descent algorithm for minimising this cost function is

$$\begin{aligned}
\mathbf{h}(n) &= \mathbf{h}(n-1) - \frac{\mu}{2}\nabla J(\mathbf{h}(n-1)) \\
\mathbf{h}(n) &= \mathbf{h}(n-1) - \frac{\mu}{2}E\left\{-2\left(d(n) - \mathbf{u}(n-1)^{\mathrm{T}}\mathbf{h}(n-1)\right)\mathbf{u}(n-1)\right\}
\end{aligned}$$

The LMS uses a noisy estimate of the gradient and the update rule is

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mu\mathbf{u}(n-1)\left(d(n) - \mathbf{u}(n-1)^{\mathrm{T}}\mathbf{h}(n-1)\right).$$

(c.ii) The LMS converges provided $0 < \mu < 2/\lambda_{\max}(\mathbf{R})$ where $\mathbf{R} = E\left(\mathbf{u}(n)\mathbf{u}(n)^{\mathrm{T}}\right)$

The limit point is $\mathbf{R}^{-1}\mathbf{p}$ where $\mathbf{p} = E\left(\mathbf{u}(n-1)d(n)\right).$

$$\mathbf{u}(n) = \mathbf{x}(n) + \mathbf{v}(n)$$
$$\mathbf{u}(n)\mathbf{u}(n)^{\mathrm{T}} = \mathbf{x}(n)\mathbf{x}(n)^{\mathrm{T}} + \mathbf{v}(n)\mathbf{v}(n)^{\mathrm{T}} + \mathrm{ct}$$

where ct are the cross terms which will have zero expectation. Thus limit point is $\mathbf{R}^{-1}\mathbf{p}$ where

$$\mathbf{R} = E\left\{\mathbf{x}(n)\mathbf{x}(n)^{\mathrm{T}}\right\} + \sigma_v^2\mathbf{I}, \qquad \mathbf{p} = E\left(\mathbf{u}(n-1)d(n)\right) = E\left(\mathbf{x}(n-1)x(n)\right)$$

Examiner's comments: Attempted by 90% of candidates. The most popular and straightforward question, well-answered by most candidates. Part (c)-i was surprisingly difficult for many. Only a few candidates managed to identify the input and desired signals correctly; most forgot that x(n) is not available.

Question 2. (a.i) In the absence of measurements, set $\widehat{X}_0 = 0$. Then $E(\widehat{X}_1) = G_1 E(Y_1) = G_1 E(X) = 0$. Similarly, if $E(\widehat{X}_{n-1}) = 0$ then $E(\widehat{X}_n) = 0$.

(a.ii) For the error $E_n = \widehat{X}_n - X$.

$$\widehat{X}_n - X = G_n \left(X + V_n - \widehat{X}_{n-1} \right) + \widehat{X}_{n-1} - X$$
$$E_n = G_n \left(V_n - E_{n-1} \right) + E_{n-1}$$

Square it and take the expectation to get:

$$\begin{aligned} E_n^2 &= G_n^2 \left(V_n - E_{n-1} \right)^2 + E_{n-1}^2 \\ &+ 2G_n \left(V_n - E_{n-1} \right) E_{n-1} \end{aligned}$$

$$\begin{aligned} E\left\{E_n^2\right\} &= G_n^2 E\left\{V_n^2 + E_{n-1}^2 - 2V_n E_{n-1}\right\} + E\left\{E_{n-1}^2\right\} \\ &+ 2G_n E\left\{\left(V_n - E_{n-1}\right) E_{n-1}\right\} \\ &= G_n^2 \left(\sigma_v^2 + E\left\{E_{n-1}^2\right\}\right) + E\left\{E_{n-1}^2\right\} \\ &- 2G_n E\left\{E_{n-1}^2\right\} \end{aligned}$$

The last line follows from the stated assumption on $\{V_n\}$. Let $\sigma_n^2 = E\{E_n^2\}$. Differentiating the right-hand side with respect to G_n and equating to 0 to solve for G_n yields:

$$G_n = \frac{\sigma_{n-1}^2}{\sigma_v^2 + \sigma_{n-1}^2}.$$

(b.i) Use $G_n = \sigma_0^2 / \left(n\sigma_0^2 + \sigma_v^2\right)$. We can work backwards:

$$\begin{aligned} \widehat{X}_n &= G_n Y_n + (1 - G_n)\widehat{X}_{n-1} \\ &= G_n Y_n + \left(1 - \frac{\sigma_0^2}{n\sigma_0^2 + \sigma_v^2}\right) \widehat{X}_{n-1} \\ &= G_n Y_n + \left(\frac{(n-1)\sigma_0^2 + \sigma_v^2}{n\sigma_0^2 + \sigma_v^2}\right) \widehat{X}_{n-1} \\ &= G_n Y_n + \left(\frac{G_n}{G_{n-1}}\right) \widehat{X}_{n-1} \end{aligned}$$

We see that $1 - G_n = G_n / G_{n-1}$. Thus

$$\begin{aligned} \widehat{X}_n &= G_n Y_n + \left(1 - G_n\right) \widehat{X}_{n-1} \\ &= G_n Y_n + \left(\frac{G_n}{G_{n-1}}\right) \left(G_{n-1} Y_{n-1} + (1 - G_{n-1})\widehat{X}_{n-2}\right) \\ &= G_n Y_n + \left(G_n Y_{n-1} + \frac{G_n}{G_{n-1}} \frac{G_{n-1}}{G_{n-2}} \widehat{X}_{n-2}\right) \\ &= G_n Y_n + \left(G_n Y_{n-1} + \frac{G_n}{G_{n-2}} \widehat{X}_{n-2}\right) \end{aligned}$$

We thus see that

$$\widehat{X}_n = G_n (Y_n + Y_{n-1} + \ldots + Y_1)$$

(b.ii) The variance of the sample mean estimate is

$$E\left\{\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right\} = E\left\{\left(X + \frac{1}{n}\sum_{i=1}^{n} V_i\right)^2\right\} = \sigma_0^2 + \frac{\sigma_v^2}{n}.$$

The variance of the Kalman filter estimate, noting $E(\widehat{X}_n) = 0$, is

$$E\left\{\left(G_n \sum_{i=1}^{n} Y_i\right)^2\right\} = \left(\frac{\sigma_0^2}{n\sigma_0^2 + \sigma_v^2}\right)^2 \left(n^2\sigma_0^2 + n\sigma_v^2\right) < \sigma_0^2 + \frac{\sigma_v^2}{n}$$

since

$$\frac{\sigma_0^2}{n\sigma_0^2 + \sigma_v^2} = \frac{1}{n + \sigma_v^2/\sigma_0^2} < \frac{1}{n}.$$

Usually the mean square error and not the variance is used to compare estimators. Think of the constant estimate $\hat{X}_n = 0$ for all n. This is clearly unbiased and has zero variance but also clearly useless since it does not use the data.

Examiner's comments: Attempted by 71% of candidates. The derivation of the gain in (a)-ii was unnecessarily long in many attempts. The variance of the Kalman estimate in (b)-ii (note not the mean square error) was challenging for many.

Question 3. (a)

$$\begin{aligned}\Pr(W_{k+1}=j \mid X_k=i) &= \Pr(X_k+W_{k+1}=j \mid X_k=i)\\ &= \Pr(W_{k+1}=j-i)\\ &= f(j-i).\end{aligned}$$

(b.i) The required joint pmf/pdf is

$$p(x_1,y_1,\ldots,x_n,y_n) = p(y_1,\ldots,y_n \mid x_1,\ldots,x_n)p(x_1,\ldots,x_n)$$
$$p(x_1,\ldots,x_n) = P_{0,x_1}P_{x_1,x_2}\ldots P_{x_{n-1},x_n}$$
$$p(y_1,\ldots,y_n \mid x_1,\ldots,x_n) = p(y_1 \mid x_1)\cdots p(y_n \mid x_n)$$
$$p(y_i \mid x_i) = \frac{1}{\sqrt{2\pi}d}\exp\left(-\frac{1}{2d^2}\left[y_i-x_i\right]^2\right)$$

(b.ii) This calculation is a two-step procedure. The first is the prediction step, which is

$$p(x_{n+1} \mid y_1,\ldots,y_n) = \sum_{x_n=-\infty}^{\infty} p(x_n \mid y_1,\ldots,y_n)P_{x_n,x_{n+1}}$$

The second is the update step which is

$$p(x_{n+1} \mid y_1,\ldots,y_{n+1}) = \frac{p(x_{n+1} \mid y_1,\ldots,y_n)\exp\left(-\frac{1}{2d^2}\left[y_{n+1}-x_{n+1}\right]^2\right)}{\sum_{x_{n+1}=\infty}^{\infty} p(x_{n+1} \mid y_1,\ldots,y_n)\exp\left(-\frac{1}{2d^2}\left[y_{n+1}-x_{n+1}\right]^2\right)}$$

(c.i) The weight is

$$w_n^i = p(y_1 \mid X_1^i)\times\cdots\times p(y_n \mid X_n^i).$$

Only this weight will make the estimate unbiased.

(c.ii) The importance sampling estimate is

$$\frac{\sum_{x_1=-\infty}^{\infty}\cdots\sum_{x_n=-\infty}^{\infty} H(x_1,\ldots,x_n)p(y_1,\ldots,y_n \mid x_1,\ldots,x_n)p(x_1,\ldots,x_n,)}{\sum_{x_1=-\infty}^{\infty}\cdots\sum_{x_n=-\infty}^{\infty} p(y_1,\ldots,y_n \mid x_1,\ldots,x_n)p(x_1,\ldots,x_n)}$$
$$= \frac{\sum_{i=1}^{N} H(X_{1:n}^i)w_n^i}{\sum_{i=1}^{N} w_n^i}$$

(c.iii) First note that

$$\sum_{x_{n+1}=-\infty}^{\infty} h(x_{n+1})p(x_{n+1} \mid y_{1:n+1})$$
$$= \sum_{x_1=-\infty}^{\infty}\cdots\sum_{x_{n+1}=-\infty}^{\infty} H(x_1,\ldots,x_{n+1})p(x_1,\ldots,x_{n+1} \mid y_{1:n+1})$$

where $H(x_1, \ldots, x_{n+1}) = h(x_{n+1})$. The importance sampling estimate is

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

where

$$w_{n+1}^i = w_n^i \times p(y_{n+1} \mid X_{n+1}^i)$$

and X_{n+1}^i is a sample from $p(x_{n+1} \mid X_n^i) = P_{X_n^i, x_{n+1}}$. Note thay we are given samples $X_{1:n}^i$ from $p(x_1, \ldots, x_n,)$ and need to extend each of these to a sample from

$$p(x_1, \ldots, x_{n+1}) = p(x_1, \ldots, x_n) p(x_{n+1} \mid x_n) = p(x_1, \ldots, x_n) P_{x_n, x_{n+1}}.$$

(d) Just need to calculate $E(H(X_{1:n}^{J_1}))$ which is

$$E(H(X_{1:n}^{J_1})) = E\left[\sum_{j=1}^N H(X_{1:n}^j) \Pr(J_1 = j)\right]$$
$$= E\left[\sum_{j=1}^N H(X_{1:n}^j) \Pr(J_1 = j)\right]$$
$$= E\left[\sum_{j=1}^N H(X_{1:n}^j) \frac{w_n^j}{\sum_{i=1} w_n^i}\right]$$

and the inner term is the importance sampling estimate from the previous part.

Examiner's comments: Attempted by 89% of candidates. Very well answered question. The exception being part (d). Many were not able to prove that resampling is unbiased. This is a conditional expectation and only a few could articulate the steps correctly.

Question 4. (a)

$$\begin{aligned} & p(X = i \mid Y_1 = y_1, \ldots, Y_k = y_k) \\ & = \frac{p(i)p(y_1 \mid i) \cdots p(y_k \mid i)}{\sum_{j=1}^{\infty} p(j)p(y_1 \mid j) \cdots p(y_k \mid j)} \\ p_k(i) & = \frac{i^{-2} \exp\left(-0.5\,(y_1 - i)^2\right) \cdots \exp\left(-0.5\,(y_k - i)^2\right)}{\sum_{j=1}^{\infty} j^{-2} \exp\left(-0.5\,(y_1 - j)^2\right) \cdots \exp\left(-0.5\,(y_k - j)^2\right)} \end{aligned}$$

(b.i) Estimate the numerator and denominator of $\sum_i p_k(i)h(i)$ separately using importance sampling as follows:

$$\frac{N^{-1} \sum_{n=1}^{N} \frac{X_n^{-2}}{q(X_n)} p(y_1 \mid X_n) \cdots p(y_k \mid X_n) h(X_n^2)}{N^{-1} \sum_{n=1}^{N} \frac{X_n^{-2}}{q(X_n)} p(y_1 \mid X_n) \cdots p(y_k \mid X_n)}$$

(b.ii) The marginal density of the observations is

$$\begin{aligned} p(y_1, \ldots, y_k) &= \sum_{i=1}^{\infty} p(i)p(y_1 \mid i) \cdots p(y_k \mid i) \\ &= \sum_{i=1}^{\infty} \frac{c}{i^2} p(y_1 \mid i) \cdots p(y_k \mid i). \end{aligned}$$

When c in unknown we write $p(y_1, \ldots, y_k)$ as the ratio

$$\frac{\sum_{i=1}^{\infty} i^{-2} p(y_1 \mid i) \cdots p(y_k \mid i)}{\sum_{i=1}^{\infty} i^{-2}}$$

and estimate the numerator and denominator separately using samples form the pmf $q(i)$ as follows:

$$\frac{N^{-1} \sum_{n=1}^{N} \frac{X_n^{-2}}{q(X_n)} p(y_1 \mid X_n) \cdots p(y_k \mid X_n)}{N^{-1} \sum_{n=1}^{N} \frac{X_n^{-2}}{q(X_n)}}.$$

(c.i) The importance sampling estimate of $p(y_1, \ldots, y_k)$. When c is known then the numerical value of $p(i)$ is known (for any i.) The importance sampling estimate is thus

$$\frac{1}{N} \sum_{n=1}^{N} \frac{p(X_n)}{q(X_n)} p(y_1 \mid X_n) \cdots p(y_k \mid X_n).$$

The expected value of the corresponding importance sampling estimate is

$$\begin{aligned}
&E\left[\frac{1}{N}\sum_{n=1}^{N}\frac{p(X_n)}{q(X_n)}p(y_1 \mid X_n)\cdots p(y_k \mid X_n)\right] \\
&= \frac{1}{N}\sum_{n=1}^{N} E\left[\frac{p(X_n)}{q(X_n)}p(y_1 \mid X_n)\cdots p(y_k \mid X_n)\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=-\infty}^{\infty}\left[\frac{p(i)}{q(i)}p(y_1 \mid i)\cdots p(y_k \mid i)q(i)\right]\right) \\
&= p(y_1,\ldots,y_k)
\end{aligned}$$

and is thus unbiased.

(c.ii) Define W_n and U_n as

$$W_n = \frac{p(X_n)}{q(X_n)}, \qquad U_n = p(y_1 \mid X_n)\cdots p(y_k \mid X_n)$$

Then $W_1U_1,\ldots,W_NU_N$ are all independent each with mean $p(y_1,\ldots,y_k)$. The variance of $N^{-1}(W_1U_1+\cdots+W_nU_n)$ is

$$\begin{aligned}
N^{-1}\mathrm{var}(W_1U_1) &= N^{-1}\left(E\left[(W_1U_1)^2\right] - p(y_1,\ldots,y_k)^2\right) \\
E\left[(W_1U_1)^2\right] &= \sum_{i=-\infty}^{\infty} q(i)\frac{p(i)}{q(i)}\frac{p(i)}{q(i)}p(y_1 \mid i)^2\cdots p(y_k \mid i)^2 \\
&= p(y_1,\ldots,y_k)^2\sum_{i=-\infty}^{\infty} q(i)\frac{p_k(i)}{q(i)}\frac{p_k(i)}{q(i)} \\
\mathrm{var}(W_1U_1) &= p(y_1,\ldots,y_k)^2\left(-1+\sum_{i=-\infty}^{\infty} q(i)\frac{p_k(i)}{q(i)}\frac{p_k(i)}{q(i)}\right)
\end{aligned}$$

(c.iii) The optimal choice for $q(i)$ is $p_k(i)$ and the variance will be zero.

$$\begin{aligned}
p_k(i) =& \frac{p(i)\exp\left(-0.5\,(y_1-i)^2\right)\cdots\exp\left(-0.5\,(y_k-i)^2\right)}{\sum_{j=-\infty}^{\infty}p(j)\exp\left(-0.5\,(y_1-j)^2\right)\cdots\exp\left(-0.5\,(y_k-j)^2\right)} \\
=& \frac{p(i)\exp\left(-\frac{0.5}{k^{-1}}\left(\frac{\bar{y}}{k}-i\right)^2\right)}{\sum_{j=-\infty}^{\infty}p(j)\exp\left(-\frac{0.5}{k^{-1}}\left(\frac{\bar{y}}{k}-j\right)^2\right)}
\end{aligned}$$

Examiner's comments: Attempted by 50% of candidates. Part (b)-ii was not answered well with many failing to use importance sampling to estimate the unknown constant c. The calculation of the variance was spot on in many attempts.