4F7-STATISTICAL SIGNAL ANALYSIS

Solutions to the Examples Paper

Question 1: Let

$$\pi(x) = p(y \mid x)p(x)$$

where $p(x)$ is the Gaussian probability density function with mean 0 and variance 1. $p(y \mid x)$ is the Gaussian probability density function with mean x and variance 1.

$$\frac{1}{N}\sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y - Z_j)^2\right] h(Z_j)$$

is an estimate of $\int h(x)\pi(x)dx$ and

$$\frac{\sum_{j=1}^{N} \exp\left[-0.5(y - Z_j)^2\right] h(Z_j)}{\sum_{j=1}^{N} \exp\left[-0.5(y - Z_j)^2\right]}$$

is an estimate of

$$\int h(z)p(x \mid y)dx = \frac{\int h(x)\pi(x)dx}{\int \pi(x)dx}.$$

The expectation is

$$
\begin{aligned}
&\mathbb{E}\left\{ \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y - Z_j)^2\right] h(Z_j) \right\} \\
&= \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}} \mathbb{E}\left\{ \exp\left[-0.5(y - Z_j)^2\right] h(Z_j) \right\} \\
&= \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}} \int \exp\left[-0.5(y - x)^2\right] h(x) p(x) dx \\
&= \int h(x) \pi(x) dx.
\end{aligned}
$$

The ratio of estimates is biased since for random variables A and B, $\mathbb{E}(A/B) \neq \mathbb{E}(A)/\mathbb{E}(B)$.

Question 2: Let $J_1, \ldots, J_N$ be discrete valued random variables, $J_i \in \{1, \ldots, N\}$, with joint conditional probability mass function

$$
\begin{aligned}
&\Pr\left(J_1 = j_1, \ldots, J_N = j_N \mid Z_1 = z_1, \ldots, Z_N = z_N\right) \\
&= \Pr\left(J_1 = j_1 \mid Z_1 = z_1, \ldots, Z_N = z_N\right) \cdots \Pr\left(J_N = j_N \mid Z_1 = z_1, \ldots, Z_N = z_N\right).
\end{aligned}
$$

That is, given the values of $Z_1, \ldots, Z_N$, the random variables $J_1, \ldots, J_N$ are independent. Furthermore, let

$$
\Pr\left(J_i = j \mid Z_1 = z_1, \ldots, Z_N = z_N\right) = \frac{\exp\left[-0.5(y - z_j)^2\right]}{\sum_{i=1}^{N} \exp\left[-0.5(y - z_i)^2\right]}.
$$

(a) Random variables $J_1, \ldots, J_N$ are the outputs of a multinomial resampling algorithm for the weighted samples $\{(Z_j, w_j)\}_{j=1}^{N}$ where the weight w_j is $w_j = \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y - Z_j)^2\right]$. The weight could have been equally defined to be $c\exp\left[-0.5(y-Z_j)^2\right]$ for any common constant c since the constant cancels out after the normalisation of the weights.

(b)

$$\begin{aligned}
&\mathbb{E}\left\{h(Z_{J_1}) \mid Z_1 = z_1, \ldots, Z_N = z_N\right\} \\
&= \sum_{j=1}^{N} h(z_j) \Pr\left(J_1 = j \mid Z_1 = z_1, \ldots, Z_N = z_N\right) \\
&= \sum_{j=1}^{N} h(z_j) \frac{\exp\left[-0.5(y-z_j)^2\right]}{\sum_{i=1}^{N} \exp\left[-0.5(y-z_i)^2\right]}.
\end{aligned}$$

(c) Note that after resampling all particles are given the same weight $\frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y-Z_i)^2\right]\right)$. This example aims to prove that resampling does not introduce a bias. To evaluate $\mathbb{E}\left\{h(Z_{J_1})\frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y-Z_i)^2\right]\right)\right\}$, use the conditioning property:

$$\begin{aligned}
&\mathbb{E}\left\{h(Z_{J_1})\frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y-Z_i)^2\right]\right)\right\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{h(Z_{J_1})\frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y-Z_i)^2\right]\right)\middle| Z_1, \ldots, Z_N\right\}\right].
\end{aligned}$$

The inner conditional expectation is

$$\begin{aligned}
&\mathbb{E}\left\{h(Z_{J_1})\frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y-Z_i)^2\right]\right)\middle| Z_1, \ldots, Z_N\right\} \\
&= \frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-0.5(y-Z_i)^2\right]\right) \mathbb{E}\left\{h(Z_{J_1}) \middle| Z_1, \ldots, Z_N\right\}
\end{aligned}$$

where the simplification uses the fact that the term $\frac{1}{N}\left(\sum_{i=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left[-0.5(y-Z_i)^2\right]\right)$ is no longer random when $Z_1,\ldots,Z_N$ are given. Thus

$$
\begin{aligned}
&\frac{1}{N}\left(\sum_{i=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left[-0.5(y-Z_i)^2\right]\right)\mathbb{E}\left\{h(Z_{J_1})\middle|\, Z_1,\ldots,Z_N\right\}\\
&=\frac{1}{N}\left(\sum_{i=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left[-0.5(y-Z_i)^2\right]\right)\sum_{j=1}^{N}h(Z_j)\frac{\exp\left[-0.5(y-Z_j)^2\right]}{\sum_{i=1}^{N}\exp\left[-0.5(y-Z_i)^2\right]}\\
&=\frac{1}{N}\frac{1}{\sqrt{2\pi}}\frac{\left(\sum_{i=1}^{N}\exp\left[-0.5(y-Z_i)^2\right]\right)}{\sum_{i=1}^{N}\exp\left[-0.5(y-Z_i)^2\right]}\sum_{j=1}^{N}h(Z_j)\exp\left[-0.5(y-Z_j)^2\right]\\
&=\frac{1}{N}\frac{1}{\sqrt{2\pi}}\sum_{j=1}^{N}h(Z_j)\exp\left[-0.5(y-Z_j)^2\right].
\end{aligned}
$$

The outer expectation is

$$
\begin{aligned}
\mathbb{E}\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left[-0.5(y-Z_j)^2\right]h(Z_j)\right\}&=\frac{1}{N}\sum_{j=1}^{N}\int h(x)\pi(x)dx\\
&=\int h(x)\pi(x)dx.
\end{aligned}
$$

Let $p(x_0,\ldots,x_n \mid y_0,\ldots,y_n)$ be the conditional probability density function of a hidden Markov model with state transition probability density function $f(x_k,x_{k+1})$ and observation probability density function $g(x_k,y_k)$. Assume $X_0\sim p(x_0)$.

Let $\pi_n(x_{0:n})=p(x_{0:n},y_{0:n})$. Let $X_{0:n}^i\sim q_n(x_0,\ldots,x_n)$, $i=1,\ldots,N$, be independent samples from a proposal probability density function $q_n(x_{0:n})$ and let $w_n^i=\pi_n(X_{0:n}^i)/q_n(X_{0:n}^i)$.

Question 3: Write down the multinomial resampling algorithm for the weighted samples $\{(X_{0:n}^i,w_n^i)\}_{i=1}^N$.

Repeat the multinomial resampling algorithm on page 47 of the lecture notes.

Question 4: Let J denote a particle index produced by the multinomial resampling algorithm. Show that $\mathbb{E}\left\{h_n(X_{0:n}^J)W_n/N\right\} = \int h_n(x_{0:n})\pi_n(x_{0:n})dx_{0:n}$ where $W_n = \sum_{j=1}^N w_n^j$.

This fact has just been established in Question 2c for a simplified example. The same argument using the conditional expectation will be used once more, which is

$$\begin{aligned}
&\mathbb{E}\left\{h_n(X_{0:n}^J)W_n/N\right\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{ h_n(X_{0:n}^J)W_n/N \right| X_{0:n}^1, \ldots, X_{0:n}^N\right\}\right]
\end{aligned}$$

The inner conditional expectation is

$$\begin{aligned}
&\mathbb{E}\left\{ h_n(X_{0:n}^J)W_n/N \right| X_{0:n}^1, \ldots, X_{0:n}^N\right\} \\
&= \frac{W_n}{N}\mathbb{E}\left\{ h_n(X_{0:n}^J) \right| X_{0:n}^1, \ldots, X_{0:n}^N\right\}
\end{aligned}$$

since W_n is not random once the values $X_{0:n}^1, \ldots, X_{0:n}^N$ are known. Since

$$\mathbb{E}\left\{ h_n(X_{0:n}^J) \right| X_{0:n}^1, \ldots, X_{0:n}^N\right\} = \frac{\sum_{j=1}^N h_n(X_{0:n}^j)w_n^j}{\sum_{i=1}^N w_n^i} = \frac{\sum_{j=1}^N h_n(X_{0:n}^j)w_n^j}{W_n},$$

$$\begin{aligned}
&\mathbb{E}\left\{ h_n(X_{0:n}^J)W_n/N \right| X_{0:n}^1, \ldots, X_{0:n}^N\right\} \\
&= \frac{W_n}{N}\mathbb{E}\left\{ h_n(X_{0:n}^J) \right| X_{0:n}^1, \ldots, X_{0:n}^N\right\} \\
&= \frac{1}{N}\sum_{j=1}^N h_n(X_{0:n}^j)w_n^j.
\end{aligned}$$

We can now evaluate the outer expectation,

$$\begin{aligned}
\mathbb{E}\left\{h_n(X^J_{0:n})W_n/N\right\} &= \mathbb{E}\left[\mathbb{E}\left\{ h_n(X^J_{0:n})W_n/N \,\middle|\, X^1_{0:n},\ldots,X^N_{0:n}\right\}\right] \\
&= \mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} h_n(X^j_{0:n})w^j_n\right] \\
&= \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left[h_n(X^j_{0:n})w^j_n\right] \\
&= \frac{1}{N}\sum_{j=1}^{N}\int h_n(x_{0:n})\pi_n(x_{0:n})dx_{0:n} \\
&= \int h_n(x_{0:n})\pi_n(x_{0:n})dx_{0:n}.
\end{aligned}$$

Question 5: Write down the particle filter algorithm when the proposal probability density function $q_n(x_{0:n})$ is

$$q_n(x_{0:n}) = p(x_0)f(x_0,x_1)\cdots f(x_{n-1},x_n).$$

Repeat the algorithm on page 51 of the lecture notes for the choice of $q_n(x_{0:n})$ above.

The particle filter's estimate of $p(y_{0:n})$ is the average of the weights at time n:

$$\begin{aligned}
\frac{W_n}{N} = \sum_{i=1}^{N} \frac{w_n^i}{N} &= \sum_{i=1}^{N} \frac{W_{n-1}}{N} \ \frac{1}{N} g(X_n^i, y_n) \\
&= \frac{W_{n-1}}{N} \sum_{i=1}^{N} \frac{u_n^i}{N} \\
&\vdots \\
&= \frac{W_0}{N} \left(\sum_{i=1}^{N} \frac{u_1^i}{N} \right) \cdots \left(\sum_{i=1}^{N} \frac{u_{n-1}^i}{N} \right) \left(\sum_{i=1}^{N} \frac{u_n^i}{N} \right)
\end{aligned}$$

where

$$\frac{W_0}{N} = \frac{1}{N} \sum_{i=1}^{N} g(X_0^i, y_0).$$

The important point here is that the estimate is the product of the average of the weights at time 0 with the average of the incremental weights from time 1 to n.

The particle filter's estimate of $\int h_n(x_{0:n}) p_n(x_{0:n} \mid y_{0:n}) dx_{0:n}$ is

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

Question 6: At time 0 the particle filter produces an unbiased estimate of the integral $\int h_0(x_0)\pi_0(x_0)dx_0$ for any function $h_0(x_0)$. The particle filter then advances this importance sampling estimate of $\pi_0(x_0)$ to an importance sampling estimate of $\pi_1(x_0, x_1)$ by first resampling and then extending each sample. We have just verified that resampling does not introduce a bias. So the weighted resampled particles still produces an unbiased estimate of $\int h_0(x_0)\pi_0(x_0)dx_0$. After extension, the particles produce an unbiased estimate of $\int\int h_1(x_0,x_1)\pi_0(x_0,x_1)dx_0dx_1$ for any function $h_1(x_{0,}x_1)$. Extrapolating this argument, it follows that the particle filter produces an unbiased estimate of the integral $\int h_k(x_{0:k})\pi_k(x_{0:k})dx_{0:k}$ for any time k and any function $h_k(x_{0:k})$.

Since

$$\begin{aligned}&p(y_{0:n})\\&=\int p(x_{0:n},y_{0:n})dx_{0:n},\end{aligned}$$

the estimate of $p(y_{0:n})$ is obtained by setting the function h_n to be $h_n(x_{0:n})=1$ for all $x_{0:n}$. Thus the estimate of $p(y_{0:n})$ is $N^{-1}\sum_{i=1}^{N}w_n^i=W_n/N$, which is unbiased.

Consider the following hidden Markov model. Let

$$X_k=aX_{k-1}+\sqrt{b}W_k,\quad k=0,1,\ldots$$

where W_k are independent and identically distributed $\mathcal{N}(0,1)$. Let $X_{-1}=x_{-1}=0$. The observation process Y_k, $k=0,1,\ldots$ is integer valued, $Y_k\in\{0,1,\ldots\}$ and follows a Poisson distribution with rate $c\exp(X_k)$,

$$\Pr\left(Y_k=y\mid X_k=x_k\right)=\frac{e^{-c\exp(x_k)}\left(c\exp(x_k)\right)^y}{y!}.$$

Let the probability mass function for Y_k given $X_k = x_k$ be $g(x_k, y_k)$, i.e. $g(x_k, y_k) = \Pr(Y_k = y_k \mid X_k = x_k)$.

Question 7: Find $\log f(x_{k-1}, x_k)$ and $\log g(x_k, y_k)$ and show that this hidden Markov model belongs to the exponential family.

$$\begin{aligned}
\log f(x_{k-1}, x_k) &= \log\left(\frac{1}{\sqrt{2\pi b}}\right) + -\frac{1}{2b}\left(x_k - a x_{k-1}\right)^2 \\
&= -\frac{1}{2}\log(2\pi b) - \frac{1}{2b}\left(x_k^2 - 2a x_{k-1}x_k + a^2 x_{k-1}^2\right) \\
&= \left(-\frac{1}{2}x_k^2\frac{1}{b} + x_{k-1}x_k\frac{a}{b} - \frac{1}{2}x_{k-1}^2\frac{a^2}{b}\right) - \frac{1}{2}\log(b) - \frac{1}{2}\log(2\pi) \\
&= \begin{bmatrix} \frac{1}{b} & \frac{a}{b} & \frac{a^2}{b} \end{bmatrix} \begin{bmatrix} -\frac{1}{2}x_k^2 \\ x_{k-1}x_k \\ -\frac{1}{2}x_{k-1}^2 \end{bmatrix} + \frac{1}{2}\log(\frac{1}{b}) - \frac{1}{2}\log(2\pi)
\end{aligned}$$

$$\log g(x_k, y_k) = \begin{bmatrix} -c, & \log c \end{bmatrix} \begin{bmatrix} \exp(x_k) \\ y_k \end{bmatrix} + y_k x_k - \log y_k!$$

When a hidden Markov model belongs to the exponential family, $\log f_\theta(x_{k-1}, x_k) + \log g_\theta(x_k, y_k)$ can be expressed as

$$H(x_{k-1}, x_k, y_k) + \left(\psi(\theta)^T S(x_{k-1}, x_k, y_k) - m(\theta)\right).$$

This is indeed the case for our model:

$$\psi(\theta) = \left[\frac{1}{b}, \frac{a}{b}, \frac{a^2}{b}, -c, \log c\right]^T$$

$$S(x_{k-1}, x_k, y_k) = \begin{bmatrix} -\frac{1}{2}x_k^2 \\ x_{k-1}x_k \\ -\frac{1}{2}x_{k-1}^2 \\ \exp(x_k) \\ y_k \end{bmatrix}$$

$$H(x_{k-1}, x_k, y_k) = -\frac{1}{2}\log(2\pi) + y_k x_k - \log y_k!$$

and $m(\theta) = -\frac{1}{2}\log(\frac{1}{b})$.

Question 8: Assume constants a and b are known and only c is to be learnt from the data record $y_0, \ldots, y_n$. Write down the intermediate function

$$Q_n(c, c') = \int \log p_{c'}(x_{0:n}, y_{0:n}) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n}$$

of the expectation-maximisation algorithm.

$$\begin{aligned}
&\log p_{c'}(x_{0:n}, y_{0:n}) \\
&= \log p_{c'}(y_{0:n} \mid x_{0:n}) + \log p(x_{0:n}) \\
&= \sum_{k=0}^{n} \log g_{c'}(x_k, y_k) + \log p(x_{0:n}) \\
&= \left(\sum_{k=0}^{n} \begin{bmatrix} -c', & \log c' \end{bmatrix} \begin{bmatrix} \exp(x_k) \\ y_k \end{bmatrix} + y_k x_k - \log y_k! \right) + \log p(x_{0:n})
\end{aligned}$$

Note that $\log p(x_{0:n})$ is not a function of parameter c' and thus can be safely ignored: the expression for $Q_n(c, c')$ with this term ignored is

$$
\begin{aligned}
&Q_n(c,c') - \int \log p(x_{0:n}) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} \\
&= \int \log p_{c'}(y_{0:n} \mid x_{0:n}) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} \\
&= \int \left(\sum_{k=0}^{n} \begin{bmatrix} -c, & \log c \end{bmatrix} \begin{bmatrix} \exp(x_k) \\ y_k \end{bmatrix} + y_k x_k - \log y_k! \right) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} \\
&= \left(\begin{bmatrix} -c', & \log c' \end{bmatrix} \begin{bmatrix} \int \left\{ \sum_{k=0}^{n} \exp(x_k) \right\} p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} \\ \sum_{k=0}^{n} y_k \end{bmatrix} \right) \\
&+ \int \left(\sum_{k=0}^{n} y_k x_k - \log y_k! \right) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&Q_n(c,c') \\
&= -c' \int \left\{ \sum_{k=0}^{n} \exp(x_k) \right\} p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} + (\log c') \sum_{k=0}^{n} y_k \\
&+ \text{terms not a function of } c'
\end{aligned}
$$

Question 9: Find the value c' that maximises $Q_n(c,c')$.

Differentiating gives

$$
\frac{d}{dc'} Q_n(c,c') = -\int \left\{ \sum_{k=0}^{n} \exp(x_k) \right\} p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} + \frac{1}{c'} \sum_{k=0}^{n} y_k
$$

and the stationary point is

$$
c' = \frac{\sum_{k=0}^{n} y_k}{\int \left\{ \sum_{k=0}^{n} \exp(x_k) \right\} p_c(x_{0:n} \mid y_{0:n}) dx_{0:n}}.
$$

Not all stationary points are maxima, so check the second derivative:

$$\frac{d}{dc'}\frac{d}{dc'}Q_n(c,c') = -\left(\frac{1}{c'}\right)^2 \sum_{k=0}^{n} y_k < 0$$

for all c' since each $y_k \geq 0$.

Question 10: Find the gradient $d \log p_c(y_{0:n})/dc$.

$$\begin{aligned}
\log p_c(y_{0:n}) =& \log \int p_c(x_{0:n}, y_{0:n}) dx_{0:n} \\
\frac{d}{dc} \log p_c(y_{0:n}) =& \frac{1}{p_c(y_{0:n})} \frac{d}{dc} p_c(y_{0:n}) \\
=& \frac{1}{p_c(y_{0:n})} \int \frac{d}{dc} p_c(x_{0:n}, y_{0:n}) dx_{0:n} \\
=& \frac{1}{p_c(y_{0:n})} \int \frac{\frac{d}{dc} p_c(x_{0:n}, y_{0:n})}{p_c(x_{0:n}, y_{0:n})} p_c(x_{0:n}, y_{0:n}) dx_{0:n} \\
=& \frac{1}{p_c(y_{0:n})} \int \frac{d}{dc} \log p_c(x_{0:n}, y_{0:n}) p_c(x_{0:n}, y_{0:n}) dx_{0:n} \\
=& \int \frac{d}{dc} \log p_c(x_{0:n}, y_{0:n}) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n}.
\end{aligned}$$

Since $\log p_c(x_{0:n}, y_{0:n}) = \log p_c(y_{0:n} \mid x_{0:n}) + \log p(x_{0:n})$ and $p(x_{0:n})$ is not a function of c, it has no contribution to $\frac{d}{dc} \log p_c(x_{0:n}, y_{0:n})$.

Thus

$$
\begin{aligned}
&\frac{d}{dc}\log p_c(x_{0:n}, y_{0:n}) \\
&= \frac{d}{dc}\log p_c(y_{0:n} \mid x_{0:n}) \\
&= \sum_{k=0}^{n} \frac{d}{dc}\log g_c(x_k, y_k) \\
&= \sum_{k=0}^{n} - \exp(x_k) + \frac{y_k}{c}.
\end{aligned}
$$

Combining all the expressions gives

$$
\begin{aligned}
\frac{d}{dc}\log p_c(y_{0:n}) &= \int \left(\sum_{k=0}^{n} - \exp(x_k) + \frac{y_k}{c} \right) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n} \\
&= \frac{1}{c}\sum_{k=0}^{n} y_k - \int \left(\sum_{k=0}^{n} \exp(x_k) \right) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n}
\end{aligned}
$$

Write down the gradient ascent algorithm for maximising $\log p_c(y_{0:n})$ and explain how a particle filter may be used to implement it.

The gradient ascent algorithm is

$$
c^{i+1} = c^i + \gamma_i \left(\frac{d}{dc}\log p_c(y_{0:n}) \right)_{c=c^i}
$$

where c^{i+1} is a change of c^i in the direction of ascent of $\log p_c(y_{0:n})$. γ_i is the step-size sequence, either constant step-size, $\gamma^i = \gamma$ for all i or a decreasing step-size sequence. (See lecture notes for all details.)

To estimate the gradient, use the particle filter to compute the integral $\int \left(\sum_{k=0}^{n} \exp(x_k)\right) p_c(x_{0:n} \mid y_{0:n}) dx_{0:n}$, i.e. run a particle filter initialised with parameter $c = c^i$ until time n and used the particles to get the estimate

$$\frac{\sum_{j=1}^{N} \left(\sum_{k=0}^{n} \exp(X_k^j)\right) w_n^j}{\sum_{j=1}^{N} w_n^j}.$$

The estimate of the gradient $\frac{d}{dc} \log p_c(y_{0:n})$ at $c = c^i$ is thus

$$\frac{1}{c^i} \sum_{k=0}^{n} y_k - \frac{\sum_{j=1}^{N} \left(\sum_{k=0}^{n} \exp(X_k^j)\right) w_n^j}{\sum_{j=1}^{N} w_n^j}.$$

S.S. Singh, Department of Engineering, University of Cambridge, Cambridge, CB1 7AT, UK

Email address: sss40@cam.ac.uk