Mathematical Background Notes for Package "HiddenMarkov"

David Harte

26 July 2006

These notes give a very brief background to some relationships that are used in the R package "HiddenMarkov". This package fits simple discrete time hidden Markov models (see Harte, 2005). R is a comprehensive statistical programming language managed by the R Development Core Team (2003).

Contents

1 Markov Chain

$\{C_i; i = 1, \cdots, n\}$ has m states $\{1, \cdots, m\}$. It satisfies the *Markov Property*:

$$\begin{aligned} \Pr\{C_i \,|\, C_{i-1}, \cdots, C_1\} &= \Pr\{C_i \,|\, C_{i-1}\} \\ &= \Pr\{C_i = k \,|\, C_{i-1} = j\} \\ &= \gamma_{jk}^{(i)} \,. \end{aligned}$$

If $\gamma_{jk}^{(i)} = \gamma_{jk}, \ \forall i$ and $j, k = 1, \cdots, m$, then $\{C_i\}$ is *homogeneous*.

Note: we use the subscript i to denote the discrete time points, and j and k to denote the Markov states.

Now assume that $\{C_i\}$ is homogeneous. Let $\Gamma = (\gamma_{jk})$ be an $m \times m$ transition matrix. Let $\delta_j^{(i)} = \Pr\{C_i = j\}$, and

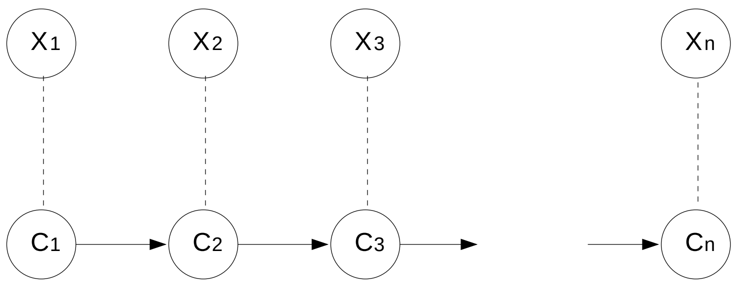
$$\delta^{(i)} = (\delta_1^{(i)}, \delta_2^{(i)}, \cdots, \delta_m^{(i)}) \,,$$

then

$$\delta^{(i)} = \delta^{(i-1)}\Gamma = \delta^{(i-2)}\Gamma^2 = \delta^{(i-3)}\Gamma^3 \,.$$

The chain is *stationary* if $\delta^{(i)} = \delta \quad \forall i$, i.e. $\delta = \delta\Gamma$.

2 Hidden Markov Model



Denote the history of the process until time i as $X^{(i)}$.

Has *conditional independence*

$$\Pr\{X_i \,|\, X^{(i-1)}, C^{(i)}\} = \Pr\{X_i \,|\, C_i\} \,.$$

When X_i is a continuous random variable, replace the probability function with the density function.

Let

$$p_{ij} = \Pr\{X_i = x_i \,|\, C_i = j\} \,,$$

and

$$D_i = \text{diag}(p_{i1}, p_{i2}, \cdots, p_{im}) \,.$$

Further, let Λ be the set of parameters relevant to the observed probability distribution p_{ij}. We denote the set of model parameters $(\delta, \Gamma, \Lambda)$ collectively as Θ.

2.1 Forward and Backward Probabilities

The *forward* probabilities are

$$\alpha_{ij} = \Pr\{X_1 = x_1, \cdots, X_i = x_i, C_i = j\}$$

for $i = 1, \cdots, n$ and $j = 1, \cdots, m$. They are calculated in a "forward" recursive manner. So

$$\alpha_{1j} = \Pr\{X_1 = x_1, C_1 = j\} = \Pr\{X_1 = x_1 \,|\, C_1 = j\} \Pr\{C_1 = j\} = \delta_j^{(1)} p_{1j}$$

then

$$\begin{aligned}
\alpha_{2j} &= \Pr\{X_1 = x_1, X_2 = x_2, C_2 = j\} \\
&= \sum_{k=1}^m \Pr\{X_1 = x_1, X_2 = x_2, C_1 = k, C_2 = j\} \\
&= \sum_{k=1}^m \Pr\{X_1 = x_1, X_2 = x_2 \,|\, C_1 = k, C_2 = j\} \Pr\{C_1 = k, C_2 = j\} \\
&= \sum_{k=1}^m \Pr\{X_1 = x_1 \,|\, C_1 = k\} \Pr\{X_2 = x_2 \,|\, C_2 = j\} \Pr\{C_2 = j \,|\, C_1 = k\} \Pr\{C_1 = k\} \\
&= \sum_{k=1}^m \alpha_{1k} \gamma_{kj} p_{2j} \\
&= \sum_{k=1}^m \delta_k^{(1)} p_{1k} \gamma_{kj} p_{2j} \,,
\end{aligned}$$

and so

$$(\alpha_{21}, \cdots, \alpha_{2m}) = \delta^{(1)} D_1 \Gamma D_2 \,.$$

Similarly, it can be shown that

$$(\alpha_{i1}, \cdots, \alpha_{im}) = \delta^{(1)} D_1 (\Gamma D_2) \cdots (\Gamma D_i) \,.$$

The *backward* probabilities are

$$\beta_{ij} = \Pr\{X_{i+1} = x_{i+1}, \cdots, X_n = x_n \,|\, C_i = j\}$$

for $i = 1, \cdots, n-1$ and $j = 1, \cdots, m$. They are calculated in a "backward" recursive manner. Initially we set

$$(\beta_{n1}, \cdots, \beta_{nm}) = (1, \cdots, 1)_{1\times m}.$$

Then

$$\begin{aligned}\beta_{(n-1)j} &= \Pr\{X_n = x_n \,|\, C_{n-1} = j\} \\ &= \Pr\{X_n = x_n, C_{n-1} = j\} / \Pr\{C_{n-1} = j\} \\ &= \sum_{k=1}^{m} \Pr\{X_n = x_n, C_{n-1} = j, C_n = k\} / \Pr\{C_{n-1} = j\} \\ &= \sum_{k=1}^{m} \Pr\{X_n = x_n \,|\, C_{n-1} = j, C_n = k\} \Pr\{C_{n-1} = j, C_n = k\} / \Pr\{C_{n-1} = j\} \\ &= \sum_{k=1}^{m} \Pr\{X_n = x_n \,|\, C_n = k\} \Pr\{C_n = k \,|\, C_{n-1} = j\}\,,\end{aligned}$$

and so

$$(\beta_{(n-1)1}, \cdots, \beta_{(n-1)m})' = \Gamma D_n 1'\,.$$

Similarly,

$$(\beta_{i1}, \cdots, \beta_{im})' = (\Gamma D_{i+1})(\Gamma D_{i+2}) \cdots (\Gamma D_n) 1'\,.$$

Given estimates of the model parameters Θ, the $n \times m$ matrices $A = (\alpha_{ij})$ and $B = (\beta_{ij})$ can be calculated in a recursive manner.

2.2 Likelihood of HMM

Let $1' = (1, \cdots, 1)_{1\times m}$. Note that

$$\begin{aligned}\Pr\{X_i = x_i\} &= \sum_{j=1}^{m} \Pr\{X_i = x_i \,|\, C_i = j\} \Pr\{C_i = j\} \\ &= \delta^{(i)} D_i 1'\,,\end{aligned}$$

and

$$\begin{aligned}&\Pr\{X_i = x_i, X_{i+1} = x_{i+1}\} \\ &= \sum_{k_i=1}^{m} \sum_{k_{i+1}=1}^{m} \Pr\{X_i = x_i, X_{i+1} = x_{i+1} \,|\, C_i = k_i, C_{i+1} = k_{i+1}\} \Pr\{C_i = k_i, C_{i+1} = k_{i+1}\} \\ &= \sum_{k_i=1}^{m} \sum_{k_{i+1}=1}^{m} \Pr\{X_i = x_i \,|\, C_i = k_i\} \Pr\{X_{i+1} = x_{i+1} \,|\, C_{i+1} = k_{i+1}\} \Pr\{C_i = k_i\} \\ &\qquad \Pr\{C_{i+1} = k_{i+1} \,|\, C_i = k_i\} \\ &= \delta^{(i)} D_i \Gamma D_{i+1} 1'\,,\end{aligned}$$

and also

$$\Pr\{X_i = x_i, X_{i+\ell} = x_{i+\ell}\} = \delta^{(i)} D_i \Gamma^{\ell} D_{i+\ell} 1'\,.$$

Similarly

$$\begin{aligned}L = \Pr\{X^{(n)} = x^{(n)}\} &= \Pr\{X_1 = x_1, \cdots, X_n = x_n\} \\ &= \delta^{(1)} D_1 \Gamma D_2 \Gamma D_3 \cdots \Gamma D_n 1' \\ &= \delta^{(1)} D_1 (\Gamma D_2)(\Gamma D_3) \cdots (\Gamma D_n) 1'\,.\end{aligned}$$

If stationary, $\delta^{(1)}$ can be replaced with $\delta = \delta\Gamma$, creating a recursive pattern ΓD_i for $i = 1, \cdots, n$.

Note the relationship with the *forward* and *backward* probabilities, i.e. for $i = 1, \cdots, n$,

$$L = (\alpha_{i1}, \cdots, \alpha_{im})(\beta_{i1}, \cdots, \beta_{im})' .$$

We want to estimate all parameters in $\Theta = (\delta, \Gamma, \Lambda)$ by maximising L. To do this, we consider the *complete data likelihood*.

2.3 Complete Data Likelihood

$$\begin{aligned}
L_c &= \Pr\{X_1 = x_1, \cdots, X_n = x_n, C_1 = c_1, \cdots, C_n = c_n\} \\
&= \Pr\{X_1 = x_1, \cdots, X_n = x_n \,|\, C_1 = c_1, \cdots, C_n = c_n\} \\
&\qquad \Pr\{C_1 = c_1, \cdots, C_n = c_n\} \\
&= \Pr\{X_1 = x_1 \,|\, C_1 = c_1\} \Pr\{C_1 = c_1\} \\
&\qquad \prod_{i=2}^{n} \Pr\{X_i = x_i \,|\, C_i = c_i\} \Pr\{C_i = c_i \,|\, C_{i-1} = c_{i-1}\} \\
&= \delta_{c_1}^{(1)} \gamma_{c_1 c_2} \gamma_{c_2 c_3} \cdots \gamma_{c_{n-1} c_n} \prod_{i=1}^{n} \Pr\{X_i = x_i \,|\, C_i = c_i\}
\end{aligned}$$

Now let

$$\begin{aligned}
p_{ij} &= \Pr\{X_i = x_i \,|\, C_i = j\} \\
u_{ij} &= \begin{cases} 1 & \text{if } C_i = j \\ 0 & \text{otherwise} \end{cases} \\
v_{ijk} &= \begin{cases} 1 & \text{if } C_{i-1} = j \text{ and } C_i = k \\ 0 & \text{otherwise} . \end{cases}
\end{aligned}$$

Then

$$\log L_c = \sum_{j=1}^{m} u_{1j} \log \delta_j^{(1)} + \sum_{j=1}^{m} \sum_{k=1}^{m} \left(\sum_{i=2}^{n} v_{ijk} \right) \log \gamma_{jk} + \sum_{j=1}^{m} \sum_{i=1}^{n} u_{ij} \log p_{ij} .$$

3 Baum-Welch Algorithm (EM)

Recall that

$$L_c = \Pr\{X^{(n)} = x^{(n)}, C^{(n)} = c^{(n)}\} ,$$

so that

$$\underbrace{L_c}_{\substack{\text{maximise}\\ \text{in M-step}}} = \underbrace{\Pr\{C^{(n)} = c^{(n)} \,|\, X^{(n)} = x^{(n)}\}}_{\text{estimate in E-step}} \Pr\{X^{(n)} = x^{(n)}\}\,.$$

3.1 Outline of Procedure

1. Guess initial values for $\widehat{\Theta}$.
2. Start Loop.
3. *E-Step*: estimate u_{ij} and v_{ijk} given $\widehat{\Theta}$ (i.e. current estimate of Θ), by taking their conditional *expectations*, i.e.:

   $$\begin{aligned} \widehat{u}_{ij} &= \mathrm{E}[u_{ij} \,|\, \widehat{\Theta}] \\ &= \Pr\{C_i = j \,|\, X^{(n)} = x^{(n)}, \widehat{\Theta}\} \\ &= \widehat{\alpha}_{ij}\widehat{\beta}_{ij}/\widehat{L} \end{aligned}$$

   and

   $$\begin{aligned} \widehat{v}_{ijk} &= \mathrm{E}[v_{ijk} \,|\, \widehat{\Theta}] \\ &= \Pr\{C_{i-1} = j, C_i = k \,|\, X^{(n)} = x^{(n)}, \widehat{\Theta}\} \\ &= \widehat{\gamma}_{jk}\widehat{\alpha}_{i-1,j}\widehat{p}_{ik}\widehat{\beta}_{ik}/\widehat{L}\,. \end{aligned}$$

4. *M-Step*: estimate new values for $\widehat{\Theta}$ by *maximising* L_c; see §3.2, §3.3, and §3.4.
5. If $\widehat{\Theta}$ not converged, return to (2).
6. Stop.

If the Markov chain is non-stationary, the *M-step* can be performed by maximising each term in L_c separately.

3.2 First Term of L_c

Want to maximise

$$\sum_{j=1}^{m} u_{1j} \log \delta_j^{(1)}$$

subject to

$$\sum_{j=1}^{m} \delta_j^{(1)} = 1\,.$$

Let

$$F = \sum_{j=1}^{m} u_{1j} \log \delta_j^{(1)} + \theta \left(1 - \sum_{j=1}^{m} \delta_j^{(1)} \right)$$

where θ is a Lagrange multiplier. Then

$$\frac{\partial F}{\partial \delta_j^{(1)}} = \frac{u_{1j}}{\delta_j^{(1)}} - \theta$$

so that $\theta = u_{1j}/\delta_j^{(1)}$ for all j, hence

$$\widehat{\delta}_j^{(1)} = u_{1j} \, .$$

3.3 Second Term of L_c

Similarly as above, let

$$F = \sum_{j=1}^{m} \sum_{k=1}^{m} \left(\sum_{i=2}^{n} v_{ijk} \right) \log \gamma_{jk} + \sum_{j=1}^{m} \theta_j \left(1 - \sum_{k=1}^{m} \gamma_{jk} \right)$$

where $\theta_1, \cdots, \theta_m$ are Lagrange multipliers. Thus

$$\frac{\partial F}{\partial \gamma_{jk}} = -\theta_j + \frac{1}{\gamma_{jk}} \sum_{i=2}^{n} v_{ijk} \, ,$$

hence letting $-\theta_j \gamma_{jk} + \sum_{i=2}^{n} v_{ijk} = 0$, we get

$$\sum_{k=1}^{m} \left(-\theta_j \gamma_{jk} + \sum_{i=2}^{n} v_{ijk} \right) = 0 \, .$$

Since $\sum_{k=1}^{m} \gamma_{jk} = 1$, then

$$\theta_j = \sum_{k=1}^{m} \sum_{i=2}^{n} v_{ijk} \, ,$$

so that

$$\widehat{\gamma}_{jk} = \frac{\sum_{i=2}^{n} v_{ijk}}{\sum_{k=1}^{m} \sum_{i=2}^{n} v_{ijk}} \, .$$

3.4 Third Term of L_c

Maximisation of the last term, i.e.

$$\sum_{j=1}^{m} \sum_{i=1}^{n} u_{ij} \log p_{ij}$$

depends on the probability distribution of the observed process, i.e. $p_{ij} = \Pr\{X_i = x_i \,|\, C_i = j\}$. The set of parameters is denoted by Λ.

The following subsections give details for specific distributions.

3.4.1 Poisson Distribution

In this case

$$p_{ij} = \Pr\{X_i = x_i \,|\, C_i = j\} = \frac{\lambda_j^{x_i}}{x_i!} \exp(-\lambda_j)\,.$$

Let

$$\begin{aligned} F &= \sum_{j=1}^m \sum_{i=1}^n u_{ij} \log p_{ij} \\ &= \sum_{j=1}^m \sum_{i=1}^n u_{ij} \left[x_i \lambda_j - \log(x_i!) - \lambda_j \right] \,, \end{aligned}$$

and so

$$\frac{\partial F}{\partial \lambda_j} = \frac{1}{\lambda_j} \sum_{i=1}^n u_{ij}(x_i - 1)\,,$$

hence

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^n u_{ij} x_i}{\sum_{i=1}^n u_{ij}}\,.$$

3.4.2 Exponential Distribution

In this case

$$p_{ij} = f_{X_i}(x_i \,|\, C_i = j) = \lambda_j \exp(-\lambda_j x_i)\,.$$

Let

$$\begin{aligned} F &= \sum_{j=1}^m \sum_{i=1}^n u_{ij} \log p_{ij} \\ &= \sum_{j=1}^m \sum_{i=1}^n u_{ij} \left[\log \lambda_j - \lambda_j x_i \right] \,, \end{aligned}$$

and so

$$\frac{\partial F}{\partial \lambda_j} = \sum_{i=1}^n u_{ij} \left(\frac{1}{\lambda_j} - x_i \right) \,,$$

hence

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^n u_{ij}}{\sum_{i=1}^n u_{ij} x_i}\,.$$

3.4.3 Binomial Distribution

In this case

$$p_{ij} = \Pr\{X_i = x_i \,|\, C_i = j\} = \binom{n_i}{x_i} \pi_j^{x_i} (1 - \pi_j)^{n_i - x_i}\,,$$

and so

$$\widehat{\pi}_j = \frac{\sum_{i=1}^n u_{ij} x_i}{\sum_{i=1}^n u_{ij} n_i}\,.$$

3.4.4 Gaussian Distribution

In this case

$$p_{ij} = f_{X_i}(x_i \,|\, C_i = j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(\frac{-1}{2\sigma_j^2}(x_i - \mu_j)^2 \right) ,$$

and so

$$\widehat{\mu}_j = \frac{\sum_{i=1}^n u_{ij} x_i}{\sum_{i=1}^n u_{ij}}$$

and

$$\widehat{\sigma}_j = \sqrt{\frac{\sum_{i=1}^n u_{ij}(x_i - \widehat{\mu}_j)^2}{\sum_{i=1}^n u_{ij}}} .$$

3.4.5 Gamma Distribution

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} \exp(-\lambda x)$$

$$\begin{aligned} F &= \frac{1}{n}\sum_{i=1}^n \log f(x_i) \\ &= \frac{1}{n}\sum_{i=1}^n [a\log\lambda - \log\Gamma(a) + (a-1)\log x_i - \lambda x_i] \\ &= a\log\lambda - \log\Gamma(a) + (a-1)\overline{\log x} - \lambda\overline{x} \end{aligned}$$

$$\begin{aligned} \frac{\partial F}{\partial \lambda} &= \frac{a}{\lambda} - \overline{x} \\ \frac{\partial F}{\partial a} &= \log\lambda - \Psi(a) + \overline{\log x} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 F}{\partial \lambda^2} &= \frac{-a}{\lambda^2} \\ \frac{\partial^2 F}{\partial a^2} &= -\Psi'(a) \\ \frac{\partial^2 F}{\partial a \partial \lambda} &= \frac{\partial^2 F}{\partial \lambda \partial a} = \frac{1}{\lambda} \end{aligned}$$

$$\begin{pmatrix} \lambda' \\ a' \end{pmatrix} = \begin{pmatrix} \lambda \\ a \end{pmatrix} - \begin{pmatrix} \frac{-a}{\lambda^2} & \frac{1}{\lambda} \\ \frac{1}{\lambda} & -\Psi'(a) \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial F}{\partial \lambda} \\ \frac{\partial F}{\partial a} \end{pmatrix}$$

The two sufficient statistics $\overline{x}$ and $\overline{\log x}$ become, for $j = 1, \cdots, m$,

$$\frac{\sum_{i=1}^n u_{ij} x_i}{\sum_{i=1}^n u_{ij}} \quad \text{and} \quad \frac{\sum_{i=1}^n u_{ij} \log x_i}{\sum_{i=1}^n u_{ij}} .$$

3.4.6 Beta Distribution

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

$$\begin{aligned} F &= \frac{1}{n}\sum_{i=1}^{n} \log f(x_i) \\ &= \log\Gamma(a+b) - \log\Gamma(a) - \log\Gamma(b) + (a-1)\overline{\log x} \\ &\qquad +(b-1)\overline{\log(1-x)} \end{aligned}$$

$$\begin{aligned} \frac{\partial F}{\partial a} &= \Psi(a+b) - \Psi(a) + \overline{\log x} \\ \frac{\partial F}{\partial b} &= \Psi(a+b) - \Psi(a) + \overline{\log(1-x)} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 F}{\partial a^2} &= \Psi'(a+b) - \Psi'(a) \\ \frac{\partial^2 F}{\partial b^2} &= \Psi'(a+b) - \Psi'(b) \\ \frac{\partial^2 F}{\partial a \partial b} &= \frac{\partial^2 F}{\partial b \partial a} = \Psi'(a+b) \end{aligned}$$

$$\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} \Psi'(a+b) - \Psi'(a) & \Psi'(a+b) \\ \Psi'(a+b) & \Psi'(a+b) - \Psi'(b) \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial F}{\partial a} \\ \frac{\partial F}{\partial b} \end{pmatrix}$$

The two sufficient statistics $\overline{\log x}$ and $\overline{\log(1-x)}$ become, for $j = 1, \cdots, m$,

$$\frac{\sum_{i=1}^n u_{ij} \log x_i}{\sum_{i=1}^n u_{ij}} \quad \text{and} \quad \frac{\sum_{i=1}^n u_{ij} \log(1-x_i)}{\sum_{i=1}^n u_{ij}} .$$

3.4.7 Log Normal Distribution

In this case

$$p_{ij} = f_{X_i}(x_i \,|\, C_i = j) = \frac{1}{\sqrt{2\pi}\sigma_j x_i} \exp\left(\frac{-1}{2\sigma_j^2} (\log x_i - \mu_j)^2 \right) ,$$

and so

$$\begin{aligned} \mathrm{E}[\log X_i \,|\, C_i = j] &= \mu_j \,, \\ \mathrm{Var}[\log X_i \,|\, C_i = j] &= \sigma_j^2 \,, \\ \mathrm{E}[X_i \,|\, C_i = j] &= \exp(\mu_j + \sigma_j^2/2)\,, \text{ and} \\ \mathrm{Var}[X_i \,|\, C_i = j] &= \exp(2\mu_j + \sigma_j^2)(\exp(\sigma_j^2) - 1)\,. \end{aligned}$$

Further

$$\widehat{\mu}_j = \frac{\sum_{i=1}^n u_{ij} \log x_i}{\sum_{i=1}^n u_{ij}}$$

and

$$\widehat{\sigma}_j = \sqrt{\frac{\sum_{i=1}^n u_{ij}(\log x_i - \widehat{\mu}_j)^2}{\sum_{i=1}^n u_{ij}}}\,.$$

3.4.8 Logistic Distribution

Like the beta and gamma distributions, a Newton iterative procedure is used here too. The required first and second derivatives can be found in Rao & Hamed (2000, §9.1.2). Here the location parameter is denoted by m and the scale parameter by a. Note that there are a couple of errors:

In Equation 9.1.10, n should be N; and Equation 9.1.11 should be

$$y_i = 1 + \exp\left(\frac{-(x_i - m)}{a}\right)\,.$$

Equation 9.1.19 should be

$$\frac{\partial^2}{\partial m^2} \log L = \frac{2}{a^2} \sum_{i-1}^{N} (y_i^{-2} - y_i^{-1})\,.$$

4 Miscellaneous

4.1 Pseudo Residuals

We follow the method outlined by Zucchini (2005). Let $X^{(-i)} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$, i.e. denotes the observed process except for the point X_i. For each $i = 1, \cdots, n$ we calculate

$$\begin{aligned} \psi_i &= \Pr\{X_i \leq x_i \,|\, X^{(-i)} = x^{(-i)}\} \\ &= \frac{\Pr\{X_i \leq x_i, X^{(-i)} = x^{(-i)}\}}{\Pr\{X^{(-i)} = x^{(-i)}\}} \\ &= \frac{\delta^{(1)} D_1 (\Gamma D_2) \cdots (\Gamma D_{i-1})(\Gamma D_i')(\Gamma D_{i+1})(\Gamma D_{i+2}) \cdots (\Gamma D_n) 1'}{\delta^{(1)} D_1 (\Gamma D_2) \cdots (\Gamma D_{i-1})(\Gamma I)(\Gamma D_{i+1})(\Gamma D_{i+2}) \cdots (\Gamma D_n) 1'}\,, \end{aligned}$$

where D_i' is an $m \times m$ diagonal matrix with elements $\Pr\{X_i \le x_i \,|\, C_i = j\}$ for $j = 1, \cdots, m$, and I is the identity matrix. This is achieved by using the forward and backward probabilities.

The pseudo residuals are then $z_i = \Phi^{-1}(\psi_i)$, where Φ denotes the standard normal distribution function. If the observation sequence has been sampled from the assumed model, then the z_i's should have an approximate standard normal distribution.

If the distribution of the observation variables is discrete the following correction is made. Also calculate $\psi_i' = \Pr\{X_i \le x_i - 1 \,|\, X^{(-i)} = x^{(-i)}\}$, then

$$z_i = \Phi^{-1}\left(\frac{\psi_i + \psi_i'}{2}\right) .$$

5 References

Elliott, R.J.; Aggoun, L. & Moore, J.B. (1994). *Hidden Markov Models: Estimation and Control.* Springer-Verlag, New York.

Harte, D.S. (2005). *Package "HiddenMarkov": Discrete Time Hidden Markov Models.* Statistics Research Associates, Wellington. URL: www.statsresearch.co.nz/software.html.

MacDonald, I.L. & Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman and Hall/CRC, Boca Raton. ISBN: 0-412-55850-5

R Development Core Team. (2003). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, ISBN 3-900051-00-3. URL: www.r-project.org/.

Rao, A.R. & Hamed, K.H. (2000). *Flood Frequency Analysis.* CRC, Boca Raton. ISBN: 0-412-55280-9

Zucchini, W. (2005). *Hidden Markov Models Short Course, 3–4 April 2005.* Macquarie University, Sydney.