Capítulo 4

Problem 5

5. Let $\{X_t\}$ be an HMM on m states.

(a) Suppose the state-dependent distributions are binomial. More precisely, assume that

$$\Pr(X_t = x \mid C_t = j) = \binom{n_t}{x} p_j^x \, (1 - p_j)^{n_t - x}.$$

Find the value for p_j that will maximize the third term of equation (4.12). (This is needed in order to carry out the M step of EM for a binomial–HMM.)

(b) Now suppose instead that the state-dependent distributions are exponential, with means $1/\lambda_j$. Find the value for λ_j that will maximize the third term of equation (4.12).

Solución

(a) Ecuación (4.12): Complete-Data Log-Likelihood.

$$\log \Pr\left(\mathbf{x}^{(T)}, C^{(T)}\right) = \sum_{j=1}^{m} u_j(1) \log \delta_j \; + \; \sum_{j=1}^{m} \sum_{k=1}^{m} \left(\sum_{t=2}^{T} v_{jk}(t) \right) \log \gamma_{jk} \; + \; \sum_{j=1}^{m} \sum_{t=1}^{T} u_j(t) \log p_j(x_t)$$

Indicadores:

$$u_j(t) = 1 \iff C_t = j \quad (t = 1, 2, \ldots, T), \qquad v_{jk}(t) = 1 \iff C_{t-1} = j, \, C_t = k \quad (t = 2, \ldots, T).$$

En el paso E del algoritmo EM se reemplaza $u_j(t)$ por $\hat{u}_j(t) = \Pr(C_t = j \mid \mathbf{x}^{(T)})$, es decir $\hat{u}_j(t) = \alpha_t(j)\beta_t(j)/L_T$

*Para distribución **binomial** dependiente del estado:*

$$p_j(x_t) = \binom{n_t}{x_t} p_j^{x_t} (1 - p_j)^{n_t - x_t}.$$

El tercer término (para el estado j) queda

$$\ell_j \; = \; \sum_{t=1}^{T} \hat{u}_j(t) \, \log\left(p_j(x_t) \right).$$

Maximizar *(para un j fijo), sustituyendo $p_j(x_t)$:*

$$\ell_j = \sum_{t=1}^{T} \hat{u}_j(t) \left[\log \binom{n_t}{x_t} + x_t \log p_j + (n_t - x_t) \log(1 - p_j) \right].$$

Notas: $\hat{u}_j(t)$ *y* $\log \binom{n_t}{x_t}$ *son constantes en p_j.*
Derivada:

$$\frac{\partial \ell_j}{\partial p_j} = \sum_{t=1}^{T} \hat{u}_j(t) \left(\frac{x_t}{p_j} - \frac{n_t - x_t}{1 - p_j} \right) = \sum_{t=1}^{T} \hat{u}_j(t) \frac{x_t - (n_t p_j)}{p_j(1 - p_j)} = 0.$$

$$\Rightarrow \sum_{t=1}^{T} \hat{u}_j(t) \, x_t = p_j \sum_{t=1}^{T} \hat{u}_j(t) \, n_t.$$

Por lo tanto,

$$\boxed{p_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t) \, x_t}{\sum_{t=1}^{T} \hat{u}_j(t) \, n_t}}$$

Prueba de la segunda derivada:

$$\frac{\partial^2 \ell_j}{\partial p_j^2} = \sum_{t=1}^{T} \hat{u}_j(t) \left(-\frac{x_t}{p_j^2} - \frac{n_t - x_t}{(1 - p_j)^2} \right) = -\sum_{t=1}^{T} \hat{u}_j(t) \left(\frac{x_t}{p_j^2} + \frac{n_t - x_t}{(1 - p_j)^2} \right) < 0,$$

$\Rightarrow$ *es un máximo.*

(b) Caso exponencial: $p_j(x_t) = \lambda_j e^{-\lambda_j x_t}$, $x_t \geq 0$.
Entonces

$$\ell_j = \sum_{t=1}^{T} \hat{u}_j(t) \log p_j(x_t) = \sum_{t=1}^{T} \hat{u}_j(t) \big(\log \lambda_j - \lambda_j x_t \big).$$

Derivada:

$$\frac{\partial \ell_j}{\partial \lambda_j} = \sum_{t=1}^{T} \hat{u}_j(t) \left(\frac{1}{\lambda_j} - x_t \right) = 0 \quad \Longrightarrow \quad \frac{1}{\lambda_j} \sum_{t=1}^{T} \hat{u}_j(t) = \sum_{t=1}^{T} \hat{u}_j(t) \, x_t.$$

$$\boxed{\lambda_j = \frac{\sum_{t=1}^{T} \hat{u}_j(t)}{\sum_{t=1}^{T} \hat{u}_j(t) \, x_t}}$$

Prueba de la segunda derivada:

$$\frac{\partial^2 \ell_j}{\partial \lambda_j^2} = -\sum_{t=1}^{T} \frac{\hat{u}_j(t)}{\lambda_j^2} < 0 \quad \Rightarrow \quad \text{*es un máximo.*}$$

Capítulo 5

Problema 5

5. (Bivariate forecast distributions for HMMs.)

(a) Find the joint distribution of X_{T+1} and X_{T+2}, given $\mathbf{X}^{(T)}$, in as simple a form as you can.

(b) For the earthquakes data, find $\Pr(X_{T+1} \leq 10, X_{T+2} \leq 10 \mid \mathbf{X}^{(T)})$.

Solución

$\Pr(X_{T+1} = x_{T+1}, X_{T+2} = x_{T+2} \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)})$. *Usando la forma matricial de la verosimilitud y cancelando el denominador:*

$$\Pr(X_{T+1}, X_{T+2} \mid \mathbf{X}^{(T)}) \;=\; \frac{\Pr(\mathbf{X}^{(T+2)})}{\Pr(\mathbf{X}^{(T)})} \;=\; \frac{\boldsymbol{\delta}\,\mathbf{P}(x_1)\Gamma\cdots\Gamma\mathbf{P}(x_T)\,\Gamma\mathbf{P}(x_{T+1})\,\Gamma\mathbf{P}(x_{T+2})\,\mathbf{1}}{\boldsymbol{\delta}\,\mathbf{P}(x_1)\Gamma\cdots\Gamma\mathbf{P}(x_T)\,\mathbf{1}}.$$

Denotemos

$$\hat{\boldsymbol{\alpha}}_T \;=\; \boldsymbol{\delta}\,\mathbf{P}(x_1)\Gamma\cdots\Gamma\mathbf{P}(x_T), \qquad \boldsymbol{\phi}_T \;=\; \frac{\hat{\boldsymbol{\alpha}}_T}{\hat{\boldsymbol{\alpha}}_T\mathbf{1}}$$

Entonces la forma matricial *queda*

$$\boxed{\Pr(X_{T+1} = x_{T+1}, X_{T+2} = x_{T+2} \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) \;=\; \boldsymbol{\phi}_T\,\Gamma\,\mathbf{P}(x_{T+1})\,\Gamma\,\mathbf{P}(x_{T+2})\,\mathbf{1}}.$$

Notación: $\mathbf{P}(x) = \mathrm{diag}\big(\rho_1(x),\ldots,\rho_m(x)\big)$ *con* $\rho_j(x) = \Pr(X = x \mid C = j)$.

Si además definimos

$$\hat{\mathbf{E}}(1) \;=\; \boldsymbol{\phi}_T\,\Gamma \;\; (\text{vector fila}),$$

podemos escribir la forma elemento a elemento *como*

$$\Pr(X_{T+1} = x_{T+1}, X_{T+2} = x_{T+2} \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{m} \underbrace{\hat{E}_i(1)}_{(\boldsymbol{\phi}_T\Gamma)_i}\,\rho_i(x_{T+1})\,\gamma_{ij}\,\rho_j(x_{T+2}).$$

(b) *Para la distribución* bivariada acumulada $\Pr(X_{T+1} \leq x, X_{T+2} \leq y \mid \mathbf{X}^{(T)})$, *reemplazamos las matrices de emisión puntuales por acumuladas:*

$$\mathbf{F}(x) \;=\; \mathrm{diag}(F_1(x),\ldots,F_m(x)), \qquad F_j(x) = \Pr(X \leq x \mid C = j).$$

Por la misma cuenta matricial,

$$\Pr(X_{T+1} \le x,\ X_{T+2} \le y \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) \;=\; \boldsymbol{\phi}_T\,\Gamma\,\mathbf{F}(x)\,\Gamma\,\mathbf{F}(y)\,\mathbf{1}$$

Para los datos de earthquakes*, tenemos*

$$\boxed{\Pr(X_{T+1} \le 10,\ X_{T+2} \le 10 \mid \mathbf{X}^{(T)}) \;=\; \boldsymbol{\phi}_T\,\Gamma\,\mathbf{F}(10)\,\Gamma\,\mathbf{F}(10)\,\mathbf{1} \;=\; 0.0508059}.$$

Obtenido al calcular numéricamente en R.