

4

Metodologia para ajuste dos parâmetros do modelo proposto

Neste Capítulo, será apresentado o método adotado para o ajuste do modelo proposto às sequências de dados que se deseja representar.

A estrutura especial do modelo proposto possibilitou a dedução das expressões analíticas para a função de verossimilhança e para as estatísticas de interesse, o que por sua vez possibilitou o desenvolvimento de um método para a estimação dos parâmetros, usando o critério da máxima verossimilhança (ML). A estimação de parâmetros foi realizada utilizando-se as ferramentas de otimização do MatLab.

Em princípio, poderíamos arbitrar parâmetros iniciais para a estimação ML, porém como se trata uma função não linear de seis parâmetros e conforme observado nos experimentos, a existência de diversos máximos locais pode levar o processo de otimização a convergir para ótimos locais distantes do ótimo global procurado, produzindo resultados inadequados.

Esta deficiência foi minimizada pelo emprego do PSO, que por ser uma técnica exaustiva de busca é menos suscetível a convergir para pontos de máximo locais.

Cabe notar que o método PSO poderia em princípio vir a consumir considerável esforço computacional, caso se buscasse melhores resultados por meio do aumento do número de “partículas” utilizadas, o que reduziria uma das principais vantagens obtidas com o uso do modelo proposto que é o tempo reduzido de cálculo com a utilização da otimização clássica.

No entanto, adotando-se um relativamente reduzido número de “partículas”, verificou-se que, com esforço computacional consideravelmente baixo, se pode obter valores de parâmetros adequados para inicialização da ferramenta de otimização do Matlab, aumentando significativamente as chances de convergência para o valor máximo global da função de verossimilhança.

4.1. Função de Verossimilhança do modelo

Apresentaremos a dedução de uma expressão analítica da função de verossimilhança dos parâmetros do modelo proposto, numa forma que facilita o uso de métodos de otimização de funções não lineares, para sua maximização.

Um fator que contribui para o aumento de eficiência da estimação é que, ao invés de se tratar a amostra de erros bit a bit, trabalha-se com a sequência de tamanhos de *gaps* e *clusters* correspondentes.

Sem perda de generalidade, admite-se que a sequência de observações se inicia por um cluster e é composta de $K+1$ blocos, sendo cada bloco correspondente a um par *gap* e *cluster* cujos comprimentos são denotados por m_k e n_k , respectivamente. Denota-se por S_k o estado em que o modelo se encontra depois de emitido o k -ésimo bloco, como mostrado na Figura 4.2.

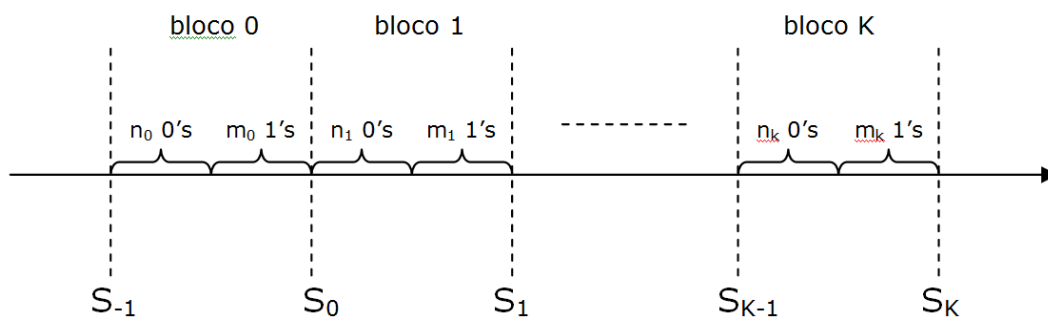


Figura 4.1 – Modelo de transição

Considerando-se as características do modelo, é fácil concluir que S_k é igual a e_0 ou e_3 , para qualquer valor de k .

Do modelo de transição da Figura 4.1, podemos definir dois vetores:

$$\underline{m}_K = (m_0, m_1, \dots, m_K) \quad \text{e} \quad \underline{n}_K = (n_0, n_1, \dots, n_K)$$

cujos elementos são os sucessivos comprimentos de clusters e comprimentos de gaps, respectivamente, retirados da sequência.

A função verossimilhança pode então ser dada por:

$$V(\underline{m}_K, \underline{n}_K) = P\left[\bigcap_{k=0}^K (1^{m_k} 0^{n_k}) \mid 1\right] = \sum_{i \in \{0,3\}} P\left[S_K = e_i; \bigcap_{k=0}^K (1^{m_k} 0^{n_k}) \mid 1\right] \quad (4.1)$$

Observe-se na equação (4.1) que o somatório representa a soma das probabilidades, considerando os dois casos de emissão de 1 dos estados e_0 e e_3 , ao final de cada bloco.

Definindo a variável auxiliar $a_t(i)$ como

$$a_t(i) = P\left[S_t = e_i; \prod_{k=0}^t (1^{m_k} 0^{n_k}) \mid 1\right] \quad (4.2)$$

Tem-se de imediato que:

$$V(\underline{m}_K, \underline{n}_K) = \sum_{i \in \{0,3\}} a_K(i) \quad (4.3)$$

Considerando-se o estado S_{t-1} , é fácil verificar que a variável $a_t(i)$ pode ser expressa como se segue:

$$a_t(i) = \sum_{j \in \{0,3\}} P\left[S_t = e_i; (1^{m_t} 0^{n_t}); S_{t-1} = e_j; \prod_{k=0}^{t-1} (1^{m_k} 0^{n_k}) \mid 1\right] \quad (4.4)$$

$$= \sum_{j \in \{0,3\}} P\left[S_{t-1} = e_j; \prod_{k=0}^{t-1} (1^{m_k} 0^{n_k}) \mid 1\right] \cdot P\left[S_t = e_i; (1^{m_t} 0^{n_t}) \mid S_{t-1} = e_j; \prod_{k=0}^{t-1} (1^{m_k} 0^{n_k}); 1\right]$$

Obtém-se assim uma recursão forma recursiva para $a_t(i)$, expressa a seguir:

$$a_t(i) = \sum_{j \in \{0,3\}} a_{t-1}(j) \cdot P\left[S_t = e_i; (1^{m_t} 0^{n_t}) \mid S_{t-1} = e_j\right] \quad (4.5)$$

Mostra-se no Apêndice A, que a função de verossimilhança acima apresentada também pode ser reescrita em forma matricial como:

$$V(\underline{m}_K, \underline{n}_K) = [1 \quad 1] \cdot \left\{ \prod_{k=0}^K P(m_k, n_k) \right\} \cdot \begin{bmatrix} \bar{P} \\ 1 - \bar{P} \end{bmatrix}, \quad (4.6)$$

$$\text{sendo } \bar{P} = \frac{1}{1 + \left(\frac{\mu}{\lambda}\right) \cdot \left(\frac{1 - \delta - \lambda}{1 - \alpha - \beta}\right)} \quad (4.7)$$

e $P(m,n)$ é uma matriz quadrada de ordem 2 expressa como

$$P(m,n) = \begin{bmatrix} P_{00}(m,n) & P_{03}(m,n) \\ P_{30}(m,n) & P_{33}(m,n) \end{bmatrix}, \quad (4.8)$$

cujos elementos são expressos a seguir:

$$P_{ij}(m,n) = \begin{cases} \mu \cdot \delta^{n-1} \cdot \lambda \cdot (1 - \mu)^{m-1} & i = 0, j = 0 \\ \mu \cdot \delta^{n-1} \cdot (1 - \delta - \lambda) \cdot \alpha^{m-1} & i = 3, j = 0 \\ (1 - \alpha - \beta) \cdot \delta^{n-L} \cdot \lambda \cdot (1 - \mu)^{m-1} & i = 0, j = 3, n \geq L \\ 0 & i = 0, j = 3, n < L \\ (1 - \alpha - \beta) \cdot \delta^{n-L} \cdot (1 - \delta - \lambda) \cdot \alpha^{m-1} & i = 3, j = 3, n \geq L \\ \beta \cdot \rho^{n-1} \cdot \alpha^{m-1} & i = 3, j = 3, n = L - 1 \\ \beta \cdot \rho^{n-1} \cdot (1 - \rho) \cdot \alpha^{m-1} & i = 3, j = 3, n < L - 1 \end{cases}. \quad (4.9)$$

4.2. Metodologia de otimização

O método de otimização do Matlab implementado pela função *fmincon* foi utilizado para maximizar a função de verossimilhança acima, na sua dependência dos parâmetros contínuos do modelo proposto $\alpha, \beta, \delta, \lambda, \mu, \rho$, ou seja, para valores pré-fixados do parâmetro inteiro L .

Como esta ferramenta do Matlab lida diretamente com problemas de minimização, estabeleceu-se o seguinte problema de otimização:

$$\underset{\theta \in R_6}{\text{Min}} -V(\underline{m}_K, \underline{n}_K | \underline{\theta}), \quad \text{onde } \underline{\theta} = [\alpha, \beta, \delta, \lambda, \mu, \rho] , \quad (4.10)$$

sujeito às restrições abaixo:

$$\begin{aligned} 0 &\leq \alpha, \beta, \delta, \lambda, \mu, \rho \leq 1 \\ 0 &\leq \delta + \lambda \leq 1 \\ 0 &\leq \alpha + \beta \leq 1 \end{aligned} \quad (4.11)$$

Este procedimento de otimização foi realizado diversas vezes, variando-se o parâmetro L em uma faixa pré-determinada de valores, para em seguida buscar também o melhor valor deste parâmetro, no sentido de máxima verossimilhança, e assim completar o processo de estimação dos parâmetros do modelo.

Deve ser ressaltado que esta operação de busca do melhor L se torna simples tendo-se em vista que este é um parâmetro inteiro e positivo. Por outro lado, o processamento relacionado à otimização nos parâmetros contínuos, para cada valor de L , consome tempo reduzido, devido à compactação e recursividade no cálculo da função verossimilhança. Note-se que estas características da função de verossimilhança se devem à estrutura especial do modelo proposto e também à representação dos dados em termos de sequências de tamanhos de *gaps* e de *clusters*.

Como mencionado anteriormente, um ponto crítico no método acima descrito foi a escolha dos valores dos parâmetros a serem utilizados como partida do algoritmo de otimização, que acabou sendo obtida através da técnica PSO.