

4 Novas Aplicações do Princípio Evolucionário

No capítulo 2 revisou-se a estratégia Bayesiana de aprendizado dos estados do MEE, sob esta perspectiva os estados são calculados geralmente como média ou a moda da densidade *a posteriori* dos estados. Infelizmente, esta densidade *a posteriori* nem sempre é possível obtê-la analiticamente. Assim, o processo de aprendizado dos estados reduz-se a um problema de aproximação numérica da densidade *a posteriori*, os quais podem ser resolvidos usando métodos de simulação estocástica. Entretanto, no caso em que os estados são calculados simplesmente como a moda da densidade *a posteriori*, então o processo de aprendizado reduz-se a um problema de otimização, o qual pode ser resolvido usando técnicas de algoritmos genéticos, como mostrado no capítulo 3.

Tendo por base as idéias expostas no capítulo 3, este trabalho aborda o campo de simulações estocásticas sob a ótica da genética Mendeliana e do princípio evolucionário da “sobrevivência dos mais aptos”. Sob esta visão, o conjunto de amostras que aproxima a densidade *a posteriori*, pode ser vista como uma população de indivíduos que tentam sobreviver num ambiente Darwiniano, sendo o indivíduo mais forte, aquele que possui maior probabilidade. Esta analogia é estendida a seguir ao apresentar-se uma interpretação do princípio Darwiniano ao problema estimação dos estados do MEE, denominada de estratégia evolucionária de aprendizado.

4.1. A Estratégia Evolucionária de Aprendizado

A estratégia ou paradigma evolucionário usa o conhecimento prévio disponível sobre o fenômeno para formular uma população $\{\Theta_{0x}^{(i)}; i = 1, \dots, N\}$ de indivíduos ou *chromossomas*, os quais evoluem ao longo do tempo. Cada indivíduo $\Theta_{0x}^{(i)}$ representa uma solução potencial do problema de estimação seqüencial dos estados do MEE, isto é, representa a evolução dos estados do MEE ao longo do

tempo. Além disso, o conhecimento prévio serve também para formular uma *função de avaliação* $f(\mathbf{Y}_{1:t} | \Theta_{0:t})$, que relaciona a população de indivíduos com as observações do fenômeno, fornecendo a cada indivíduo sua *aptidão* ao problema. Neste contexto, as estimações dos estados são obtidas através de um processo evolutivo guiado pelo princípio Darwiniano da “sobrevivência dos mais aptos”, aplicado sobre o espaço crescente dos estados $\Theta_{0:t} = \{\theta_0, \theta_1, \dots, \theta_t\}$ e sobre o espaço crescente das observações do fenômeno $\mathbf{Y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$, com $\theta_i \in \mathcal{R}^n$ e $\mathbf{y}_i \in \mathcal{R}^m$ considerados como variáveis aleatórias contínuas e multidimensionais.

Por outro lado, a modelagem probabilística Bayesiana do fenômeno apresentada no capítulo 2, pode ser usada pela estratégia evolucionária como o conhecimento prévio disponível sobre o fenômeno, assim a *função de avaliação* $f(\mathbf{Y}_{1:t} | \Theta_{0:t})$ seria equivalente a densidade de probabilidade da verosimilhança $p(\mathbf{Y}_{1:t} | \Theta_{0:t})$. Entretanto, a diferença com a estratégia Bayesiana está em que a *densidade de importância* $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$ é concebido inspirado em conceitos da genética Mendeliana e do princípio Darwiniano da “sobrevivência dos mais aptos”. A seguir apresenta-se a análise teórica de convergência e a implementação computacional de um algoritmo de aprendizado *online* evolucionário denominado neste trabalho de *Filtro de Partículas Genéticas* (FPG).

4.2. O Filtro de Partículas Genéticas

A diferença entre o *Filtro de Partículas Bootstrap* (FPB), na qual a densidade *a priori* $p(\Theta_{0:t})$ adotava-se como a *densidade de importância* $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$, o FPG propõe adotar uma densidade $p(\Theta_{0:t}^* | \Theta_{0:t})$ que seja condicionada à densidade *a priori*. Assim, $p(\Theta_{0:t}^* | \Theta_{0:t})$ como qualquer outro processo estocástico, pode ser decomposto em dois elementos [27]:

$$\pi(\Theta_{0:t} | \mathbf{Y}_{1:t}) = p(\Theta_{0:t}^* | \Theta_{0:t}) = g(\Theta_{0:t}^* | \Theta_{0:t})\alpha(\Theta_{0:t}^* | \Theta_{0:t}) \quad (47)$$

um *núcleo de transição* $g(\Theta_{0:t}^* | \Theta_{0:t})$ e uma *probabilidade de aceitação* $\alpha(\Theta_{0:t}^* | \Theta_{0:t})$. O objetivo do *núcleo de transição* é modificar ligeiramente a

trajetória convencional $\Theta_{0:t} = \{\theta_0, \theta_1, \dots, \theta_t\}$ dos estados, governada por $p(\theta_t | \theta_{t-1})$, para uma trajetória alternativa $\Theta_{0:t}^* = \{\theta_0, \theta_1, \dots, \theta_t^*\}$ a qual é aceita ou rejeitada (que equivale a preservar a trajetória convencional), segundo a *probabilidade de aceitação*, que tem como objetivo preservar a densidade *a posteriori* $p(\Theta_{0:t} | \mathbf{Y}_{1:t})$ ou a densidade *preditiva* $p(\Theta_{0:t} | \mathbf{Y}_{1:t-1})$ como a *densidade estacionária* de $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$, garantindo desta forma a convergência de $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$ para $p(\Theta_{0:t} | \mathbf{Y}_{1:t})$ ou $p(\Theta_{0:t} | \mathbf{Y}_{1:t-1})$ respectivamente.

Entretanto, sob a ótica da genética Mendeliana, $g(\Theta_{0:t}^* | \Theta_{0:t})$ pode ser visto como um *operador genético*, uma vez que seu papel é criar novos indivíduos $\{\Theta_{0:t}^{*(i)}; i = 1, \dots, N\}$ a partir de $\{\Theta_{0:t}^{(i)}; i = 1, \dots, N\}$. Já sob a ótica da “sobrevivência dos mais aptos”, $\alpha(\Theta_{0:t}^* | \Theta_{0:t})$ pode ser visto como um *esquema de seleção*, uma vez que seu papel é aceitar ou rejeitar os novos indivíduos.

Assim, sem perda de generalidade o *operador genético* $g(\Theta_{0:t}^* | \Theta_{0:t})$, pode ser expresso, através da seguinte equação matricial.

$$\{\Theta_{0:t}^{*(i)}\}_{i=1}^N = \mathbf{Q} \{\Theta_{0:t}^{(i)}\}_{i=1}^N + \varepsilon \mathbf{R} \quad (48)$$

onde ε representa um ruído branco, \mathbf{Q} é uma matriz quadrada não-nula de ordem N e \mathbf{R} é uma matriz constante (usualmente a matriz identidade \mathbf{I}). Tipicamente, a matriz \mathbf{Q} opera somente nos indivíduos da população atual $\{\theta_t^{(i)}\}_{i=1}^N$ em t , e preserva de forma intata as trajetória destes indivíduos acontecidas no passado até $t - 1$, ou seja $\{\Theta_{0:t-1}^{*(i)}\}_{i=1}^N = \{\Theta_{0:t-1}^{(i)}\}_{i=1}^N$. Assim, a equação (47) aplicado a população atual é equivalente a:

$$\begin{bmatrix} \Theta_t^{*(1)} \\ \Theta_t^{*(2)} \\ \vdots \\ \Theta_t^{*(N)} \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN} \end{bmatrix} \begin{bmatrix} \Theta_t^{(1)} \\ \Theta_t^{(2)} \\ \vdots \\ \Theta_t^{(N)} \end{bmatrix} + \varepsilon \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{bmatrix} \quad (49)$$

Esta equação matricial define a distribuição de probabilidade $\pi_g(\Theta_t^* | \Theta_t)$ para a nova população de indivíduos $\{\theta_t^{*(i)}\}_{i=1}^N$ em t . Entretanto, $\pi_g(\Theta_t^* | \Theta_t)$ é uma distribuição marginal em t de $\pi_g(\Theta_{0:t}^* | \Theta_{0:t})$, que define a distribuição de

probabilidade de $\{\Theta_{0:t}^{*(i)}; i = 1, \dots, N\}$, que representa a evolução completa no tempo da nova população de indivíduos.

Por outro lado, as diversas equações matriciais representadas pela equação (49) podem ser classificadas em dois tipos: (a) equações matriciais de *mutação* e (b) equações matriciais de *cruzamento*. As equações matriciais de *mutação* possuem a propriedade da matriz \mathbf{Q} ser uma matriz diagonal, com isto, a nova população de indivíduos $\{\theta_t^{*(1)}, \theta_t^{*(2)}, \dots, \theta_t^{*(N)}\}$ são formados, alterando isoladamente cada indivíduo da população $\{\theta_t^{(1)}, \theta_t^{(2)}, \dots, \theta_t^{(N)}\}$ atual. Um exemplo típico deste caso é quando $\mathbf{Q} = \mathbf{I}$, isto é:

$$\theta_t^{*(i)} = \theta_t^{(i)} + \varepsilon \quad (50)$$

esta equação denominada de “passeio aleatório” define a distribuição de probabilidade $\theta_t^{*(i)} \sim \pi_m(\cdot | \theta_t^{(i)})$ da nova população de indivíduos em t .

As equações matriciais de *cruzamento* possuem a propriedade da matriz \mathbf{Q} ser qualquer matriz à exceção da matriz diagonal e a matriz nula, com isto, a nova população de indivíduos $\{\theta_t^{*(1)}, \theta_t^{*(2)}, \dots, \theta_t^{*(N)}\}$ são formados, combinando dois ou mais indivíduos da população $\{\theta_t^{(1)}, \theta_t^{(2)}, \dots, \theta_t^{(N)}\}$ atual. Por exemplo:

$$\begin{bmatrix} \theta_t^{*(i)} \\ \theta_t^{*(j)} \end{bmatrix} = \begin{bmatrix} 1-\gamma & \gamma \\ \gamma & 1-\gamma \end{bmatrix} \begin{bmatrix} \theta_t^{(i)} \\ \theta_t^{(j)} \end{bmatrix} + \varepsilon \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (51)$$

onde $\gamma \in [0,1)$. Esta equação define também a distribuição de probabilidades $(\theta_t^{*(i)}, \theta_t^{*(j)}) \sim \pi_c(\cdot | \theta_t^{(i)}, \theta_t^{(j)})$ da nova população de indivíduos em t .

Com relação ao *esquema de seleção* $\alpha(\Theta_{0:t}^* | \Theta_{0:t})$, o critério mais comumente citada na literatura de simulações estocásticas é o critério de Metropolis *et al.* [30] e Hastings [31] que tem por objetivo, no presente problema, preservar a densidade *a posteriori* $p(\Theta_{0:t} | \mathbf{Y}_{1:t})$ ou a densidade *preditiva* $p(\Theta_{0:t} | \mathbf{Y}_{1:t-1})$ como a *densidade estacionária* de $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$, garantindo desta forma a convergência de $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$ para $p(\Theta_{0:t} | \mathbf{Y}_{1:t})$ ou $p(\Theta_{0:t} | \mathbf{Y}_{1:t-1})$ respectivamente. Para o caso em que a densidade *a posteriori* $p(\Theta_{0:t} | \mathbf{Y}_{1:t})$ seja adotada como *densidade estacionária* de $\pi(\Theta_{0:t} | \mathbf{Y}_{1:t})$, então $\alpha(\Theta_{0:t}^* | \Theta_{0:t})$ é expresso por:

$$\alpha(\Theta_{0:t}^* | \Theta_{0:t}) = \min \left\{ 1, \frac{p(\Theta_{0:t}^* | \mathbf{Y}_{1:t}) \pi_g(\Theta_{0:t}^* | \Theta_{0:t})}{p(\Theta_{0:t} | \mathbf{Y}_{1:t}) \pi_g(\Theta_{0:t} | \Theta_{0:t}^*)} \right\} \quad (52)$$

Considerando que $\Theta_{0:t}^* = \{\theta_0, \theta_1, \dots, \theta_t^*\} = \{\Theta_{0:t-1}^*, \theta_t^*\}$, então pode ser aplicado a propriedade $p(\mathbf{a}, \mathbf{b} | \mathbf{c}) = p(\mathbf{a} | \mathbf{b}, \mathbf{c})p(\mathbf{b} | \mathbf{c})$ em $\pi_g(\Theta_{0:t}^* | \Theta_{0:t})$ sendo que $\mathbf{a} = \theta_t^*$, $\mathbf{b} = \Theta_{0:t-1}^*$ e $\mathbf{c} = \Theta_{0:t}$. Procedendo analogamente para $\pi_g(\Theta_{0:t} | \Theta_{0:t}^*)$ obtém-se:

$$\alpha(\Theta_{0:t}^* | \Theta_{0:t}) = \min \left\{ 1, \frac{p(\Theta_{0:t}^* | \mathbf{Y}_{1:t}) \pi_g(\theta_t^* | \Theta_{0:t-1}^*, \Theta_{0:t}) \pi_g(\Theta_{0:t-1}^* | \Theta_{0:t})}{p(\Theta_{0:t} | \mathbf{Y}_{1:t}) \pi_g(\theta_t | \Theta_{0:t-1}, \Theta_{0:t}^*) \pi_g(\Theta_{0:t-1} | \Theta_{0:t}^*)} \right\} \quad (53)$$

Sendo $\Theta_{0:t-1}^* = \Theta_{0:t-1}$ então $\pi_g(\Theta_{0:t-1}^* | \Theta_{0:t}) = \pi_g(\Theta_{0:t-1} | \Theta_{0:t}^*)$ uma vez que qualquer trajetória acontecida no passado até $t-1$, não é afetado no futuro em t , logo:

$$\alpha(\Theta_{0:t}^* | \Theta_{0:t}) = \min \left\{ 1, \frac{p(\Theta_{0:t}^* | \mathbf{Y}_{1:t}) \pi_g(\theta_t^* | \Theta_{0:t-1}^*, \Theta_{0:t})}{p(\Theta_{0:t} | \mathbf{Y}_{1:t}) \pi_g(\theta_t | \Theta_{0:t-1}, \Theta_{0:t}^*)} \right\} \quad (54)$$

Segundo a equação (49), a nova população de indivíduos é formado a partir da população atual em t , em consequência não depende de sua trajetória histórica acontecida no passado, desta forma:

$$\alpha(\Theta_{0:t}^* | \Theta_{0:t}) = \min \left\{ 1, \frac{p(\Theta_{0:t}^* | \mathbf{Y}_{1:t}) \pi_g(\theta_t^* | \theta_t)}{p(\Theta_{0:t} | \mathbf{Y}_{1:t}) \pi_g(\theta_t | \theta_t^*)} \right\} \quad (55)$$

Aplicando o teorema de Bayes em $p(\Theta_{0:t}^* | \mathbf{Y}_{1:t})$ e $p(\Theta_{0:t} | \mathbf{Y}_{1:t})$, obtém-se:

$$\alpha(\Theta_{0:t}^* | \Theta_{0:t}) = \min \left\{ 1, \frac{p(\mathbf{Y}_{1:t} | \Theta_{0:t}^*) p(\Theta_{0:t}^*) \pi_g(\theta_t^* | \theta_t)}{p(\mathbf{Y}_{1:t} | \Theta_{0:t}) p(\Theta_{0:t}) \pi_g(\theta_t | \theta_t^*)} \right\} \quad (56)$$

Logo, substituindo adequadamente as equações (22), (23) em (56), resulta:

$$\alpha(\Theta_{0:t}^* | \Theta_{0:t}) = \min \left\{ 1, \frac{p(\mathbf{y}_t | \theta_t^*) p(\theta_t^* | \theta_{t-1}) \pi_g(\theta_t^* | \theta_t)}{p(\mathbf{y}_t | \theta_t) p(\theta_t | \theta_{t-1}) \pi_g(\theta_t | \theta_t^*)} \right\} \quad (57)$$

A aplicação na prática da equação (57) seria da seguinte forma: para o caso em que a população de indivíduos candidatos $\{\theta_t^{*(i)}\}_{i=1}^N$ fossem gerados através da *mutação* dos indivíduos $\{\theta_t^{(i)}\}_{i=1}^N$ da população atual, então:

$$\theta_t^{(i)} = \begin{cases} \theta_t^{*(i)} & \text{se } u \leq \rho_m \\ \theta_t^{(i)} & \text{caso contrário} \end{cases} \quad (58)$$

sendo $u \sim U(0,1)$ e ρ_m expresso por

$$\rho_m = \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{*(i)})p(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_{t-1}^{(i)})\pi_m(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_t^{(i)})}{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{(i)})p(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{t-1}^{(i)})\pi_m(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_t^{*(i)})} \quad (59)$$

a equação (58) indica que se $u \leq \rho_m$, então o indivíduo candidato substitui ao indivíduo atual $\boldsymbol{\theta}_t^{(i)} \leftarrow \boldsymbol{\theta}_t^{*(i)}$, em caso contrario o indivíduo atual é preservado $\boldsymbol{\theta}_t^{(i)} \leftarrow \boldsymbol{\theta}_t^{(i)}$. Entretanto, para o caso em que a população de indivíduos candidatos $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$ sejam gerados através do *cruzamento* dos indivíduos $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ da população atual, então:

$$(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) = \begin{cases} (\boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)}) & \text{se } u \leq \rho_c \\ (\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) & \text{caso contrário} \end{cases} \quad (60)$$

com $u \sim U(0,1)$ e ρ_c expresso por:

$$\rho_c = \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{*(i)})p(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_{t-1}^{(i)})}{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{(i)})p(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{t-1}^{(i)})} \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{*(j)})p(\boldsymbol{\theta}_t^{*(j)} | \boldsymbol{\theta}_{t-1}^{(j)})}{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{(j)})p(\boldsymbol{\theta}_t^{(j)} | \boldsymbol{\theta}_{t-1}^{(j)})} \frac{\pi_c(\boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)} | \boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)})}{\pi_c(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)} | \boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)})} \quad (61)$$

a equação (60) indica que se $u \leq \rho_c$, então os indivíduos candidatos substituem aos indivíduos atuais $(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) \leftarrow (\boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)})$, em caso contrario os indivíduos atuais são preservados $(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) \leftarrow (\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)})$.

Por outro lado, para o caso em que a densidade de previsão $p(\boldsymbol{\Theta}_{0:t} | \mathbf{Y}_{1:t-1})$ seja adotada como a *densidade estacionária* de $\pi(\boldsymbol{\Theta}_{0:t} | \mathbf{Y}_{1:t})$, o esquema de seleção $\alpha(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})$ de Metropolis-Hastings é realizado de forma análoga ao caso anterior. Isto é:

$$\begin{aligned} \alpha(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t}) &= \min \left\{ 1, \frac{p(\boldsymbol{\Theta}_{0:t}^* | \mathbf{Y}_{1:t-1})\pi_g(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})}{p(\boldsymbol{\Theta}_{0:t} | \mathbf{Y}_{1:t-1})\pi_g(\boldsymbol{\Theta}_{0:t} | \boldsymbol{\Theta}_{0:t}^*)} \right\} \\ &= \min \left\{ 1, \frac{p(\mathbf{Y}_{1:t-1} | \boldsymbol{\Theta}_{0:t}^*)p(\boldsymbol{\Theta}_{0:t}^*)\pi_g(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})}{p(\mathbf{Y}_{1:t-1} | \boldsymbol{\Theta}_{0:t})p(\boldsymbol{\Theta}_{0:t})\pi_g(\boldsymbol{\Theta}_{0:t} | \boldsymbol{\Theta}_{0:t}^*)} \right\} \\ &= \min \left\{ 1, \frac{p(\boldsymbol{\theta}_t^* | \boldsymbol{\theta}_{t-1})\pi_g(\boldsymbol{\theta}_t^* | \boldsymbol{\theta}_t)}{p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})\pi_g(\boldsymbol{\theta}_t | \boldsymbol{\theta}_t^*)} \right\} \end{aligned} \quad (62)$$

A aplicação na prática da equação (62) é realizado de forma análoga à equação (57), com a exceção de ρ_m e ρ_c , os quais são definidos por:

$$\rho_m = \frac{p(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_{t-1}^{(i)}) \pi_m(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_t^{(i)})}{p(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{t-1}^{(i)}) \pi_m(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_t^{(i)})} \quad (63)$$

$$\rho_c = \frac{p(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_{t-1}^{(i)}) p(\boldsymbol{\theta}_t^{*(j)} | \boldsymbol{\theta}_{t-1}^{(j)}) \pi_c(\boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)} | \boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)})}{p(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{t-1}^{(i)}) p(\boldsymbol{\theta}_t^{(j)} | \boldsymbol{\theta}_{t-1}^{(j)}) \pi_c(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)} | \boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)})} \quad (64)$$

Para os dois critérios de seleção de Metropolis-Hastings apresentados nesta seção, a expressão de $\alpha(\boldsymbol{\Theta}_{0t}^* | \boldsymbol{\Theta}_{0t})$ torna-se complicada de ser avaliada analiticamente na medida que são combinados vários indivíduos, como é o caso do operador genético de *cruzamento*. Além disso, quando a matriz \mathbf{Q} possui elementos aleatórios, por exemplo $\gamma \sim \mathbf{U}(0,1)$ na equação (51), a densidade $\pi_c(\cdot | \cdot)$ da equação (61) e (64) deve ser substituído equivalentemente pela densidade $\tilde{\pi}_c(\cdot | \cdot) = \int \pi_c(\cdot | \cdot; \gamma) p(\gamma) d\gamma$, a qual elimina a influencia de γ no operador genético. Entretanto, esta integral pode inviabilizar na prática a execução do critério de seleção de Metropolis-Hastings.

A idéia fundamental que está por trás do critério de Metropolis-Hastings é tratar de eliminar os indivíduos com baixa *aptidão*, localizados em regiões de baixa probabilidade de $p(\boldsymbol{\Theta}_{0t} | \mathbf{Y}_{1t})$ ou $p(\boldsymbol{\Theta}_{0t} | \mathbf{Y}_{1t-1})$, e tratar de multiplicar os indivíduos com *aptidão* alta. Esta idéia em essência segue o princípio da “sobrevivência dos mais aptos” aplicado a uma população “local” formada pelos indivíduos genitores e seus filhos. Assim, o critério de Metropolis-Hastings pode ser visto como um *esquema de seleção* “local”, em consequência, o *esquema de seleção* “global”, seria a aplicação do princípio da “sobrevivência dos mais aptos” a todos os indivíduos da população de genitores e os indivíduos da população filha. Em analogia com os AG, os indivíduos genitores são formados, selecionando os indivíduos mais aptos da população atual. Logo, é criado a população filha (indivíduos candidatos) com base na população de genitores, através da execução dos operadores genéticos de *cruzamento* e *mutação*. Em seguida, calcula-se a *aptidão normalizada* $\tilde{w}_t^{(i)}$ (equivalente ao *peso de importância normalizado*) dos indivíduos candidatos, expresso pela equação (29). Finalmente, executa-se qualquer esquema de *amostragem com reposição* [20-22], tal como *seleção multinomial*, *seleção por roleta*, *seleção estocástica universal*, etc, sobre a população dos indivíduos candidatos. Também é possível reamostrar N indivíduos da população formada pelos genitores e dos filhos simultaneamente.

Baseado nos resultados teóricos obtidos acima, pode ser derivado o seguinte algoritmo computacional, denominado neste trabalho de *Filtro de Partículas Genéticas* (FPG).

Procedimento Filtro de Partículas Genéticas

Início

$$t \leftarrow 1$$

Iniciar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$

Avaliar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$

Selecionar $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$

Enquanto (condição de parada não é satisfeita) Fazer

Início

$$t \leftarrow t + 1$$

Evoluir $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^N$

Avaliar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$

Selecionar $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$

Fim

Fim

onde $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ representa as amostras distribuídas de acordo com $p(\boldsymbol{\Theta}_t | \mathbf{Y}_{1:t})$ e $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$ são as amostras distribuídas de acordo com $\pi(\boldsymbol{\Theta}_t | \mathbf{Y}_{1:t})$.

A seguir detalha-se cada um dos procedimentos do algoritmo do FPG para o caso em que a densidade *a posteriori* $p(\boldsymbol{\Theta}_{0:t} | \mathbf{Y}_{1:t})$ seja adotada como a *densidade de estacionária* de $\pi(\boldsymbol{\Theta}_t | \mathbf{Y}_{1:t})$, e considerando também um *esquema de seleção* global. Entretanto, para o caso em que a densidade *a preditiva* $p(\boldsymbol{\Theta}_{0:t} | \mathbf{Y}_{1:t-1})$ seja adotada como a *densidade de estacionária* de $\pi(\boldsymbol{\Theta}_t | \mathbf{Y}_{1:t})$, e para o caso de um *esquema de seleção* local, alguns estágios do algoritmos FPG sofrem ligeiras modificações.

Iniciar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, indivíduos $\boldsymbol{\theta}_1^{*(i)} = \boldsymbol{\theta}_0^{(i)}$ sendo $\boldsymbol{\theta}_0^{(i)} \sim \pi(\boldsymbol{\theta}_0)$.

Avaliar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, indivíduos, avaliar a *aptidão normalizada*

$$\tilde{\omega}_t^{(i)} = \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{*(i)})}{\sum_{j=1}^N p(\mathbf{y}_t | \boldsymbol{\theta}_t^{*(j)})} \quad (65)$$

Selecionar $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$:

Formar a população de genitores $\{\tilde{\boldsymbol{\theta}}_t^{(i)}, N^{-1}\}_{i=1}^N$ através de um processo de reamostragem de $\{\boldsymbol{\theta}_t^{*(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$ de acordo com sua *aptidão normalizada*.

Para $i = 1, \dots, N/2$, $j = N - i + 1$, $u \sim \mathbf{U}(0,1)$

$$(\ddot{\boldsymbol{\theta}}_t^{(i)}, \ddot{\boldsymbol{\theta}}_t^{(j)}) = \begin{cases} \sim \pi_c(\cdot | \tilde{\boldsymbol{\theta}}_t^{(i)}, \tilde{\boldsymbol{\theta}}_t^{(j)}) & \text{if } u \leq \rho_c \\ (\tilde{\boldsymbol{\theta}}_t^{(i)}, \tilde{\boldsymbol{\theta}}_t^{(j)}) & \text{otherwise} \end{cases} \quad (66)$$

Para $i = 1, \dots, N$, $u \sim \mathbf{U}(0,1)$

$$\ddot{\boldsymbol{\theta}}_t^{(i)} = \begin{cases} \sim \pi_m(\cdot | \ddot{\boldsymbol{\theta}}_t^{(i)}) & \text{if } u \leq \rho_m \\ \ddot{\boldsymbol{\theta}}_t^{(i)} & \text{otherwise} \end{cases} \quad (67)$$

Amostrar com substituição N indivíduos $\{\boldsymbol{\theta}_t^{(i)}, N^{-1}\}_{i=1}^N$ a partir de $\{\ddot{\boldsymbol{\theta}}_t^{(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$ de acordo com sua *aptidão normalizada*.

Evoluir $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, indivíduos $\boldsymbol{\theta}_t^{(i)} \sim p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}^{(i)})$

Fixar $\boldsymbol{\Theta}_{0t}^{*(i)} = \{\boldsymbol{\Theta}_{0t}^{(i)}, \boldsymbol{\theta}_t^{*(i)}\}$.

Note-se que na equação (72) e (73), $\rho_c \in [0,1)$ representa a probabilidade de *cruzamento* e $\rho_m \in [0,1)$ a probabilidade de *mutação*, os quais são determinados previamente no caso de ser utilizado o *esquema de seleção global*, e são determinados pelas equações (61) e (59), respectivamente, no caso de ser utilizado o *esquema de seleção local*. Note-se também, que na equação (71) não aparece $\tilde{\omega}_{t-1}^{(i)}$ uma vez que os indivíduos a serem evoluídos $\{\boldsymbol{\Theta}_{0t-1}^{(i)}; i = 1, \dots, N\}$ possuem pesos uniformes após do estágio *Selecionar*, no tempo $t - 1$. Além disso, não é

necessário armazenar todas as trajetórias dos indivíduos o tempo 0 até o tempo t , uma vez que a cada instante de tempo está-se interessado em estimar $p(\boldsymbol{\theta}_t | \mathbf{Y}_{1:t})$.

A figura 2 mostra uma representação gráfica do algoritmo, considerando a trajetória de $N = 10$ indivíduos.

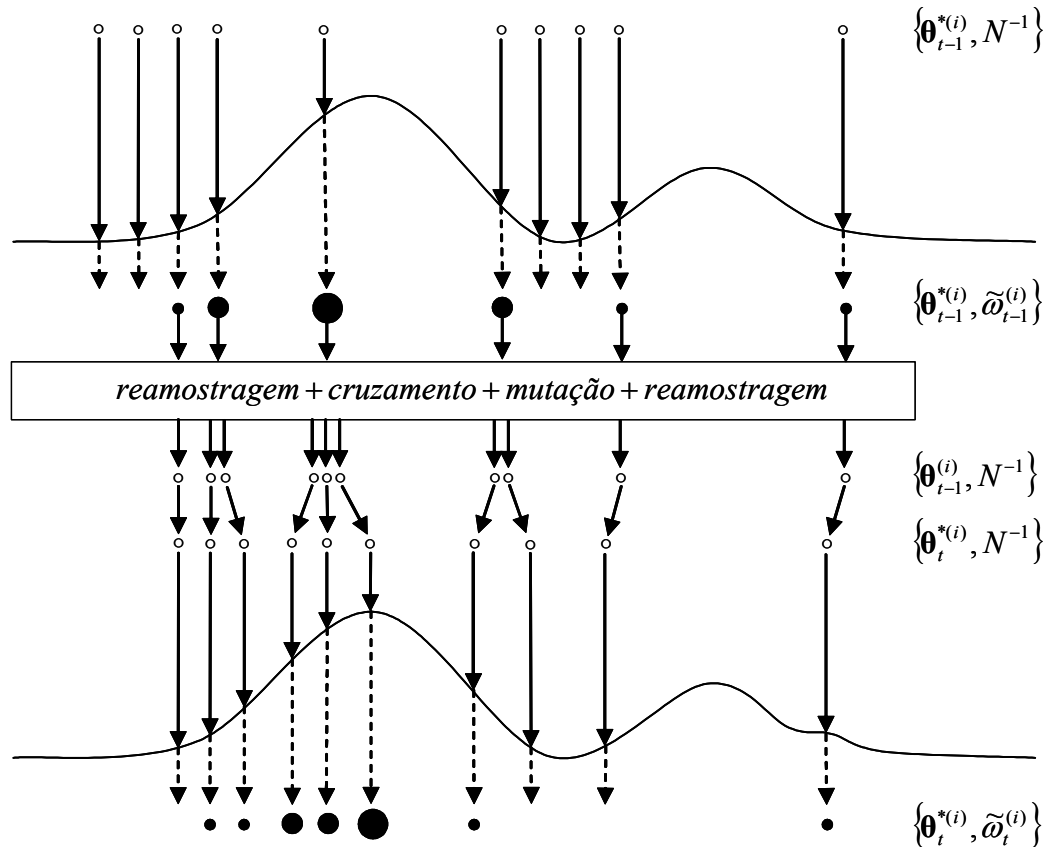


Figura 2 - Representação gráfica do algoritmo do FPG

Neste exemplo o FPG começa no tempo $t - 1$ com indivíduos $\{\boldsymbol{\theta}_{t-1}^{*(i)}, N^{-1}\}$ tendo *aptidões* com valores uniformes, que representam uma aproximação de $p(\boldsymbol{\theta}_{t-1} | \mathbf{Y}_{1:t-2})$. Em seguida, a *aptidão* de cada indivíduo é avaliada usando a informação no tempo $t - 1$. Como resultado obtém-se indivíduos $\{\boldsymbol{\theta}_{t-1}^{*(i)}, \tilde{\omega}_{t-1}^{(i)}\}$ ponderados, que representam uma aproximação de $p(\boldsymbol{\theta}_{t-1} | \mathbf{Y}_{1:t-1})$. Posteriormente, no estágio *Selecionar*, são escolhidos os indivíduos genitores os quais são evoluídos pelos operadores de *cruzamento* e *mutação* e em seguida aplica-se novamente uma reamostragem para selecionar os indivíduos mais aptos, obtendo-se novamente indivíduos $\{\boldsymbol{\theta}_{t-1}^{(i)}, N^{-1}\}$ com *aptidão* com valores uniformes, que ainda fornecem uma aproximação de $p(\boldsymbol{\theta}_{t-1} | \mathbf{Y}_{1:t-1})$. Finalmente, no estágio *Evoluir*, os indivíduos são evoluídos no tempo através da equação de estado e em

seguida são avaliadas usando a informação no tempo t , obtendo-se como resultado indivíduos $\{\boldsymbol{\theta}_t^{*(i)}, \tilde{\omega}_t^{(i)}\}$ ponderados, que representam uma aproximação de $p(\boldsymbol{\theta}_t | \mathbf{Y}_{1:t-1})$.

Observando o algoritmo do FPG e o algoritmo FPB do capítulo 2, vê-se que suas implementações são similares à exceção do estágio *Selecionar* onde são introduzidos os operadores genéticos. A literatura mostra que Higuchi [51] foi um dos primeiros a introduzir os operadores genéticos no FPB com o objetivo de incrementar a diversidade das partículas após o estágio *Selecionar*. Entretanto, a diferença entre a proposta deste trabalho e a proposta de Higuchi é que esta considera a representação binária dos *cromossomas*, a qual não é apropriada para problemas numéricos de alta precisão [39] onde é recomendável usar uma representação em ponto flutuante como é realizado no FPG. Além disso, observando o algoritmo de FPG e algoritmo PE do capítulo 3, vê-se claramente que o FPG possui um pseudo código parecido ao dos AGs convencionais, inclusive os estágios *Iniciar*, *Avaliar* e *Selecionar* são conceitualmente equivalentes. A diferença está no estágio *Evoluir*, pois o FPG possui a informação adicional de evolução dos estados $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ do MEE, o que não acontece com os AGs, isto porque o FPG tem por objetivo amostrar densidades de probabilidades variantes no tempo e o AG tem por objetivo otimizar uma função de custo cuja definição não muda no tempo. Entretanto, pode ser desenvolvido um caso especial do FPG, denominado neste trabalho de *Filtro Evolutivo (FE)*, que atue exatamente como um AG convencional. Em conseqüência, a análise teórica de convergência do FE, pode ser considerado uma fundamentação teórica dos AG baseado numa representação de ponto flutuante dos *cromossomas*. Este será o assunto abordado na próxima seção.

4.3. O Filtro Evolutivo

No capítulo 2 revisou-se a estratégia Bayesiana de aprendizado dos estados do MEE, sob esta perspectiva os estados são calculados geralmente como média ou a moda da densidade *a posteriori* dos estados. Entretanto, no caso em que os estados fossem calculados simplesmente como a moda da densidade *a posteriori*,

então o processo de aprendizado reduz-se a um problema de otimização que pode ser expresso num contexto de maximização ou minimização.

A equação (40) define o problema de otimização num contexto de maximização, na qual θ^{MM} representa a moda global da densidade *a posteriori* de interesse representado de forma genérica por $p(\theta)$. Entretanto a equação (41) define o problema de otimização num contexto de minimização, na qual θ^{\min} representa o argumento onde a função $\mathbf{h}(\theta) \geq 0, \forall \theta \in \mathfrak{R}^n$ possui seu mínimo valor global. Em ambos os casos a transformação monotônica $\mathbf{h}(\theta) = -\ln[p(\theta)]$ garante que a solução da equação (40) ou da equação (41) seja a mesma. Assim, $\theta^{\min} \equiv \theta^{MM}$.

No capítulo 3, foram utilizadas técnicas de AGs na solução deste problema de otimização, a qual foi abordado num contexto de minimização. Entretanto, a seguir será abordado este problema num contexto de maximização.

Observando o pseudo código PE dos AGs, vê-se claramente que o algoritmo PE segue uma estratégia seqüencial na procura de θ^{MM} . Isto sugere modelar o AG usando a teoria de MCS [15, 52,53] descrita no capítulo 2. Assim, o AG pode ser modelado como um processo estocástico $\pi(\Theta_{0:t})$, definido num espaço dimensional crescente $\Theta_{0:t} = \{\theta_0, \theta_1, \dots, \theta_t\}$ com o tempo. Além disso, a função objetivo $p(\theta)$ a ser maximizado pode ser usado para definir um outro processo estocástico $p(\Theta_{0:t})$ com a condição que suas densidades marginais $p(\theta_0), p(\theta_1), \dots, p(\theta_t)$ sejam “invariantes no tempo” e todas iguais a $p(\theta)$. Neste contexto, o objetivo do AG $\pi(\Theta_{0:t})$ é aproximar-se suavemente da densidade objetivo $p(\Theta_{0:t})$, conforme o transcurso do tempo, e com isto estimar algumas quantidades desconhecidas de $p(\Theta_{0:t})$, tal como sua moda global θ^{MM} . Note que nesta modelagem do AG, o índice t representa uma iteração do AG a qual é modelado através de um “tempo fictício”.

Uma forma de monitorar a convergência de $\pi(\Theta_{0:t})$ para $p(\Theta_{0:t})$ é através da distância *Chi-square*, definida como a variância $\text{var}_\pi[\omega(\Theta_{0:t})]$ de $\omega(\Theta_{0:t})$, definida por:

$$\omega(\Theta_{0:t}) = \frac{p(\Theta_{0:t})}{\pi(\Theta_{0:t})} \quad (68)$$

Tipicamente, $\omega(\Theta_{0:t})$ é denominada como o *peso de importância* na comunidade de MCS; já na comunidade de AG é denominada de *aptidão*. Contudo, grandes valores da distância *Chi-square* indicam uma aproximação deficiente de $\pi(\Theta_{0:t})$ com relação a $p(\Theta_{0:t})$. Na prática, a distância *Chi-square* usualmente não pode ser calculada diretamente, uma vez que é difícil simular amostras de $p(\Theta_{0:t})$. Felizmente uma aproximação baseada no *coeficiente de variação*, $C^2(\omega_t)$, foi sugerida em [54],

$$C^2(\tilde{\omega}) = \frac{1}{N} \sum_{i=1}^N \{N\tilde{\omega}_t^{(i)} - 1\} \quad (69)$$

onde $\tilde{\omega}_t^{(i)}$ é denominado de *peso de importância normalizado* ou de *aptidão normalizada* tanto pela comunidade de MCS quanto pela de AG; e definida por:

$$\tilde{\omega}_t^{(i)} = \frac{\omega(\Theta_{0:t}^{(i)})}{\sum_{j=1}^N \omega(\Theta_{0:t}^{(j)})} \quad (70)$$

Com base neste conjunto de amostras ponderadas $\{\Theta_{0:t}^{(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$, a densidade $p(\Theta_{0:t})$ pode ser aproximada empiricamente por:

$$\hat{p}(\Theta_{0:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\Theta_{0:t}^{(i)}}(\Theta_{0:t}) \quad (71)$$

onde $\delta_{\Theta_{0:t}^{(i)}}(\Theta_{0:t})$ denota a massa do delta-Dirac localizado em $\Theta_{0:t}^{(i)}$. Uma forma eficiente de estimar $\tilde{\omega}_t^{(i)}$ ao longo do tempo é através do algoritmo de *amostragem por importância* [54-58], a qual aproveita a estrutura recursiva da equação (68)

$$\begin{aligned} \omega(\Theta_{0:t}) &= \frac{p(\Theta_0)p(\Theta_1|\Theta_0)\cdots p(\Theta_t|\Theta_0,\dots,\Theta_{t-1})}{\pi(\Theta_0)\pi(\Theta_1|\Theta_0)\cdots \pi(\Theta_t|\Theta_0,\dots,\Theta_{t-1})} \\ &= \frac{p(\Theta_{0:t-1})p(\Theta_t|\Theta_0,\dots,\Theta_{t-1})}{\pi(\Theta_{0:t-1})\pi(\Theta_t|\Theta_0,\dots,\Theta_{t-1})} \\ &= \omega(\Theta_{0:t-1}) \frac{p(\Theta_t|\Theta_0,\dots,\Theta_{t-1})}{\pi(\Theta_t|\Theta_0,\dots,\Theta_{t-1})} \end{aligned} \quad (72)$$

Substituindo a equação acima na equação (70), o *peso de importância normalizado* satisfaz a seguinte forma recursiva:

$$\tilde{\omega}_t^{(i)} \propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1})} \quad (73)$$

Relembrando que o processo estocástico $p(\boldsymbol{\Theta}_{0:t})$ é invariante no tempo, então $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1}) = p(\boldsymbol{\theta})$. Entretanto, o processo evolutivo $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1})$ precisa ser cuidadosamente formulado de forma a garantir a convergência de $\pi(\boldsymbol{\Theta}_{0:t})$ para $p(\boldsymbol{\Theta}_{0:t})$. Isto pode ser feito se $\pi(\boldsymbol{\Theta}_{0:t})$ é definido como uma cadeia Markoviana $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1}) = \pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ *irredutível, aperiódica* e que satisfaça à seguinte *condição de reversibilidade*:

$$p(\boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) = p(\boldsymbol{\theta}_{t-1})\pi(\boldsymbol{\theta}_{t-1} | \boldsymbol{\theta}_t) \quad (74)$$

Com isto, se cumpre que:

$$\int p(\boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})d\boldsymbol{\theta}_t = \int p(\boldsymbol{\theta}_{t-1})\pi(\boldsymbol{\theta}_{t-1} | \boldsymbol{\theta}_t)d\boldsymbol{\theta}_t = \pi(\boldsymbol{\theta}_{t-1}) \quad (75)$$

mostrando que o processo Markoviano $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ possui $p(\boldsymbol{\theta})$ como sua *densidade estacionária*, garantindo assim a convergência do AG $\pi(\boldsymbol{\Theta}_{0:t})$ para o processo estocástico $p(\boldsymbol{\Theta}_{0:t})$. Seguindo as idéias da seção anterior, $\pi(\boldsymbol{\Theta}_{0:t})$ como qualquer outro processo Markoviano, pode ser decomposto por [27]:

$$\pi(\boldsymbol{\Theta}_{0:t}) = g(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})\alpha(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t}) \quad (76)$$

um *operador genético* $g(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})$ e um *esquema de seleção* $\alpha(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})$. O objetivo do *operador genético* é modificar ligeiramente a trajetória convencional $\boldsymbol{\Theta}_{0:t} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t\}$ dos indivíduos da população, para uma trajetória alternativa $\boldsymbol{\Theta}_{0:t}^* = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t^*\}$ a qual é aceita ou rejeitada segundo o *esquema de seleção*, que tem como objetivo preservar a densidade $p(\boldsymbol{\Theta}_{0:t})$ como *densidade estacionária* de $\pi(\boldsymbol{\Theta}_{0:t})$.

A análise teórica do *operador genético* $g(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})$ é idêntica à da seção anterior. Já o *esquema de seleção* $\alpha(\boldsymbol{\Theta}_{0:t}^* | \boldsymbol{\Theta}_{0:t})$ pode ser executado de forma local ou global, preservando em qualquer destes casos a invariância de $p(\boldsymbol{\Theta}_{0:t})$ em $\pi(\boldsymbol{\Theta}_{0:t})$. Para o caso de um *esquema de seleção* local pode ser adaptado adequadamente o critério de Metropolis *et al.* [30] e Hastings [31] da seguinte forma

$$\begin{aligned}
\alpha(\Theta_{0:t}^* | \Theta_{0:t}) &= \min \left\{ 1, \frac{p(\Theta_{0:t}^*) \pi_g(\Theta_{0:t}^* | \Theta_{0:t})}{p(\Theta_{0:t}) \pi_g(\Theta_{0:t} | \Theta_{0:t}^*)} \right\} \\
&= \min \left\{ 1, \frac{p(\Theta_0) p(\Theta_1 | \Theta_0) \dots p(\Theta_t^* | \Theta_{t-1}) \pi_g(\Theta_{0:t}^* | \Theta_{0:t})}{p(\Theta_0) p(\Theta_1 | \Theta_0) \dots p(\Theta_t | \Theta_{t-1}) \pi_g(\Theta_{0:t} | \Theta_{0:t}^*)} \right\} \\
&= \min \left\{ 1, \frac{p(\Theta_t^* | \Theta_{t-1}) \pi_g(\Theta_t^* | \Theta_{0:t-1}^*, \Theta_{0:t}) \pi_g(\Theta_{0:t-1}^* | \Theta_{0:t})}{p(\Theta_t | \Theta_{t-1}) \pi_g(\Theta_t | \Theta_{0:t-1}, \Theta_{0:t}^*) \pi_g(\Theta_{0:t-1} | \Theta_{0:t}^*)} \right\} \\
&= \min \left\{ 1, \frac{p(\Theta_t^*) \pi_g(\Theta_t^* | \Theta_t)}{p(\Theta_t) \pi_g(\Theta_t | \Theta_t^*)} \right\}
\end{aligned} \tag{77}$$

A aplicação da equação anterior para os exemplos dos operadores genéticos de *mutação* e de *cruzamento*, definidos pelas equações (50) e (51), são realizados de forma análoga à seção anterior, com a diferença que ρ_m, ρ_c são definidos por:

$$\rho_m = \frac{p(\theta_t^{(i)}) \pi_m(\theta_t^{(i)} | \theta_t^{*(i)})}{p(\theta_t^{*(i)}) \pi_m(\theta_t^{*(i)} | \theta_t^{(i)})} \tag{78}$$

$$\rho_c = \frac{p(\theta_t^{(i)}) p(\theta_t^{(j)}) \pi_c(\theta_t^{(i)}, \theta_t^{(j)} | \theta_t^{*(i)}, \theta_t^{*(j)})}{p(\theta_t^{*(i)}) p(\theta_t^{*(j)}) \pi_c(\theta_t^{*(i)}, \theta_t^{*(j)} | \theta_t^{(i)}, \theta_t^{(j)})} \tag{79}$$

Entretanto, para o caso de um *esquema de seleção* global, ρ_m, ρ_c são denominados como a probabilidade de *cruzamento* $\rho_c \in [0,1)$ e a probabilidade de *mutação* $\rho_m \in [0,1)$. Após de determinar previamente os valores de ρ_m, ρ_c , os indivíduos $\theta_t^{(i)}$ da população atual são evoluídos pelos operadores genéticos formando a população de indivíduos $\theta_t^{*(i)}$ candidatos (filhos). Logo, calcula-se a *aptidão normalizada* $\tilde{\omega}_t^{(i)}$ dos indivíduos candidatos através da equação (73); em finalmente executa-se qualquer esquema de *amostragem com reposição* [20-22], tal como *seleção multinomial*, *seleção por roleta*, *seleção estocástica universal*, etc.

Com base nos resultados teóricos desenvolvidos nesta seção, pode ser derivado o seguinte algoritmo computacional, denominado neste trabalho de *Filtro Evolutivo* (FE).

Procedimento Filtro Evolutivo

Início

$$t \leftarrow 1$$

Iniciar $\{\theta_t^{*(i)}\}_{i=1}^N$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ de $\{\theta_t^{*(i)}\}_{i=1}^N$

Enquanto (condição de parada não é satisfeita) Fazer

Início

$$t \leftarrow t + 1$$

Evoluir $\{\theta_t^{*(i)}\}_{i=1}^N$ de $\{\theta_{t-1}^{(i)}\}_{i=1}^N$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ de $\{\theta_t^{*(i)}\}_{i=1}^N$

Fim

Fim

onde $\{\theta_t^{(i)}\}_{i=1}^N$ representa as amostras distribuídas de acordo com $p(\theta_t)$, e $\{\theta_t^{*(i)}\}_{i=1}^N$ as amostras distribuídas de acordo com $\pi(\theta_t)$.

Os detalhes da implementação do algoritmo do FE, considerando um *esquema de seleção* global são explicados a seguir:

Iniciar $\{\theta_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, indivíduos $\theta_1^{*(i)} = \theta_0^{(i)}$ sendo $\theta_0^{(i)} \sim \pi(\theta_0)$.

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, indivíduos, avaliar a *aptidão normalizada*

$$\tilde{\omega}_t^{(i)} = \frac{p(\theta_t^{*(i)})}{\sum_{j=1}^N p(\theta_t^{*(j)})} \quad (80)$$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ de $\{\theta_t^{*(i)}\}_{i=1}^N$:

Amostrar com substituição N indivíduos $\{\theta_t^{(i)}, N^{-1}\}_{i=1}^N$ a partir dos indivíduos

candidatos $\{\theta_t^{*(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$ de acordo com sua *aptidão normalizada*.

Evoluir $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, indivíduos $\boldsymbol{\theta}_t^{(i)} = \boldsymbol{\theta}_{t-1}^{(i)}$.

Para $i = 1, \dots, N/2$, $j = N - i + 1$, $u \sim \mathbf{U}(0,1)$

$$(\tilde{\boldsymbol{\theta}}_t^{(i)}, \tilde{\boldsymbol{\theta}}_t^{(j)}) = \begin{cases} \sim \pi_c(\cdot | \boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) & \text{if } u \leq \rho_c \\ (\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) & \text{otherwise} \end{cases} \quad (81)$$

Para $i = 1, \dots, N$, $u \sim \mathbf{U}(0,1)$

$$\boldsymbol{\theta}_t^{*(i)} = \begin{cases} \sim \pi_m(\cdot | \tilde{\boldsymbol{\theta}}_t^{(i)}) & \text{if } u \leq \rho_m \\ \tilde{\boldsymbol{\theta}}_t^{(i)} & \text{otherwise} \end{cases} \quad (82)$$

Fixar $\boldsymbol{\Theta}_{0t}^{*(i)} = \{\boldsymbol{\Theta}_{0t}^{(i)}, \boldsymbol{\theta}_t^{*(i)}\}$.

Note que na equação (80) não aparece $\tilde{\boldsymbol{\omega}}_{t-1}^{(i)}$ uma vez que os indivíduos a serem evoluídos $\{\boldsymbol{\Theta}_{0t-1}^{(i)}; i = 1, \dots, N\}$ possuem pesos uniformes após do estágio *Selecionar*, no tempo $t - 1$. Além disso, não é necessário armazenar todas as trajetórias dos indivíduos o tempo 0 até o tempo t , uma vez que a cada instante de tempo está-se interessado em estimar as densidades marginais $p(\boldsymbol{\theta}_t)$.

A convergência de $\pi(\boldsymbol{\Theta}_{0t})$ para $p(\boldsymbol{\Theta}_{0t})$ evoluído pelo algoritmo FE tem sido largamente estudada pela teoria MCS (veja detalhes em [15,59,60]). Porém, usando a lei dos grandes números, isto é, $N \rightarrow \infty$ e $t \rightarrow \infty$, é esperado que uma das máximas modas $\boldsymbol{\theta}_t^{MM}$ das densidades marginais seja exatamente a moda global $\boldsymbol{\theta}^{MM}$ de $p(\boldsymbol{\Theta}_{0t})$ e, como $\boldsymbol{\theta}^{\min} \equiv \boldsymbol{\theta}^{MM}$, então este mesmo ponto é o mínimo valor global da função $\mathbf{h}(\boldsymbol{\theta})$. Por outro lado, percebe-se que o pseudo código do algoritmo do FE não é exatamente igual ao pseudo código do algoritmo PE dos AGs convencionais mostrado no capítulo 3. Entretanto, é fácil reconhecer que o algoritmo FE segue também o princípio Darwiniano da “sobrevivência dos mais aptos”. Neste contexto o algoritmo FE é simplesmente uma nova versão de pseudo código de AG, como muitos outros existentes na literatura [61-63]. A principal vantagem da versão FE é que a cada instante de tempo a população está distribuída de acordo com $p(\boldsymbol{\theta}_t)$ enquanto que na versão PE a população está distribuída de acordo com $\pi(\boldsymbol{\theta}_t)$. Além desta ligeira diferença, a convergência do

algoritmo PE para a moda global θ^{MM} de $p(\Theta_{0:t})$ é garantida pelos seguintes dois argumentos: (i) os dois algoritmos (FE e PE) executam as mesmas equações e (ii) o algoritmo PE como qualquer AG convencional usa uma regra de seleção *elitista* pela qual é selecionado ou preservado o melhor indivíduo $\theta^{elitista}$ evoluído pelo algoritmo. Logo conclui-se que:

$$\theta^{elitista} = \theta^{MM} = \theta^{\min} \quad (83)$$

provando desta maneira a convergência dos AGs em espaços contínuos. Entretanto, o FE é uma AG especial, já que além de sua propriedade de convergência a θ^{MM} , o FE converge também a densidade $p(\theta)$. Em consequência, o FE é um amostrador de densidades de probabilidade “estacionárias”. Logo, o FE é um algoritmo de aprendizado *batch* evolucionário, equivalente ao método MCCM que é um algoritmo de aprendizado *batch* Bayesiano.

Com base na equivalência existente entre o FPG e os métodos MCS, assim como do FE com os métodos MCCM, pode ser concluído a existência de uma equivalência entre o princípio evolucionário e teorema de Bayes a qual é mostrado na figura 3.

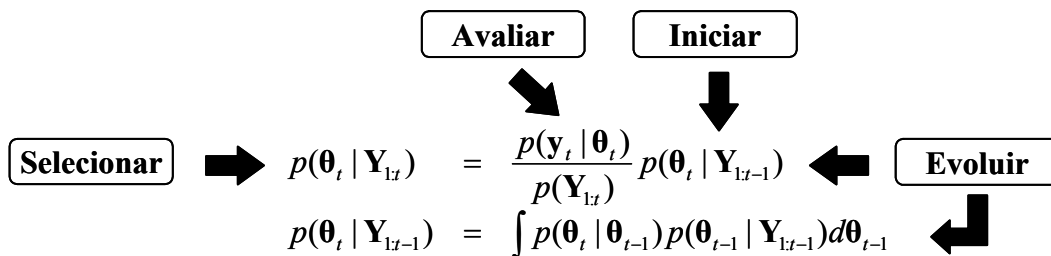


Figura 3 - Equivalência entre o princípio evolucionário e o teorema de Bayes

Com base nesta equivalência, o teorema de Bayes pode ser visto como um formalismo teórico do princípio evolucionário e, em contrapartida, o princípio evolucionário pode ser visto uma forma de implementação discreta do teorema de Bayes. Independentemente da perspectiva adotada, os algoritmos de aprendizado *batch/online*, Bayesianos-Evolucionários apresentam ampla generalidade pois permitem analisar sinais complexos de problemas reais, caracterizados por comportamentos não-lineares, não-gaussianos e não-estacionários. Além disso, esta equivalência confirma que o princípio evolucionário forma um novo mecanismo de inferência, denominada neste trabalho de *inferência evolucionária*.

4.4. Resumo

Este capítulo apresentou as contribuições desta tese, isto é, o estabelecimento de uma equivalência entre o teorema de Bayes e o princípio evolucionário, permitindo, assim, o desenvolvimento de um novo mecanismo de cálculo dos estados do MEE, denominado de inferência evolucionária. Também foram apresentados o FPG e do FE, que são algoritmos de aprendizado *online* e *batch* evolucionários, respectivamente. Além disso, foi mostrado que o FE é, em essência, um AG pois, além da sua capacidade de convergência às distribuições de probabilidade, o FE converge também a sua moda global. Consequentemente, a fundamentação teórica desenvolvida para o FE demonstra, analiticamente, a convergência dos AGs em espaços contínuos. Além disso, a equivalência entre o teorema de Bayes e do princípio evolucionário, permitiu o estabelecimento de uma equivalência entre o FPG e o método MCS, uma vez que ambos métodos são usados na amostragem de distribuições de probabilidade variantes no tempo. Também, foi estabelecido a equivalência entre o FE e o método MCCM, uma vez que ambos são usados na amostragem de distribuições de probabilidade estacionárias.

No próximo capítulo são apresentadas aplicações do FPG e FE em problemas de filtragem dos estados de um sistema não-linear unidimensional, problemas de otimização de funções e problemas de aprendizado de hiperparâmetros de um sistema de volatilidade estocástica. O objetivo destas aplicações é avaliar o desempenho e robustez do FPG e do FE.