

# Taxa de Mutação Efetiva no Modelo de Moran

**Telles Timóteo da Silva\***

Depto de Física e Matemática, UFSJ,  
36420-000, Ouro Branco, MG  
E-mail: timoteo@ufsj.edu.br

**Resumo:** *Neste trabalho desenvolvemos um modelo de Moran que inclui flutuação temporal no parâmetro de mutação através de uma cadeia de Markov. Investigamos o comportamento temporal assintótico de tal modelo, comparando-o com o modelo de Moran sem flutuação. Conseguimos, desta forma, atribuir um valor para a taxa de mutação efetiva, na dependência dos parâmetros do modelo com flutuação.*

**Palavras-chave:** *Modelo de Moran, Mutação, Cadeia de Markov Estacionária*

## 1 Introdução

A genética de populações estuda a evolução do conjunto de genes de uma população de indivíduos, quando sujeitos às forças que tendem a modificar o conteúdo genético desta população [2]. Os modelos matemáticos são elaborados para tentar explicar e prever a forma como se dá esta evolução. Dentre os modelos discretos em genética populacional destacam-se o modelo de Wright-Fisher [4] e o modelo de Moran [6]. O modelo de Wright-Fisher se aplica a situações em que as gerações da população em estudo não se sobrepõem. Já o modelo de Moran tenta captar o efeito de sobreposição de gerações, pois consiste na reposição de apenas um indivíduo a cada geração.

Neste trabalho desenvolvemos um modelo de Moran que inclui flutuação temporal no parâmetro de mutação através de uma cadeia de Markov. Investigamos o comportamento temporal assintótico de tal modelo, comparando-o com o modelo de Moran sem flutuação. Conseguimos, desta forma, atribuir um valor para a taxa de mutação efetiva, na dependência dos parâmetros do modelo com flutuação.

O texto está dividido da seguinte forma. Na Seção 2 apresentamos o modelo de Moran clássico (sem flutuação), com mutação, e suas principais características relevantes para este trabalho. A seguir (Seção 3), desenvolvemos o modelo de Moran com flutuação na taxa de mutação. A análise do modelo e a obtenção da taxa de mutação efetiva é feita na Seção 4. Uma possível aplicação do resultado obtido está delineada na Seção 5.

## 2 Modelo de Moran

Considere uma população haplóide fixa de  $N$  indivíduos. Analisando um locus genético que admite apenas dois alelos  $A$  e  $a$ , seja  $f(t)$  a frequência com que o gene  $A$  ocorre no tempo  $t$ . Suponha que um indivíduo da população é escolhido aleatoriamente para se reproduzir (isto é, gerar uma cópia de si) e que um indivíduo é escolhido para morrer. Além disso, os indivíduos estão sujeitos a sofrerem mutação: o gene  $A$  se transforma em  $a$  ou o gene  $a$  se transforma em

---

\*Trabalho parcialmente financiado pelo CNPq Processo No. 314729/2009-7 e FAPEMIG Processo No. 00253-10.

A com probabilidade  $\mu$ . Então temos as seguintes probabilidades de transição:

$$\begin{aligned} u_{i,i-1} &= \mathcal{P}\left(f(t+1) = \frac{i-1}{N} \mid f(t) = \frac{i}{N}\right) = \frac{i[\mu i + (1-\mu)(N-i)]}{N^2} \\ u_{i,i+1} &= \mathcal{P}\left(f(t+1) = \frac{i+1}{N} \mid f(t) = \frac{i}{N}\right) = \frac{(N-i)[(1-\mu)i + \mu(N-i)]}{N^2} \\ u_{i,i} &= \mathcal{P}\left(f(t+1) = \frac{i}{N} \mid f(t) = \frac{i}{N}\right) = 1 - u_{i,i-1} - u_{i,i+1} \\ u_{i,j} &= 0, \text{ se } |i-j| \geq 2 \end{aligned}$$

Uma medida efetiva do estado do processo num dado instante de tempo é a heterozigosidade  $H(t)$  definida como sendo a probabilidade de dois genes tomados ao acaso com reposição serem distintos, sendo assim dada por

$$H(t) = 2f(t)[1 - f(t)].$$

No modelo de Moran com mutação, onde  $\mu$  é diferente de zero, nenhum dos genes se extingue na população, e portanto  $H(t)$  é uma medida sempre não-nula.

Podemos calcular o valor esperado condicional da frequência do gene  $A$  no tempo  $t+1$ , dada a frequência no tempo  $t$ , i.e.,  $E[f(t+1)|f(t)]$ , obtendo

$$E[f(t+1)|f(t)] = f(t) \left(1 - \frac{2\mu}{N}\right) + \frac{\mu}{N}$$

e sendo fácil de ver que

$$\lim_{t \rightarrow \infty} E[f(t)] = \frac{1}{2}, \tag{1}$$

pois o modelo é simétrico com respeito à permuta entre os dois tipos de genes alelos.

Da mesma forma, podemos calcular o valor esperado condicional de  $H(t+1)$  dado  $f(t)$ , para o qual obtemos

$$E[H(t+1)|f(t)] = \left(1 - \frac{4\mu}{N} - \frac{2-4\mu}{N^2}\right) H(t) + \frac{2\mu}{N} \left(1 - \frac{1}{N}\right)$$

além disso, temos o limite

$$\lim_{t \rightarrow \infty} E[H(t)] = \frac{2\mu}{N} \left(1 - \frac{1}{N}\right) \left[ \frac{1}{\frac{2}{N^2} + \frac{4\mu}{N} \left(1 - \frac{1}{N}\right)} \right] = \frac{1}{\frac{1}{\mu(N-1)} + 2}. \tag{2}$$

Note que quando  $N$  tende a infinito na expressão acima, obtemos que a heterozigosidade tende a  $1/2$ , o que era de se esperar intuitivamente. Porém para populações cujo contingente populacional é finito e pequeno, o valor assintótico da média da heterozigosidade dado pelo limite na expressão (2) diminui à medida que diminui a taxa de mutação, como ilustra a Figura 1.

Nosso objetivo neste trabalho é desenvolver um modelo de Moran que inclua flutuação temporal no parâmetro de mutação. Buscaremos, então, investigar se tal modelo se comporta assintoticamente como algum modelo com taxa de mutação constante. A base de comparação é o comportamento assintótico temporal da medida de heterozigosidade da população. No modelo com flutuação que vamos estudar, a taxa de mutação varia segundo uma cadeia de Markov.

### 3 Modelo de Moran com flutuação na taxa de mutação

Seja  $\delta(t)$  uma cadeia de Markov estacionária,  $\delta(t) \in \{1, 2\}$ , com matriz de transição:

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} 1-q & q \\ r & 1-r \end{bmatrix}$$

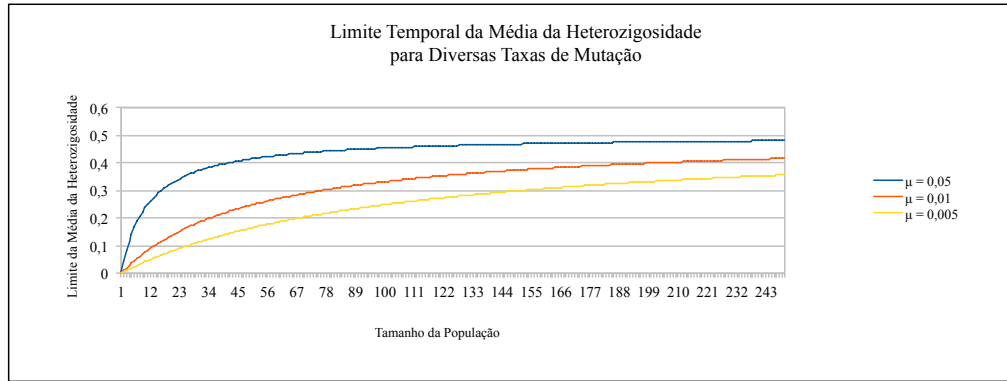


Figura 1: Valor assintótico da média da heterozigidade para diferentes taxas de mutação à medida que aumenta o tamanho da população.

e distribuição inicial estacionária

$$\pi = [\pi_1 \ \pi_2] = \left[ \frac{r}{q+r} \quad \frac{q}{q+r} \right].$$

A cadeia  $\delta(t)$  é tomada independente de  $f(t)$ .

Observe que

$$\mathbf{P}^t = \begin{bmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{bmatrix} = \frac{1}{r+q} \begin{bmatrix} r+q(1-r-q)^t & q-q(1-r-q)^t \\ r-r(1-r-q)^t & q+r(1-r-q)^t \end{bmatrix}$$

e assim

$$\lim_{t \rightarrow \infty} \mathbf{P}^t = \frac{1}{r+q} \begin{bmatrix} r & q \\ r & q \end{bmatrix}$$

no caso em que  $0 < r+q < 2$ .

Seja  $\mu(t)$  uma seqüência tal que  $\mu(t) = \mu_i$ , se  $\delta(t) = i$ , para  $i = 1, 2$ , com  $0 \leq \mu_i \leq 1$ . Teremos portanto um **modelo de Moran onde a taxa de mutação que atua sobre a população no instante  $t$  é escolhida dentre dois possíveis parâmetros  $\mu_1$  ou  $\mu_2$  segundo o estado da cadeia de Markov  $\delta(t)$ .**

Note que, por causa do papel simétrico desempenhado pelos alelos  $A$  e  $a$ , continuamos a ter ainda que

$$\lim_{t \rightarrow \infty} E[f(t)] = \frac{1}{2},$$

como na expressão (1). Desta forma o processo  $f(t)$  não fornece uma base para comparação entre modelos. Por isso buscamos uma base de comparação utilizando a medida de heterozigidade.

Defina

$$\begin{aligned} \gamma(t) &= -2 \frac{\mu(t)}{N} \left( 1 - \frac{1}{N} \right) \\ \alpha(t) &= 1 - \frac{2}{N^2} + 2\gamma_\mu(t) \end{aligned}$$

Do modelo de Moran temos agora que o valor esperado da heterozigidade  $\mathcal{H}(t+1)$  condicionado a  $f(t)$  e ao estado da cadeia de Markov  $\delta(t)$  é dado por

$$E[\mathcal{H}(t+1)|f(t), \delta(t)] = \mathcal{H}(t)\alpha(t) - \gamma(t).$$

Agora, tomando o valor esperado condicional de  $\mathcal{H}(t+1)$  com relação aos valores de  $f(0)$  e  $\delta(0)$  no instante inicial temos, pela estacionariedade da cadeia de Markov  $\delta(t)$ ,

$$E[\mathcal{H}(t+1)|f(0), \delta(0)] = E \left[ \mathcal{H}(0) \prod_{k=0}^t \alpha(k) - \gamma(s) \sum_{s=0}^{t-1} \left[ \prod_{k=s+1}^t \alpha(k) \right] - \gamma(t) \mid f(0), \delta(0) \right]. \quad (3)$$

Uma referência para o método que utilizamos é o trabalho de M. Iizuka [5].

Definindo

$$g_i(t) = E \left[ \prod_{k=0}^t \alpha(k) \mid \delta(0) = i \right]$$

temos, para  $i \neq j$ ,

$$g_i(t+2) = g_i(t+1)[(1-q)\alpha_i + (1-r)\alpha_j] + [q+r-1]g_i(t)\alpha_i\alpha_j \tag{4}$$

onde  $\alpha_i = \alpha(0)|_{\delta(0)=i}$ .

Também definimos, para  $t \geq 1, 0 \leq s \leq t-1$ :

$$m_i(t-s) = -E \left\{ \gamma(s) \prod_{k=s+1}^t \alpha(k) \mid f_A(0), \delta(0) = i \right\}$$

donde vem

$$m_i(t-s) = -p_{i1}(s) \{ (1-q)g_1(t-s-1) + qg_2(t-s-1) \} \gamma_1(s) - p_{i2}(s) \{ rg_1(t-s-1) + (1-r)g_2(t-s-1) \} \gamma_2(s)$$

onde  $\gamma_i(s) = \gamma(s)|_{\delta(s)=i}$ .

Assim a equação (3) fica

$$E[\mathcal{H}(t+1)|f(0), \delta(0) = i] = g_i(t)\mathcal{H}(0) + \sum_{s=0}^{t-1} m_i(t-s) - \gamma_1(t)p_{i1}(t) - \gamma_2(t)p_{i2}(t) \tag{5}$$

donde segue

$$E[\mathcal{H}(t+1)] = \sum_{i=1}^2 \left\{ g_i(t)\mathcal{H}(0) + \sum_{s=0}^{t-1} m_i(t-s) - \gamma_1(t)p_{i1}(t) - \gamma_2(t)p_{i2}(t) \right\} \pi_i$$

A análise do comportamento de  $E[\mathcal{H}(t+1)|f(0), \delta(0) = i]$  depende, primeiramente, da solução da equação (4), que é dada por

$$g_i(t) = c_i^+ \lambda_+^t + c_i^- \lambda_-^t$$

onde

$$\begin{aligned} \lambda_+ &= \frac{a + \sqrt{a^2 + 4b}}{2} \\ \lambda_- &= \frac{a - \sqrt{a^2 + 4b}}{2} \\ c_i^+ &= \frac{\alpha_i}{2} \left[ 1 + \frac{2}{\sqrt{a^2 + 4b}} \left( \alpha_1 p_{i1} + \alpha_2 p_{i2} - \frac{a}{2} \right) \right] \\ c_i^- &= \frac{\alpha_i}{2} \left[ 1 - \frac{2}{\sqrt{a^2 + 4b}} \left( \alpha_1 p_{i1} + \alpha_2 p_{i2} - \frac{a}{2} \right) \right] \\ a &= (1-q)\alpha_1 + (1-r)\alpha_2 \\ b &= \alpha_1 \alpha_2 (q+r-1). \end{aligned}$$

Observe que  $\lim_{t \rightarrow \infty} g_i(t) = 0$ .

Calculando somas de progressões geométricas e passando o limite, obtemos que

$$\begin{aligned} \lim_{t \rightarrow \infty} \sum_{s=0}^{t-1} m_i(t-s) &= \frac{1}{r+q} \left[ \left( \frac{c_1^+}{\lambda_+ - 1} + \frac{c_1^-}{\lambda_- - 1} \right) (r\gamma_1 p_{11} + q\gamma_2 p_{21}) \right. \\ &\quad \left. + \left( \frac{c_2^+}{\lambda_+ - 1} + \frac{c_2^-}{\lambda_- - 1} \right) (r\gamma_1 p_{12} + q\gamma_2 p_{22}) \right] \\ \lim_{t \rightarrow \infty} \gamma_1 p_{i1}(t) &= \gamma_1 \frac{r}{r+q} \\ \lim_{t \rightarrow \infty} \gamma_2 p_{i2}(t) &= \gamma_2 \frac{q}{r+q} \end{aligned}$$

Tomando o limite na equação (5) e utilizando as parcelas anteriores, obtemos

$$\begin{aligned} \lim_{t \rightarrow \infty} E[\mathcal{H}(t)|f(0), \delta(0) = i] &= \frac{1}{r+q} \left[ \left( \frac{c_1^+}{\lambda_+ - 1} + \frac{c_1^-}{\lambda_- - 1} \right) (r\gamma_1 p_{11} + q\gamma_2 p_{21}) \right. \\ &\quad \left. + \left( \frac{c_2^+}{\lambda_+ - 1} + \frac{c_2^-}{\lambda_- - 1} \right) (r\gamma_1 p_{12} + q\gamma_2 p_{22}) - r\gamma_1 - q\gamma_2 \right] \end{aligned}$$

Após simplificação, chegamos a

$$\lim_{t \rightarrow \infty} E[\mathcal{H}(t)|f(0), \delta(0) = i] = \frac{1}{1-a-b} \left\{ 2\gamma_1\gamma_2(1-r-q) + (r\gamma_1 + q\gamma_2) \left[ \frac{2}{N^2} - \frac{2}{(r+q)N^2} - 1 \right] \right\}. \tag{6}$$

Em particular, o valor limite independe do estado inicial da cadeia de Markov  $\delta(t)$ .

## 4 Análise

Note que se tivermos  $\mu_1 = \mu_2$ , então  $\gamma_1$  será igual a  $\gamma_2$ ,  $\alpha_1$  se igualará a  $\alpha_2$  e, portanto, do limite (6) obteremos, novamente, o limite (2) como caso particular.

Caso  $\mu_1 \neq \mu_2$ , podemos atribuir um valor para a **taxa de mutação efetiva** ao igualarmos os limites (2) e (6). A taxa de mutação efetiva é uma taxa de mutação constante  $\mu^*$  que produz assintoticamente o mesmo efeito sobre a heterozigidade de uma população com taxa de mutação flutuante. Igualando (2) e (6) obtemos

$$\frac{1}{\frac{1}{\mu^*(N-1)} + 2} = \frac{1}{1-a-b} \left\{ 2\gamma_1\gamma_2(1-r-q) + (r\gamma_1 + q\gamma_2) \left[ \frac{2}{N^2} - \frac{2}{(r+q)N^2} - 1 \right] \right\}$$

e explicitando  $\mu^*$  temos

$$\mu^* = \frac{\frac{4\mu_1\mu_2}{N} \left(1 - \frac{1}{N}\right) (1-r-q) + (r\mu_1 + q\mu_2) \left[1 - \frac{2}{N^2} + \frac{2}{(r+q)N^2}\right]}{(r+q) \left[1 - \frac{2}{N^2} + \frac{2}{(r+q)N^2}\right] + \frac{4}{N} \left(1 - \frac{1}{N}\right) \left[\mu_2(1-r) + \mu_1(1-q) - \frac{r\mu_1 + q\mu_2}{r+q}\right]} \tag{7}$$

Para  $N \rightarrow \infty$  obtemos ainda

$$\lim_{N \rightarrow \infty} \mu^* = \frac{r\mu_1 + q\mu_2}{r+q}, \tag{8}$$

isto é,  $\mu^*$  se aproxima da média ponderada das taxas de mutação.

A Figura 2 ilustra o comportamento da taxa de mutação efetiva para diversos valores de tamanho de população. Observe que a curva para  $N = 100$  é quase indistinguível da curva para  $N \rightarrow \infty$ .

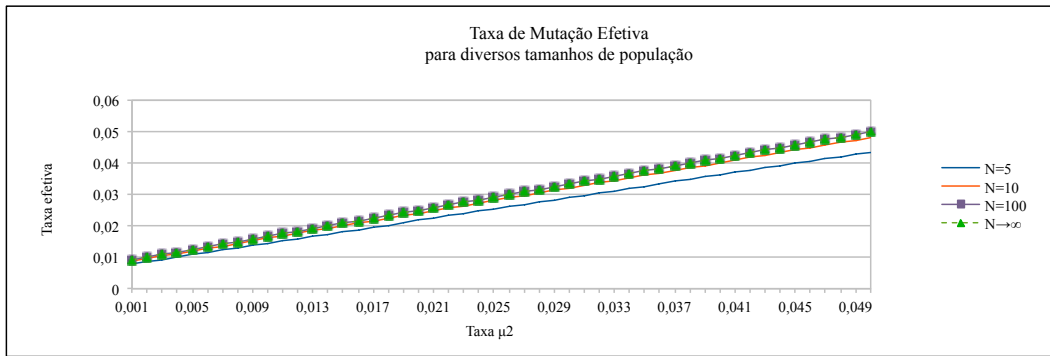


Figura 2: Taxa de mutação efetiva  $\mu^*$  para diversos valores de tamanho de população, calculada com base nas expressões (7) e (8), em função da taxa de mutação  $\mu_2$ . Parâmetros utilizados  $\mu_1 = 0,05$ ,  $r = 0,1$  e  $q = 0,5$ .

## 5 Aplicação

Taxas de mutação de determinados sítios de DNA são utilizadas para se estimar eventos de diferenciação genética [3]. Modelos filogenéticos, por exemplo, utilizam o número de substituições (mutações) de nucleotídeos que ocorrem entre sítios de diferentes indivíduos, sejam da mesma ou de espécies distintas, para estimar árvores evolutivas. Sabendo-se o número de substituições e a taxa com que ocorrem é possível estimar há quanto tempo ocorreu a diferenciação [4]. Porém, este *relógio molecular*, que mede a passagem de tempo através de alterações no DNA, não necessariamente é regular em suas batidas. O processo de substituição de nucleotídeos é influenciado pelo tipo de mutação que ocorre, quando ela ocorre, se a mutação vai se fixar ou se será perdida [1].

Ao longo da história natural de uma população, portanto, as taxas de mutação podem flutuar devido a mudanças no habitat.

O modelo proposto no presente trabalho dá um pequeno passo no sentido de levar em conta aspectos reais e relevantes do mundo natural, já que consegue lidar com uma característica importante para a estimação mais acurada de árvores filogenéticas, qual seja, a possibilidade de que o par *relógio molecular - taxa de mutação* não seja constante ao longo do tempo. Além disso, o modelo permite intuir que, ao se observar a taxa de mutação por uma janela temporal estreita, pode-se incorrer no erro de sub ou sobre estimar o tempo de diferenciação de espécies. Decorre que a taxa de mutação *efetiva* deve ser utilizada preferencialmente nas aplicações, desde que seja possível estimá-la em populações reais.

## 6 Conclusão

Neste trabalho construímos um modelo de Moran cuja taxa de mutação varia no tempo segundo uma cadeia de Markov. Pudemos, então, atribuir um valor para a *taxa de mutação efetiva* com base nos parâmetros do modelo. O modelo de Moran tem um apelo mais realístico para aplicações práticas, pois sub-entende a noção de *sobreposição de gerações*, fenômeno que é observado no mundo natural. Porém o modelo de Wright-Fisher, apesar de não lidar com a sobreposição de gerações, também é muito utilizado em aplicações. No que concerne aos aspectos essenciais do presente trabalho, é factível se construir um modelo de Wright-Fisher com flutuação do parâmetro de mutação. Outras possíveis generalizações também podem ser elaboradas, como se considerar mais de dois estados na cadeia de Markov, ou se trabalhar com modelos que incluam outros fatores de alteração do conteúdo genético, como a seleção natural.

## Referências

- [1] L. Bromham, “Reading the Story in DNA: a beginner’s guide to molecular evolution”, Oxford University Press, Oxford, 2008.
- [2] R. Bürger, “The mathematical theory of selection, recombination, and mutation”, John Wiley & Sons, Chichester, 2000.
- [3] L. L. Cavalli-Sforza, “Genes, Povos e Línguas”, Companhia das Letras, São Paulo, 2003.
- [4] W. J. Ewens, “Mathematical Population Genetics: I. theoretical introduction”, 2nd, Springer-Verlag, Berlin, 2004.
- [5] M. Iizuka, The Effective Size of Fluctuating Populations, *Theoretical Population Biology*, 59 (2001) 281–286.
- [6] P. A. P. Moran, Random processes in genetics, *Proc. Camb. Phil. Soc.*, 54 (1958) 60–71.