

Modelos BioMatemáticos

<http://correio.fc.ul.pt/~mcg/aulas/biopop/>

Pedro J.N. Silva

Sala 4.1.16

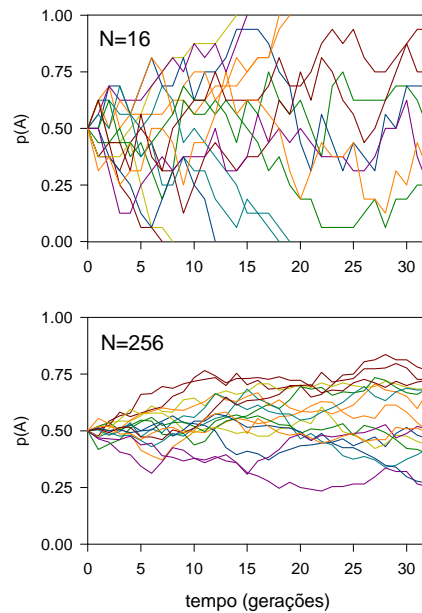
Departamento de Biologia Vegetal
Faculdade de Ciências da Universidade de Lisboa

Pedro.Silva@fc.ul.pt

Genética Populacional

Esquema

- Introdução à genética
- Lei de Hardy-Weinberg – gene autossómico
- Lei de Hardy-Weinberg – gene ligado ao sexo
- Efeitos evolutivos da mutação
- Efeitos evolutivos da finidade populacional



Consequências

1. As frequências genéticas das populações finitas variam aleatoriamente ao longo das gerações, processo a que chamamos deriva genética.
2. Em virtude desta variação, a frequência de um alelo pode chegar a 0 ou a 1; no primeiro caso esse alelo perde-se, no segundo fixa-se (e perdem-se todos os outros).
3. Populações inicialmente idênticas tendem a ficar cada vez mais diferentes.

Populações de um único indivíduo diplóide

Começamos pelo caso mais extremo e simples: uma população de um único indivíduo (de uma espécie diplóide de reprodutores sazonais com gerações separadas).

O genótipo do indivíduo pode ser aa , Aa ou AA , pelo que esta população pode ter 0, 1 ou 2 alelos A .

Podemos caracterizar o estado da população pelo número de alelos A : 0, 1, ou 2.

Populações de um único indivíduo diplóide

O indivíduo reproduz-se por autofecundação (que remédio!), e só sobrevive um descendente, que forma a geração seguinte (mantendo a grandeza populacional constante, e igual a 1).

Não podemos determinar com certeza o genótipo desse descendente, mas podemos estudar a probabilidade de ele ser AA , Aa ou aa .

Populações de um único indivíduo diplóide

Comecemos pela probabilidade de o descendente ser Aa, ignorando a possibilidade de mutações.

Se o progenitor for aa o descendente não pode ser Aa (a probabilidade é nula); se o progenitor for Aa, o descendente é também Aa com probabilidade $1/2$; se o progenitor for AA o descendente não pode ser Aa (a probabilidade é nula).

Populações de um único indivíduo diplóide

A probabilidade de o descendente ser Aa é então zero vezes a probabilidade de o progenitor ser aa, mais $1/2$ vezes a probabilidade de o progenitor ser Aa, mais zero vezes a probabilidade de o progenitor ser AA (por aplicação directa do teorema da probabilidade total).

Populações de um único indivíduo diplóide

Podemos dizer o mesmo com menos repetições (depois de alguma preparação).

Seja X_t o número de alelos A na população na geração t, e ρ_t o vector cuja i-ésima componente é a probabilidade de haver i alelos A na mesma geração:

$$\rho_t(i) = \text{Prob}[X_t = i]$$

Populações de um único indivíduo diplóide

Neste caso temos uma população de um indivíduo diplóide, e portanto $i=0, 1, 2$ (correspondendo ao indivíduo ser aa, Aa ou AA) e ρ_t tem portanto três elementos (a probabilidade de haver 0, 1 ou 2 alelos A na população, respectivamente):

$$\rho_t = \text{Prob}[X_t = 0, X_t = 1, X_t = 2]$$

Se, por exemplo, soubermos que o indivíduo é heterozigoto, temos

$$\rho_t = (0 \ 1 \ 0)$$

Populações de um único indivíduo diplóide

A probabilidade de o descendente ser Aa é então:

$$\rho_1(1) = 0\rho_0(0) + \frac{1}{2}\rho_0(1) + 0\rho_0(2)$$

A probabilidade de o descendente ser Aa é zero vezes a probabilidade de o progenitor ser aa, mais 1/2 vezes a probabilidade de o progenitor ser Aa, mais zero vezes a probabilidade de o progenitor ser AA.

Populações de um único indivíduo diplóide

Acabámos de ver qual a probabilidade de haver 1 alelo A na geração seguinte, *i.e.*, de o descendente ser heterozigoto (*1 alelo A* significa aqui *exactamente 1 alelo A*).

Mais geralmente, podemos perguntar qual a probabilidade de o descendente ser aa, ou de ser Aa, ou de ser AA – ou, o que é o mesmo, a probabilidade de haver 0, 1, ou 2 alelos A na população ao fim de uma geração.

Populações de um único indivíduo diplóide

Se na geração inicial o indivíduo for aa, a probabilidade de o descendente ser aa é 1, e as probabilidades de ser Aa e AA são 0; se o inicial for heterozigoto as probabilidades são 1/4 1/2 1/4, para os três genótipos aa, Aa e AA; e se o inicial for AA temos 0 0 1, respectivamente.

Podemos escrever tudo isto sob a forma de uma tabela de duas entradas, com o número de alelos A da população na geração t indicado à esquerda, e o número de alelos A na geração t+1 indicado em cima, e as probabilidades respectivas no corpo da tabela.

Populações de um único indivíduo diplóide

		Geração t+1		
		0	1	2
Geração t	0	1	0	0
	1	1/4	1/2	1/4
	2	0	0	1

Em cada célula temos a probabilidade de haver 0, 1, ou 2 alelos A numa geração, condicional a haver 0, 1, ou 2 alelos A na geração anterior. Por exemplo (como já vimos), se houver 1 alelo A na geração t, a probabilidade de haver 2 alelos A na geração t+1 é 1/4.

Populações de um único indivíduo diplóide

		Geração t+1		
		0	1	2
Geração t	0	1	0	0
	1	1/4	1/2	1/4
	2	0	0	1

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

Depois de percebermos a organização dos valores, não precisamos mais da tabela completa, e podemos escrever o mesmo de forma mais concisa, usando uma matriz de 3x3 elementos.

Populações de um único indivíduo diplóide

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

Todos os elementos da matriz são não-negativos, e a soma de cada linha é 1. A matrizes com estas propriedades chama-se matrizes estocásticas.

Populações de um único indivíduo diplóide

Assim, \mathbf{P} é uma matriz de probabilidades de transição, que nos indica a probabilidade de a população passar de um estado a outro, em duas gerações sucessivas.

Por exemplo, a probabilidade de ter inicialmente 1 alelo A e passar a ter 2 na geração seguinte é P_{12} e, de um modo geral, a probabilidade de passar do estado i ao estado j é P_{ij} (os índices indicam o número inicial e final de alelos A da população, por esta ordem):

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{pmatrix}$$

Populações de um único indivíduo diplóide

A probabilidade de haver 0 alelos A ao fim de uma geração é

$$\rho_1(0) = 1\rho_0(0) + \frac{1}{4}\rho_0(1) + 0\rho_0(2)$$

a probabilidade de haver 1 é (como já tínhamos visto)

$$\rho_1(1) = 0\rho_0(0) + \frac{1}{2}\rho_0(1) + 0\rho_0(2)$$

e a de haver 2 é

$$\rho_1(2) = 0\rho_0(0) + \frac{1}{4}\rho_0(1) + 1\rho_0(2)$$

Populações de um único indivíduo diplóide

$$\rho_1(0) = P_{00}\rho_0(0) + P_{10}\rho_0(1) + P_{20}\rho_0(2)$$

$$\rho_1(1) = P_{01}\rho_0(0) + P_{11}\rho_0(1) + P_{21}\rho_0(2)$$

$$\rho_1(2) = P_{02}\rho_0(0) + P_{12}\rho_0(1) + P_{22}\rho_0(2)$$

Estas três equações podem ser escritas como o produto do vector ρ_0 pela matriz \mathbf{P} :

$$\boldsymbol{\rho}_1 = \boldsymbol{\rho}_0 \mathbf{P}$$

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

Populações de um único indivíduo diplóide

$$\boldsymbol{\rho}_1 = \boldsymbol{\rho}_0 \mathbf{P}$$

Para estudarmos a evolução da população ao longo do tempo podemos usar esta equação como de costume, já que é válida para quaisquer duas gerações sucessivas.

Ao fim de duas gerações temos

$$\boldsymbol{\rho}_2 = \boldsymbol{\rho}_1 \mathbf{P} = \boldsymbol{\rho}_0 \mathbf{P}\mathbf{P} = \boldsymbol{\rho}_0 \mathbf{P}^2$$

Populações de um único indivíduo diplóide

$$\rho_2 = \rho_1 \mathbf{P} = \rho_0 \mathbf{P}\mathbf{P} = \rho_0 \mathbf{P}^2$$

Matematicamente, a matriz \mathbf{P}^2 representa o produto de \mathbf{P} por si própria.

Biologicamente, os seus elementos indicam-nos a probabilidade de a população ter j alelos A , condicional a ter i alelos A duas gerações atrás.

De um modo geral, ao fim de um número qualquer de gerações, temos

$$\rho_t = \rho_0 \mathbf{P}^t$$

Populações de um único indivíduo diplóide

$$\rho_t = \rho_0 \mathbf{P}^t$$

Esta equação representa portanto a evolução de uma população de um único indivíduo ao longo do tempo.

Dado que a população tem um número finito de indivíduos, as frequências genéticas futuras não estão determinadas (*i.e.*, não podemos saber qual o valor das frequências genotípicas e alélicas nas gerações seguintes), mas podemos saber a probabilidade de as frequências tomarem cada um dos seus valores possíveis, dada por ρ_t .

Populações de um único indivíduo diplóide

$$\rho_t = \rho_0 \mathbf{P}^t$$

A ρ_t chamamos a distribuição de frequências alélicas (na geração t).

t	Número de alelos A		
	0	1	2
0	0	1	0
1	0.25	0.5	0.25
2	0.375	0.25	0.375
3	0.4375	0.125	0.4375
4	0.46875	0.0625	0.46875
5	0.484375	0.03125	0.484375
6	0.4921875	0.015625	0.4921875
7	0.49609375	0.0078125	0.49609375
8	0.498046875	0.00390625	0.498046875
9	0.4990234375	0.001953125	0.4990234375
10	0.49951171875	0.0009765625	0.49951171875

Populações de um único indivíduo diplóide

$$\rho_t = \rho_0 \mathbf{P}^t$$

Esta equação refere-se a populações de um único indivíduo diplóide.

E se a população tiver mais de um indivíduo – temos de recomeçar do princípio? Não!

Esta representação da evolução populacional, sob forma matricial, é válida para qualquer grandeza populacional.

De facto, podemos usar exactamente as mesmas equações sem qualquer alteração, só precisamos de dar mais alguma atenção à matriz de transição \mathbf{P} .

Populações de qualquer grandeza

É altura de formalizarmos um pouco o que temos feito e, em particular explicitar os nossos pressupostos.

Como no estudo anterior de populações infinitas, fazemos primeiro bastantes simplificações (as quais poderiam ser levantadas mais tarde).

Para começar, mantemos todos os pressupostos do modelo de Hardy-Weinberg, à excepção do último.

Em particular, consideramos uma população panmítica conceptualmente isolada, e ignoramos os efeitos da mutação e da selecção natural.

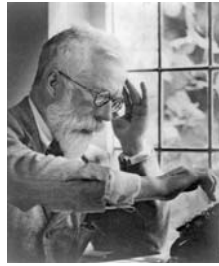
Populações de qualquer grandeza

Além disso, consideramos que a nossa população modelo é de uma espécie monóica, com reprodução sazonal e gerações separadas, de grandeza constante, constituída por N indivíduos diplóides.

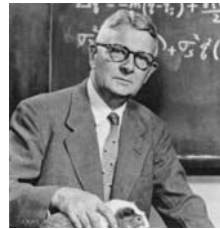
Assumimos ainda o facto de um indivíduo ter um descendente em nada afecta a probabilidade de ter outros.

Este modelo foi proposto e analisado por R.A. Fisher e S. Wright (e por muitos outros, depois).

Populações de qualquer grandeza



R.A. Fisher, 1952



Sewall Wright, 1954

Populações de qualquer grandeza

Consideremos então um gene autossômico com dois alelos, A e a.

O número total de alelos da população é $2N$.

O número de alelos A é X , e a sua frequência é $p = X/2N$; X pode variar entre 0 (todos os alelos são a) e $2N$ (são todos A).

O número de alelos a, e a respectiva frequência, podem ser obtidos a partir dos valores para A: o número de alelos a é $2N-X$, e a sua frequência é $q = (2N-X)/2N = 1-p$.

Populações de qualquer grandeza

Estamos agora prontos a modelar o processo de formação de uma nova geração, nesta população finita.

A forma mais fácil de o fazer é imaginar que todos os indivíduos da população produzem um número muito grande de gâmetas (independente do seu genótipo), dos quais $2N$ são amostrados aleatoriamente (dois a dois) para formar uma nova geração.

A matriz de transição

A probabilidade de a nova geração conter j ($0, 1, \dots, 2N$) alelos A, condicional a haver i ($0, 1, \dots, 2N$) alelos A na geração parental, é dada por uma matriz quadrada de ordem $2N+1$, \mathbf{P} , cujos elementos são

$$\begin{aligned} P_{ij} &= \text{Prob}[X_1 = j \mid X_0 = i] \\ &= \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \\ &= \frac{2N!}{j!(2N-j)!} p^j q^{2N-j} \end{aligned}$$

A matriz de transição

$$P_{ij} = \text{Prob}[X_1 = j \mid X_0 = i] = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} = \frac{2N!}{j!(2N-j)!} p^j q^{2N-j}$$

É fácil ver que este processo não tem memória: na passagem da geração 2 para a 3 só intervêm as frequências alélicas da geração 2, e não as da geração 1, ou de qualquer outra geração anterior.

Além disso, as probabilidades de transição não variam com o tempo: por exemplo, se uma população tiver três alelos A, a probabilidade de passar a ter 2 é independente da idade da população.

A matriz de transição

$$P_{ij} = \text{Prob}[X_1 = j \mid X_0 = i] = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} = \frac{2N!}{j!(2N-j)!} p^j q^{2N-j}$$

A frequência de um alelo (donde, também a do outro) varia assim como se flutuasse, ou andasse à deriva, sem qualquer tendência a "corrigir" desvios ocorridos em gerações anteriores, pelo que esses desvios tendem a acumular-se.

A matriz de transição

Como estamos a considerar que não há mutação e a população é isolada, os estados terminais (0 e $2N$ em número de alelos, 0 e 1 em frequências) são absorventes: se uma população entrar num destes estados, jamais poderá daí escapar, e as frequências alélicas não variam mais.

A perda de variabilidade é, portanto, irreversível.

As frequências intermédias, pelo contrário, correspondem a estados transientes: uma população pode ocupar um estado intermédio numa geração, na geração seguinte ocupar outro, e mais tarde voltar ao mesmo e voltar a sair, etc.

A matriz de transição

As frequências terminais podem ser atingidas a partir de qualquer frequência intermédia (numa só geração).

Assim, em cada geração há uma probabilidade finita de a população se tornar monomórfica ao nível do gene em estudo.

A matriz de transição

No limite, a população torna-se inevitavelmente monomórfica, quaisquer que sejam as frequências iniciais e a grandeza populacional (desde que finita).

Enquanto a deriva ocorre em todas as populações finitas, a fixação casual de genes só é irreversível em populações isoladas em que não haja mutação.

Em qualquer outra situação, a fixação de um gene é temporária, já que a migração ou a mutação podem repor a variabilidade genética da população.

Evolução da população

Tal como antes, representemos a distribuição de frequências alélicas como um vector de $2N+1$ elementos:

$$\rho_t(i) = \text{Prob}[X_t = i], \quad i = 0, 1, \dots, 2N$$

onde X_t é o número de alelos A na geração t.

A probabilidade de ter i alelos A na geração 1 pode ser obtida calculando o produto deste vector na geração 0 pela matriz de transição **P**:

$$\boldsymbol{\rho}_1 = \boldsymbol{\rho}_0 \mathbf{P}$$

Evolução da população

Não há nada de especial acerca das gerações 0 e 1 que nos impeça de continuar este processo do mesmo modo, para calcular a distribuição de frequências alélicas em qualquer outra geração. Assim, podemos escrever

$$\rho_2 = \rho_1 \mathbf{P}$$

o que nos permite acompanhar a distribuição de frequências alélicas ao longo do tempo, calculando cada geração a partir da anterior.

Evolução da população

Não há nada de especial acerca das gerações 0 e 1 que nos impeça de continuar este processo do mesmo modo, para calcular a distribuição de frequências alélicas em qualquer outra geração. Assim, podemos escrever

$$\rho_2 = \rho_1 \mathbf{P}$$

ou, de um modo geral,

$$\rho_{t+1} = \rho_t \mathbf{P}$$

e, ao fim de um número qualquer de gerações,

$$\rho_t = \rho_0 \mathbf{P}^t$$

Evolução da população

$$P_{ij} = \text{Prob}[X_1 = j \mid X_0 = i] = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} = \frac{2N!}{j!(2N-j)!} p^j q^{2N-j}$$

$$\boldsymbol{\rho}_{t+1} = \boldsymbol{\rho}_t \mathbf{P} \quad \boldsymbol{\rho}_t = \boldsymbol{\rho}_0 \mathbf{P}^t$$

Matematicamente, a matriz \mathbf{P}^t representa \mathbf{P} multiplicada por si própria t vezes; biologicamente, os seus elementos indicam-nos a probabilidade de a população ter j alelos A , condicional a ter i alelos A t gerações atrás.

Estas equações representam matematicamente o modelo de Fisher-Wright.

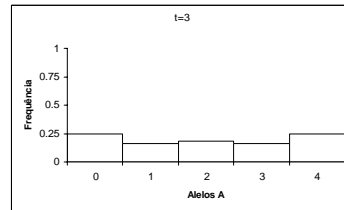
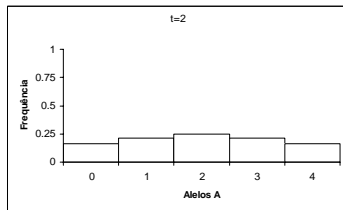
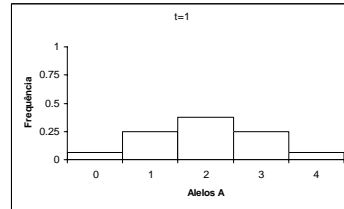
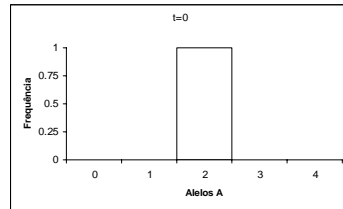
Evolução da população

$$P_{ij} = \text{Prob}[X_1 = j \mid X_0 = i] = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} = \frac{2N!}{j!(2N-j)!} p^j q^{2N-j}$$

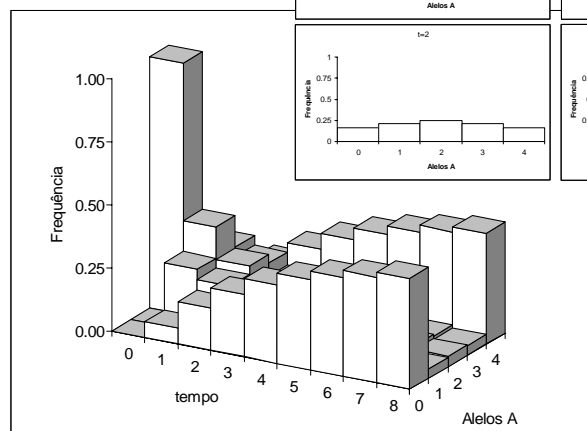
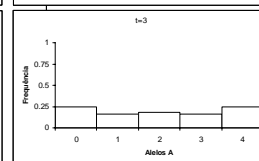
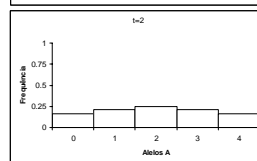
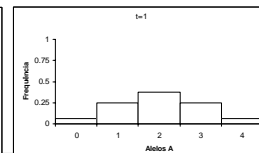
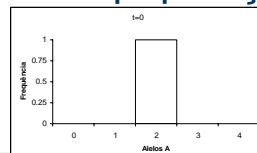
$$\boldsymbol{\rho}_{t+1} = \boldsymbol{\rho}_t \mathbf{P} \quad \boldsymbol{\rho}_t = \boldsymbol{\rho}_0 \mathbf{P}^t$$

Podemos utilizar estas equações para obter a distribuição de frequências alélicas (*i.e.*, a probabilidade de a população ter 0, 1, ..., $2N$ alelos A ; ou a proporção esperada de populações com 0, 1, ..., $2N$ alelos A) ao longo do tempo, acompanhando assim a evolução esperada da população.

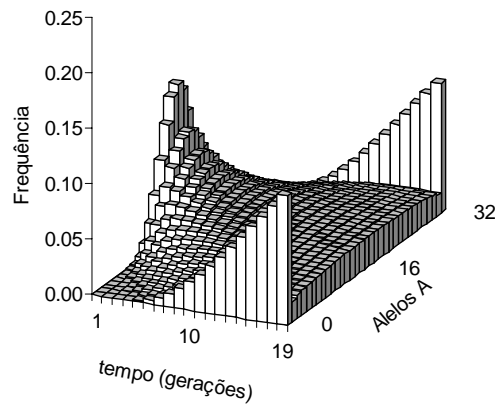
Evolução da população



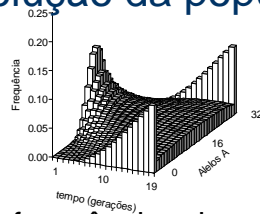
Evolução da população



Evolução da população



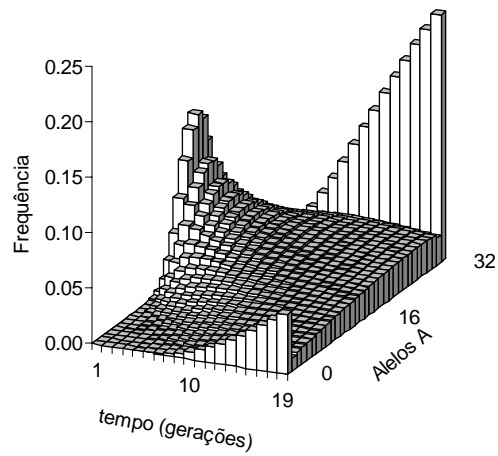
Evolução da população



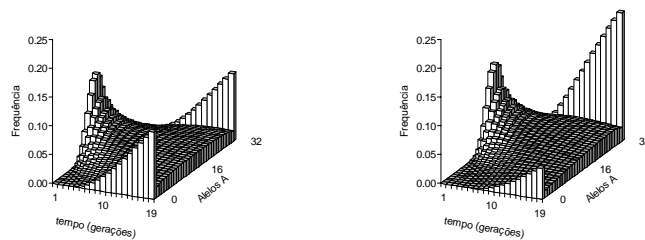
Neste caso, as frequências dos estados intermédios tendem a uniformizar-se e a tender para zero, e as frequências dos estados terminais, iguais ao longo do tempo, tendem ambas para $1/2$ (as frequências alélicas iniciais).

No caso de as frequências alélicas iniciais serem diferentes de $1/2$, a população tende ainda a ficar monomórfica, mas as probabilidades de fixação de cada um dos alelos já não são iguais.

Evolução da população



Evolução da população



Em qualquer caso

- as frequências dos estados intermédios tendem a uniformizar-se e a reduzir-se até zero
- a soma das frequências dos estados terminais tende para 1

Evolução da população

O produto $\rho_0 \mathbf{P}^i$ pode ser escrito usando os valores e vectores próprios da matriz \mathbf{P}' (a transposta de \mathbf{P}).

Neste modelo, os valores próprios de \mathbf{P}' são dados por

$$\lambda_i = \binom{2N}{i} \frac{i!}{(2N)^i}$$

Portanto, $\lambda_0=1$, $\lambda_1=1$, $\lambda_2=1-1/2N$, e $\lambda_2 > \lambda_3 > \dots > \lambda_{2N} > 0$.

Evolução da população

A estes valores próprios correspondem $2N+1$ vectores próprios linearmente independentes, $v_0, v_1, v_2, \dots, v_{2N}$.

$$v_0' = [1 \ 0 \ \dots \ 0 \ 0], \quad v_1' = [0 \ 0 \ \dots \ 0 \ 1]$$

Logo, qualquer que seja ρ_0 , existem (e são únicos!) $(c_0, c_1, \dots, c_{2N})$ tais que

$$\rho_0 = c_0 v_0' + c_1 v_1' + c_2 v_2' + \dots + c_{2N} v_{2N}'$$

Evolução da população

Assim, podemos escrever o produto $\rho_0 P^t$ como

$$\begin{aligned}\rho_0 P^t &= c_0 \lambda_0^t v_0' + c_1 \lambda_1^t v_1' + c_2 \lambda_2^t v_2' + \dots + c_{2N} \lambda_{2N}^t v_{2N}' \\ &= c_0 v_0' + c_1 v_1' + c_2 \lambda_2^t v_2' + \dots + c_{2N} \lambda_{2N}^t v_{2N}'\end{aligned}$$

Lembrando que $0 < \lambda_i < 1$ $i=2, \dots, 2N$,

$$\lim_{t \rightarrow \infty} \rho_0 P^t = c_0 v_0' + c_1 v_1' = [c_0 \quad 0 \quad \dots \quad 0 \quad c_1]'$$

Em qualquer caso

- as frequências dos estados intermédios tendem a uniformizar-se e a reduzir-se até zero
- a soma das frequências dos estados terminais tende para 1

Evolução da população

Além disso, como $1 > \lambda_2 > \dots > \lambda_{2N} > 0$, para t suficientemente grande, o produto $\rho_0 P^t$ pode ser bem aproximado por

$$\rho_0 P^t \cong c_0 v_0' + c_1 v_1' + c_2 \left(1 - \frac{1}{2N}\right)^t v_2'$$

O maior valor próprio não unitário, $\lambda_2 = 1 - 1/2N$, tem assim um papel essencial na evolução de populações finitas, já que determina a distribuição de frequências alélicas para t grande (finito).