



UNIVERSITÀ DEGLI STUDI DI PADOVA

---

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

# Modelli Stocastici di Genetica delle Popolazioni

**Relatore:**

Dott. Samir Suweis

**Correlatore:**

Dott.ssa Anna Tovo

**Laureando:** Carlo Polato

**Matricola:** 1123510

---

Anno Accademico 2017/2018



*Ai miei genitori.*



## Sommario

I sistemi ecologici sono sistemi complessi: i diversi elementi che li costituiscono interagiscono tra loro su scale (di lunghezza e temporali) diverse, dando origine a fenomeni collettivi non trivialmente spiegabili come sovrapposizione lineare della dinamica delle singole specie.

Questi sistemi mostrano sorprendenti comportamenti universali (ossia indipendenti dai dettagli del sistema osservato). Questi pattern possono essere spiegati utilizzando un approccio basato su processi stocastici ed idee della meccanica statistica, che vanno sotto il nome di teoria neutrale.

Analoghe proprietà emergenti sono anche osservabili nell'organizzazione dei genomi. Negli ultimi anni sta crescendo la disponibilità di dati di metagenomica, in cui interi ecosistemi vengono sequenziati. Nasce quindi la possibilità di legare i pattern della microscala (genomica) con quelli alla macroscala (ecologia). In questa tesi si affronterà lo studio della dinamica di popolazione di geni, in particolare si analizzeranno i modelli di Wright-Fisher e di Moran e la loro relazione con l'ecologia.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Modello di Wright-Fisher</b>	<b>3</b>
2.1	Modello di Wright-Fisher con Mutazione . . . . .	5
2.2	Simulazioni Wright-Fisher . . . . .	6
<b>3</b>	<b>Modello di Moran</b>	<b>9</b>
3.1	Scomposizione in autovettori destri e sinistri . . . . .	9
3.2	Sviluppo in serie degli autovettori . . . . .	10
3.3	Modello di Moran con Mutazione . . . . .	13
3.4	Simulazioni Moran . . . . .	14
<b>4</b>	<b>Voter Model e Teoria Neutrale</b>	<b>17</b>
4.1	Voter Model . . . . .	17
4.2	<i>Soluzione Stazionaria</i> . . . . .	19
<b>5</b>	<b>Conclusioni</b>	<b>21</b>
<b>A</b>	<b>Fissazione con Mutazione</b>	<b>23</b>
<b>B</b>	<b>Teoremi di Cannings</b>	<b>25</b>
B.1	Applicazione al Modello di Wright-Fisher senza mutazione . . . . .	26
B.2	Applicazione al Modello di Wright-Fisher con mutazione . . . . .	27
B.3	Applicazione al Modello di Moran senza mutazione . . . . .	28
B.4	Applicazione al Modello di Moran con mutazione . . . . .	28
<b>C</b>	<b>Verifica degli autovettori</b>	<b>31</b>





# Capitolo 1

## Introduzione

L'evoluzione di sistemi complessi è spesso descritta attraverso modelli stocastici, i quali non sono impiegati unicamente nell'ambito della fisica, ma hanno giocato un ruolo fondamentale anche nella comprensione della dinamica delle popolazioni [1, 2]. La disciplina più studiata è la genetica delle popolazioni, ovvero lo studio della frequenza dei geni all'interno di una popolazione di individui che si riproducono. Il primo modello di questo tipo è stato concepito da Wright [3] e Fisher [4] indipendentemente, mentre una sua interessante variante è stata proposta in seguito da Moran [5]. Entrambi si basano sul presupposto ideale di considerare una popolazione di dimensione fissata, per la quale la riproduzione avviene casualmente senza che nessun allele sia privilegiato. La controparte in ecologia è data dalla cosiddetta teoria neutrale [6], che attraverso processi di nascita e morte (e migrazione) descrive la dinamica delle popolazioni delle specie in ecosistemi biodiversi quali le barriere coralline [7] o le foreste tropicali [8]. In questo lavoro abbiamo ricavato in modo formale le proprietà di popolazioni di geni alla stazionarietà, relative ai modelli di Wright-Fisher e Moran. Abbiamo poi valutato le differenze e le somiglianze tra questi due modelli, verificando anche i risultati ottenuti analiticamente mediante simulazioni numeriche.

Infine, abbiamo introdotto un sistema stocastico utilizzato per descrivere la biodiversità in ecosistemi, noto come "voter model", e abbiamo mostrato il parallelismo tra i modelli usati in genomica ed in ecologia, esplicitando anche le relazioni tra le terminologie usate nei due ambiti.



## Capitolo 2

# Modello di Wright-Fisher

Supponiamo di avere una popolazione di  $N$  individui aploidi (ovvero le cui cellule presentano un corredo cromosomico costituito da geni che si presentano in unica copia) che alla generazione  $t$  si riproducono, ciascuno dando vita ad un numero causale di figli che formano la generazione  $t + 1$ . Dato che gli individui sono aploidi vi è una corrispondenza 1 a 1 tra individui e geni. Concentriamoci su un particolare gene, che si può trovare in due soli stati (alleli) che chiameremo  $A$  e  $B$ . Il modello di Wright-Fisher modella la dinamica di popolazione attraverso i seguenti passaggi:

- si sceglie casualmente un individuo dalla popolazione iniziale;
- si fa una copia dell'individuo;
- se la variante del modello lo prevede, gli si dà la possibilità di mutare;
- si inserisce la copia nella nuova popolazione;
- si ripete il procedimento tante volte quanti sono gli individui della popolazione.

Chiaramente dati questi presupposti le generazioni non si sovrappongono, in quanto ad ogni generazione muoiono  $N$  geni e ne nascono altri  $N$ . Notiamo che il modello di Wright-Fisher si può mappare in un modello probabilistico per l'estrazione con reinserimento di biglie colorate da un'urna. Se nella generazione  $t$  ci sono  $n'$  alleli  $A$  e  $N - n'$  alleli  $B$ , la probabilità che la generazione  $t + 1$  contenga  $n$  alleli  $A$  è data dalla distribuzione binomiale

$$p_{nn'} = \binom{N}{n} p_{1,n'}^n p_{2,n'}^{N-n} \quad (2.1)$$

dove

$$p_{1,n'} = \frac{n'}{N} \quad \text{e} \quad p_{2,n'} = \left(1 - \frac{n'}{N}\right) \quad (2.2)$$

rappresentano la probabilità di scegliere un  $A$  e la probabilità di scegliere un  $B$  rispettivamente. Quindi la probabilità di transizione da uno stato iniziale con  $n'$  alleli  $A$  al nuovo stato con  $n$  alleli  $A$  è data da

$$p_{nn'} = \binom{N}{n} \left(\frac{n'}{N}\right)^n \left(1 - \frac{n'}{N}\right)^{N-n}. \quad (2.3)$$

Ora siamo interessati a calcolare la probabilità di trovare  $n$  alleli  $A$  nella generazione  $t$  dati  $n_0$  alleli  $A$  come condizione iniziale, che denoteremo con  $P(n, t)$ . Dato che la costruzione della generazione  $t + 1$  dipende unicamente dalla generazione  $t$  il processo è Markoviano. Possiamo quindi definire un vettore probabilità

$$\vec{P}(t) = \begin{pmatrix} P(0, t) \\ P(1, t) \\ P(2, t) \\ \vdots \end{pmatrix} \quad (2.4)$$

e una matrice di transizione

$$\mathcal{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (2.5)$$

in termini dei quali l'evoluzione del sistema è:  $P(n, t + 1) = \sum_{n'} p_{nn'} P(n', t)$  o, in forma vettoriale:

$$\vec{P}(t + 1) = \begin{pmatrix} P(0, t + 1) \\ P(1, t + 1) \\ P(2, t + 1) \\ \vdots \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} P(0, t) \\ P(1, t) \\ P(2, t) \\ \vdots \end{pmatrix} = \mathcal{P} \vec{P}(t). \quad (2.6)$$

Dato che la somma delle componenti del vettore probabilità  $\vec{P}(t)$  deve risultare pari ad 1, così come la somma delle componenti delle colonne della matrice di transizione  $\mathcal{P}$ , abbiamo le due condizioni

$$\sum_n P(n, t) = 1 \quad \forall t, \quad \sum_n p_{nn'} = 1 \quad \forall n'. \quad (2.7)$$

Le matrici che soddisfano questa condizione sono dette *matrici stocastiche*. Adesso possiamo esprimere facilmente lo stato della generazione  $t$  in termini della condizione iniziale:

$$\vec{P}(t) = \mathcal{P} \vec{P}(t - 1) = \mathcal{P} \mathcal{P} \vec{P}(t - 2) = \cdots = \mathcal{P}^t \vec{P}(0). \quad (2.8)$$

Per grandi  $t$  il comportamento è dominato dai più grandi (in valore assoluto) tra gli autovalori di  $\mathcal{P}$  e dai corrispondenti autovettori. Gli autovalori del modello di Wright-Fisher sono stati calcolati per la prima volta da Feller in [9], e sono<sup>1</sup>:

$$\lambda_j = \binom{N}{j} \frac{j!}{N^j} = \frac{N(N-1)\cdots(N-j+1)}{N^j}. \quad (2.9)$$

Nel limite di  $t \rightarrow \infty$  il sistema può tendere ad uno stato stazionario, che chiameremo  $\vec{P}^*$ . Se esiste si ha che  $\mathcal{P} \vec{P}^* = \vec{P}^*$ , ovvero  $\vec{P}^*$  è autovettore destro di  $\mathcal{P}$  con autovalore

<sup>1</sup>Nell'appendice (B.1) è mostrata una derivazione alternativa di questo risultato

1. In altri termini lo stato finale è composto da alleli tutti di tipo  $A$  o tutti di tipo  $B$ . Questo corrisponde allo stato stazionario:

$$\vec{P}^* = \begin{pmatrix} 1 - \Pi \\ 0 \\ \vdots \\ 0 \\ \Pi \end{pmatrix}, \quad (2.10)$$

dove  $\Pi$  è la probabilità che l'allele  $A$  diventi quello dominante alla stazionarietà (ovvero che l'allele  $A$  si "fissi" nella popolazione). Un'alta facile deduzione per questo modello è che il numero atteso di alleli  $A$ ,  $n(t)$ , non cambia da una generazione a un'altra. Infatti:

$$\langle n(t+1) \rangle = \sum_n n P(n, t+1) = \sum_n n \sum_{n'} p_{nn'} P(n', t) = \sum_{n'} n' P(n', t) = \langle n(t) \rangle, \quad (2.11)$$

dove nel penultimo passaggio si è sfruttato il fatto che:  $\sum_n n p_{nn'} = N p_1 = N \frac{n'}{N} = n'$ , risultato noto della distribuzione binomiale. In particolare  $\langle n(t) \rangle = \langle n(0) \rangle = n_0$  e questo risultato può essere utilizzato per determinare  $\Pi$ . Infatti per  $t \rightarrow \infty$  solo due stati sono possibili:  $n = N$  (con probabilità  $\Pi$ ) e  $n = 0$  (con probabilità  $1 - \Pi$ ), perciò si ha:

$$n_0 = \lim_{t \rightarrow \infty} \langle n(t) \rangle = N\Pi + 0(1 - \Pi) \Rightarrow n_0 = N\Pi \Rightarrow \Pi = \frac{n_0}{N}. \quad (2.12)$$

## 2.1 Modello di Wright-Fisher con Mutazione

Nella sezione precedente abbiamo considerato il modello senza mutazione. Consideriamo ora il caso in cui un allele  $A$  può mutare con probabilità  $u$  in un allele  $B$  e un allele  $B$  può mutare con probabilità  $v$  in un allele  $A$ . Come già spiegato in precedenza, il processo di mutazione viene considerato dopo quello di riproduzione, e quindi le probabilità  $p_1$  e  $p_2$  vengono modificate nel seguente modo:

$$p_{1,n'} = (1 - u) \frac{n'}{N} + v \left(1 - \frac{n'}{N}\right) \quad p_{2,n'} = u \frac{n'}{N} + (1 - v) \left(1 - \frac{n'}{N}\right) \quad (2.13)$$

mentre la probabilità di transizione diventa

$$p_{nn'} = \binom{N}{n} \left[ (1 - u) \frac{n'}{N} + v \left(1 - \frac{n'}{N}\right) \right]^n \left[ u \frac{n'}{N} + (1 - v) \left(1 - \frac{n'}{N}\right) \right]^{N-n}. \quad (2.14)$$

L'aggiunta della mutazione implica però che il numero atteso di alleli  $A$  non è più conservato. Infatti:

$$\langle n(t+1) \rangle = \sum_n n P(n, t+1) = \sum_n n \sum_{n'} p_{nn'} P(n', t) = \sum_n n p_{nn'} \sum_{n'} P(n', t),$$

ma dato che  $\sum_n n p_{nn'} = N p_1 = (1 - u) n' + v (N - n')$  si ha che

$$\begin{aligned} \langle n(t+1) \rangle &= \sum_{n'} [(1 - u) n' + v (N - n')] P(n', t) = \\ &= (1 - u) \langle n(t) \rangle + v (N - \langle n(t) \rangle). \end{aligned} \quad (2.15)$$

Come mostrato nell'appendice (A) si può ricavare che per grandi tempi

$$\frac{\langle n(t+1) \rangle}{N} \xrightarrow{t \rightarrow \infty} \frac{v}{u+v}. \quad (2.16)$$

Perciò nè l'allele A nè l'allele B possono diventare fissati, essendoci sempre la possibilità che uno dei due venga creato attraverso mutazione anche qualora non ve ne fossero presenti nella popolazione. In altri termini la mutazione permette la coesistenza nella popolazione di geni con entrambi gli alleli.

Anche in questa variante del modello si possono ricavare analiticamente gli autovalori<sup>2</sup>:

$$\lambda_j = (1-u-v)^j \binom{N}{j} \frac{j!}{N^j} = (1-u-v)^j \frac{N(N-1)\cdots(N-j+1)}{N^j}. \quad (2.17)$$

## 2.2 Simulazioni Wright-Fisher

In figura (2.1) riportiamo i risultati ottenuti simulando numericamente il modello di Wright-Fisher senza mutazione. Possiamo vedere che, dopo un transiente iniziale, il sistema raggiunge lo stato stazionario in cui tutti gli alleli sono di un solo tipo. Al contrario in figura (2.2) si può vedere come, per il caso con mutazione, le condizioni iniziali dopo pochi passi temporali diventino influenti, in quanto i valori dei due alleli si stabilizzano attorno alle linee verdi rappresentanti i rapporti allelici di stazionarietà predetti dalla teoria.

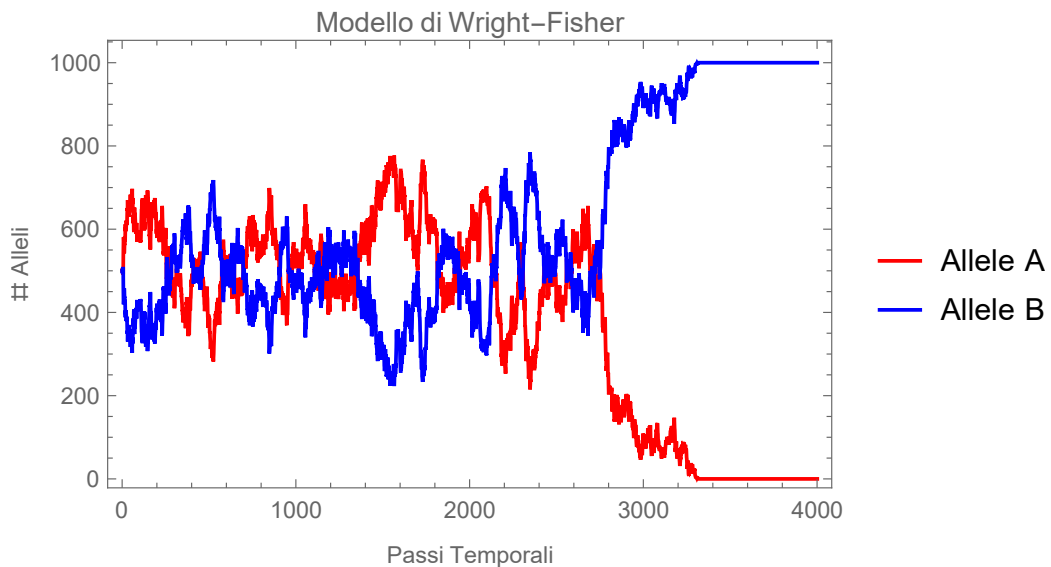


Figura 2.1: Simulazione della fissazione per il modello di Wright-Fisher senza mutazione, con percentuale di alleli A di partenza  $p = 0.5$ .

<sup>2</sup>Rimandiamo la derivazione di questo risultato in appendice (B.2)

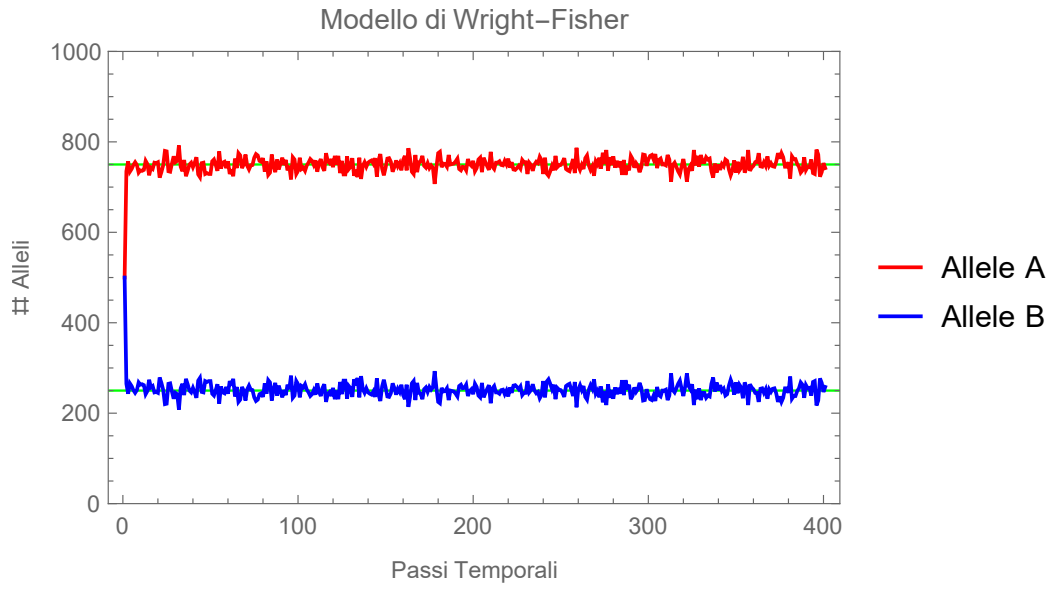


Figura 2.2: Simulazione della fissazione per il modello di Wright-Fisher con mutazione, con percentuale di alleli  $A$  di partenza  $p = 0.5$  e probabilità di mutazione  $u = 0.25$  e  $v = 0.75$ .

---





## Capitolo 3

# Modello di Moran

Una variante del modello di Wright-Fisher è quella proposta da Moran, dove ogni passo consiste nel far morire un gene e sostituirlo con un allele  $A$  con probabilità  $p_1$  o con un allele  $B$  con probabilità  $p_2$ . Queste probabilità sono calcolate considerando il sistema prima della morte del gene, perciò come per il Wright-Fisher sono  $p_{1,n} = \frac{n}{N}$  e  $p_{2,n} = (1 - \frac{n}{N})$ . Il modello di Moran a differenza di quello di Wright-Fisher ha generazioni che si sovrappongono. Infatti dopo  $N$  passi non è detto che tutti i geni siano stati rimpiazzati una volta, anzi è molto probabile che qualcuno non sia stato rimpiazzato affatto e che altri siano sopravvissuti per poco. A causa dei problemi di definizione di "generazione" solitamente si usa questo termine per indicare un singolo passo. Questa definizione, pur essendo controintuitiva, è ben posta. Quindi in una generazione il numero di alleli  $A$  può aumentare di 1 (se muore un  $B$  e nasce un  $A$ ), rimanere costante (se muore un  $A$  e nasce un  $A$  oppure se muore un  $B$  e nasce un  $B$ ) o diminuire di 1 (se muore un  $A$  e nasce un  $B$ ). Perciò gli unici elementi non nulli della matrice di transizione sono:

$$p_{(n+1)n} = \left(1 - \frac{n}{N}\right) p_{1,n} = \left(1 - \frac{n}{N}\right) \left(\frac{n}{N}\right), \quad (3.1)$$

$$p_{nn} = \left(\frac{n}{N}\right) p_{1,n} + \left(1 - \frac{n}{N}\right) p_{2,n} = \left(\frac{n}{N}\right)^2 + \left(1 - \frac{n}{N}\right)^2, \quad (3.2)$$

$$p_{(n-1)n} = \left(\frac{n}{N}\right) p_{2,n} = \left(\frac{n}{N}\right) \left(1 - \frac{n}{N}\right). \quad (3.3)$$

### 3.1 Scomposizione in autovettori destri e sinistri

Prima di procedere con l'analisi del modello di Moran è necessario introdurre un teorema tratto dal libro [10], fondamentale per l'esposizione successiva.

**Teorema 3.1** (Scomposizione in autovettori destri e sinistri). *Se  $\mathcal{P}$  è una matrice  $M \times M$  e i suoi  $M$  autovalori  $\lambda_1, \dots, \lambda_M$  sono distinti, allora detti  $\vec{v}^{(1)}, \dots, \vec{v}^{(M)}$  gli autovettori destri e  $U$  la matrice le cui colonne sono i suddetti vettori, si ha che:*

- *l'inversa di  $U$  ha per righe gli autovettori sinistri di  $\mathcal{P}$ ,  $\vec{\pi}^{(1)}, \dots, \vec{\pi}^{(M)}$ , normalizzati in modo da soddisfare  $\vec{\pi}^{(i)} \vec{v}^{(i)} = 1, \quad \forall i,$*
- $\mathcal{P}^t = \sum_{i=1}^M \lambda_1^t \vec{v}^{(i)} \vec{\pi}^{(i)}.$

**Dimostrazione**

Partiamo col dimostrare il primo punto. Dalla definizione di autovettore sinistro  $\vec{\pi}^{(j)}\mathcal{P} = \lambda_j\vec{\pi}^{(j)}$  e di autovettore destro  $\mathcal{P}\vec{v}^{(i)} = \lambda_i\vec{v}^{(i)}$ , otteniamo  $\vec{\pi}^{(j)}\mathcal{P}\vec{v}^{(i)} = \lambda_j\vec{\pi}^{(j)}\vec{v}^{(i)}$  ma anche  $\vec{\pi}^{(j)}\mathcal{P}\vec{v}^{(i)} = \lambda_i\vec{\pi}^{(j)}\vec{v}^{(i)}$ , quindi:  $(\lambda_i - \lambda_j)\vec{\pi}^{(j)}\vec{v}^{(i)} = 0$  che significa che  $\forall i \neq j$  si ha  $\vec{\pi}^{(j)}\vec{v}^{(i)} = 0$ .

Ora consideriamo il prodotto tra la matrice con gli autovettori sinistri come righe e quella con gli autovettori destri come colonne:

$$\begin{aligned} \begin{pmatrix} \vec{\pi}^{(1)} \\ \vec{\pi}^{(2)} \\ \vdots \\ \vec{\pi}^{(M)} \end{pmatrix} (\vec{v}^{(1)}, \vec{v}^{(2)}, \dots, \vec{v}^{(M)}) &= \begin{pmatrix} \vec{\pi}^{(1)}\vec{v}^{(1)} & \vec{\pi}^{(1)}\vec{v}^{(2)} & \dots & \vec{\pi}^{(1)}\vec{v}^{(M)} \\ \vec{\pi}^{(2)}\vec{v}^{(1)} & \vec{\pi}^{(2)}\vec{v}^{(2)} & \dots & \vec{\pi}^{(2)}\vec{v}^{(M)} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{\pi}^{(M)}\vec{v}^{(1)} & \vec{\pi}^{(M)}\vec{v}^{(2)} & \dots & \vec{\pi}^{(M)}\vec{v}^{(M)} \end{pmatrix} = \\ &= \begin{pmatrix} \vec{\pi}^{(1)}\vec{v}^{(1)} & 0 & \dots & 0 \\ 0 & \vec{\pi}^{(2)}\vec{v}^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \vec{\pi}^{(M)}\vec{v}^{(M)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad (3.4) \end{aligned}$$

dove si è usato il fatto che  $\vec{\pi}^{(j)}\vec{v}^{(i)} = 0 \forall i \neq j$  e nell'ultimo passaggio il fatto che  $\vec{\pi}^{(i)}\vec{v}^{(i)} = 1 \forall i$ .

Dato che la matrice inversa è unica e  $U = (\vec{v}^{(1)}, \dots, \vec{v}^{(M)})$  allora

$$U^{-1} = \begin{pmatrix} \vec{\pi}^{(1)} \\ \vec{\pi}^{(2)} \\ \vdots \\ \vec{\pi}^{(M)} \end{pmatrix}. \quad (3.5)$$

Passiamo adesso alla dimostrazione del secondo punto. Dato che  $\mathcal{P}$  è diagonalizzabile possiamo scrivere  $\mathcal{P} = UDU^{-1}$  da cui  $\mathcal{P}^t = UDU^{-1}UDU^{-1} \dots UDU^{-1} = UD^tU^{-1}$ , dove nell'ultimo passaggio si è sfruttato il fatto che  $U^{-1}U = \mathbb{1}$ . Perciò  $\mathcal{P}^t = UD^tU^{-1} = \sum_{i=1}^M UD^{t(i)}U^{-1}$ , dove  $D^{t(i)}$  è la matrice diagonale con unico elemento non nullo  $D_{ii}^{t(i)} = \lambda_i^t$ .  $\square$

Il teorema (3.1) può essere generalizzato nel teorema seguente, la cui dimostrazione è del tutto simile alla precedente.

**Teorema 3.2** (Generalizzazione del teorema (3.1)). *Se  $\mathcal{P}$  è una matrice  $M \times M$  e i suoi  $M$  autovalori  $\lambda_1, \dots, \lambda_M$  sono tali per cui per ognuno di loro molteplicità algebrica e geometrica coincidono, allora considerando una base di autovettori destri e sinistri ortonormale la tesi del teorema (3.1) è ancora valida.*

## 3.2 Sviluppo in serie degli autovettori

Gli autovalori e gli autovettori del modello di Moran senza mutazione sono stati calcolati per la prima volta da Hannan nell'appendice di [5], in seguito da Watterson in [11] usando i polinomi di Chebyshev e successivamente da Gani in

[12] con un'altra tecnica e sono<sup>1</sup>:  $\lambda_j = 1 - \frac{j(j-1)}{N^2}$ , da cui si ricava  $\lambda_0 = \lambda_1 = 1$  e  $\lambda_2 = 1 - \frac{2}{N^2}$ . Gli autovettori destri e sinistri, invece, associati a tali autovalori sono, rispettivamente<sup>2</sup>:

$$\begin{aligned} \vec{v}^{(0)} &= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \vec{v}^{(1)} &= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} & \vec{v}^{(2)} &= \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{2} \end{pmatrix} \\ \vec{\theta}^{(0)} &= (1, \dots, 1) & \vec{\theta}^{(1)} &= (0, 1, \dots, N) \\ \vec{\theta}^{(2)} &= (0, N-1, 2(N-2), \dots, n(N-n), \dots, 0) \end{aligned}$$

Ora per soddisfare le ipotesi del teorema (3.2) bisogna effettuare un cambio di base ad una ortonormale  $\vec{\pi}^{(i)}, \vec{v}^{(j)}$ , in modo che  $\vec{\pi}^{(i)}\vec{v}^{(j)} = \delta_{ij} \quad \forall i, j$ . Cerchiamo quindi dei vettori  $\vec{\pi}^{(i)}$  che siano combinazione lineare dei  $\vec{\theta}^{(i)}$ . Per esempio da  $\vec{\pi}^{(1)}\vec{v}^{(1)} = 1$  otteniamo  $\vec{\pi}^{(1)} = (0, \frac{1}{N}, \dots, 1)$ . Per poi soddisfare  $\vec{\pi}^{(0)}\vec{v}^{(1)} = 0$  dobbiamo trovare la giusta combinazione lineare tra i due.

$$\left[ \alpha(1, \dots, 1) + \beta \left( 0, \frac{1}{N}, \dots, 1 \right) \right] \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \alpha + \beta = 0$$

Scegliendo  $\alpha = 1$  e  $\beta = -1$  otteniamo  $\vec{\pi}^{(0)} = (1, \dots, \frac{1}{N}, 0)$ , da cui  $\vec{\pi}^{(0)}\vec{v}^{(0)} = 1$  e  $\vec{\pi}^{(1)}\vec{v}^{(0)} = 0$ . Infine per garantire  $\vec{\pi}^{(2)}\vec{v}^{(2)} = 1$ , dato che  $\sum_{n=0}^N n(N-n) = \frac{N(N^2-1)}{6}$  otteniamo  $\vec{\pi}^{(2)} = \frac{6}{N(N^2-1)}\vec{\theta}^{(2)}$ . Riassumendo:

$$\begin{aligned} \vec{v}^{(0)} &= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \vec{v}^{(1)} &= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} & \vec{v}^{(2)} &= \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{2} \end{pmatrix} \\ \vec{\pi}^{(0)} &= \left( 1, \dots, \frac{1}{N}, 0 \right) & \vec{\pi}^{(1)} &= \left( 0, \frac{1}{N}, \dots, 1 \right) \\ \vec{\pi}^{(2)} &= \frac{6}{N(N^2-1)} (0, N-1, 2(N-2), \dots, n(N-n), \dots, 0) \end{aligned}$$

Dato che il processo del modello di Moran è di tipo Markoviano se ne ricava che  $\vec{P}(t) = \mathcal{P}\vec{P}(t-1) = \mathcal{P}\mathcal{P}\vec{P}(t-2) = \dots = \mathcal{P}^t\vec{P}(0)$ , dove  $\vec{P}(t)$  è il vettore

<sup>1</sup>Nell'appendice (B.3) è mostrata una derivazione alternativa di questo risultato.

<sup>2</sup>Nell'appendice (C) si verifica senza derivare questo risultato.

probabilità al tempo  $t$  e il vettore probabilità al tempo  $t = 0$  è dato da:

$$\vec{P}(0) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

Dove l'unico valore non nullo è nella  $n_0 + 1$ -esima posizione. Perciò sfruttando il teorema (3.2) ricaviamo  $\vec{P}(t) = \mathcal{P}^t \vec{P}(0) = \sum_{i=1}^M \lambda_1^t \vec{v}^{(i)} \vec{\pi}^{(i)} \vec{P}(0)$ . Calcolando i prodotti vettore riga-vettore colonna  $\vec{\pi}^{(i)} \vec{P}(0)$ :

$$\vec{\pi}^{(0)} \vec{P}(0) = \left(1, \dots, \frac{1}{N}, 0\right) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \frac{N - n_0}{N} = 1 - \frac{n_0}{N},$$

$$\vec{\pi}^{(1)} \vec{P}(0) = \left(0, \frac{1}{N}, \dots, 1\right) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \frac{n_0}{N},$$

$$\begin{aligned} \vec{\pi}^{(2)} \vec{P}(0) &= \frac{6}{N(N^2 - 1)} (0, N - 1, 2(N - 2), \dots, n(N - n), \dots, 0) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \\ &= \frac{6n_0(N - n_0)}{N(N^2 - 1)}. \end{aligned}$$

Ora possiamo scrivere lo sviluppo in serie del vettore probabilità in funzione degli

autovettori:

$$\begin{aligned} \vec{P}(t) &\approx \left(1 - \frac{n_0}{N}\right) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} (1)^t + \frac{n_0}{N} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} (1)^t + \frac{6n_0(N-n_0)}{N(N^2-1)} \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{2} \end{pmatrix} \left(1 - \frac{2}{N^2}\right)^t \\ &= \begin{pmatrix} 1 - \frac{n_0}{N} \\ 0 \\ \vdots \\ 0 \\ \frac{n_0}{N} \end{pmatrix} + \frac{6n_0(N-n_0)}{N(N^2-1)} \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{2} \end{pmatrix} \left(1 - \frac{2}{N^2}\right)^t. \end{aligned} \quad (3.6)$$

### 3.3 Modello di Moran con Mutazione

Quando prendiamo in considerazione anche la mutazione, le probabilità  $p_1$  e  $p_2$  vengono modificate nel seguente modo:

$$p_{1,n} = (1-u) \frac{n}{N} + v \left(1 - \frac{n}{N}\right) \quad p_{2,n} = u \frac{n}{N} + (1-v) \left(1 - \frac{n}{N}\right) \quad (3.7)$$

e la probabilità di transizione diventa

$$p_{(n+1)n} = \left(1 - \frac{n}{N}\right) p_{1,n} = (1-u) \left(1 - \frac{n}{N}\right) \left(\frac{n}{N}\right) + v \left(1 - \frac{n}{N}\right)^2 \quad (3.8)$$

$$\begin{aligned} p_{nn} &= \left(\frac{n}{N}\right) p_{1,n} + \left(1 - \frac{n}{N}\right) p_{2,n} = \\ &= (1-u) \left(\frac{n}{N}\right)^2 + (1-v) \left(1 - \frac{n}{N}\right)^2 + (u+v) \left(\frac{n}{N}\right) \left(1 - \frac{n}{N}\right) \end{aligned} \quad (3.9)$$

$$p_{(n-1)n} = \left(\frac{n}{N}\right) p_{2,n} = (1-v) \left(\frac{n}{N}\right) \left(1 - \frac{n}{N}\right) + u \left(\frac{n}{N}\right)^2. \quad (3.10)$$

L'aggiunta della mutazione, come per il modello di Wright-Fisher, implica che il numero atteso di alleli  $A$  per generazione non è più conservato. L'equazione di partenza ha la stessa forma della (2.15), con la differenza che la probabilità di mutazione, essendo per generazione, differisce di un fattore  $N$  fra i due modelli. Si ha quindi:

$$\langle n(t+1) \rangle = \langle n(t) \rangle + v \left(1 - \frac{\langle n(t) \rangle}{N}\right) - u \left(\frac{\langle n(t) \rangle}{N}\right),$$

che dividendo per  $N$  diventa:

$$\begin{aligned} \frac{\langle n(t+1) \rangle}{N} &= a_{t+1} = \frac{\langle n(t) \rangle}{N} + \frac{v}{N} \left(1 - \frac{\langle n(t) \rangle}{N}\right) - \frac{u}{N} \left(\frac{\langle n(t) \rangle}{N}\right) = \\ &= a_t + \frac{v}{N} (1 - a_t) - \frac{u}{N} a_t = a_t \left(1 - \frac{u}{N} - \frac{v}{N}\right) + \frac{v}{N}. \end{aligned}$$

Ora, ripercorrendo il ragionamento fatto per il modello di Wright-Fisher, come mostrato nell'appendice (A), si può ricavare che per grandi tempi  $a_t \xrightarrow[t \rightarrow \infty]{} \frac{v}{u+v}$ . Perciò nè l'allele A nè l'allele B possono diventare fissati, in quanto c'è sempre la possibilità che un allele di un certo tipo venga creato attraverso mutazione anche se al momento non ce ne sono presenti nella popolazione di quel tipo. Anche in questo caso si possono ricavare analiticamente gli autovalori<sup>3</sup>:

$$\lambda_j = 1 - \frac{j}{N}(u+v) - \frac{j(j-1)}{N^2}(1-u-v). \quad (3.11)$$

### 3.4 Simulazioni Moran

In figura (3.1) riportiamo i risultati ottenuti simulando numericamente il modello di Moran senza mutazione, analogamente a quanto fatto per il modello di Wright-Fisher nel paragrafo (2.2). Possiamo vedere che, dopo un transiente iniziale, il sistema raggiunge lo stato stazionario in cui tutti gli alleli sono di un solo tipo. Al contrario in figura (3.2) si può vedere come, per il caso con mutazione, le condizioni iniziali dopo pochi passi temporali diventino influenti, in quanto i valori dei due alleli si stabilizzano attorno alle linee verdi rappresentanti i rapporti allelici di stazionarietà predetti dalla teoria.

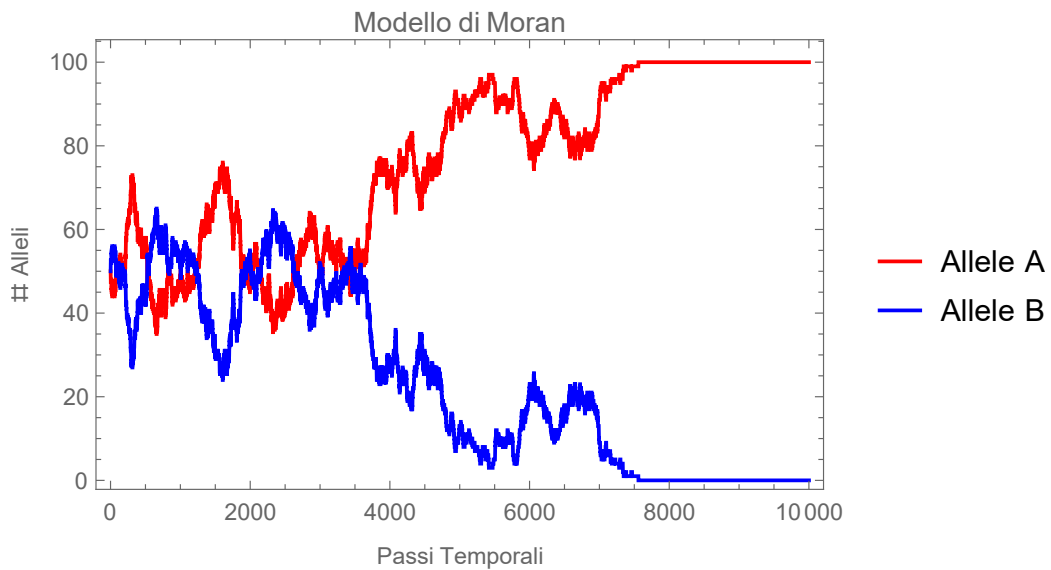


Figura 3.1: Simulazione della fissazione per il modello di Moran senza mutazione, con percentuale di alleli A di partenza  $p = 0.5$ .

<sup>3</sup>Rimandiamo la derivazione di questo risultato in appendice (B.4)

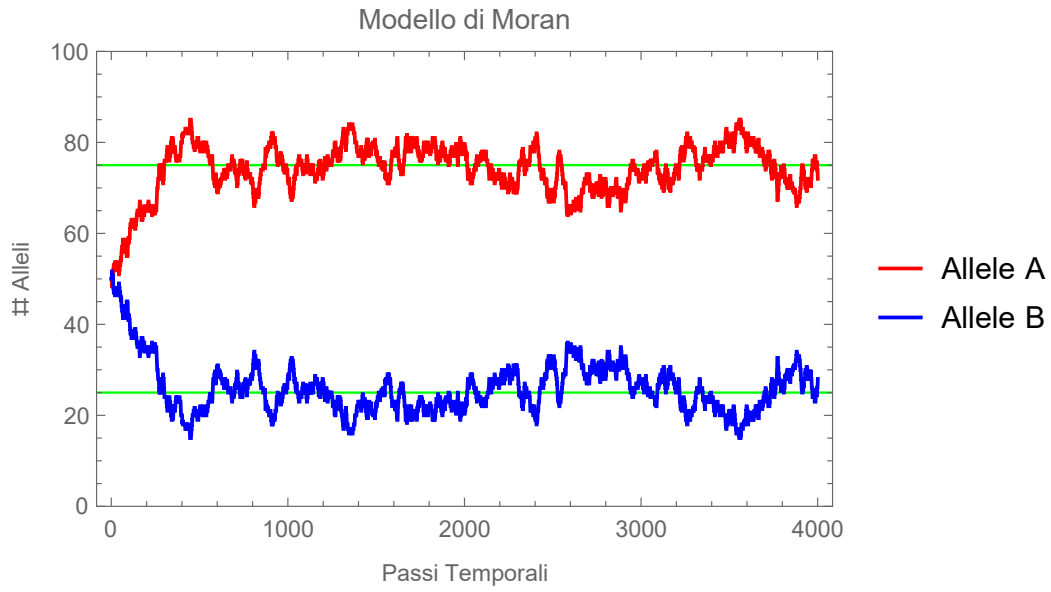


Figura 3.2: Simulazione della fissazione per il modello di Moran con mutazione, con percentuale di alleli  $A$  di partenza  $p = 0.5$  e probabilità di mutazione  $u = 0.25$  e  $v = 0.75$ .

---





## Capitolo 4

# Voter Model e Teoria Neutrale

Nei modelli di genomica delle popolazioni trattati precedentemente abbiamo considerato organismi aploidi che si riproducevano a sessualmente (cioè da un solo genitore). Questo ci ha permesso di identificare un gene con un individuo e viceversa. Una modellizzazione di questo tipo trova un immediato paragone nel campo dell'ecologia, associando i concetti *gene*  $\leftrightarrow$  *individuo* e *allele*  $\leftrightarrow$  *specie*. Inoltre, esistono variazioni dei modelli di Wright-Fisher e di Moran che implementano i concetti ecologici di *isola* e *deme*. Per esempio in alcune situazioni una popolazione può essere suddivisa in sottogruppi separati geograficamente (isole), tali per cui l'estinzione di un allele (specie) in una di essi non significa l'estinzione di tutta la popolazione. La popolazione di un'isola viene detta deme e viene influenzata dagli altri demi attraverso il processo di migrazione, nel quale alcuni individui provenienti da isole diverse prendono il posto di quelli morti. Il parallelismo con l'ecologia può essere esteso con le corrispondenze *popolazione*  $\leftrightarrow$  *metacomunità*, *mutazione*  $\leftrightarrow$  *speciazione* e *migrazione*  $\leftrightarrow$  *immigrazione*. Per comprendere meglio il confronto tra genomica ed ecologia bisogna introdurre uno dei più importanti modelli demografici in ecologia, il *voter model*.

### 4.1 Voter Model

Questo modello suppone che gli individui possano appartenere a varie specie (più di due). Tuttavia per capire l'andamento di una *comunità locale* (l'equivalente del deme) ci concentreremo solamente su una specie e tratteremo indistintamente tutte le altre. Questa approssimazione, detta *teoria neutrale*, è l'ipotesi fondamentale su cui è costruito il modello: trascuriamo quindi ogni possibile vantaggio o svantaggio che potrebbe portare una specie a riprodursi maggiormente o a migrare più facilmente rispetto alle altre. La dinamica di una comunità locale è governata dagli eventi di nascita e morte generalizzati (che includono speciazione, immigrazione ed emigrazione). Quando un individuo muore, viene sostituito dalla prole di uno degli altri componenti della comunità locale, oppure da un immigrato proveniente dalla metacomunità, come si può vedere in figura (4.1).

Ogni individuo della comunità locale ha la stessa probabilità, rispetto agli altri, di rimpiazzare quello morto, perché andremo a considerare il modello nella cosiddetta versione *non spaziale*. Questa approssimazione è ragionevole se siamo

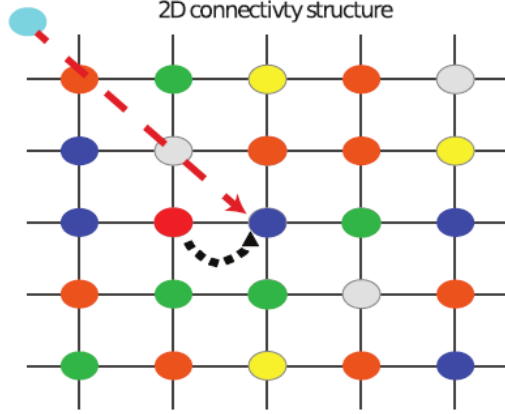


Figura 4.1: Processo con cui un individuo morto viene rimpiazzato nel voter model.

interessati a descrivere foreste composte da alberi, le cui spore vengono trasportate dal vento così lontane da rendere valida la precedente condizione di equiprobabilità. Chiamiamo  $b_{n,k}$  e  $d_{n,k}$  le variabili che rappresentano le probabilità di nascita e morte, rispettivamente, nella specie  $k$ -esima quando questa abbia  $n$  individui, e imponiamo le condizioni al contorno  $b_{-1,k} = d_{0,k} = 0$ . Denotiamo con  $p_{n,k}(t)$  la probabilità che la  $k$ -esima specie contenga  $n$  individui al tempo  $t$ . Si può mostrare<sup>1</sup> che l'evoluzione temporale di  $p_{n,k}(t)$  è regolata dalla *master equation*:

$$\frac{dp_{n,k}(t)}{dt} = p_{n+1,k}(t) d_{n+1,k} + p_{n-1,k}(t) b_{n-1,k} - p_{n,k}(t) (b_{n,k} + d_{n,k}). \quad (4.1)$$

Per calcolare i parametri  $b_{n,k}$  e  $d_{n,k}$  basta considerare le due regole dinamiche che governano questo processo stocastico:

1. Con probabilità  $1 - m$  si scelgono casualmente due individui dalla comunità locale. Se appartengono alla stessa specie, non succede nulla. Altrimenti, in maniera equiprobabile, uno degli individui viene rimpiazzato con la prole dell'altro. In altre parole i due individui servono come candidati per estinzione e figliatura.
2. Con probabilità  $m$  si sceglie casualmente un individuo dalla comunità locale. Lo si sostituisce con un nuovo individuo di un'altra specie, scelto con probabilità proporzionale all'abbondanza di quella specie nella metacomunità. Questo evento corrisponde alla morte dell'individuo scelto nella comunità locale seguita dall'arrivo di un immigrato dalla metacomunità.

$$b_{n,k} = (1 - m) \frac{n}{J} \frac{J - n}{J - 1} + m \frac{\mu_k}{J_M} \left(1 - \frac{n}{J}\right), \quad (4.2)$$

$$d_{n,k} = (1 - m) \frac{n}{J} \frac{J - n}{J - 1} + m \left(1 - \frac{\mu_k}{J_M}\right) \frac{n}{J}, \quad (4.3)$$

<sup>1</sup>Si veda per esempio [2, 8, 13].

dove  $\mu_k$  è l'abbondanza della specie  $k$ -esima nella metacomunità e  $J_M$  è la popolazione totale della metacomunità. Il membro di destra dell'equazione (4.2) consiste di due termini. Il primo corrisponde alla regola (1) con una nascita nella specie  $k$ -esima accompagnata dalla morte di un individuo appartenente ad una delle altre specie nella comunità locale. Il secondo termine corrisponde alla regola (2) e tiene conto di un aumento della popolazione della specie  $k$ -esima a causa dell'immigrazione dalla metacomunità. L'immigrazione è proporzionale all'abbondanza relativa  $\frac{\mu_k}{J_M}$  della specie  $k$ -esima nella metacomunità. L'interpretazione dell'equazione (4.3) è analoga. Se approssimiamo i parametri di nascita e morte all'ordine  $\mathcal{O}\left(\frac{n}{J}\right)$  possiamo semplificarli ottenendo:

$$b_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \frac{\mu_k}{J_M} \left(1 - \frac{n}{J}\right) \xrightarrow{J \gg n} (1-m) \frac{n}{J} + m \frac{\mu_k}{J_M} \left(1 - \frac{n}{J}\right),$$

$$d_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \left(1 - \frac{\mu_k}{J_M}\right) \frac{n}{J} \xrightarrow{J \gg n} (1-m) \frac{n}{J} + m \left(1 - \frac{\mu_k}{J_M}\right) \frac{n}{J}.$$

## 4.2 Soluzione Stazionaria

Consideriamo un caso particolare del modello precedente, quello in cui la specie considerata non sia presente nella metacomunità ( $\mu_k = 0$ ). Conseguentemente le probabilità di nascita e morte diventano:

$$b_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} \xrightarrow{J \gg n} (1-m) \frac{n}{J}, \quad (4.4)$$

$$d_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \frac{n}{J} \xrightarrow{J \gg n} (1-m) \frac{n}{J} + m \frac{n}{J} = \frac{n}{J}. \quad (4.5)$$

A questo punto possiamo chiederci se sia possibile derivare analiticamente una soluzione stazionaria. Partendo dall'equazione (4.1) e imponendo l'indipendenza dal tempo della soluzione stazionaria, otteniamo:

$$\frac{dP^*(n)}{dt} = 0 = J(n+1) - J(n) \quad (4.6)$$

con

$$J(n) = d_n P^*(n) - b_{n-1} P^*(n-1). \quad (4.7)$$

Notiamo che  $P^*(n)$ , essendo uno stato stazionario, non dipende da  $t$  e inoltre, per l'ipotesi di teoria neutrale, non dipende nemmeno dalla specie  $k$ . Ometteremo quindi l'indice  $k$  anche per le probabilità di nascita e morte. Questo significa che:

$$J(0) = d_0 P^*(0) - b_{-1} P^*(-1) = 0, \quad \text{allora}$$

$$0 = J(n+1) - J(n) \Rightarrow \sum_{z=0}^{n-1} [J(z+1) - J(z)] = J(n) - J(0) = J(n)$$

$$J(n) = 0 \Rightarrow 0 = d_n P^*(n) - b_{n-1} P^*(n-1) \Rightarrow P^*(n) = \frac{b_{n-1}}{d_n} P^*(n-1).$$

Perciò

$$P^*(n) = P^*(0) \prod_{z=1}^n \frac{b_{z-1}}{d_z}, \quad (4.8)$$

dove  $P^*(0)$  si ricava dalla condizione di normalizzazione  $\sum_n P^*(n) = 1$ . Ricordando le equazioni (4.4) e (4.5) si ricava:

$$P^*(n) = -\frac{(1-m)^n}{n \log(m)}. \quad (4.9)$$

Questa funzione è monotona e decrescente molto rapidamente. Perciò è improbabile che lo stato stazionario della specie considerata abbia un'abbondanza relativa elevata, mentre verosimilmente sarà decisamente bassa.

## Capitolo 5

# Conclusioni

In questo lavoro abbiamo analizzato principalmente 3 modelli: Wright-Fisher, Moran e voter. Il Wright-Fisher e il Moran hanno due caratteristiche principali comuni:

- per entrambi è possibile ricavare analiticamente gli autovalori delle matrici di transizione,
- lo stato di fissazione per grandi tempi dipende unicamente dalle probabilità di mutazione e non dalle condizioni iniziali.

Di contro, per il modello di Moran, il fatto che il processo di riproduzione avvenga per un individuo alla volta, anziché per gruppi di  $N$ , permette di confrontarlo più facilmente con il voter model. Inoltre è possibile ricavare più informazioni sullo stato di fissazione, in quanto sono noti anche gli autovettori della matrice di transizione. Infatti la scomposizione in autovettori per il modello di Moran ci permette di sapere quale sia lo stato di fissazione più probabile, esclusi i due banali che si presentano anche per il Wright-Fisher. Il voter model rispetto a quello di Moran ha dei presupposti completamente diversi, anche se l'ipotesi di teoria neutrale facilita il confronto con quest'ultimo. A differenza dei modelli di genetica, infatti, nel modello ecologico non si costruiscono delle matrici di transizione all'interno del contesto delle catene di Markov, nel voter model la dinamica delle abbondanze delle specie viene descritta attraverso una master equation. Tuttavia può essere instaurato un forte parallelismo tra i linguaggi della genomica e dell'ecologia, ed i ragionamenti che permettono di ricavare i coefficienti di nascita e morte per il voter model sono simili a quelli con cui si ricavano le probabilità  $p_1$  e  $p_2$  per i primi due modelli. In ultima istanza però, per quanto il voter model e il modello di Moran siano concettualmente identici, il formalismo con cui vengono trattati è diverso, perciò si ottengono risultati molto differenti. Inoltre nel caso di mutazione l'equazione (3.6) non è più valida, quindi si hanno molte meno informazioni rispetto al corrispondente caso con speciazione descritto dall'equazione (4.9). Questo lavoro mette in luce come dei modelli concettualmente semplici, possano dare origine ad una varietà molto ampia di risultati. Questo è ancora più accattivante se si considera che stiamo entrando sempre di più nell'era dei "Big Data", perciò coloro che vorranno studiare questi ed altri modelli per le loro interessanti capacità predittive, potranno contare su una grande quantità di dati, sia per la genomica che per l'ecologia, per verificare l'affidabilità di questi modelli.



## Appendice A

# Fissazione con Mutazione

Deriviamo di seguito formalmente l'equazione (2.16) che descrive il limite di fissazione per grandi tempi in caso di mutazione per i modelli di Wright-Fisher e Moran. Partiamo da  $\langle n(t+1) \rangle = (1-u)\langle n(t) \rangle + v(N - \langle n(t) \rangle)$

$$\begin{aligned}\frac{\langle n(t+1) \rangle}{N} = a_{t+1} &= (1-u)\frac{\langle n(t) \rangle}{N} + v\left(1 - \frac{\langle n(t) \rangle}{N}\right) = \\ &= (1-u)a_t + v(1-a_t) = (1-u-v)a_t + v.\end{aligned}$$

Assumiamo adesso che  $a_t = \frac{v}{u+v} + c_t$  da cui  $a_{t+1} = \frac{v}{u+v} + c_{t+1}$ , perciò

$$\begin{aligned}a_{t+1} &= (1-u-v)a_t + v = (1-u-v)c_t + (1-u-v)\frac{v}{u+v} + v \\ \frac{v}{u+v} + c_{t+1} &= (1-u-v)c_t + \frac{v}{u+v} - \frac{(u+v)v}{u+v} + v \Rightarrow \\ \Rightarrow c_{t+1} &= (1-u-v)c_t \Rightarrow c_t = (1-u-v)^t c_0 \\ a_t &= \frac{v}{u+v} + c_t = \frac{v}{u+v} + (1-u-v)^t c_0 = \\ &= \frac{v}{u+v} + (1-u-v)^t \left[ a_0 - \frac{v}{u+v} \right] = \frac{v}{u+v} [1 - (1-u-v)^t] + a_0 (1-u-v)^t.\end{aligned}$$

Ora notiamo che  $0 < 1-u-v < 1$  perché la probabilità totale di mutazione  $u+v$  deve essere compresa fra 0 e 1 cioè vale  $0 < u+v < 1$  da cui  $a_t \xrightarrow{t \rightarrow \infty} \frac{v}{u+v}$  perché  $(1-u-v)^t \xrightarrow{t \rightarrow \infty} 0$  se  $0 < 1-u-v < 1$ .





## Appendice B

# Teoremi di Cannings

Il primo tentativo di generalizzare questi modelli è dovuto a Felsenstein che in [14] generalizza i modelli di Wright-Fisher, Moran ed altri per il caso senza mutazione. Dopodiché Cannings in [15] espone un metodo per comprendere tutte le varianti sopracitate includendo anche i casi con mutazione, infine Gladstien in [16] mostra che gli approcci di Felsenstein e Cannings sono equivalenti, confutando anche un teorema proposto da Cannings con un controesempio e fornendo il giusto teorema. Noi di seguito utilizzeremo l'approccio riportato in [17] che segue il lavoro di Cannings proprio perché si tratta del modo più semplice per ricavare gli autovalori dei vari modelli. Partiamo dai teoremi originali di Cannings.

**Teorema B.1** (Primo teorema di Cannings). *Consideriamo una popolazione di geni di dimensione fissata  $N$  che si riproduce ai tempi  $t = 0, 1, 2, \dots$  e supponiamoli divisi in due tipi:  $A$  e  $B$ , tali per cui  $X_t$  è il numero di geni  $A$  al tempo  $t$ . Definiamo  $w_i$  la variabile che assume valori 1 o 0 a seconda che il gene  $i$ -esimo al tempo  $t$  continui a esistere al tempo  $t + 1$ -esimo e  $z_i$  la variabile che indica quanti sono i componenti della prole del gene  $i$ -esimo al tempo  $t$ . Allora  $X_{t+1} = \sum_{i=1}^{X_t} (w_i + z_i)$ , inoltre detto  $y_i = w_i + z_i$ , se  $p_{ij} = \text{Prob}\{X_{t+1} = j | X_t = i\} \quad \forall i, j = 0, 1, \dots, N$  gli autovalori della matrice  $\mathcal{P} = \{p_{ij}\}$  sono*

$$\lambda_0 = 1, \quad \lambda_j = \mathbb{E}(y_1 y_2 \cdots y_j), \quad j = 1, 2, \dots, N \quad (\text{B.1})$$

**Teorema B.2** (Secondo teorema di Cannings). *Manteniamo la notazione del teorema precedente, inoltre supponendo i geni  $z_i$  di tipo  $A$ , definiamo  $z_{i1}$  come il numero di geni  $z_i$  di tipo  $A$  non mutati e  $z_{i2}$  come il numero di geni  $z_i$  di tipo  $B$  mutati in geni di tipo  $A$ . Allora  $X_{t+1} = \sum_{i=1}^{X_t} (w_i + z_{i1}) + \sum_{i=X_t+1}^N z_{i2}$ , inoltre detto  $y_i = w_i + z_{i1}$ , se  $p_{ij} = \text{Prob}\{X_{t+1} = j | X_t = i\} \quad \forall i, j = 0, 1, \dots, N$  gli autovalori della matrice  $\mathcal{P} = \{p_{ij}\}$  sono*

$$\lambda_0 = 1, \quad \lambda_j = \mathbb{E} \left[ \prod_{i=1}^j (y_i - z_{i2}) \right] = \mathbb{E} \left[ \prod_{i=1}^j (w_i + z_{i1} - z_{i2}) \right], \quad j = 1, 2, \dots, N \quad (\text{B.2})$$

## B.1 Applicazione al Modello di Wright-Fisher senza mutazione

Prendiamo in considerazione il modello di Wright-Fisher, dove gli  $N$  geni sono descritti dalle  $N$  variabili casuali  $(y_1, \dots, y_N)$  che hanno una distribuzione multinomiale con parametri  $(N; p_1, \dots, p_N)$ , con  $p_N = 1 - \sum_{i=1}^{N-1} p_i$ . Partiamo col mostrare che  $p(y_1, \dots, y_j) = \binom{N}{\vec{y}} p_1^{y_1} \dots p_j^{y_j} \left(1 - \sum_{i=1}^j p_i\right)^{N-n_j}$  dove  $n_j = \sum_{i=1}^j y_i$  e  $\binom{N}{\vec{y}} = \frac{N!}{y_1! \dots y_j! (N-n_j)!}$

$$\begin{aligned}
p(y_1, \dots, y_j) &= \sum_{y_{j+1}=0}^{N-n_j} \dots \sum_{y_{N-1}=0}^{N-n_{N-2}} p(y_1, \dots, y_j, y_{j+1}, \dots, y_{N-1}, N - n_{N-1}) = \\
&= \sum_{y_{j+1}=0}^{N-n_j} \dots \sum_{y_{N-1}=0}^{N-n_{N-2}} \frac{N!}{y_1! \dots y_{N-1}! (N - n_{N-1})!} \times \\
&\quad \times p_1^{y_1} \dots p_{N-1}^{y_{N-1}} \left(1 - \sum_{i=1}^{N-1} p_i\right)^{N-n_{N-1}} = \\
&= \sum_{y_{j+1}=0}^{N-n_j} \dots \sum_{y_{N-2}=0}^{N-n_{N-3}} \frac{N! p_1^{y_1} \dots p_{N-2}^{y_{N-2}}}{y_1! \dots y_{N-2}!} \times \\
&\quad \times \sum_{y_{N-1}=0}^{N-n_{N-2}} \frac{1}{y_{N-1}! (N - n_{N-1})!} p_{N-1}^{y_{N-1}} \left(1 - \sum_{i=1}^{N-1} p_i\right)^{N-n_{N-1}} = \\
&= \sum_{y_{j+1}=0}^{N-n_j} \dots \sum_{y_{N-2}=0}^{N-n_{N-3}} \frac{N! p_1^{y_1} \dots p_{N-2}^{y_{N-2}}}{y_1! \dots y_{N-2}! (N - n_{N-2})!} \times \\
&\quad \times \sum_{y_{N-1}=0}^{N-n_{N-2}} \frac{(N - n_{N-2})!}{y_{N-1}! (N - n_{N-1})!} p_{N-1}^{y_{N-1}} \left(1 - \sum_{i=1}^{N-1} p_i\right)^{N-n_{N-1}} = \\
&= \sum_{y_{j+1}=0}^{N-n_j} \dots \sum_{y_{N-2}=0}^{N-n_{N-3}} \frac{N! p_1^{y_1} \dots p_{N-2}^{y_{N-2}}}{y_1! \dots y_{N-2}! (N - n_{N-2})!} \left(1 - \sum_{i=1}^{N-2} p_i\right)^{N-n_{N-2}} =
\end{aligned}$$

Dove nell'ultimo passaggio si è sfruttato il teorema binomiale notando che

$$\begin{aligned}
&\sum_{y_{N-1}=0}^{N-n_{N-2}} \frac{(N - n_{N-2})!}{y_{N-1}! (N - n_{N-1})!} p_{N-1}^{y_{N-1}} \left(1 - \sum_{i=1}^{N-1} p_i\right)^{N-n_{N-1}} = \\
&= \sum_{y_{N-1}=0}^{N-n_{N-2}} \frac{(N - n_{N-2})!}{y_{N-1}! (N - n_{N-2} - y_{N-1})!} p_{N-1}^{y_{N-1}} \left(1 - \sum_{i=1}^{N-1} p_i\right)^{N-n_{N-2}-y_{N-1}} = \\
&= \left(p_{N-1} + 1 - \sum_{i=1}^{N-1} p_i\right)^{N-n_{N-2}} = \left(1 - \sum_{i=1}^{N-2} p_i\right)^{N-n_{N-2}}
\end{aligned}$$

Se iteriamo questo procedimento sfruttando ogni volta il teorema binomiale otteniamo

$$= \frac{N! p_1^{y_1} \cdots p_j^{y_j}}{y_1! \cdots y_j! (N - n_j)!} \left(1 - \sum_{i=1}^j p_i\right)^{N-n_j} = \binom{N}{\vec{y}} p_1^{y_1} \cdots p_j^{y_j} \left(1 - \sum_{i=1}^j p_i\right)^{N-n_j}$$

Ora coerentemente con il modello di Wright-Fisher imponiamo  $p_1 = \cdots = p_j = \frac{1}{N}$  perciò sfruttando il teorema (B.1) otteniamo

$$\begin{aligned} \lambda_j &= \mathbb{E}(y_1 y_2 \cdots y_j) = \sum_{\substack{y_1=0, \dots, y_j=0 \\ y_1 + \dots + y_j = N}}^N y_1 \cdots y_j \frac{N!}{y_1! \cdots y_j! (N - \sum_{i=1}^j y_i)!} \times \\ &\quad \times \left(\frac{1}{N}\right)^{\sum_{i=1}^j y_i} \left(1 - \frac{j}{N}\right)^{N - \sum_{i=1}^j y_i} = \\ &= \left(\frac{1}{N}\right)^j \sum_{\substack{y_1=0, \dots, y_j=0 \\ y_1 + \dots + y_j = N}}^N \frac{N!}{(y_1 - 1)! \cdots (y_j - 1)! (N - j - \sum_{i=1}^j (y_i - 1))!} \times \\ &\quad \times \left(\frac{1}{N}\right)^{\sum_{i=1}^j (y_i - 1)} \left(1 - \frac{j}{N}\right)^{N - j - \sum_{i=1}^j (y_i - 1)} = \end{aligned}$$

perché  $N - j - \sum_{i=1}^j (y_i - 1) = N - j - \sum_{i=1}^j y_i + \sum_{i=1}^j 1 = N - j - \sum_{i=1}^j y_i + j = N - \sum_{i=1}^j y_i$ . Ora introducendo la nuova variabile  $x_i = y_i - 1$  e considerando che gli estremi delle sommatorie cambiano in maniera diversa da quanto ci aspetteremmo a causa della condizione  $y_1 + \cdots + y_j = N$  otteniamo:

$$\begin{aligned} &\frac{N(N-1) \cdots (N-j+1)}{N^j} \sum_{\substack{x_1=0, \dots, x_j=0 \\ x_1 + \dots + x_j = N-j}}^{N-j} \frac{(N-j)!}{x_1! \cdots x_j! (N-j - \sum_{i=1}^j x_i)!} \times \\ &\quad \times \left(\frac{1}{N}\right)^{\sum_{i=1}^j x_i} \left(1 - \frac{j}{N}\right)^{N-j - \sum_{i=1}^j x_i} \end{aligned}$$

Perciò si ha:

$$\lambda_j = \frac{N(N-1) \cdots (N-j+1)}{N^j} \tag{B.3}$$

## B.2 Applicazione al Modello di Wright-Fisher con mutazione

Per calcolare gli autovalori del modello di Wright-Fisher con mutazione sfruttiamo il teorema (B.2) e ricordiamo l'equazione (B.3) che fornisce gli autovalori nel caso senza mutazione.

$$\begin{aligned} \lambda_j &= \mathbb{E} \left[ \prod_{i=1}^j (w_i + z_{i1} - z_{i2} | z_i) \right] = \mathbb{E} \left[ \prod_{i=1}^j (z_{i1} - z_{i2} | z_i) \right] = \mathbb{E} \left[ \prod_{i=1}^j (1 - u - v) z_i \right] = \\ &= (1 - u - v)^j \mathbb{E} \left[ \prod_{i=1}^j z_i \right] = (1 - u - v)^j \frac{N(N-1) \cdots (N-j+1)}{N^j} \end{aligned}$$

Dove il primo passaggio è dato dal fatto che, nel modello di Wright-Fisher, ogni generazione tutti i geni muoiono, perciò  $y_i \equiv 0$ . Il secondo passaggio invece è giustificato dal fatto che dato  $z_i, z_{i1}$  e  $z_{i2}$  hanno distribuzione multinomiale con parametri  $1 - u$  e  $v$  rispettivamente. Infatti  $u$  è il tasso di mutazione da  $A$  a  $B$  e  $v$  quello da  $B$  ad  $A$ .

### B.3 Applicazione al Modello di Moran senza mutazione

Per calcolare gli autovalori del modello di Moran senza mutazione sfrutteremo il teorema (B.1), prima di usarlo però calcoliamo i valori possibili del prodotto  $y_1 y_2 \cdots y_j$  che sono:

$$y_1 y_2 \cdots y_j = \begin{cases} 0 & \text{se uno dei geni muore e nessuno si riproduce} \\ 2 & \text{se uno dei geni si riproduce e nessuno muore} \\ 1 & \text{altrimenti} \end{cases}$$

corrispondentemente le relative probabilità sono

$$\begin{aligned} \text{Prob}(y_1 y_2 \cdots y_j = 0) &= \frac{j}{N} \frac{N-1}{N} = j \frac{N-1}{N^2} \\ \text{Prob}(y_1 y_2 \cdots y_j = 2) &= \frac{j}{N} \frac{N-j}{N} = j \frac{N-j}{N^2} \\ \text{Prob}(y_1 y_2 \cdots y_j = 0) &= 1 - \text{Prob}(y_1 y_2 \cdots y_j = 0) - \text{Prob}(y_1 y_2 \cdots y_j = 2) = \\ &= 1 - j \frac{N-1}{N^2} - j \frac{N-j}{N^2} = 1 - j \frac{2N-j-1}{N^2} \end{aligned}$$

Ora procediamo con il calcolo degli autovalori

$$\begin{aligned} \lambda_j &= 0 \left[ j \frac{N-1}{N^2} \right] + 1 \left[ 1 - j \frac{2N-j-1}{N^2} \right] + 2 \left[ j \frac{N-j}{N^2} \right] = \\ &= 1 - \frac{2Nj - j^2 - j - 2Nj + 2j^2}{N^2} = 1 - \frac{j(j+1)}{N^2} \end{aligned}$$

### B.4 Applicazione al Modello di Moran con mutazione

Per calcolare gli autovalori del modello di Moran con mutazione sfruttiamo nuovamente il teorema (B.2):

$$\begin{aligned} \lambda_j &= \mathbb{E} \left[ \prod_{i=1}^j (w_i + z_{i1} - z_{i2}) \right] = \mathbb{E} \left[ \left( \prod_{i=1}^{k_1} w_i \right) \left( \prod_{i=k_1+1}^{k_1+k_2} z_{i1} \right) \left( \prod_{i=k_1+k_2+1}^j z_{i2} \right) \right] = \\ &= (1-u)^{(k_1+k_2)-k_1} (v)^{j-(k_1+k_2)} \mathbb{E} \left[ \left( \prod_{i=1}^{k_1} w_i \right) \left( \prod_{i=k_1+1}^j z_i \right) \right] = \\ &= (1-u)^{k_2} (v)^{j-k_1-k_2} \mathbb{E} \left[ \left( \prod_{i=1}^{k_1} w_i \right) \left( \prod_{i=k_1+1}^j z_i \right) \right] = \mathbb{E} \left[ \prod_{i=1}^j (w_i + (1-u-v) z_i) \right]. \end{aligned}$$

Ora supponiamo che le  $w_i$  siano indipendenti dalle  $z_i$  cioè che

$$\text{Prob}(w = w^*, z = z^*) = \text{Prob}(w = w^*) \text{Prob}(z = z^*)$$

Da cui otteniamo

$$\lambda_j = \sum_{k=0}^j \binom{j}{k} \mathbb{E} \left[ \prod_{i=1}^k z_i \right] \mathbb{E} \left[ \prod_{i=k+1}^j w_i \right] (1-u-v)^k$$

Che rimane un'equazione sensata solo se imponiamo che

$$\mathbb{E} \left[ \prod_{i=1}^0 z_i \right] = 1 \quad \text{e} \quad \mathbb{E} \left[ \prod_{i=j+1}^j w_i \right] = 1$$

Ora imponiamo una condizione peculiare del modello di Moran, ovvero  $z_p z_q = 0 \quad \forall p \neq q$ . Quindi gli unici termini della sommatoria che sopravvivono sono quelli con  $k=0$  e  $k=1$ .

$$\begin{aligned} \lambda_j &= \binom{j}{0} \mathbb{E} \left[ \prod_{i=1}^0 z_i \right] \mathbb{E} \left[ \prod_{i=1}^j w_i \right] (1-u-v)^0 + \binom{j}{1} \mathbb{E} \left[ \prod_{i=1}^1 z_i \right] \mathbb{E} \left[ \prod_{i=2}^j w_i \right] (1-u-v)^1 = \\ &= \mathbb{E} \left[ \prod_{i=1}^j w_i \right] + j \mathbb{E}[z_i] \mathbb{E} \left[ \prod_{i=2}^j w_i \right] (1-u-v) \end{aligned}$$

Ora consideriamo che  $\mathbb{E}[z_i] = \frac{1}{N}$  perché solo 1 gene su  $N$  genera prole. Inoltre  $\mathbb{E} \left[ \prod_{i=1}^j w_i \right] = 1 - \frac{j}{N}$  e  $\mathbb{E} \left[ \prod_{i=2}^j w_i \right] = 1 - \frac{j-1}{N}$  dato che  $w_i = 0$  se l' $i$ -esimo gene muore e le probabilità che uno muoia sono  $\frac{j}{N}$  e  $\frac{j-1}{N}$  con probabilità complementari  $1 - \frac{j}{N}$  e  $1 - \frac{j-1}{N}$ . Perciò

$$\begin{aligned} \lambda_j &= 1 - \frac{j}{N} + \frac{j}{N} \left( 1 - \frac{j-1}{N} \right) (1-u-v) = \\ &= 1 - \frac{j}{N} + \frac{j}{N} (1-u-v) - \frac{j(j-1)}{N^2} (1-u-v) = 1 - \frac{j}{N} (u+v) - \frac{j(j-1)}{N^2} (1-u-v) \end{aligned}$$



## Appendice C

### Verifica degli autovettori

Verifichiamo adesso che gli autovettori lo siano effettivamente per la matrice di transizione del modello di Moran senza mutazione. Partiamo dagli autovettori destri in particolare con  $\vec{v}^{(0)}$

$$\begin{pmatrix} 1 & (\frac{1}{N})(1 - \frac{1}{N}) & 0 & 0 & \dots & 0 \\ 0 & (\frac{1}{N})^2 + (1 - \frac{1}{N})^2 & (\frac{2}{N})(1 - \frac{2}{N}) & 0 & \dots & 0 \\ 0 & (\frac{1}{N})(1 - \frac{1}{N}) & (\frac{2}{N})^2 + (1 - \frac{2}{N})^2 & (\frac{3}{N})(1 - \frac{3}{N}) & \dots & 0 \\ 0 & 0 & (\frac{2}{N})(1 - \frac{2}{N}) & (\frac{3}{N})^2 + (1 - \frac{3}{N})^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Dove l'uguaglianza è dato dal fatto che l'unico elemento non nullo del vettore è il primo, perciò viene selezionata solo la prima colonna della matrice. Proseguiamo con  $\vec{v}^{(1)}$

$$\begin{pmatrix} 1 & (\frac{1}{N})(1 - \frac{1}{N}) & 0 & 0 & \dots & 0 \\ 0 & (\frac{1}{N})^2 + (1 - \frac{1}{N})^2 & (\frac{2}{N})(1 - \frac{2}{N}) & 0 & \dots & 0 \\ 0 & (\frac{1}{N})(1 - \frac{1}{N}) & (\frac{2}{N})^2 + (1 - \frac{2}{N})^2 & (\frac{3}{N})(1 - \frac{3}{N}) & \dots & 0 \\ 0 & 0 & (\frac{2}{N})(1 - \frac{2}{N}) & (\frac{3}{N})^2 + (1 - \frac{3}{N})^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Dove l'uguaglianza è dato dal fatto che l'unico elemento non nullo del vettore è l'ultimo, perciò viene selezionata solo l'ultima colonna della matrice. Proseguiamo

con  $\vec{v}^{(2)}$

$$\begin{pmatrix} 1 & (\frac{1}{N})(1 - \frac{1}{N}) & 0 & 0 & \cdots & 0 \\ 0 & (\frac{1}{N})^2 + (1 - \frac{1}{N})^2 & (\frac{2}{N})(1 - \frac{2}{N}) & 0 & \cdots & 0 \\ 0 & (\frac{1}{N})(1 - \frac{1}{N}) & (\frac{2}{N})^2 + (1 - \frac{2}{N})^2 & (\frac{3}{N})(1 - \frac{3}{N}) & \cdots & 0 \\ 0 & 0 & (\frac{2}{N})(1 - \frac{2}{N}) & (\frac{3}{N})^2 + (1 - \frac{3}{N})^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{2} \end{pmatrix} =$$

$$= \left(1 - \frac{2}{N^2}\right) \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{2} \end{pmatrix}$$

Dove per provare l'uguaglianza basta fare il conto per la prima riga della matrice e poi per una generica. Per la prima abbiamo  $-\frac{N-1}{2} + 1 + (\frac{1}{N})(1 - \frac{1}{N}) = -\frac{N-1}{2} + \frac{1}{N} - \frac{1}{N^2} = \frac{-N^3 + N^2 + 2N - 2}{2N^2} = \frac{(N^2-2)(1-N)}{2N^2} = (1 - \frac{2}{N^2}) (-\frac{N-1}{2})$  Mentre per una riga generica abbiamo  $\frac{X}{N}(1 - \frac{X}{N}) + (\frac{X+1}{N})^2 + (1 - \frac{X+1}{N})^2 + \frac{X+2}{N}(1 - \frac{X+2}{N}) = \frac{XN - X^2}{N^2} + \frac{X^2 + 2X + 1}{N^2} + \frac{N^2 + X^2 + 1 - 2NX - 2N + 2X}{N^2} + \frac{NX + 2N - X^2 - 4X - 4}{N^2} = \frac{N^2 - 2}{N^2} = 1 - \frac{2}{N^2}$

Ora passiamo a verificare gli autovettori sinistri, partiamo con  $\vec{\pi}^{(0)}$

$$\left(1, \dots, \frac{1}{N}, 0\right) \begin{pmatrix} 1 & (\frac{1}{N})(1 - \frac{1}{N}) & 0 & 0 & \cdots & 0 \\ 0 & (\frac{1}{N})^2 + (1 - \frac{1}{N})^2 & (\frac{2}{N})(1 - \frac{2}{N}) & 0 & \cdots & 0 \\ 0 & (\frac{1}{N})(1 - \frac{1}{N}) & (\frac{2}{N})^2 + (1 - \frac{2}{N})^2 & (\frac{3}{N})(1 - \frac{3}{N}) & \cdots & 0 \\ 0 & 0 & (\frac{2}{N})(1 - \frac{2}{N}) & (\frac{3}{N})^2 + (1 - \frac{3}{N})^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \left(1, \dots, \frac{1}{N}, 0\right).$$

Per una colonna generica si ha:  $\frac{N-X+1}{N} [\frac{X}{N}(1 - \frac{X}{N})] + \frac{N-X}{N} [(\frac{X}{N})^2 + (1 - \frac{X}{N})^2] + \frac{N-X-1}{N} [\frac{X}{N}(1 - \frac{X}{N})] = \frac{N-X}{N} [(\frac{X}{N})^2 + (1 - \frac{X}{N})^2] + \frac{2N-2X}{N} [\frac{X}{N}(1 - \frac{X}{N})] = \frac{N-X}{N} [(\frac{X}{N})^2 + (1 - \frac{X}{N})^2 + 2\frac{X}{N}(1 - \frac{X}{N})] = \frac{N-X}{N} [(\frac{X}{N} + 1 - \frac{X}{N})^2] = \frac{N-X}{N}$ .

Passiamo adesso a  $\vec{\pi}^{(1)}$

$$\left(0, \frac{1}{N}, \dots, 1\right) \begin{pmatrix} 1 & (\frac{1}{N})(1 - \frac{1}{N}) & 0 & 0 & \cdots & 0 \\ 0 & (\frac{1}{N})^2 + (1 - \frac{1}{N})^2 & (\frac{2}{N})(1 - \frac{2}{N}) & 0 & \cdots & 0 \\ 0 & (\frac{1}{N})(1 - \frac{1}{N}) & (\frac{2}{N})^2 + (1 - \frac{2}{N})^2 & (\frac{3}{N})(1 - \frac{3}{N}) & \cdots & 0 \\ 0 & 0 & (\frac{2}{N})(1 - \frac{2}{N}) & (\frac{3}{N})^2 + (1 - \frac{3}{N})^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \left(0, \frac{1}{N}, \dots, 1\right)$$

Per una colonna generica si ha:  $\frac{X-1}{N} [\frac{X}{N}(1 - \frac{X}{N})] + \frac{X}{N} [(\frac{X}{N})^2 + (1 - \frac{X}{N})^2] +$



$$\begin{aligned} \frac{X+1}{N} \left[ \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] &= \frac{X}{N} \left[ \left( \frac{X}{N} \right)^2 + \left( 1 - \frac{X}{N} \right)^2 \right] + \frac{2X}{N} \left[ \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] = \\ &= \frac{X}{N} \left[ \left( \frac{X}{N} \right)^2 + \left( 1 - \frac{X}{N} \right)^2 + 2 \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] = \frac{X}{N} \left[ \left( \frac{X}{N} + 1 - \frac{X}{N} \right)^2 \right] = \frac{X}{N}. \end{aligned}$$

Infine concludiamo la verifica con

$$\begin{aligned} &\frac{6}{N(N^2-1)} (0, N-1, 2(N-2), \dots, n(N-n), \dots, 0) \times \\ &\times \begin{pmatrix} 1 & \left(\frac{1}{N}\right)\left(1-\frac{1}{N}\right) & 0 & 0 & \dots & 0 \\ 0 & \left(\frac{1}{N}\right)^2 + \left(1-\frac{1}{N}\right)^2 & \left(\frac{2}{N}\right)\left(1-\frac{2}{N}\right) & 0 & \dots & 0 \\ 0 & \left(\frac{1}{N}\right)\left(1-\frac{1}{N}\right) & \left(\frac{2}{N}\right)^2 + \left(1-\frac{2}{N}\right)^2 & \left(\frac{3}{N}\right)\left(1-\frac{3}{N}\right) & \dots & 0 \\ 0 & 0 & \left(\frac{2}{N}\right)\left(1-\frac{2}{N}\right) & \left(\frac{3}{N}\right)^2 + \left(1-\frac{3}{N}\right)^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} = \\ &= \left( 1 - \frac{2}{N^2} \right) \frac{6}{N(N^2-1)} (0, N-1, 2(N-2), \dots, n(N-n), \dots, 0) \end{aligned}$$

Per una colonna generica si ha:

$$\begin{aligned} &(X-1)(N-X+1) \left[ \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] + X(N-X) \left[ \left( \frac{X}{N} \right)^2 + \left( 1 - \frac{X}{N} \right)^2 \right] + \\ &+ (X+1)(N-X-1) \left[ \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] = X(N-X) \left[ \left( \frac{X}{N} \right)^2 + \left( 1 - \frac{X}{N} \right)^2 \right] + \\ &+ 2X(N-X) \left[ \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] + (X-N+X-1+N-X-X-1) \left[ \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] = \\ &= X(N-X) \left[ \left( \frac{X}{N} \right)^2 + \left( 1 - \frac{X}{N} \right)^2 + 2 \frac{X}{N} \left( 1 - \frac{X}{N} \right) \right] - 2 \frac{X(N-X)}{N^2} = \\ &= X(N-X) \left[ \left( \frac{X}{N} + 1 - \frac{X}{N} \right)^2 \right] - 2 \frac{X(N-X)}{N^2} = X(N-X) \left[ 1 - \frac{2}{N^2} \right] \end{aligned}$$

Dove nel primo passaggio si è usata la proprietà distributiva notando che:

$$\begin{aligned} &(X-1)((N-X)+1) + (X+1)((N-X)-1) = \\ &2X(N-X) + (X-N+X-1+N-X-X-1). \end{aligned}$$



# Bibliografia

- [1] R. A. Blythe and A. J. McKane, "Stochastic models of evolution in genetics, ecology and linguistics," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 07, p. P07018, 2007.
- [2] S. Azaele, S. Suweis, J. Grilli, I. Volkov, J. R. Banavar, and A. Maritan, "Statistical mechanics of ecological systems: Neutral theory and beyond," *Reviews of Modern Physics*, vol. 88, no. 3, p. 035003, 2016.
- [3] S. Wright, "Evolution in mendelian populations," *Genetics*, vol. 16, no. 2, p. 97, 1931.
- [4] R. A. Fisher, "The genetical theory of natural selection, 272 pp," 1930.
- [5] P. A. P. Moran, "Random processes in genetics," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 54, pp. 60–71, Cambridge University Press, 1958.
- [6] S. P. Hubbell, "The unified neutral theory of biodiversity and biogeography (mpb-32)(monographs in population biology)," 2001.
- [7] I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan, "Patterns of relative species abundance in rainforests and coral reefs," *Nature*, vol. 450, no. 7166, p. 45, 2007.
- [8] I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan, "Neutral theory and relative species abundance in ecology," *Nature*, vol. 424, no. 6952, p. 1035, 2003.
- [9] W. Feller *et al.*, "Diffusion processes in genetics," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, The Regents of the University of California, 1951.
- [10] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge University Press, 2013.
- [11] G. A. Watterson, "Markov chains with absorbing states: a genetic example," *The Annals of Mathematical Statistics*, pp. 716–729, 1961.
- [12] J. Gani, "On the stochastic matrix in a genetic model of moran," *Biometrika*, vol. 48, no. 1/2, pp. 203–206, 1961.
- [13] C. Gardiner, *Stochastic methods*, vol. 4. springer Berlin, 2009.

- [14] J. Felsenstein, "The rate of loss of multiple alleles in finite haploid populations," *Theoretical Population Biology*, vol. 2, no. 4, pp. 391–403, 1971.
- [15] C. Cannings, "The latent roots of certain markov chains arising in genetics: a new approach, i. haploid models," *Advances in Applied Probability*, vol. 6, no. 2, pp. 260–290, 1974.
- [16] K. Gladstien, "The characteristic values and vectors for a class of stochastic matrices arising in genetics," *SIAM Journal on Applied Mathematics*, vol. 34, no. 4, pp. 630–642, 1978.
- [17] W. J. Ewens, *Mathematical Population Genetics 1: Theoretical Introduction*, vol. 27. Springer Science & Business Media, 2004.