Diffusion approximation of a multilocus model with assortative mating

S. Lemaire, Université Paris-Sud

October 19, 2011

Joint work with A. Etheridge, Oxford University
`http://hal.archives-ouvertes.fr/hal-00560285`
To appear in Electron. J. Probab.

Introduction

Question : What are the consequences of assortative mating on the evolution of the genetic composition of a population?

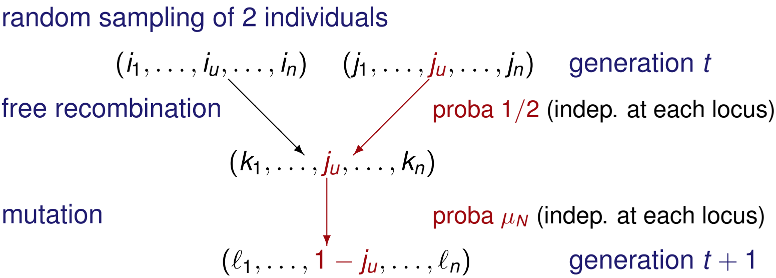
Outline :

1. Serva & Peliti (1991) random mating model and Higgs & Derrida (1992) modifications
2. Description of a Moran model with assortative mating
3. Approximation by a diffusion
4. Properties of the limiting diffusion
5. Extensions of the model

Random mating model (Serva & Peliti, 1991)

Wright-Fisher type model

- Population of size N
- Diallelic model with n loci: an individual is represented by its type $\boldsymbol{i} = (i_1, \ldots, i_n) \in \{0,1\}^n$
- Random pairing: each individual α has 2 parents α_1 and α_2 randomly chosen from the previous generation.



random sampling of 2 individuals

$(i_1, \ldots, i_u, \ldots, i_n)$ $(j_1, \ldots, j_u, \ldots, j_n)$ generation t

free recombination

proba $1/2$ (indep. at each locus)

$(k_1, \ldots, j_u, \ldots, k_n)$

mutation

proba μ_N (indep. at each locus)

$(\ell_1, \ldots, 1 - j_u, \ldots, \ell_n)$ generation $t+1$

Genetic distance (Serva & Peliti, 1991)

Genetic distance between 2 individuals of type ***i*** and ***j***

$$d_n(\boldsymbol{i},\boldsymbol{j}) = \sum_{k=1}^{n} |i_k - j_k| \text{ (Hamming distance)}$$

Genetic distance (Serva & Peliti, 1991)

Genetic distance between 2 individuals of type ***i*** and ***j***

$$d_n(\boldsymbol{i},\boldsymbol{j}) = \sum_{k=1}^{n} |i_k - j_k| \text{ (Hamming distance)}$$

$D_{t,n,N}$: Hamming distance between 2 individuals randomly sampled in the population at time t.

Genetic distance (Serva & Peliti, 1991)

Genetic distance between 2 individuals of type ***i*** and ***j***

$$d_n(\boldsymbol{i},\boldsymbol{j}) = \sum_{k=1}^{n} |i_k - j_k| \text{ (Hamming distance)}$$

$D_{t,n,N}$: Hamming distance between 2 individuals randomly sampled in the population at time t.

Asymptotic behavior.
Assume that $\mu_N = \frac{\mu}{N}$. Let t, n and N converge to $+\infty$.

- $E(\frac{1}{n}D_{t,n,N}) \to \lambda := \frac{2\mu}{1+4\mu}$
- $\mathrm{Var}(\frac{1}{n}D_{t,n,N}) \to 0$.

Genetic distance (Serva & Peliti, 1991)

Genetic distance between 2 individuals of type $\boldsymbol{i}$ and $\boldsymbol{j}$

$$d_n(\boldsymbol{i},\boldsymbol{j}) = \sum_{k=1}^{n} |i_k - j_k| \text{ (Hamming distance)}$$

$D_{t,n,N}$: Hamming distance between 2 individuals randomly sampled in the population at time t.

Asymptotic behavior.
Assume that $\mu_N = \frac{\mu}{N}$. Let t, n and N converge to $+\infty$.

- $E(\frac{1}{n}D_{t,n,N}) \to \lambda := \frac{2\mu}{1+4\mu}$
- $\mathrm{Var}(\frac{1}{n}D_{t,n,N}) \to 0$.

► If reproduction is possible only between 2 individuals of type $\boldsymbol{i}$ and $\boldsymbol{j}$ so that $\frac{1}{n}d_n(\boldsymbol{i},\boldsymbol{j}) < \lambda_0$ and if $\lambda_0 < \lambda$ the population will never reach its natural equilibrium.

Species-Formation model (Higgs & Derrida, 1992)

Modification of the previous model:

- Population of size N
- Diallelic model with n loci: an individual is represented by its type $\boldsymbol{i} = (i_1, \ldots, i_n) \in \{0, 1\}^n$
- Random pairing: each individual α has 2 parents α_1 and α_2 randomly chosen from the previous generation.
- Free recombination: each allele of α is inherited from either α_1 or α_2 with probability $(1/2, 1/2)$ independently at each locus.
- a mutation occurs with probability μ_N independently at each locus.

Species-Formation model (Higgs & Derrida, 1992)

- Population of size N
- Diallelic model with n loci: an individual is represented by its type $\boldsymbol{i} = (i_1, \ldots, i_n) \in \{0, 1\}^n$
- assortative pairing: each individual α has 2 parents α_1 and α_2 chosen from the previous generation
  - α_1 is chosen at random
  - α_2 is chosen only from those individuals having a Hamming distance smaller than $n\lambda_0$.

  If impossible, α_1 is discarded.
- Free recombination: each allele of α is inherited from either α_1 or α_2 with probability $(1/2, 1/2)$ independently at each locus.
- a mutation occurs with probability μ_N independently at each locus.

Species-Formation model (Higgs & Derrida, 1992)

The introduction of a cut off value λ_0 leads to a partition into isolated subgroups they called 'species'.

Species: a member of a species A have

- Hamming distance smaller than $n\lambda_0$ with at least one member of A,
- Hamming distance greater than or equal to $n\lambda_0$ with each member of another species.

Species-Formation model (Higgs & Derrida, 1992)

The introduction of a cut off value λ_0 leads to a partition into isolated subgroups they called 'species'.

Species: a member of a species A have

- Hamming distance smaller than $n\lambda_0$ with at least one member of A,
- Hamming distance greater than or equal to $n\lambda_0$ with each member of another species.

Observations by simulations

- A continuous process of divisions into 'species' and extinctions.

Assortative mating Moran model

Modifications:

- Moran model: at each step, 2 individuals α_1 and α_2 are chosen in the population; α is removed and is replaced by an offspring of (α_1, α_2).

Assortative mating Moran model

Modifications:

- Moran model: at each step, 2 individuals α_1 and α_2 are chosen in the population; α is removed and is replaced by an offspring of (α_1, α_2).
- Assortment: let $s_{N,0}, \ldots, s_{N,n} \geq 0$.
  - α_1 is chosen at random. Let $\boldsymbol{i}$ denote its type.
  - each individual of type $\boldsymbol{j}$ have a probability $\propto \left(1 + s_{N,d_n(\boldsymbol{i},\boldsymbol{j})}\right)$ to be α_2.

Assortative mating Moran model

Modifications:

- **Moran model:** at each step, 2 individuals α_1 and α_2 are chosen in the population; α is removed and is replaced by an offspring of (α_1, α_2).
- **Assortment:** let $s_{N,0}, \dots, s_{N,n} \geq 0$.
  - α_1 is chosen at random. Let $\boldsymbol{i}$ denote its type.
  - each individual of type $\boldsymbol{j}$ have a probability $\propto \big(1 + s_{N,d_n(\boldsymbol{i},\boldsymbol{j})}\big)$ to be α_2.

▶ positive assortative mating if $s_{N,0} \geq \dots \geq s_{N,n}$.

Assortative mating Moran model

Modifications:

- **Moran model:** at each step, 2 individuals α_1 and α_2 are chosen in the population; α is removed and is replaced by an offspring of (α_1, α_2).
- **Assortment:** let $s_{N,0}, \ldots, s_{N,n} \geq 0$.
  - α_1 is chosen at random. Let $\boldsymbol{i}$ denote its type.
  - each individual of type $\boldsymbol{j}$ have a probability $\propto \left(1 + s_{N,d_n(\boldsymbol{i},\boldsymbol{j})}\right)$ to be α_2.
  - ► positive assortative mating if $s_{N,0} \geq \ldots \geq s_{N,n}$.

Other features are unchanged:

- free recombination
- mutation with probability μ_N independently at each locus.

Transition probabilities of the Markov chain $(Z_t^{(N)})_t$

$Z_t^{(N)}(\boldsymbol{k})$: proportion of individuals of type $\boldsymbol{k} \in \{0,1\}^n$ at time t.
$Z_t^{(N)}$:=$\{Z_t^{(N)}(\boldsymbol{k}),\ \boldsymbol{k} \in \{0,1\}^n\}$ composition of the population at time t

- $(Z_t^{(N)})_t$ is a Markov chain

Assume that $Z_t^{(N)} = z$.
In one step, one individual of type $\boldsymbol{i}$ is removed and replaced by one individual of type $\boldsymbol{j} \neq \boldsymbol{i}$ with probability:

$$\sum_{\boldsymbol{k},\ell\in\{0,1\}^n} z(\boldsymbol{i}) \frac{(1+s_{N,d(\boldsymbol{i},\boldsymbol{k})})z(\boldsymbol{k})}{1+\sum_{\boldsymbol{h}} s_{N,d(\boldsymbol{i},\boldsymbol{h})}z(\boldsymbol{h})} 2^{-d(\boldsymbol{i},\boldsymbol{k})} \mathbf{1}_{\{\ell_u\in\{i_u,k_u\}\forall u\}} \mu_N^{d(\boldsymbol{\ell},\boldsymbol{j})}(1-\mu_N)^{n-d(\boldsymbol{\ell},\boldsymbol{j})}.$$

Transition probabilities of the Markov chain $(Z_t^{(N)})_t$

$Z_t^{(N)}(\boldsymbol{k})$: proportion of individuals of type $\boldsymbol{k} \in \{0,1\}^n$ at time t.
$Z_t^{(N)}$:=$\{Z_t^{(N)}(\boldsymbol{k}),\ \boldsymbol{k} \in \{0,1\}^n\}$ composition of the population at time t

- $(Z_t^{(N)})_t$ is a Markov chain

Assume that $Z_t^{(N)} = z$.
In one step, one individual of type $\boldsymbol{i}$ is removed and replaced by one individual of type $\boldsymbol{j} \neq \boldsymbol{i}$ with probability:

$$\sum_{\boldsymbol{k},\boldsymbol{\ell} \in \{0,1\}^n} z(\boldsymbol{i}) \frac{(1+s_{N,d(\boldsymbol{i},\boldsymbol{k})})z(\boldsymbol{k})}{1+\sum_{\boldsymbol{h}} s_{N,d(\boldsymbol{i},\boldsymbol{h})} z(\boldsymbol{h})} 2^{-d(\boldsymbol{i},\boldsymbol{k})} \mathbf{1}_{\{\ell_u \in \{i_u,k_u\} \forall u\}} \mu_N^{d(\boldsymbol{\ell},\boldsymbol{j})} (1-\mu_N)^{n-d(\boldsymbol{\ell},\boldsymbol{j})}.$$

Transition probabilities of the Markov chain $(Z_t^{(N)})_t$

$Z_t^{(N)}(\boldsymbol{k})$: proportion of individuals of type $\boldsymbol{k} \in \{0,1\}^n$ at time t.
$Z_t^{(N)}$:=$\{Z_t^{(N)}(\boldsymbol{k}),\ \boldsymbol{k} \in \{0,1\}^n\}$ composition of the population at time t

- $(Z_t^{(N)})_t$ is a Markov chain

Assume that $Z_t^{(N)} = z$.
In one step, one individual of type $\boldsymbol{i}$ is removed and replaced by one individual of type $\boldsymbol{j} \neq \boldsymbol{i}$ with probability:

$$\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\{0,1\}^n} z(\boldsymbol{i}) \frac{(1+s_{N,d(\boldsymbol{i},\boldsymbol{k})})z(\boldsymbol{k})}{1+\sum_{\boldsymbol{h}} s_{N,d(\boldsymbol{i},\boldsymbol{h})} z(\boldsymbol{h})} 2^{-d(\boldsymbol{i},\boldsymbol{k})} \mathbf{1}_{\{\ell_u \in \{i_u,k_u\} \forall u\}} \mu_N^{d(\boldsymbol{\ell},\boldsymbol{j})} (1-\mu_N)^{n-d(\boldsymbol{\ell},\boldsymbol{j})}.$$

Transition probabilities of the Markov chain $(Z_t^{(N)})_t$

$Z_t^{(N)}(\boldsymbol{k})$: proportion of individuals of type $\boldsymbol{k} \in \{0,1\}^n$ at time t.
$Z_t^{(N)}$:=$\{Z_t^{(N)}(\boldsymbol{k}),\ \boldsymbol{k} \in \{0,1\}^n\}$ composition of the population at time t

- $(Z_t^{(N)})_t$ is a Markov chain

Assume that $Z_t^{(N)} = z$.
In one step, one individual of type $\boldsymbol{i}$ is removed and replaced by one individual of type $\boldsymbol{j} \neq \boldsymbol{i}$ with probability:

$$\sum_{\boldsymbol{k},\ell\in\{0,1\}^n} z(\boldsymbol{i})\frac{(1+s_{N,d(\boldsymbol{i},\boldsymbol{k})})z(\boldsymbol{k})}{1+\sum_{\boldsymbol{h}} s_{N,d(\boldsymbol{i},\boldsymbol{h})}z(\boldsymbol{h})}2^{-d(\boldsymbol{i},\boldsymbol{k})}\mathbf{1}_{\{\ell_u\in\{i_u,k_u\}\forall u\}}\mu_N^{d(\ell,\boldsymbol{j})}(1-\mu_N)^{n-d(\ell,\boldsymbol{j})}.$$

Transition probabilities of the Markov chain $(Z_t^{(N)})_t$

$Z_t^{(N)}(\boldsymbol{k})$: proportion of individuals of type $\boldsymbol{k} \in \{0,1\}^n$ at time t.
$Z_t^{(N)}$:=$\{Z_t^{(N)}(\boldsymbol{k}),\ \boldsymbol{k} \in \{0,1\}^n\}$ composition of the population at time t

- $(Z_t^{(N)})_t$ is a Markov chain

Assume that $Z_t^{(N)} = z$.
In one step, one individual of type $\boldsymbol{i}$ is removed and replaced by one individual of type $\boldsymbol{j} \neq \boldsymbol{i}$ with probability:

$$\sum_{\boldsymbol{k},\ell\in\{0,1\}^n} z(\boldsymbol{i})\frac{(1+s_{N,d(\boldsymbol{i},\boldsymbol{k})})z(\boldsymbol{k})}{1+\sum_{\boldsymbol{h}} s_{N,d(\boldsymbol{i},\boldsymbol{h})}z(\boldsymbol{h})}2^{-d(\boldsymbol{i},\boldsymbol{k})}\mathbf{1}_{\{\ell_u\in\{i_u,k_u\}\forall u\}}\mu_N^{d(\boldsymbol{\ell},\boldsymbol{j})}(1-\mu_N)^{n-d(\boldsymbol{\ell},\boldsymbol{j})}.$$

Diffusion approximation

Hyp: $\mu_N = \frac{\mu}{N}$ and $s_{N,\ell} = \frac{s_\ell}{N}\ \forall \ell \in \{0, \dots, n\}$

Diffusion approximation

Hyp: $\mu_N = \frac{\mu}{N}$ and $s_{N,\ell} = \frac{s_\ell}{N}$ $\forall \ell \in \{0,\dots,n\}$

The one-locus case Let $Z_t^{(N)}$ be the proportion of individuals with 0-allele at time $t \in \mathbb{N}$.

- $E_z(Z_1^{(N)} - z) = \frac{1}{N^2}\Big((1-2z)\mu + z(1-z)(1/2-z)(s_1-s_0)\Big) + O(1/N^3)$
- $E_z((Z_1^{(N)} - z)^2) = \frac{1}{N^2} z(1-z) + O(1/N^3)$
- $E_z((Z_1^{(N)} - z)^4) = O(1/N^4)$

Diffusion approximation

If $Z_0^{(N)} \Rightarrow Z_0$, then $(Z^{(N)}_{[N^2 t]})_t \overset{D_{[0,1]}([0,\infty))}{\Longrightarrow} Z$ diffusion with generator $\mathcal{G}_1$:
$\mathcal{G}_1 = \frac{1}{2}x(1-x)\frac{d^2}{dx^2} + ((1-2x)\mu + (s_1-s_0)(1/2-x)x(1-x))\frac{d}{dx}$
$\mathcal{D}(\mathcal{G}_1) = C^2([0,1])$.

Wright-Fisher diffusion with selection and mutation

$$\mathcal{G}_1 = \tfrac{1}{2}x(1-x)\frac{d^2}{dx^2} + \Big((1-2x)\mu + (s_1 - s_0)(1/2 - x)x(1-x)\Big)\tfrac{d}{dx}$$
$$\mathcal{D}(\mathcal{G}_1) = C^2([0,1]).$$

Another one-locus discrete model

- Diploid population of size N (identified as a $2N$ haploid population)
- Wright-Fisher model with
  - viability coefficients:

$$w_{0,1}^{(N)} = w_{1,0}^{(N)} = 1 + \frac{s_1}{4N} \quad \text{and} \quad w_{0,0}^{(N)} = w_{1,1}^{(N)} = 1 + \frac{s_0}{4N}.$$

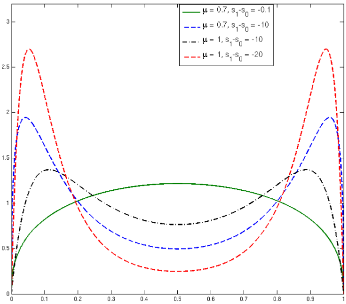
  - mutation probability: $\frac{\mu}{2N}$.

► Approximation of the 0-allele frequency at time $[2Nt]$ by a diffusion with generator $\mathcal{G}_1$.

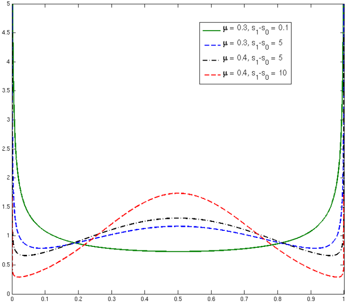
Properties of the one-locus diffusion

$$\mathcal{G}_1 = \frac{1}{2}x(1-x)\frac{d^2}{dx^2} + ((1-2x)\mu + (s_1-s_0)(1/2-x)x(1-x))\frac{d}{dx}$$

- If $\mu > 0$, the diffusion has a reversible measure with density $g_{\mu,s}(x) = C_{\mu,s}\big(x(1-x)\big)^{2\mu-1}\exp((s_1-s_0)x(1-x))$.
- If $0 < \mu < 1/2$ then 0 and 1 are regular boundaries,
- If $\mu \geq 1/2$ then 0 and 1 are entrance boundaries.



$\mu > 1/2$: global maximum at $1/2$ iff $s_1 - s_0 \geq 4 - 8\mu$.



$0 < \mu < 1/2$: global minimum at $1/2$ iff $s_1 - s_0 \leq 4 - 8\mu$.

The linkage disequilibrium

Hyp: free recombination , $\mu_N = \frac{\mu}{N}$ and $s_{N,\ell} = \frac{s_\ell}{N}$ $\forall \ell \in \{0, \dots, n\}$.

- separation of timescales.

The linkage disequilibrium

Hyp: free recombination , $\mu_N = \frac{\mu}{N}$ and $s_{N,\ell} = \frac{s_\ell}{N}$ $\forall \ell \in \{0, \dots, n\}$.

- separation of timescales.

Example in the case of two loci:

If $Z_t^{(N)} = \begin{pmatrix} Z_t^{(N)}(00) \\ Z_t^{(N)}(10) \\ Z_t^{(N)}(01) \\ Z_t^{(N)}(11) \end{pmatrix}$ then $E_z(Z_1^{(N)} - z) = \frac{1}{2N} D(z) \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} + O(N^{-2})$

where $D(z) = z(01)z(10) - z(11)z(00)$ (linkage disequilibrium)

The linkage disequilibrium

Hyp: free recombination , $\mu_N = \frac{\mu}{N}$ and $s_{N,\ell} = \frac{s_\ell}{N}$ $\forall \ell \in \{0, \dots, n\}$.

- separation of timescales.

Example in the case of two loci:

If $Z_t^{(N)} = \begin{pmatrix} Z_t^{(N)}(00) \\ Z_t^{(N)}(10) \\ Z_t^{(N)}(01) \\ Z_t^{(N)}(11) \end{pmatrix}$ then $E_z(Z_1^{(N)} - z) = \frac{1}{2N} D(z) \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} + O(N^{-2})$

where $D(z) = z(01)z(10) - z(11)z(00)$ (linkage disequilibrium)

0-allele frequencies at each locus:

$x(1) = z(00) + z(01)$ and $x(2) = z(00) + z(10)$

The linkage disequilibrium

Hyp: free recombination , $\mu_N = \frac{\mu}{N}$ and $s_{N,\ell} = \frac{s_\ell}{N}$ $\forall \ell \in \{0, \dots, n\}$.

- separation of timescales.

Example in the case of two loci:

If $Z_t^{(N)} = \begin{pmatrix} Z_t^{(N)}(00) \\ Z_t^{(N)}(10) \\ Z_t^{(N)}(01) \\ Z_t^{(N)}(11) \end{pmatrix}$ then $E_z(Z_1^{(N)} - z) = \frac{1}{2N} D(z) \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} + O(N^{-2})$

where $D(z) = z(01)z(10) - z(11)z(00)$ (linkage disequilibrium)

0-allele frequencies at each locus:

$x(1) = z(00) + z(01)$ and $x(2) = z(00) + z(10)$

- $D(z) = x(1)x(2) - z(00)$

The two-locus case

In first approximation, the motion is governed by:

$$\frac{d}{dt}z(t) = \frac{1}{2}D(z(t))\begin{pmatrix}1\\-1\\-1\\1\end{pmatrix}$$

- $\frac{d}{dt}D(z(t)) = -\frac{1}{2}D(z(t))$.

The two-locus case

In first approximation, the motion is governed by:

$$\frac{d}{dt}z(t) = \frac{1}{2}D(z(t))\begin{pmatrix}1\\-1\\-1\\1\end{pmatrix}$$

- ► $\frac{d}{dt}D(z(t)) = -\frac{1}{2}D(z(t))$.

► The genotypic frequencies rapidly converges to a product distribution.

► At a slower rate, the 0-allelic frequencies converge to a 2-dimensional diffusion.

The two-locus case

In first approximation, the motion is governed by:

$$\frac{d}{dt}z(t) = \frac{1}{2}D(z(t))\begin{pmatrix}1\\-1\\-1\\1\end{pmatrix}$$

- $\frac{d}{dt}D(z(t)) = -\frac{1}{2}D(z(t))$.

- The genotypic frequencies rapidly converges to a product distribution.
- At a slower rate, the 0-allelic frequencies converge to a 2-dimensional diffusion.

Proof: application of the diffusion approximation theorem stated by Ethier & Nagylaki (Adv. Appl. Prob. 80) to:

$$\begin{cases} X_t^{(N)}(1) = Z_t^{(N)}(00) + Z_t^{(N)}(10) \\ X_t^{(N)}(2) = Z_t^{(N)}(00) + Z_t^{(N)}(01) \\ Y_t^{(N)} = X_t^{(N)}(1)X_t^{(N)}(2) - Z_t^{(N)}(00) \end{cases}$$

Convergence theorem in the two-locus case

New coordinates: $x(1) = z(00) + z(10)$, $x(2) = z(00) + z(01)$ and $y = x(1)x(2) - z(00)$.

1. Conditions on $Y_t^{(N)}$: uniformly on z.
   - $NE_z(Y_1^{(N)} - y) = c(x, y) + o(1)$,
   - $NE_z((Y_1^{(N)} - y)^2) = o(1)$,

   where c is a C^2 function, s. t. $c(x, 0) = 0$ and the ODE

   $$\frac{d}{dt}u(t, x, y) = c(x, u(t, x, y)), \quad u(0, x, y) = y.$$

   have a solution for every (t, x, y) that converges uniformly to 0 as $t \to +\infty$

Convergence theorem in the two-locus case

New coordinates: $x(1) = z(00) + z(10)$, $x(2) = z(00) + z(01)$ and $y = x(1)x(2) - z(00)$.

1. Conditions on $Y_t^{(N)}$: uniformly on z.
   - $NE_z(Y_1^{(N)} - y) = c(x, y) + o(1)$,
   - $NE_z((Y_1^{(N)} - y)^2) = o(1)$,

   where c is a C^2 function, s. t. $c(x, 0) = 0$ and the ODE

   $$\frac{d}{dt}u(t, x, y) = c(x, u(t, x, y)), \quad u(0, x, y) = y.$$

   have a solution for every (t, x, y) that converges uniformly to 0 as $t \to +\infty$

   Here $c(x, y) = -1/2y$

Convergence theorem in the two-locus case

New coordinates: $x(1) = z(00) + z(10)$, $x(2) = z(00) + z(01)$ and $y = x(1)x(2) - z(00)$.

1. Conditions on $Y_t^{(N)}$: ok
2. Conditions on $X_t^{(N)} = (X_t^{(N)}(1), X_t^{(N)}(2))$: uniformly on z, for $i, j \in \{1, 2\}$
    - $N^2 E_z(X_1^{(N)}(i) - x(i)) = b_i(x, y) + o(1)$,
    - $N^2 E_z((X_1^{(N)}(i) - x(i))(X_1^{(N)}(j) - x(j))) = a_{i,j}(x, y) + o(1)$,
    - $N^4 E_z((X_1^{(N)}(i) - x(i))^4) = o(1)$.

   where b_i and $a_{i,j}$ are continuous functions.

Convergence theorem in the two-locus case

New coordinates: $x(1) = z(00) + z(10)$, $x(2) = z(00) + z(01)$ and $y = x(1)x(2) - z(00)$.

1. Conditions on $Y_t^{(N)}$: ok
2. Conditions on $X_t^{(N)} = (X_t^{(N)}(1), X_t^{(N)}(2))$: uniformly on z, for $i, j \in \{1, 2\}$
   - $N^2 E_z(X_1^{(N)}(i) - x(i)) = b_i(x, y) + o(1)$,
   - $N^2 E_z((X_1^{(N)}(i) - x(i))(X_1^{(N)}(j) - x(j))) = a_{i,j}(x, y) + o(1)$,
   - $N^4 E_z((X_1^{(N)}(i) - x(i))^4) = o(1)$.

   where b_i and $a_{i,j}$ are continuous functions.
3. Condition on the generator:

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{n} a_{i,j}(x,0)\frac{\partial^2}{\partial_{x_i}\partial_{x_j}} + \sum_{i=1}^{n} b_i(x,0)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{L}) = C^2([0,1]^2),$$

generates a strongly continuous semigroup on $C([0,1]^2)$ corresponding to a diffusion process X.

Convergence theorem in the two-locus case

New coordinates: $x(1) = z(00) + z(10)$, $x(2) = z(00) + z(01)$ and $y = x(1)x(2) - z(00)$.

1. Conditions on $Y_t^{(N)}$: ok
2. Conditions on $X_t^{(N)} = (X_t^{(N)}(1), X_t^{(N)}(2))$: ok
3. Condition on the generator:

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{n} a_{i,j}(x,0)\frac{\partial^2}{\partial_{x_i}\partial_{x_j}} + \sum_{i=1}^{n} b_i(x,0)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{L}) = C^2([0,1]^2),$$

generates a strongly continuous semigroup on $C([0,1]^2)$ corresponding to a diffusion process X.
► application of a result of Cerrai & Clément (Bull.Sc. Math., 2004)

Convergence theorem in the two-locus case

New coordinates: $x(1) = z(00) + z(10)$, $x(2) = z(00) + z(01)$ and $y = x(1)x(2) - z(00)$.

1. Conditions on $Y_t^{(N)}$: ok
2. Conditions on $X_t^{(N)} = (X_t^{(N)}(1), X_t^{(N)}(2))$: ok
3. Condition on the generator: ok

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{n} a_{i,j}(x,0)\frac{\partial^2}{\partial_{x_i}\partial_{x_j}} + \sum_{i=1}^{n} b_i(x,0)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{L}) = C^2([0,1]^2),$$

generates a strongly continuous semigroup on $C([0,1]^2)$ corresponding to a diffusion process X.

Convergence results:

(i) If $(t_N)_N \to +\infty$, $Y_{[Nt_N]}^{(N)} \Rightarrow 0$.

(ii) If $X_0^{(N)} \Rightarrow X_0$ then $\{X_{[N^2 t]}^{(N)}, t \geq 0\} \overset{D}{\Longrightarrow} X$,

General case of n loci

Notations: for a nonempty set L of $[\![1, n]\!]$,

- $X_t^{(N)}(L) = \sum_{\boldsymbol{j} \in \{0,1\}^n,\ \boldsymbol{j}_L \equiv 0} Z_t^{(N)}(\boldsymbol{j})$: frequency of individuals having the allele 0 on each locus $i \in L$ at time t,
- $Y_t^{(N)}(L) = \prod_{i \in L} X_t^{(N)}(\{i\}) - X_t^{(N)}(L)$: linkage disequilibrium for the set of loci L at time t.

$Z_t^{(N)}$ is a product distribution iff $Y_t^{(N)}(L) = 0\ \forall L \subset [\![1, n]\!]$ s. t. $|L| \geq 2$

General case of n loci

Notations: for a nonempty set L of $[\![1, n]\!]$,

- $X_t^{(N)}(L) = \sum_{\boldsymbol{j}\in\{0,1\}^n,\ \boldsymbol{j}_L\equiv 0} Z_t^{(N)}(\boldsymbol{j})$: frequency of individuals having the allele 0 on each locus $i \in L$ at time t,
- $Y_t^{(N)}(L) = \prod_{i\in L} X_t^{(N)}(\{i\}) - X_t^{(N)}(L)$: linkage disequilibrium for the set of loci L at time t.

Results

- If $(t_N)_N \to +\infty$ then $\{Y_{[Nt_N]}^{(N)}(L), |L| \geq 2\} \Rightarrow 0$
- Let $X_t^{(N)} = (X_t^{(N)}(\{1\}), ..., X_t^{(N)}(\{n\}))$.
  If $X_0^{(N)} \Rightarrow X_0$ then $\{X_{[N^2 t]}^{(N)}, t \geq 0\} \overset{D}{\Longrightarrow} X$ where X is a diffusion on $[0,1]^n$ with generator

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n).$$

Generator of the n-locus diffusion

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i,x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

with $b_i(x) = (1-2x_i)\mu + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ and

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset [\![1,n]\!]\setminus\{i\},\\ |A|=k}} \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

Generator of the n-locus diffusion

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i,x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

with $b_i(x) = (1-2x_i)\mu + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ and

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset[\![1,n]\!]\setminus\{i\},\\ |A|=k}}\prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

Behaviour at the boundaries

$0 < \mu < 1/2$: 0 and 1 are attainable for each coordinate of the diffusion.

$$\mathbb{P}[\exists t > 0,\ X_t(i) = 0] > 0 \text{ and } \mathbb{P}[\exists t > 0,\ X_t(i) = 1] > 0 \quad \forall i \in [\![1,n]\!].$$

$\mu \geq 1/2$: 0 and 1 are inaccessible for each coordinate of the diffusion.

$$\mathbb{P}[\exists t > 0,\ X_t(i) \in \{0,1\}] = 0 \text{ and } \mathbb{P}[\lim_{t\to+\infty} X_t(i) \in \{0,1\}] = 0 \quad \forall i \in [\![1,n]\!].$$

Generator of the n-locus diffusion

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i,x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

with $b_i(x) = (1-2x_i)\mu + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ and

$$P_{i,s}(x) = \sum_{k=0}^{n-1} (s_{k+1} - s_k) \sum_{\substack{A\subset [\![1,n]\!]\setminus\{i\},\\ |A|=k}} \prod_{j\in A} (2x_j(1-x_j)) \prod_{\ell\in [\![1,n]\!]\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell)).$$

Properties

- $P_{i,s}(x)$ is a symmetric polynomial function of the $n-1$ variables $\rho_j = x_j(1-x_j)$ for $j \neq i$.
- $P_{i,s}(x)$ is an increasing function of $s_{k+1} - s_k$.
- The n coordinates of the diffusion evolve as n independent one-locus diffusions $\Leftrightarrow s_{\ell+1} - s_\ell = s_1 - s_0 \ \forall \ell$.

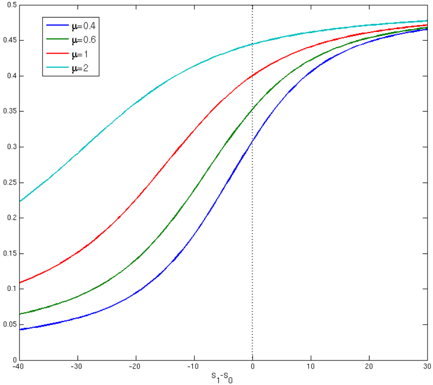
Genetic distance in the independent case

Hyp: $s_{\ell+1} - s_\ell = s_1 - s_0 \quad \forall \ell \in [\![1, n]\!]$

Genetic distance in the independent case

Hyp: $s_{\ell+1} - s_\ell = s_1 - s_0 \quad \forall \ell \in [\![1, n]\!]$

► In the independent case, under the stationary measure, the Hamming distance D_n between two individuals randomly chosen follows the binomial distribution $\mathcal{B}(n, h(s_1 - s_0))$ where h is an increasing function and $h(0) = \frac{2\mu}{1+4\mu}$.



Reversible measure of the *n*-locus diffusion

If $\mu > 0$, the diffusion has a reversible measure with density:

$$g_{n,\mu,s}(x) = C \exp\left((2\mu - 1) \sum_{i=1}^{n} \ln(x_i(1 - x_i)) + H_n(x) \right)$$

$$\text{with } H_n(x) = 1/2 \sum_{\ell=1}^{n} (s_\ell - s_0) \sum_{\substack{L \subset [\![1,n]\!], \\ |L| = \ell}} \prod_{j \in L} (2x_j(1 - x_j)) \prod_{k \notin L} (1 - 2x_k(1 - x_k)).$$

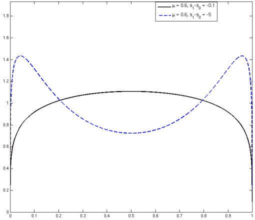
Reversible measure of the *n*-locus diffusion

If $\mu > 0$, the diffusion has a reversible measure with density:

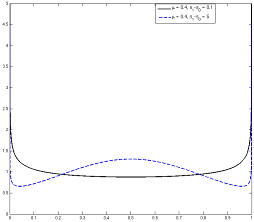
$$g_{n,\mu,s}(x) = C \exp\left((2\mu - 1)\sum_{i=1}^{n} \ln(x_i(1-x_i)) + H_n(x)\right)$$

$$\text{with } H_n(x) = 1/2 \sum_{\ell=1}^{n} (s_\ell - s_0) \sum_{\substack{L \subset [\![1,n]\!], \\ |L| = \ell}} \prod_{j \in L} (2x_j(1-x_j)) \prod_{k \notin L} (1 - 2x_k(1-x_k)).$$

Typical graphs of the density in the one-locus case depending on the sign of $s_1 - s_0 + 8\mu - 4$

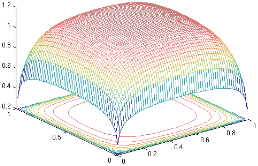


when $\mu > 1/2$



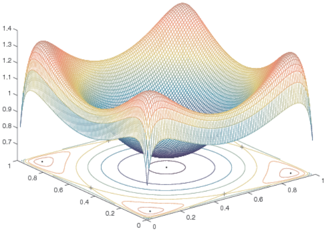
when $0 < \mu < 1/2$

Typical Graphs of the stationary density when $\mu > 1/2$



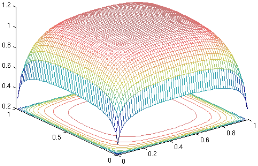
Density in the two-locus case 2 when $\mu = 0.6$, $s_1 - s_0 = -0.4$ and $s_2 - s_1 = -0.6$.

- Typical graph when $s_{\ell+1} - s_\ell \geq -(8\mu - 4)\ \forall \ell$.



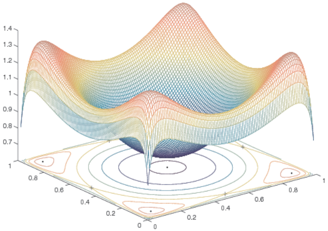
Density in the two-locus case 2 when $\mu = 0.6$, $s_1 - s_0 = -2$ and $s_2 - s_1 = -6$.

Typical Graphs of the stationary density when $\mu > 1/2$



Density in the two-locus case 2 when $\mu = 0.6$, $s_1 - s_0 = -0.4$ and $s_2 - s_1 = -0.6$.

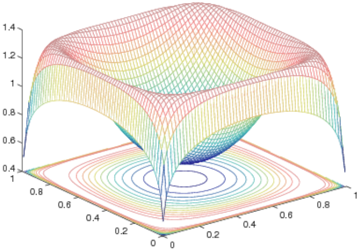
- Typical graph when $s_{\ell+1} - s_\ell \geq -(8\mu - 4)\ \forall \ell$.



Density in the two-locus case 2 when $\mu = 0.6$, $s_1 - s_0 = -2$ and $s_2 - s_1 = -6$.

- Typical graph if $2^{-(n-1)} \sum_{k=0}^{n-1} \binom{n-1}{k} (s_{k+1} - s_k) < 4 - 8\mu$, $s_n - s_{n-1} \leq s_{n-1} - s_{n-2} \leq \cdots \leq s_1 - s_0 \leq 0$ and $s_{n-1} - s_{n-2} < 0$.

A continuum of critical points



Density of the stationary measure in the 2-locus case when

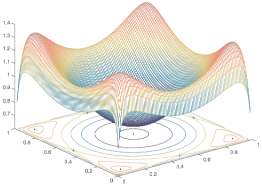
- $\mu = 0.6$,
- $s_1 - s_0 = 0$ et $s_2 - s_1 = -12$.

Description of the stationary density

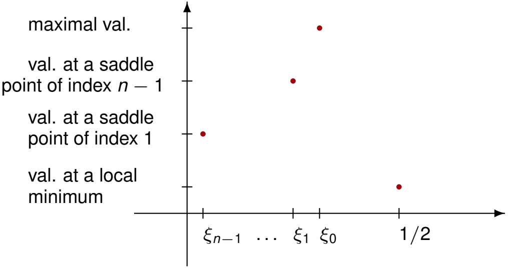
Hyp : $\mu > 1/2$ and $s_n - s_{n-1} \leq \cdots \leq s_1 - s_0$ strongly negative.

The critical points of the density $g_{n,\mu,s}$ in $]0, 1/2]^n$:

- global maximum at $(\xi_0, \dots, \xi_0)$ with $\xi_0 < 1/2$,
- saddle points of index $n - k$ at points having k coordinates equal to $1/2$ and $n - k$ coordinates equal to $\xi_k < 1/2$,
- local mimimum at $(1/2, \dots, 1/2)$



Values of the log-density at the critical points

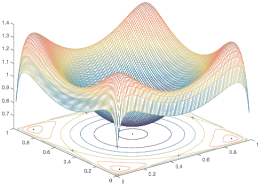


Description of the stationary density

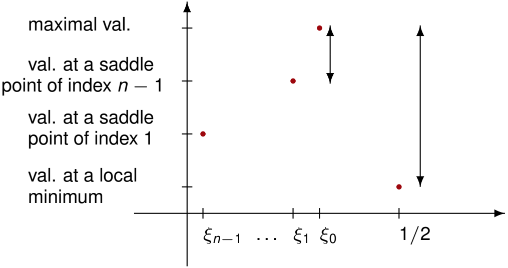
Hyp : $\mu > 1/2$ and $s_n - s_{n-1} \leq \cdots \leq s_1 - s_0$ strongly negative.

The critical points of the density $g_{n,\mu,s}$ in $]0, 1/2]^n$:

- global maximum at $(\xi_0, \ldots, \xi_0)$ with $\xi_0 < 1/2$,
- saddle points of index $n-k$ at points having k coordinates equal to $1/2$ and $n-k$ coordinates equal to $\xi_k < 1/2$,
- local mimimum at $(1/2, \ldots, 1/2)$



Values of the log-density at the critical points

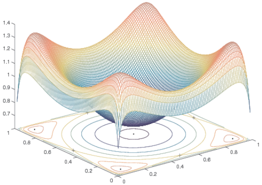


Description of the stationary density

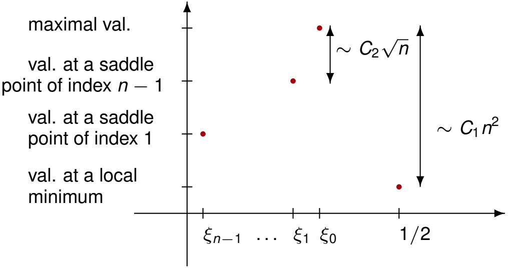
Hyp : $\mu > 1/2$ and $s_n - s_{n-1} \leq \cdots \leq s_1 - s_0$ strongly negative.

The critical points of the density $g_{n,\mu,s}$ in $]0, 1/2]^n$:

- global maximum at $(\xi_0, \ldots, \xi_0)$ with $\xi_0 < 1/2$,
- saddle points of index $n - k$ at points having k coordinates equal to $1/2$ and $n - k$ coordinates equal to $\xi_k < 1/2$,
- local mimimum at $(1/2, \ldots, 1/2)$



Values of the log-density at the critical points

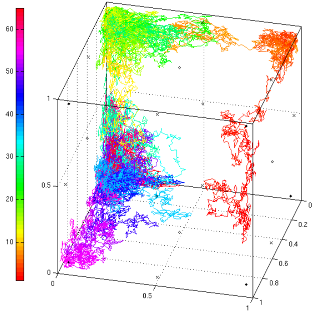


Ex : $s_k = -(bk + ck^2)$ with
$c > 0$, $b + c > 0$
$b + c + 4c(n-1) > 8\mu - 4$

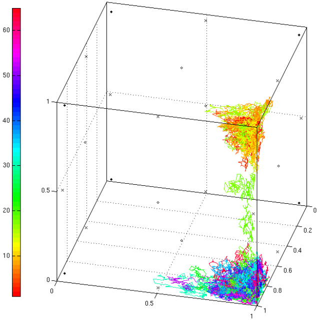
▶ $\xi_0 \sim C_3 n^{-1/2}$

Examples of trajectories in the 3-locus case

Population of size $N = 1000$, initialy of type $(0, 0, 0)$.
Evolution between iterations N^2 and $33N^2$ by step of N, change of color after $0.5N^2$ iterations



$\mu = 1$, $s_1 - s_0 = -20$,
$s_2 - s_1 = -40$, $s_3 - s_2 = -60$



$\mu = 1$, $s_1 - s_0 = -30$,
$s_2 - s_1 = -60$, $s_3 - s_2 = -90$

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

with drift

$$b_i(x) = \mu(1-2x_i) + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$$

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

with drift

$$b_i(x) = \mu(1-2x_i) + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$$

- Proba. of mutation: allele 0 $\xrightarrow{\frac{\mu_0}{N}}$ allele 1, allele 1 $\xrightarrow{\frac{\mu_1}{N}}$ allele 0.

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

with drift

$$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$$

- Proba. of mutation: allele 0 $\xrightarrow{\frac{\mu_0}{N}}$ allele 1, allele 1 $\xrightarrow{\frac{\mu_1}{N}}$ allele 0.

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$$b_i(x) = \mu_1(1-x_i) + \mu_0 x_i - (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$$

- Proba. of mutation: allele 0 $\xrightarrow{\frac{\mu_0}{N}}$ allele 1, allele 1 $\xrightarrow{\frac{\mu_1}{N}}$ allele 0.
- recombination law: for $L \subset [\![1,n]\!]$, let r_L be the probab. that the offspring inherits the genes of the 1st parent at loci $\ell \in L$ and the genes of the 2nd parent at loci $\ell \notin L$.

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$$b_i(x) = \mu_1(1-x_i) + \mu_0 x_i - (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$$

- Proba. of mutation: allele 0 $\xrightarrow{\frac{\mu_0}{N}}$ allele 1, allele 1 $\xrightarrow{\frac{\mu_1}{N}}$ allele 0.
- recombination law: for $L \subset [\![1,n]\!]$, let r_L be the probab. that the offspring inherits the genes of the 1st parent at loci $\ell \in L$ and the genes of the 2nd parent at loci $\ell \notin L$.

  Hyp: (i) $r_L = r_{[\![1,n]\!]\setminus L}$, $\forall L \subset [\![1,n]\!]$

  (ii) recombination between any pair of loci occurs with a positive probability (loose linkage)

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ with

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset[\![1,n]\!]\setminus\{i\},\\ |A|=k}}\prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ with

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset[\![1,n]\!]\setminus\{i\},\\ |A|=k}}\ \prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

Assortment coeff.: $s_{i,j}$ if i and j are the type of the 1st parent and the 2nd parent resp.

Hyp : $s_{i,j} = s_{j,i}$;
$s_{i,j}$ depends only on the loci at which i and j differ.

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ with

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset[\![1,n]\!]\setminus\{i\},\\ |A|=k}}\ \prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

Assortment coeff.: $s_{i,j}$ if i and j are the type of the 1st parent and the 2nd parent resp.

Hyp : $s_{i,j} = s_{j,i}$;

$s_{i,j}$ depends only on the loci at which i and j differ.

- 5 parameters in the 2-locus case: $s_{i,i} = s_{00,00}\ \forall i \in \{0,1\}^2$, $s_{10,00} = s_{11,01} = s_{00,10} = s_{01,11}$ and $s_{01,00} = s_{11,10} = s_{00,01} = s_{10,11}$, $s_{01,10} = s_{10,01}$ et $s_{11,00} = s_{00,11}$

Extension of the model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ with

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset[\![1,n]\!]\setminus\{i\},\\ |A|=k}}\prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

Assortment coeff.: $\boldsymbol{s_{i,j}}$ if ***i*** and ***j*** are the type of the 1st parent and the 2nd parent resp.

Hyp : $\boldsymbol{s_{i,j} = s_{j,i}}$;
$\boldsymbol{s_{i,j}}$ depends only on the loci at which ***i*** and ***j*** differ.

- 5 parameters in the 2-locus case: $s_{i,i} = s_{00,00}\ \forall i \in \{0,1\}^2$, $s_{10,00} = s_{11,01} = s_{00,10} = s_{01,11}$ and $s_{01,00} = s_{11,10} = s_{00,01} = s_{10,11}$, $s_{01,10} = s_{10,01}$ et $s_{11,00} = s_{00,11}$

$m_A(s)$: mean of $\boldsymbol{s_{i,j}}$ for types ***i*** and ***j*** that differ on the loci in A and are equal elsewhere.

Extension of the model: assortment

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$ with

$$P_{i,s}(x) = \sum_{k=0}^{n-1}(s_{k+1}-s_k)\sum_{\substack{A\subset[\![1,n]\!]\setminus\{i\},\\ |A|=k}}\ \prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell)).$$

- $P_{i,s}(x) = \sum_{A\subset[\![1,n]\!]\setminus\{i\}}(m_{A\cup\{i\}}(s) - m_A(s))\prod_{j\in A}(2x_j(1-x_j))\prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell))$

Assortment coeff.: $s_{i,j}$ if ***i*** and ***j*** are the type of the 1st parent and the 2nd parent resp.

Hyp : $s_{i,j} = s_{j,i}$
$s_{i,j}$ depends only on the loci at which ***i*** and ***j*** differ.

$m_A(s)$: mean of $s_{i,j}$ for types ***i*** and ***j*** that differ on the loci in A and are equal elsewhere.

Properties of the diffusion in the general model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$
with

$$P_{i,s}(x) = \sum_{A\subset [\![1,n]\!]\setminus\{i\}} (m_{A\cup\{i\}}(s) - m_A(s)) \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell))$$

$$m_A(s) = \frac{1}{2^n}\sum_{(i,j)\in F_A} s_{\boldsymbol{i},\boldsymbol{j}} \text{ where } F_A = \{(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{i}_u \neq \boldsymbol{j}_u \text{ iff } u \in A\}.$$

Properties

- $P_{i,s}(x)$ is a polynome in $\rho_j = x_j(1-x_j)$, $j \neq i$,

Properties of the diffusion in the general model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$
with

$$P_{i,s}(x) = \sum_{A\subset [\![1,n]\!]\setminus\{i\}} (m_{A\cup\{i\}}(s) - m_A(s)) \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in [\![1,n]\!]\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell))$$

$$m_A(s) = \frac{1}{2^n}\sum_{(i,j)\in F_A} s_{\boldsymbol{i},\boldsymbol{j}} \text{ where } F_A = \{(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{i}_u \neq \boldsymbol{j}_u \text{ iff } u \in A\}.$$

Properties

- $P_{i,s}(x)$ is a polynome in $\rho_j = x_j(1-x_j)$, $j \neq i$,
- $P_{i,s}(x)$ is an increasing function of $m_{A\cup\{i\}}(s) - m_A(s)$, $\forall A \not\ni i$,

Properties of the diffusion in the general model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$
with

$$P_{i,s}(x) = \sum_{A\subset [\![1,n]\!]\setminus\{i\}} (m_{A\cup\{i\}}(s) - m_A(s)) \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}}(1-2x_\ell(1-x_\ell))$$

$$m_A(s) = \frac{1}{2^n}\sum_{(i,j)\in F_A} s_{\boldsymbol{i},\boldsymbol{j}} \text{ where } F_A = \{(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{i}_u \neq \boldsymbol{j}_u \text{ iff } u \in A\}.$$

Properties

- $P_{i,s}(x)$ is a polynome in $\rho_j = x_j(1-x_j)$, $j \neq i$,
- $P_{i,s}(x)$ is an increasing function of $m_{A\cup\{i\}}(s) - m_A(s)$, $\forall A \not\ni i$,
- the n coordinates are independent iff $m_{A\cup\{i\}}(s) - m_A(s)$ does not depend on A.

Properties of the diffusion in the general model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$
with

$$P_{i,s}(x) = \sum_{A\subset[\![1,n]\!]\setminus\{i\}} (m_{A\cup\{i\}}(s) - m_A(s)) \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in[\![1,n]\!]\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell))$$

$$m_A(s) = \frac{1}{2^n}\sum_{(i,j)\in F_A} s_{\boldsymbol{i},\boldsymbol{j}} \text{ where } F_A = \{(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{i}_u \neq \boldsymbol{j}_u \text{ iff } u \in A\}.$$

Properties

- $P_{i,s}(x)$ is a polynome in $\rho_j = x_j(1-x_j)$, $j \neq i$,
- $P_{i,s}(x)$ is an increasing function of $m_{A\cup\{i\}}(s) - m_A(s)$, $\forall A \not\ni i$,
- the n coordinates are independent iff $m_{A\cup\{i\}}(s) - m_A(s)$ does not depend on A.
- if $\mu_0 > 0$ and $\mu_1 > 0$, the diffusion has a reversible measure.

Applications of the diffusion in the general model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$
with

$$P_{i,s}(x) = \sum_{A\subset \llbracket 1,n\rrbracket\setminus\{i\}} (m_{A\cup\{i\}}(s) - m_A(s)) \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in\llbracket 1,n\rrbracket\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell))$$

$$m_A(s) = \frac{1}{2^n}\sum_{(i,j)\in F_A} s_{\boldsymbol{i},\boldsymbol{j}} \text{ where } F_A = \{(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{i}_u \neq \boldsymbol{j}_u \text{ iff } u\in A\}.$$

Applications

- Assortment based on a phenotypic trait which is determined by n genes whose effects are similar and additive,

Applications of the diffusion in the general model

$$\mathcal{G}_n = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial^2_{x_i}} + \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial_{x_i}}, \quad \mathcal{D}(\mathcal{G}_n) = C^2([0,1]^n),$$

$$b_i(x) = \mu_1(1-x_i) - \mu_0 x_i + (1/2 - x_i)x_i(1-x_i)P_{i,s}(x)$$

with

$$P_{i,s}(x) = \sum_{A\subset [\![1,n]\!]\setminus\{i\}} (m_{A\cup\{i\}}(s) - m_A(s)) \prod_{j\in A}(2x_j(1-x_j)) \prod_{\ell\in [\![1,n]\!]\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell))$$

$$m_A(s) = \frac{1}{2^n}\sum_{(i,j)\in F_A} s_{\boldsymbol{i},\boldsymbol{j}} \text{ where } F_A = \{(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{i}_u \neq \boldsymbol{j}_u \text{ iff } u\in A\}.$$

Applications

- Assortment based on a phenotypic trait which is determined by n genes whose effects are similar and additive,
- Situation in which some loci have a greater influence on the mating choice than others.

Diffusion approximation for other discrete models

Assortating mating / viability selection:

- (Ethier & Nagylaki, 89) Two-locus Wright-Fisher model in a diploid population of size N with
  - viability selection (coeff. $1+\frac{s_{\boldsymbol{i},\boldsymbol{j}}}{4N}$ for a pair of gametes $(\boldsymbol{i},\boldsymbol{j})$);
  - low mutation rates (coeff. $\frac{\mu_0}{2N}$ and $\frac{\mu_1}{2N}$);
  - loose recombination (recombination fraction $c_N \to c$ s.t. $Nc_N \to +\infty$);

► same limiting diffusion

Diffusion approximation for other discrete models

Assortating mating / viability selection:

- (Ethier & Nagylaki, 89) Two-locus Wright-Fisher model in a diploid population of size N with
  - viability selection (coeff. $1 + \frac{s_{i,j}}{4N}$ for a pair of gametes $(\boldsymbol{i}, \boldsymbol{j})$);
  - low mutation rates (coeff. $\frac{\mu_0}{2N}$ and $\frac{\mu_1}{2N}$);
  - loose recombination (recombination fraction $c_N \to c$ s.t. $Nc_N \to +\infty$);

  ▶ same limiting diffusion
- A Moran model with selection, loose recombination and low mutation in the n-locus case:
  - in each time step, a pair of gametes is chosen so that their genotypes have a probability proportional to $(1 + \frac{s_{i,j}}{N})$ of being ($\boldsymbol{i}$, $\boldsymbol{j}$).
  - They produce a gamete which displaces a *randomly chosen* gamete in the population.

  ▶ same limiting diffusion

Remarks

- The multilocus Moran model allowed to identify a class of diffusions for allelic frequencies:

$$\frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\partial^2_{i,i} +$$

$$\sum_{i=1}^{n}\Big\{\mu_1(1-x_i)-\mu_0 x_i+(1/2-x_i)x_i(1-x_i)\sum_{L\subset [\![1,n]\!]\setminus\{i\}}\alpha_{L\cup\{i\}}\prod_{j\in L}x_j(1-x_j)\Big\}\partial_i$$

- Diffusion approximations are generally insensitive to the detailed dynamics of the underlying individual-based model.
- It is possible to add new features to the multilocus Moran model as survival selection and spatial constraint to the pairing.

S. Cerrai and P. Clément.
Well-posedness of the martingale problem for some degenerate diffusion processes occuring in dynamics of populations.
Bull. Sci. Math., 128:355–389, 2004.

M. de Aguiar, M. Baranger, E. Baptestini, and Y. Bar-Yam.
Global patterns of speciation and diversity.
Nature, 460:384–387, 2009.

S. N. Ethier and T. Nagylaki.
Diffusion approximations of Markov chains with two time scales and applications to population genetics.
Adv. Appl. Prob., 12:14–49, 1980.

S. N. Ethier and T. Nagylaki.
Diffusion approximations of the two-locus Wright-Fisher model.
J. Math Biol., 27:17–28, 1989.

P. Higgs and B. Derrida.
Genetic distance and species formation in evolving populations.
J. Mol. Evol., 35:454–465, 1992.

M. Serva and L. Peliti.
A statistical model of an evolving population with sexual reproduction.
J. Phys. A: Math. Gen., 24, 1991.