Diffusion approximations
for one-locus multi-allele kin selection, mutation and random drift in group-structured populations: A unifying approach to selection models in population genetics

Sabin Lessard, Université de Montréal
In memory of Sam Karlin

Diffusion approximations
for one-locus multi-allele kin selection, mutation and random drift in group-structured populations: A unifying approach to selection models in population genetics

Sabin Lessard, Université de Montréal
In memory of Sam Karlin

- Historical perspectives
- Population structure
- Diffusion approximation
- Inclusive fitness formulation
- Concluding remarks ... and further results!

THE
GENETICAL THEORY OF
NATURAL SELECTION

BY

R. A. FISHER, Sc.D., F.R.S.

OXFORD
AT THE CLARENDON PRESS
1930

University of Montreal

Fundamental Theorem of Natural Selection

The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time. (Fisher 1930, p.35)

Fundamental Theorem of Natural Selection

The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time. (Fisher 1930, p.35)

- Partial change in mean fitness
  (Price 1972, Ewens 1989, Lessard 1997)

Fundamental Theorem of Natural Selection

The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time. (Fisher 1930, p.35)

- ▸ Partial change in mean fitness
  (Price 1972, Ewens 1989, Lessard 1997)

- ▸ Increase of mean fitness under weak selection
  (Nagylaki 1976, 1977, 1987, 1989, 1993)

Group selection

The average adaptiveness of the species thus advances under intergroup selection, an enormously more effective process than intragroup selection. (Wright 1932)

Group selection

The average adaptiveness of the species thus advances under intergroup selection, an enormously more effective process than intragroup selection. (Wright 1932)

By all odds the most important cases of interdeme selection are those in which the character that increases the probability of survival of the deme as a unit is itself being selected against *within the population.* (Lewontin 1965)

Kin selection

Species ... should tend to evolve behaviour such that each organism appears to be attempting to maximize its inclusive fitness. (Hamilton 1964)

Kin selection

Species ... should tend to evolve behaviour such that each organism appears to be attempting to maximize its inclusive fitness. (Hamilton 1964)

$$W_J = 1 + \sum_I \rho_{J \to I} e_{J \to I}$$

for some coefficient of relatedness $\rho_{J \to I}$

Evolutionarily stable strategy

Roughly, an ESS is a strategy such that, if most of the members of a population adopt it, there is no "mutant" strategy that would give higher reproductive fitness. (Maynard Smith and Price 1973)

Evolutionarily stable strategy

Roughly, an ESS is a strategy such that, if most of the members of a population adopt it, there is no "mutant" strategy that would give higher reproductive fitness. (Maynard Smith and Price 1973)

Stable state of the replicator dynamics (Taylor & Jonker 1978)

$$\dot{x}_k = x_k((H\mathbf{x})_k - \mathbf{x} \cdot H\mathbf{x})$$

for a game matrix $H = \|h_{kl}\|$, and conversely if H symmetric.

Evolutionary stability for finite populations

For finite N, we propose that B is an ESS ... if two conditions hold: (1) selection opposes A invading B,... and (2) selection opposes A replacing B. (Nowak et al. 2004)

Evolutionary stability for finite populations

For finite N, we propose that B is an ESS ... if two conditions hold: (1) selection opposes A invading B,... and (2) selection opposes A replacing B. (Nowak et al. 2004)

$$P(\text{fixation of single } A \text{ among } N \text{ genes}) < \frac{1}{N}$$

Questions

Questions

- Does an inclusive fitness make sense in finite populations?

- What are the roles of inbreeding and group selection?

- What coefficients of relatedness come into play?

- What is the relationship with game dynamics?

Population structure

Population structure

D groups of N mating pairs of diploid individuals

L alleles $A_1, ..., A_L$ at a single locus

$z_i(t)$ frequency of ordered group type $(A_{i,1}, ..., A_{i,4N})$

in generation $t \geq 0$ for $i = 1, ..., L^{4N}$

Life cycle

Life cycle

Reproduction: infinite number of offspring

Life cycle

Reproduction: infinite number of offspring

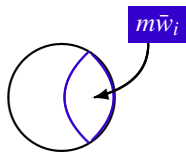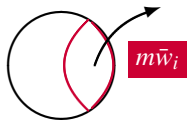Selection: viability of $A_k A_l$ in a group of type i

$$w_{kl,i} = 1 + \frac{\sigma_{kl,i}}{4ND}$$

with scaled selection coefficient $\sigma_{kl,i} = h_{kl} + \bar{v}_i$
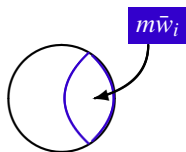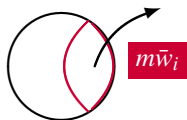
Migration:
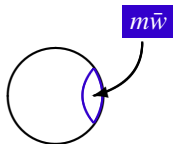
Migration:

- Proportional dispersal

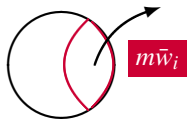## Migration:
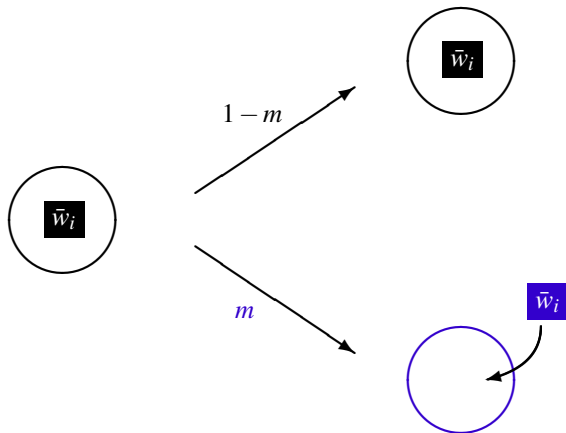
- Proportional dispersal

$m\bar{w}_i$ $m\bar{w}_i$

- Uniform dispersal

$m\bar{w}_i$ $m\bar{w}$

- Local extinction and recolonization

Mating: after migration or before migration

Mating: after migration or before migration

Mutation: from A_k to A_l with probability

$$u_{kl} = \frac{\mu_{kl}}{4ND}$$

Mating: after migration or before migration

Mutation: from A_k to A_l with probability

$$u_{kl} = \frac{\mu_{kl}}{4ND}$$

Sampling: group of type j from type i with probability

$$P_{ij}^D(\mathbf{z}(t))$$

Key Lemma

Key Lemma

When $D = \infty$, there is uniform convergence of the frequency

$$z_j(t+1) = \sum_i z_i(t) P_{ij}(\mathbf{z}(t))$$

to

$$\hat{z}_j(\mathbf{x}) = \sum_{\mathbf{r}} c_j(\mathbf{r}) x_1^{r_1} \cdot \ldots \cdot x_L^{r_L}$$

with $c_j(\mathbf{r})$ the number of ways for a group of type j to have

r_k unrelated ancestral genes A_k and x_k the frequency of A_k.

Proof: Coalescent approach

$\int$-456

THE COALESCENT

J.F.C. Kingman

Mathematical Institute, University of Oxford, England

The n-coalescent is a continuous-time Markov
chain on a finite set of states, which des-
cribes the family relationships among a
sample of n members drawn from a large haploid

Diffusion approximation

Diffusion approximation

The allele frequency process $\mathbf{X}(\lfloor 4ND\tau \rfloor)$ *converges weakly when* $D \rightarrow \infty$ *to a diffusion with infinitesimal covariances*

$$a_{kl}(\mathbf{x}, \mathbf{0}) = Cx_k(\delta_{kl} - x_l)$$

for some coefficient of diffusion C, and infinitesimal means

$$b_k(\mathbf{x}, \mathbf{0}) = \sum_l \mu_{lk} x_l - \sum_l \mu_{kl} x_k + x_k((H\mathbf{x})_k - \mathbf{x} \cdot H\mathbf{x})$$

for some fitness matrix H.

Proof: Two-time scale MC (Ethier and Nagylaki 1980)

$$E_{\mathbf{z}}(\Delta X_k) = \frac{b_k(\mathbf{x}, \mathbf{y})}{4ND} + o(D^{-1})$$

$$E_{\mathbf{z}}((\Delta X_k)(\Delta X_l)) = \frac{a_{kl}(\mathbf{x}, \mathbf{y})}{4ND} + o(D^{-1})$$

$$E_{\mathbf{z}}((\Delta X_k)^4) = o(D^{-1})$$

$$E_{\mathbf{z}}(\Delta Y_j) = c_j(\mathbf{x}, \mathbf{y}) + o(1),$$

$$\text{Var}_{\mathbf{z}}(\Delta Y_j) = o(1)$$

uniformly in $\mathbf{z}$, where $\mathbf{y} = \mathbf{z} - \hat{\mathbf{z}}$ and $c_j(\mathbf{x}, \mathbf{y}) = \sum_i z_i P_{ij}(\mathbf{z}) - z_j$

Inclusive fitness formulation

Inclusive fitness formulation

With proportional dispersal before mating:

$$H = \|(1-f_J)\sigma_{kl}^{\bullet} + f_J \sigma_{kk}^{\bullet\bullet}\|$$

where

$$\sigma_{kl}^{\bullet} = h_{kl} - (1-m)\rho_{J\to I}^{\bullet} h_{kl} + m\rho_{J\to I}^{\bullet} v_{kl}$$

$$\sigma_{kk}^{\bullet\bullet} = h_{kk} - (1-m)\rho_{J\to I}^{\bullet\bullet} h_{kk} + m\rho_{J\to I}^{\bullet\bullet} v_{kk}$$

with coefficients of relatedness $\rho_{J\to I}^{\bullet}$ and $\rho_{J\to I}^{\bullet\bullet}$ given

that J is allozygous or autozygous, respectively, and

inbreeding coefficient f_J

- Uniform dispersal before mating: *stronger group selection*

$(1-m)^2$ instead of $(1-m)$

- Uniform dispersal before mating: *stronger group selection*

  $(1-m)^2$ instead of $(1-m)$

- Proportional or uniform dispersal after mating: *smaller size*

  $C = (1 - f_J + 4m(2-m)f_J)$ instead of $C = (1 - f_J)$

- Uniform dispersal before mating: *stronger group selection*

$$(1-m)^2 \text{ instead of } (1-m)$$

- Proportional or uniform dispersal after mating: *smaller size*

$$C = (1 - f_J + 4m(2-m)f_J) \text{ instead of } C = (1-f_J)$$

- Local extinction before mating: *like proportional dispersal*

but with larger C if and only if $4Nm > 1$

Concluding remarks

Concluding remarks

- A *diffusion approximation* for selection, mutation and drift in group-structured populations as the number of groups increases can be ascertained by a two-time scale argument.

Concluding remarks

- A *diffusion approximation* for selection, mutation and drift in group-structured populations as the number of groups increases can be ascertained by a two-time scale argument.

- With interactions within groups, the selection drift functions can be expressed in terms of scaled inclusive fitnesses for allozygous and autozygous individuals, giving the entries of an *inclusive fitness matrix H*.

Concluding remarks

- A *diffusion approximation* for selection, mutation and drift in group-structured populations as the number of groups increases can be ascertained by a two-time scale argument.

- With interactions within groups, the selection drift functions can be expressed in terms of scaled inclusive fitnesses for allozygous and autozygous individuals, giving the entries of an *inclusive fitness matrix H*.

- *Competition within groups* is the result of population regulation; *competition between groups* the result of dispersal or colonization. *Inbreeding* only affects the coefficient of diffusion and the coefficients of relatedness.

- Without interaction or without inbreeding, H is symmetric like in the classical viability model for a random mating population (Kimura 1964). In general, *H is a sum of a symmetric matrix and a rank one matrix*.

- Without interaction or without inbreeding, H is symmetric like in the classical viability model for a random mating population (Kimura 1964). In general, *H is a sum of a symmetric matrix and a rank one matrix*.

- In the case of an infinite number of groups with the inverse of the intensity of selection s as unit of time as $s \to 0$ and without mutation, the deterministic dynamics is described by the *replicator equation with H as game matrix*.

- Without interaction or without inbreeding, H is symmetric like in the classical viability model for a random mating population (Kimura 1964). In general, *H is a sum of a symmetric matrix and a rank one matrix*.

- In the case of an infinite number of groups with the inverse of the intensity of selection s as unit of time as $s \to 0$ and without mutation, the deterministic dynamics is described by the *replicator equation with H as game matrix*.

- The approach leads to *sampling formulas* and it can be used to study the *evolution of cooperation*.

THE END!

Evolution of cooperation

Evolution of cooperation

▶ A diffusion approximation holds for any scaled selection coefficients depending on the individual genotype and the group type, e.g.,

$$\sigma_{kl,i} = (h_{kl}, 1 - h_{kl}) \cdot A(\bar{h}_i, 1 - \bar{h}_i)$$

with *generalized conditional coefficients of relatedness* between two or more individuals.

Evolution of cooperation

- A diffusion approximation holds for any scaled selection coefficients depending on the individual genotype and the group type, e.g.,

$$\sigma_{kl,i} = (h_{kl}, 1 - h_{kl}) \cdot A(\bar{h}_i, 1 - \bar{h}_i)$$

with *generalized conditional coefficients of relatedness* between two or more individuals.

- Is *tit-for-tat* replacing *always to defect* in the repeated Prisoner's dilemma (Nowak et al. 2004) made easier in group-structured populations? (in preparation)

Sampling formulas

Sampling formulas

▶ With groups of N haploid individuals dispersing at rate m, the equilibrium frequencies when $D = \infty$ depend on

$$\frac{N!}{k! \prod_{i=1}^{k} n_i!} \left(\frac{mN}{1-m} \right)^k \sum_{(r_{i,t})} \prod_{t=1}^{T} \left[\frac{(1-m)^{r_t} N^{[r_{t+1}]} \prod_{i=1}^{k} S_{r_{i,t} r_{i,t+1}}}{N^{r_t} - (1-m)^{r_t} N^{[r_t]}} \right]$$

the probability for a focal group to have k labeled migrant ancestors having $n_1, ..., n_k$ descendants (Lessard 2007).

Sampling formulas

► With groups of N haploid individuals dispersing at rate m, the equilibrium frequencies when $D = \infty$ depend on

$$\frac{N!}{k! \prod_{i=1}^{k} n_i!} \left(\frac{mN}{1-m} \right)^k \sum_{(r_{i,t})} \prod_{t=1}^{T} \left[\frac{(1-m)^{r_t} N^{[r_{t+1}]} \prod_{i=1}^{k} S_{r_{i,t} r_{i,t+1}}}{N^{r_t} - (1-m)^{r_t} N^{[r_t]}} \right]$$

the probability for a focal group to have k labeled migrant ancestors having $n_1, ..., n_k$ descendants (Lessard 2007).

► This extends Ewens's (1972) sampling formula for the infinitely-many-alleles model to an exact Wright-Fisher population!

THANKS!