

Introduction to Evolutive Population Dynamics

Franco Bagnoli

December 4, 2007

- 1 Introduction
- 2 Modeling
- 3 Genetic models
- 4 Game theory

The matter we deal with

Evolutionary population dynamics deals with *replicators*, i.e., agents that have an individuality (present many variations), are able to reproduce themselves with some changes (their offspring are similar to them), that undergoes selection for some limited resource (like food or space).

Example of replicators

- Living cells (mono and multicellular organisms, sexual and asexual, etc.): DNA, mitochondria,
- Viruses.
- Transposable elements, plasmids.
- Computer viruses.
- Memes (cultural replicators).
- Other chemical replicators (prions, crystals...) ?.

Structure

- An out-equilibrium “chemical” background, that sets the “rules” and furnishes the energy.
- Information contents (genotype).
- Appearance (phenotype).
- Persistence and diffusion (reproduction).
- Variability (mutations and sexual recombination).
- Selection.

Genotype in prokaryotes

- In living cells, the genotype is a sequence of basis (letters of a four-symbol alphabet) that serve either to code for a protein, for an RNA that is not further translated, to promote for such a translation, and in general for the structural shape of DNA.
- The genotype is transmitted to offspring after selection *as a whole* and after *sexual recombination*.
- However, it is better to think at it not at the level of basis, but at the level of *genes*, which are the smaller units with some definite effect, and that have a large change of not being cut by recombination (so they are small).
- Genes have effects of phenotype and can be traced generation after generation. They are the real replicators: in some sense we are just a bag of genes, and the fact that this bag replicates through the bottleneck of germinal lines, obliges genes to cooperation.

Selfish genes

- A gene has just only “goal”: to increase its copy number on population, either afetr direct replication, or by favoring other copies in the population.
- All genes that reproduce through the germinal line are obliged to cooperation for a succesful phenotype at least up to the reproduction age, further if offsprings need parental cares.
- However, this cooperation is reached after selection: a gene that is just introduced into an individual has no way of “knowing” where it is, and probable causes debilitation, cancer or other diseases (this may be the origin of immunity and graft intollerance).
- Genes that may reproduce outside the germinal line are not obliged to cooperation: viruses.
- Genes tend to be parasites, if this is harmless: replicons, transposons, etc.

Memes

- Another level of replicators can be found in our mind: it is so rich that there is space for “useless” ideas, that may reproduce and may even cooperate for fitness.
- A meme is just like a gene: a small unit of information that propagates among individuals and is subject to mutation.
- At the biological level, the corresponding “genotype” is the pattern of activation of neural areas in our brain.
- Examples are songs (better: motives), slang expressions, metaphors, behavioral habits, ideas, religion and other irrational patterns.
- Even “useless” memes contribute to fitness, mainly through sexual selection
- However, memes contribute also to survival: think of driving habits, food choices, drugs, etc. (in other times people carrying “wrong” memes could even be burned..)

Phenotype

- The phenotype is “how a bag of genes modify the environment”. Clearly, our body constitute a part of our phenotype.
- However, also the dam of a beaver is part of its phenotype: the phenotype may extend out of the body.
- The phenotype determines the fitness success: survival, social role and sexual success.
- The phenotype depends on genes in a very complex way, so given a gene, it is not an easy task to determine if it will increase or decrease the fitness of a “bag” .
- This problem is “solved” by random mutations and sexual recombination.
- The phenotype depends on age (which gene is activated at a given age) and on past experiences (and therefore on social community).

Fitness

- The fitness for a given genotype is given by the probability of producing living offsprings.
- However, due to mutation and especially to sexual recombination, the offsprings are different from the parents.
- Therefore, fitness for a gene may be different from the fitness for an individual. If possible, a gene would modify the replication machinery in order to increase its copy number in offsprings, for instance by breaking the recombination. How this “fair” mechanism survived (we know that there are genes able to modify it) is an interesting subject of investigation.

Patterns of evolution

- What is the main motor of evolution? genetic drift or coevolution?
- How did cooperation appear? isn't it ruled out by selfishness?
- How does species emerge? The phylogenetic trees reflect what actually happened?
- How can be sex so common in spite of its cost?
- How did any other pattern of life evolve? was it an accident or a necessity? Is it stable? Can be explained/modeled within our knowledge of life, or we need to introduce something new?

Role of modelization

- Many patterns may be illustrated and understood by using simple “block” models (logic reasoning): *qualitative models*.
- However, sometimes there are two or more equally plausible alternatives, that cannot be decided “a priori”. Their “convenience” may depend *quantitatively* on parameters. They could even be followed *in parallel* by part of the population.
- In this case one has to resort to *quantitative modeling*.
- In a few cases we are able to analyze the model with paper and pencils, generally assuming uniformity or in any case reducing it to a few equations.
- The preferred approach, also to confirm the analytical approximations, is to use **computer simulations**.

Computer modelization

- Computer models are based on simplifications, of course, and the stronger they are (compatible with the presence of the effect under investigation) the more general the pattern is.
- Microscopic (agent-based) simulations allow also to check the stability of a strategy, the robustness of an effect, the possibility of mixed behavior.
- After a pattern is schematically confirmed, one can add all other ingredients, and check if the effect persists.

Rules of the game

- The individual (genotype and phenotype).
- The population (an array of individuals).
- The pool (a network of interactions): may be all-to-all, spatially located on a lattice, in plain space or on a network.
- Pick up an individual and compute its probability of survival (fitness). If unfortunate, free the space...
- Otherwise, if there is available free space (and, for sexual organisms, if there is an available partner), replicate it, with mutations.
- Let it run for a long time and ... voilà: Adam and Eve...

The Genotype

- DNA: sequence of symbols, with informative (genes) and structural contents: the latter is (also) part of the phenotype.
- Gene: a piece of genetic information that has some function and persists long enough (protein-coding, structure, parasitic, ...).
- Simplification: instead of sequences, let us consider the genotype as a bag of genes (synonymous of loci). But there are interesting questions about packing information (viruses), structural constraints, etc.
- Discrete number of alleles (generally 2): the genome is an array of zeros and ones.
- $g = (g_1, g_2, \dots, g_L)$, $g_i \in \{0, 1\}$.
- The genotypic space is highly dimensional.

The Phenotype

- Phenotype: array of quantitative traits $f = (f_1, \dots, f_M)$. It may include “internal” traits or extended ones (like the beaver ponds).
- The phenotype is determined by the genotype and the history of individual (age, experiences). Let us suppose for the moment that $f = p(g)$.
- In generale f_i is a function of many genes, and any gene affect many traits (this is why genetic engineering is hard). Some trait may depend on one or very few genes (for instance “structural” genes.
- $f_i(g_j, g_k, \dots)$ may be a linear (non-epistatic) or nonlinear (epistatic) function of g_j, g_k, \dots .
- The phenotypic space may be one or high-dimensional (depends on what one is interested in).

Selection (fitness)

- Let us simply assume that the the probability of survival A depends on the phenotype $A(f)$.
- In general, it depends also on the distribution of other phenotypes in the network of contacts (hares do not like to stay near to lynxs, the latters have opposite opinions).
- The probability of passing (and spreading) the genotype to the following generation depends also on available space, which may be considered a global coupling.
- We can have evolution without selection (neutral evolution).

Fitness (or adaptive) Landscape

- Sewall Wright introduced the concept of adaptive landscape, similar to hill climbing.
- In our formalism, $A = A(g)$ and does not depend on the distribution of phenotypes.
- However, there is still global competition: Red Queen effect.
- In practice, the fitness is given by $A(g)/\langle A \rangle$.

Simple results

- For vanishing mutations the average fitness $\langle A \rangle$ is a nondecreasing function of time (Fisher theorem),
- And the asymptotic population distribution is a delta peak at the global maximum of the fitness (master sequence).
- For finite mutations, the master sequence is surrounded by a cloud of mutants with lower fitness (quasispecies).
- Broader peaks may “win” over sharper and slightly higher ones.
- Coexistence is fragile. Anyhow, coexisting strains have the same average fitness (Gause principle).
- The portions of the genome subjected to higher selective pressure are less mutable than “neutral” ones.

Implementation

Fixed size population, nonoverlapping generations, stirred (mean field).

- An individual is represented as a string of L bits g_1, \dots, g_L . If $L < 32$ (or 64), one can pack bits inside an integer.
- The environment E is an array of N individuals.
- For each generation, one has to fill up another environment E' .
 - Choose an individual i at random.
 - Compute $A = A(E(i), E)$ (may depend on the other phenotypes).
 - With probability **proportional** to A it is copied in E' , with mutations.
 - Repeat until E' is full.
 - Replace E with E' .
- A may be an arbitrary positive quantity, the survival probability is computed as $A/\langle A \rangle$.

Fitness landscape, fixed size

```
integer,parameter :: N=100, L=10, TMAX = 1000
integer :: E(N), E1(N) ! environments
double precision :: r, A, AA
integer :: i, j, t

call setup(E) ! initialize the simulation
do t=1,TMAX ! loop
  AA = sum(fitness(E))/N ! average fitness
  j=1; do while (j < N)
    i = random_choose(N)
    A = fitness(E(i)) ! compute fitness
    if (random(1) < A/AA) then
      E1(j) = mutate(E(i)); j = j+1
    end if
  end do
  E1=E
end do
```

Fitness landscape, variable size

```
integer,parameter :: N=100, L=10, TMAX = 1000
integer :: E(N) ! environment
double precision :: r, A, AA
integer :: i, j, t, nn=N/10 ! nn=size of population

call setup(E, nn) ! initialize the simulation
AA = sum(fitness(E)) ! total fitness
do t=1,TMAX ! loop
  j=1; do while (j < N)
    A = fitness(E(j)) ! compute fitness
    if (random(1) < A/(AA/nn)*(1-nn/N)) then
      nn = nn + 1; E(nn) = mutate(E(j))
      j = j + 1; AA = AA + A
    else
      E(j) = E(nn); nn = nn-1, AA = A-1
    end if
  end do
end do
end do
```

Fitness depending on age: senescence (Penna model)

- Assume that gene i is activate at age i .
- Genes $g_i = 0$ are good, $g_i = 1$ are bad.
- If three bad genes are activated the individual dies.
- At age 8 or greater, an individual can reproduce (variable population).
- Bad genes concentrate at age greater than 8, causing semelparousy.
- Introducing parental cares and death risk by childbirth, one may justify menopause.

Infinite population

In the case of infinite population, one can use the probability distribution $p(g) = \lim_{N \rightarrow \infty} \frac{\sum_i [g_i = g]}{N}$.

The evolution equation is, for discrete time intervals

$$p(g, t + 1) = \frac{A(g)}{\langle A \rangle} p(g, t) + \text{mutations}$$

or, for continuous time

$$\frac{\partial p(g, t)}{\partial t} = (A(g) - \langle A \rangle) p(g, t) + \text{mutations}$$

Evolution is a kind of pattern formation in genotypic space.

Mutations in fitness landscape

In fitness landscapes, mutations play a fundamental role:

- Individual histories are like random walks on a potential (fitness=energy, mutation=temperature).
- Random walks interact only on a global way (competition for space).
- Transitions (speciations) are dominated by search time (finding a niche).
- Searching efficiency depends on the smoothness of fitness landscape.
- Mutations may cause error threshold (disappearance of master sequence), and mutational meltdown (extinction in variable population, keeping the quasispecies structure).

Smoothness of fitness landscape

S. Kauffman (the origins of order) introduced a “general” framework for fitness landscapes:

- There are L phenotypic traits, corresponding to L genes (L is N for Kauffman)
- Each trait i depends on K genes, by a function $f_i(g_i, \dots, g_{i+k})$ (genes may be pick at random).
- The total fitness is $f = \sum_i f_i$.
- If $K = 1$, gene effects are additive (non-epistatic interactions): smooth landscape, one maximum.
- If $K = L$ and functions are random, the fitness landscape is random, and there are $2^L / (L - 1)$ maxima (no search possible).
- For small values of K maxima are correlated (they share some part of genome): recombination (sexual reproduction) is advantageous.

Evolution in a flat landscape

In a flat fitness landscape (no selection) evolution is just random walk. Kimura's neutral theory apply.

- Evolution depends strongly on the structure of mutations (genotypic space). Taking into consideration only point mutations one has a *nice* hypercubic metric structure.
- In finite populations a mutation, even if slightly deleterious, may get fixed.
- This is the main “motor” of sympatric speciation, genetic drift, founder effect.
- One can compute “easily” the time required to join two genotypes backward in time (phenotypic trees, molecular clock).
- Transpositions, viral insertions, deletions, etc. are “long jumps” in genotypic space. They may originate a “small world” phenomenon...

Competition

The fitness actually depends also on the rest of population.

$$\frac{\partial p(g)}{\partial t} = \frac{A(g, \mathbf{p})}{\langle A \rangle} p(g) + \text{mutations},$$

$$A(g, \mathbf{p}) = \exp \left(H_0(g) + \sum_{g' \simeq g} H_1^{(i)}(g, g') p(g') + \sum_{g' \neq g} H_1^{(e)}(g, g') p(g') + \dots \right).$$

- One can include simple competition terms: intraspecies ($H_1^{(i)}$) and interspecies ($H_1^{(e)}$) contributions.
- Intraspecific competition broadens the distribution, favouring dispersion.
- Interspecific competition stabilizes coexistence.
- Predation induces competition between predators and between preys.

Spatial structure

- In actual simulations, one has to consider the kind of spatial structure.
- We just focused on “stirred” environment (random coupling).
- Another possibility is using a regular lattice.
- Or using a connection graph, eventually evolving in time.

The structure of evolutive patterns

Is evolution dominated by mutations (punctuated equilibrium) or by selection?

- Does evolution walks on a fitness landscape (evolutive times determined by mutations) or is it dominated by coevolution?
- Once that one removes the parts of genotypic space that are unviable (proteins that do not fold, not working metabolic pathways...), how much space is left? Does phylogenetic trees represent a vanishing portion of the available space or not?
- Are transitions (extinctions, speciations) result of random walks (hopeful monsters), slow effects of coevolution or externally driven?

Coevolution: the Bak-Sneppen Model

```
integer, parameter :: N=100, TMAX=1000
double precision :: F(N)
integer:: i, j, t
```

```
F(N) = random(N) ! 0<F(i)<1
```

```
do t=1,TMAX
  i = argmin(F)
  F(i-1:i+1) = random(3)      ! use periodic
                              ! boundary conditions
  call track_avalanches(i)    ! Is i a neighbor
                              ! of previous change?
end do
```

Sex, finally!

- Sex (actually, recombination) increases variability without increasing mortality by mutations (that also prevent increasing the genome length due to error threshold).
- Sex/recombination is advantageous for escaping predation or parasitism.
- With diploidity (or horizontal transmission of genes) genes has more freedom to rearrange (but they have to cooperate, if interested in the survival or the organism...).
- Although sex may be expensive, it may be advantageous in variable environments.
- Sexual selection can be much more effective than fitness for speciation.

Doebeli and Dieckman model (Nature 400, 354)

- Diploid or haploid individuals
- Three sets of genes (loci): ecological character x , mating preferences m and marker trait y .
- The static fitness landscape is unimodal in x (favors intermediate x).
- Competition however, favors differentiation on x .
- Mating alleles are $+$ and $-$, m is the sum. $m > 0$ is assortative, $m = 0$ is random and $m < 0$ is disassortative.
- Two cases: mating depends on x or y .
- When mating depends on the ecological trait x , sympatric speciation occurs quickly.
- Also when mating depends on the neutral mark y , sympatric speciation occurs, but more slowly.

Cooperation (Nowak, Science **314**, 1560 (2006))

- Interspecies interaction can be divided into predation (or parasitism), cooperation and competition.
- Predation may lead to interspecies competition (your enemy is my friend).
- Cooperation is strongly related to competition: there is no “natural” cooperation (genes are selfish).
- Nevertheless, cooperation is common: genes, social insect, multicellular organisms, slime molds, even humans...
- So, what are the evolutionary basis of cooperation?
- In the following, we shall deal with strategies: they may be of genetic or cultural origin (memes...). We consider the “competition” between two strategies.

Strategies

- There is a certain number of possible states (say, two..): Cooperate (C) or defect (D).
- Cooperators pay a cost c for another individual to receive a benefit. Defectors do not pay but may profit of benefits.
- A strategy is an “automaton” rule that establish (in a deterministic or probabilistic way) the next state of an individual given the recent states of both opponents.
- Strategies are inherited and selected.

Payoff

Selection is based on a payoff table (for one of participant):

payoff	C	D
C	α	β
D	δ	γ

Let us denote with p the probability of “playing” C (and $1 - p$ the probability of playing D). The payoff is related to fitness:

$$p' = \frac{\alpha p^2 + \beta p(1 - p)}{\alpha p^2 + \beta p(1 - p) + \delta p(1 - p) + \gamma(1 - p)^2}$$

neglecting fluctuation, the asymptotic state is given by the fixed points of the previous equations.

Scenarios

A society of cooperators has a higher fitness than a society of defectors, but is in general susceptible to invasions or mutation by defectors, without strategie. Possible scenarios are:

- (D) If the only stable point is $p^* = 0$, defectors always dominates.
- (ESS) If there are two stable points, $p^* = 0$ and $p^* = 1$, but the latter has only a tiny basin, then cooperators are evolutionarily stable strategy (ESS): a single mutant cannot invade, but multiple ones can and are helped by fluctuations. $\alpha > \beta$.
- (RD) If the basin of $p^* = 1$ extends up to $p_0 = 1/2$, it is favoured by fluctuations. $\alpha + \gamma > \beta + \delta$
- (AD) The basin up to $p_0 = 2/3$ can be shown to be related to stability respect to fixation of a single mutant (Kimura theory). $\alpha + 2\gamma > \beta + 2\delta$
- (C) Finally, if the basin extends up to $p_0 \simeq 1$, cooperators always dominate (except for $p_0 = 1$).

Kin selection

Standard payoff:

payoff	C	D
C	$b - c$	$-c$
D	b	0

Benefits (may be long term like parental cares) have to be greater than costs. Defectors always wins.

Kin selection

Haldane once said: "I will jump into the river to save two brothers or eight cousins". If you cooperate with a relative that shares a fraction r of your genes, then your payoff (well, that of your genes) is augmented by a fraction r of the payoff of your opponent.

payoff	C	D
C	$(b - c)(1 + r)$	$br - c$
D	$b - rc$	0

For $\frac{c}{b} < r$, cooperation is ESS, RD, AD.

Kin selection says that cooperation among relatives (cells in multicellular organisms, social insects) derives by selfishness of genes.

Direct reciprocity

For a “one shot” game the best strategy is to defect. But Axelrod discovered by computer experiments that for repeated games between two opponents it is best to cooperate and to forgive: tit for tat or similar strategies.

In this case the parameter is the probability w of another encounter (or the expected number $1/w$ of rounds).

payoff	C	D
C	$(b - c)/(1 - w)$	$-c$
D	b	0

- ESS for $c/b < w$
- RD for $c/b < w/(2 - w)$
- AD for $c/b < w/(3 - 2w)$

Reputation (indirect reciprocity)

For humans, reputation is a valuable quantity. It is defined as the average cooperation-to defection record. It may be known with a probability q . If you know that the opponent is a defector, defect, otherwise cooperate.

payoff	C	D
C	$(b - c)$	$-c(1 - q)$
D	$b(1 - q)$	0

- ESS for $c/b < q$
- RD for $c/b < q/(2 - q)$
- AD for $c/b < q/(3 - 2q)$

Network reciprocity

Human societies are structured networks. For a given connectivity k of a node, a cooperator pays a cost c and each of the neighbors receive a benefit b .

payoff	C	D
C	$(b - c)$	$H - c$
D	$b(1 - H)$	0

with $H = \frac{(b-c)k-2c}{(k+1)(k-2)}$.

ESS, RD, AD for $c/b < 1/k$

Group selection

Group selection is based on the higher payoff of cooperation, but one has to find a mechanism for stabilizing it against defectors. The idea is that the society automatically splits into m groups of size n . Cooperators help only inside groups, and successful groups split more often.

payoff	C	D
C	$(b - c)(n + m)$	$bm - c(m + n)$
Theoretical and D	bn	0

ESS, RD, AD for $c/b < m/(m + n)$.

Conclusions

- Theoretical and computer studies cannot say “this model explains this observed effect”
- But they can check the consistency of hypothesis with an effect.
- They may also be useful in testing the robustness with respect to parameters.