Evolution with Positive Feedback between Phenotype and Inheritance

Yoav Ram* Uri Liberman† Marcus W. Feldman‡

March 31, 2017 | v3.0

Abstract

TODO

Contents

Model	2
Wright-Fisher model	2
Definitions	2
Reproduction	2
Inheritance	4
Iteration	4
Recurrence equation	5
Results	5
Constant environment	5
Proposition for $\eta = 0 \dots \dots \dots \dots \dots$	6
Proof	6
	6
Proof	6
Proposition for general case	6
Proof	6
Periodic environment	7
	8
	9
	1

 $^{^*\}mathrm{Department}$ of Biology, Stanford University, Stanford, CA 94305-5020, yoav@yoavram.com

 $^{^\}dagger School$ of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel 69978, uril@tauex.tau.ac.il

 $^{^{\}ddagger} Department$ of Biology, Stanford University, Stanford, CA 94305-5020, mfeld-man@stanford.edu; Corresponding author

Random	environments											 16
Supporting figu	ures											16
References												16

Model

Wright-Fisher model

Definitions

- N: constant population size.
- ϕ_i : phenotype of individual $i, 1 \le i \le N, \phi_i \in \{A, B\}$.
- ϵ_t : the environment at generation $t, \epsilon_t \in \{A, B\}$.
- W: individual fitness when phenotype and environment match, $\phi_i = \epsilon_t$.
- w: individual fitness when phenotype and environment do not match, $\phi_i \neq \epsilon_t$.
- ω_i : fitness of individual *i* at time t, $\omega_i = W \cdot 1_{\phi_i = \epsilon_t} + w \cdot 1_{\phi_i \neq \epsilon_t}$.
- $\bar{\omega}$: population mean fitness.
- π_i : phenotype probability, the probability that individual i becomes phenotype $A, 1 \le i \le N$.
- Π : set of phenotype probabilities in the population, $\Pi = {\pi_i}_{1 \le i \le N}$
- η : phenotypic inheritance rate, $0 \le \eta \le 1$.

Reproduction

For each offspring in the population of generation t+1 we choose a parent from the population of generation t and this choice depends on the parent relative fitness: the probability that individual i is the parent is relative to its fitness, ω_i . Therefore, reproduction is modeled by a multinomial distribution. Therefore, reproduction includes the effects of natural selection and random genetic drift.

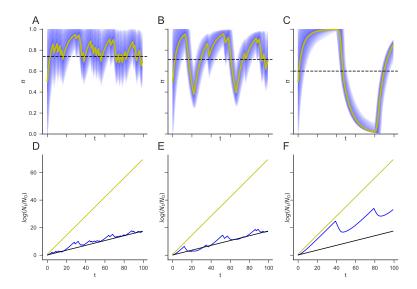


Figure 1: Reproduction of Figure 2 from Xue and Leibler (2016). **A-C**: the average π in the population over time and the distribution in blue; **D-F**: the effective population growth rate. **A,D**: the environment is randomly chosen at each generation such that P(A)=0.7 and P(B)=0.3; **B,E**: the environment flips between A and B, the duration of each is geometrically distributed with p=1/10 for A and p=1/5 for B; **C,F**: the environments flips every 40 generations. Parameters: N=100,000, # generations = 100, $\eta=0.1$, W=2, w=0.2.

Inheritance

The offspring inherits the phenotype probability of the parent with a modification – if the parent became A, then the offspring is even more likely to be A; if the parent was B, then the offspring is less likely to be A. Specifically, for parent k and offspring i:

$$\pi_i = \pi_k \cdot (1 - \eta) + \eta \cdot 1_{\phi_k = A} \tag{1}$$

Note

The notation in Eq. 1 is different from Eq. 1 in Xue and Leibler (2016), as i denotes individual, rather than phenotype. But the process is the same.

Note

The expected difference between parent and offspring phenotype probability is $E[\pi_i - \pi_k | \pi_k] = 0$.

Iteration

At each generation t, Π is updated according to the following steps. Initial values can be determined (i.e., $\forall i, \ \pi_i = 0.5$), or values can be drawn from an initial distribution (i.e., $\pi_i \sim TN(0.5, 0.05)$, TN is the truncated normal distribution). In addition, the sequence ϵ_t is independent of the iteration on Pi.

At each generation t:

- 1. **Development**: the phenotypes of all individuals are drawn from corresponding Bernoulli distributions depending on their phenotype probabilities: $P(\phi_i = A) = \pi_i$.
- 2. **Fitness**: the fitness of all individuals is set: $\omega_i = W \cdot 1_{\phi_i = \epsilon_t} + w \cdot 1_{\phi_i \neq \epsilon_t}$.
- 3. **Reproduction**: the number of offspring of each individual, b_i , is drawn from a multinomial distribution $MN(N, \{\frac{\omega_i}{\sum_i \omega_i}\}_{1 \leq i \leq N})$, such that

$$P(b_1 = x_1, \dots, b_N = x_N) = \frac{N!}{x_1! \cdot \dots \cdot x_N!} \cdot \left(\frac{\omega_1}{\sum_i \omega_i}\right)^{x_1} \cdot \dots \cdot \left(\frac{\omega_N}{\sum_i \omega_i}\right)^{x_N}$$

4. **Inheritance**: the set of phenotype probabilities of the offspring generation is updated using Eq. 1 such that for each i, Π includes exactly b_i copies of $(\pi_i \cdot (1 - \eta) + \eta \cdot 1_{\phi_i = A})$.

Note

Only development and reproduction are stochastic; natural selection and drift occur at the reproduction step.

Note

If $\eta = 0$, $\pi_i \in \{0, 1\}$, and $\epsilon_t \equiv A$, then we have a standard single locus bi-allelic selection-drift Wright-Fisher model.

Recurrence equation

We approximate the Wright-Fisher model using a recurrence equation. This approximation is most suitable when the population is composed of a single lineage – for example, when selection is extreme and there is a common ancestor.

Define x to be the probability that a random individual in the population is A. What is x', the probability that a random offspring of that individual is A?

Assuming an "infinite" population undergoing exponential growth, this depends on (i) if the parent was A or B, with probabilities x and 1-x, (ii) on the relative contribution of A and B phenotypes to the next generation in terms of fitness, and (iii) on the probability that offspring of A or B phenotypes eventually become A, according to the inheritance process (Eq. 1):

$$x' = x \cdot \frac{\omega_A}{\bar{\omega}} \cdot ((1 - \eta)x + \eta) + (1 - x) \cdot \frac{\omega_B}{\bar{\omega}} \cdot (1 - \eta)x \tag{2}$$

Where ω_A and ω_B are the fitness of phenotypes A and B in the current generation (depending on ϵ_t).

When we write a similar recurrence for the probability that an individual is B = ((1-x)' = F(1-x)) and sum the two equations, we find that $\bar{\omega} = x\omega_A + (1-x)\omega_B$ is the mean fitness.

This recurrence equation can be reorganized to:

$$x' = x \frac{x(1-\eta)(\omega_A - \omega_B) + \eta \omega_A + (1-\eta)\omega_B}{x(\omega_A - \omega_B) + \omega_B}$$
(3)

Results

Constant environment

We start with a constant environment $\epsilon_t = A \ \forall t$, such that $\omega_A = W > w = \omega_B$.

Proposition for $\eta = 0$

If $\eta = 0$ then any $x \in [0, 1]$ solves x' = x.

Proof

Set $x' = x, \eta = 0$ in Eq. 3:

$$x = x \frac{x(\omega_A - \omega_B) + \omega_B}{x(\omega_A - \omega_B) + \omega_B}$$

First, x = 0 solves this equality. Otherwise, for x > 0:

$$\Leftrightarrow x(\omega_A - \omega_B) + \omega_B = x(\omega_A - \omega_B) + \omega_B$$

which is an identity \blacksquare .

Proposition for W = w

If W = w then any $x \in [0, 1]$ solves x' = x.

Proof

Set $W = w > 0, x' = x \neq 0$ in Eq. 3:

$$x = x\eta W + (1-\eta)WW \Leftrightarrow$$

$$1 = 1\eta + (1-\eta)1 \Leftrightarrow$$

$$1 = 1$$

Proposition for general case

If $\eta > 0$ and $\omega_A > \omega_B$, then $x^* = 1$ is the only solution for x' = x.

Proof

First, we check that $x^* = 1$ solves x' = x. Set x = 1 in Eq. 3:

$$x' = \frac{1 - \eta)(\omega_A - \omega_B) + \eta \omega_A + (1 - \eta)\omega_B}{\omega_A} = \frac{\omega_A}{\omega_A} = 1$$

Next, we check that x = 0 doesn't solve x' = x by setting x = 0 in Eq. 3:

$$x' = \frac{\eta \omega_A + (1 - \eta)\omega_B}{\omega_B} > 0$$

since both the denominator and the numerator are positive.

Next, we check that any 0 < x < 1 doesn't solve x' = x. Set x' = x in Eq. 2 and substituting $\bar{\omega} = x\omega_A + (1-x)\omega_B$:

$$x = x \cdot \frac{\omega_A}{\bar{\omega}} \cdot ((1 - \eta)x + \eta) + (1 - x) \cdot \frac{\omega_B}{\bar{\omega}} \cdot (1 - \eta)x \Leftrightarrow$$

$$\bar{\omega} = \omega_A \cdot ((1 - \eta)x + \eta) + (1 - x) \cdot \omega_B \cdot (1 - \eta) \Leftrightarrow$$

$$x\omega_A + (1 - x)\omega_B = (1 - \eta)x\omega_A + \eta\omega_A + (1 - \eta)(1 - x)\omega_B \Leftrightarrow$$

$$\eta(x\omega_A + (1 - x)\omega_B) = \eta\omega_A \Leftrightarrow$$

$$x\omega_A + (1 - x)\omega_B = \omega_A \Leftrightarrow$$

$$(1 - x)\omega_A = (1 - x)\omega_B \Leftrightarrow$$

$$x = 1$$

Periodic environment

We concentrate on periodic environments in which both environments occur exactly the same number of generations in each "period". A simple example is A1B1=ABABABAB..., in which the environment switches every generation every generation from A to B and vice versa, or A2B1=AABAABAAB... in which the every two _A_s are followed by a single B. In general, AkBl denotes an environmental regime in which the period is of length k+l and composed of exactly k _A_s and l _B_s.

We simulated evolution in such environments, and Fig. 2 shows the evolution of the distribution of π in a population evolving in three such environmental regimes.

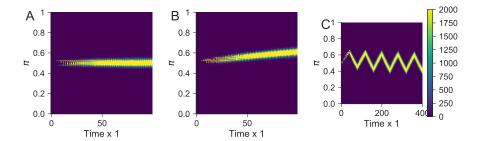


Figure 2: Distribution of π in populations evolving in periodic environments. (A) A1B1, (B) A2B1, (C) A40B40. Parameters: $N=100,000, \eta=0.01, W=1, w=0.1$.

A1B1 regime

When the environment changes every generation, we can write the following recursion, which sets $\omega_A = W, \omega_B = w$ in Eq. 3 to determine x' and and then sets $\omega_A = w, \omega_B = W$ to determine x'':

$$x' = x \frac{x(1-\eta)(W-w) + \eta W + (1-\eta)w}{x(W-w) + w}$$

$$x'' = x' \frac{x(1-\eta)(w-W) + \eta w + (1-\eta)W}{x'(w-W) + W}$$
(4)

We are looking for solutions for x'' = x, which evaluates to a quartic polynomial. Two solutions are x = 0, 1 (assign to Eq. 7 to check), but there are two more potential solutions such that

$$x'' - x = x(1 - x)G(x) = 0$$
$$G(x) = Ax^2 + Bx + C$$

Using SymPy, a Python library for symbolic mathematics, a free alternative to Wolfram MathematicaTM (Team 2014), we find all four solution of x'' - x = 0:

$$G(x) = x^{2} - \frac{W(1-\eta) - w(3-\eta)}{(2-\eta)(W-w)}x - \frac{w}{(2-\eta)(W-w)}$$
 (5)

To find the roots of G(x), recall that $W > w \ge 0, 1 \ge \eta \ge 0$, so

$$G(0) = \frac{-w}{(2-\eta)(W-w)} < 0$$

and

$$G(1) = 1 - \frac{W(1-\eta) - w(3-\eta)}{(2-\eta)(W-w)} - \frac{w}{(2-\eta)(W-w)} = \frac{W}{(2-\eta)(W-w)} > 0$$

and $\lim_{x\to\pm\infty}G(x)=+\infty$.

Therefore, one root of G(x) is negative and one, \tilde{x} , is positive and below 1. Let $\delta = \frac{-B - \sqrt{B^2 - 4AC}}{2A} - \frac{-B + \sqrt{B^2 - 4AC}}{2A}$ (where A, B, C are the coefficients of G(x), defined in Eq. 5). Then, $\delta = \frac{\sqrt{(W+w)^2 - \eta(2-\eta)(W-w)^2}}{(2-\eta)(W-w)}$. Because $\eta(2-\eta)$ is maximized at 1,

$$(W+w)^2 - \eta(2-\eta)(W-w)^2 > (W+w)^2 - (W-w)^2 = 4Ww > 0$$

so $\delta > 0$. Therefore, the positive root is:

$$\tilde{x} = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \tag{6}$$

Fig. 3 shows \tilde{x} (dashed green) compared with x from iteration of Eq. 7 (blue) and with the population mean π ($\bar{\pi}$) in Wright-Fisher simulations (orange) for several combinations of η, W, w . All iterations started with $\bar{\pi} = 0.5$; in the WF simulations, population size N is 100,000, the initial population is drawn from N(0.5, 0.05), and the results are based on 50 simulations per parameter set. Note that the x-axis shows every other generation* (end of each period). The analytic approximation is good when selection is extreme (w = 0), but overestimates $\bar{\pi}$ when selection in not extreme (w = 0.1). In both cases the initial population distribution did not affect the results (as long as it wasn't trivial, *i.e.* $\pi = 0$, see Fig. 5).

A2B1 regime

In the A2B1 regime (every two generations in the A environment are followed by a generation in environment B), an analytic approximation is not possible, as solving x''' - x = 0 requires solving a polynomial of degree 6. However, iterating the relevant recurrence equation:

$$x' = x \frac{x(1-\eta)(W-w) + \eta W + (1-\eta)w}{x(W-w) + w}$$

$$x'' = x' \frac{x'(1-\eta)(W-w) + \eta W + (1-\eta)w}{x'(W-w) + w}$$

$$x''' = x'' \frac{x''(1-\eta)(w-W) + \eta w + (1-\eta)W}{x''(w-W) + W}$$
(7)

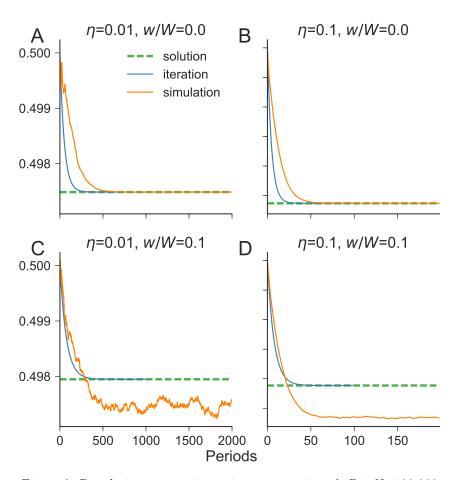


Figure 3: Population mean π in environment regime A1B1. N=100,000.

provides similar results: the equilibrium value is in good fit with Wright-Fisher simulations for extreme selection (w = 0) but over estimates the equilibrium otherwise (Fig. 4).

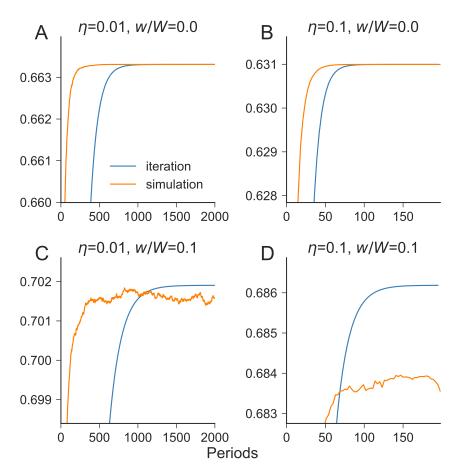


Figure 4: Population mean π in environment regime A2B1. N=100,000.

Protected polymorphisms in AkBI regime

What can we say about the more general case of k generations in environment A and l generations in B? We examine the existence of a protected polymorphism (Prout 1968), which means that none of the phenotypes become extinct even when initially rare. Environments A and B select for $\pi=1$ and $\pi=0$, respectively, and these are absorbing states: if all individuals are, for example, $\pi=0$, then they are all of phenotype B and all offspring will be $\pi=0$, too. Mathematically, we examine the stability of x=0 and x=1; if both are unstable,

then a protected polymorphism occurs. Intuitively, this will happen if neither environment occurs enough to fix it's preferred state.

We rewrite Eq. 3 as $x' = x \cdot f_A(x)$ in environment A and $x' = x \cdot f_B(x)$ in environment B, where:

$$f_A(x) = \frac{x(1-\eta)(W-w) + \eta W + (1-\eta)w}{x(W-w) + w}$$
$$f_B(x) = \frac{x(1-\eta)(w-W) + \eta w + (1-\eta)W}{x(w-W) + W}$$

We concentrate on $l \geq k$ and check if x = 0 is stable, because (i) if x = 0 is not stable when $l \geq k$ then x = 1 cannot be stable either, as selection is stronger, on the whole, towards 0; and (ii) checking the other case (stability of x = 1 when $k \geq l$) is symmetric, and can be done in the same way by writing a recurrence equation for the frequency y of phenotype B rather than A and studying the case of y = 0.

To check if x = 0 is stable, we start with a value very close to 0 and check if after a period of k+l generations the population is closer or farther from 0 compared to where it started.

For $x_0 = x(t = 0) \sim 0$, we can use a linear approximation of the form $f_A(x_0) = f_A(0) + o(x_0)$ and $f_B(x_0) = f_B(0) + o(x_0)$, where:

$$f_A(0) = 1 + \eta(\frac{W - w}{w})$$

$$f_B(0) = 1 + \eta(\frac{w - W}{W})$$

For k generations in environment A, and l generations with environment B, in any given order, we can write:

$$\begin{split} x_{k+l} &= x(t=k+l) \approx x_0 f_A^k(0) f_B^l(0) \Rightarrow \\ &\frac{x_{k+l}}{x_0} \approx f_A^k(0) f_B^l(0) \end{split}$$

so that if we start very close to zero $(x_0 \sim 0)$, the multiplicative change over the k+l generations can be approximated by $f_A^k(0)f_B^l(0)$.

If $f_A^k(0)f_B^l(0) > 1$, then x = 0 is not stable; since x = 1 is not stable either (due to $l \ge k$), then we have a protected polymorphism somewhere (0 < x(t) < 1 for any generation t). In contrast, if $f_A^k(0)f_B^l(0) < 1$, then x = 0 is stable and the protected polymorphism disappears.

Following we examine the protected polymorphism in several special and general cases.

$$W = w$$

In this case, fitness in both environments is equivalent, there is no selection, and therefore evolution is neutral. Indeed, we find that $f_A(x) = f_B(x) \equiv 1$, without an approximation.

$$\eta = 0$$

In this case, there is no feedback between phenotype and inheritance, and only drift generates genetic variance, and evolution is neutral. Indeed, we get $f_A(x) = f_B(x) = \equiv 1$.

$$\eta = 1$$

In this case, development is not stochastic, and after one generation the model becomes a standard two-type genetic model. Only genetic drift generates genetic variance, but natural selection does play a role. Indeed, we get $f_A^k(0)f_B^l(0)=\left(\frac{W}{w}\right)^{k-l}$. Since W>w, we find that $\frac{x_{k+l}}{x_0}$ is

$$\begin{cases} < 1 & , k < l \\ = 1 & , k = l \\ > 1 & , k > l \end{cases}$$

Proposition for k = l

If $k = l, W > w > 0, 1 > \eta > 0$, then $f_A^k(0) f_B^l(0) > 1$.

Proof

First,
$$f_A^k(0)f_B^l(0) = (f_A(0)f_B(0))^k > 1$$
 iff $f_A(0)f_B(0) > 1$.

To show the latter,

$$f_A(0)f_B(0) = (1 + \eta \frac{W - w}{w}) \cdot (1 + \eta \frac{w - W}{W}) = (1 - \eta + \eta \frac{W}{w}) \cdot (1 - \eta + \eta \frac{w}{W}) = (1 - \eta)^2 + \eta^2 + \eta (1 - \eta)(\frac{W}{w} + \frac{w}{W}) = (1 - 2\eta(1 - \eta) + \eta(1 - \eta)(\frac{W^2 + w^2}{Ww}) = (1 + \eta(1 - \eta)\frac{W^2 - 2Ww + w^2}{Ww} = (1 + \eta(1 - \eta)\frac{(W - w)^2}{Ww})$$

which, under the proposition conditions, is > 1.

Proposition for l > k = 1

If
$$l > 1 + (1 - \eta) \frac{W - w}{w}$$
 then $f_A(0) f_B^l(0) < 1$.

Proof

Set n = l - 1. Then,

$$n > (1 - \eta) \frac{W - w}{w} \Leftrightarrow$$

$$n \eta \frac{W - w}{W} > \eta (1 - \eta) \frac{(W - w)^2}{Ww} \Leftrightarrow$$

$$1 - n \eta \frac{w - W}{W} > 1 + \eta (1 - \eta) \frac{(W - w)^2}{Ww} \Leftrightarrow$$

$$1 > \frac{1 + \eta (1 - \eta) \frac{(W - w)^2}{Ww}}{1 - n \eta \frac{w - W}{W}}$$
(8)

Now, $W > w \Rightarrow 1 \ge \frac{W-w}{W} \ge 0$, and together with $1 \ge \eta \ge 0$ we get $0 \ge \eta \frac{w-W}{W} \ge -1$. These conditions allow us to use the following Bernoulli inequality (proof with induction):

$$(1+x)^n \le \frac{1}{1-nx}, \quad \forall x \in [-1,0], \forall n \in \mathbb{N}.$$
(9)

From the Bernoulli inequality we have:

$$\left(1 + \eta \frac{w - W}{W}\right)^n \le \frac{1}{1 - n\eta \frac{w - W}{W}} \tag{10}$$

Taken together, Eq. 8 and Eq. 10 imply that:

$$f_A(0)f_B^{n+1}(0) = \left(1 + \eta \frac{W - w}{w}\right) \left(1 + \eta \frac{w - W}{W}\right) \left(1 + \eta \frac{w - W}{W}\right)^n = \left(1 + \eta(1 - \eta)\frac{(W - w)^2}{Ww}\right) \left(1 + \eta \frac{w - W}{W}\right)^n \le \frac{1 + \eta(1 - \eta)\frac{(W - w)^2}{Ww}}{1 - \eta \eta \frac{w - W}{W}} < 1$$

Proposition for general case: $l > k \ge 1$

If
$$l > k \Big(1 + (1-\eta) \frac{W-w}{w}\Big)$$
, then $f_A^k(0) f_B^l(0) < 1$.

Proof

First, assume $\frac{l-k}{k} \in \mathbb{N}$ and set $n = \frac{l-k}{k} \Rightarrow n > (1-\eta) \frac{W-w}{w}$.

Now, using the previous proposition,

$$f_A^k(0)f_B^l(0)=f_A^k(0)f_B^{(n+1)k}(0)=(f_A(0)f_B^{n+1}(0))^k<1$$

because $\forall y > 0, k > 0 \ y < 1 \Rightarrow y^k < 1$.

Next, relax the assumption $\frac{l-k}{k} \in \mathbb{N}$; set $n = \lceil \frac{l-k}{k} \rceil > \frac{l-k}{k} > (1-\eta) \frac{W-w}{w}$, then

$$f_A^k(0)f_B^l(0) < f_A^k(0)f_B^{(n+1)k}(0) = (f_A(0)f_B^{n+1}(0))^k < 1$$

and again, the previous proposition provides the last inequality.

Proposition

If
$$l < k \left(1 + \frac{(1-\eta) \frac{W-w}{w}}{1+\eta(1-\eta) \frac{(W-w)^2}{Ww}} \right)$$
 then $f_A^k(0) f_B^l(0) > 1$.

Proof

Similar to previous proposition, but using a different Bernoulli inequality:

$$(1+x)^n \ge 1 + nx, \quad \forall x > -1, \forall n \in \mathbb{R} \setminus (0,1).$$

Random environments

TODO: - local stochastic stability - see Levikson & Karlin, 1970's; Liberman & Karlin, 1970's

Supporting figures

• Fig. 5

References

Prout, T. 1968. Sufficient conditions for multiple niche polymorphism. The American Naturalist 102:493–496.

Team, S. D. 2014. SymPy: Python library for symbolic mathematics.

Xue, B., and S. Leibler. 2016. Evolutionary learning of adaptation to varying environments through a transgenerational feedback. Proceedings of the National Academy of Sciences 201608756.

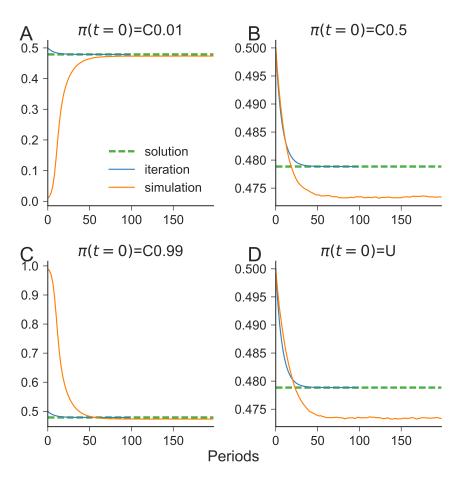


Figure 5: Population mean π in environment regime A1B1. Initial population distribution: (A) π_i =0.01; (B) π_i =0.5; (C) π_i =0.99; (D) $\pi_i \sim Uniform(0,1)$. N=100,000.