A probabilist's roadmap to the polygenic limit

Philibert Courau

IBENS, Collège de France, Wien Universität

December 18th, 2025

In front of a jury composed of:

Paul JENKINS	PU, University of Warwick	Rapporteur
Charline SMADI	DR INRAE, Université Grenoble Alpes	Rapporteuse
Luis-Miguel CHEVIN	DR CNRS, CEFE	Examinateur
Joachim HERMISSON	PU Universität Wien	Examinateur
Thierry MORA	DR CNRS, ENS	Examinateur
Amandine VÉBER	DR CNRS, Université Paris Cité	Examinatrice
Amaury LAMBERT	PU École Normale Supérieure (ENS)	Directeur
Emmanuel SCHERTZER	PU Universität Wien	Directeur

The polygenic limit

The **polygenic limit** describes a **large** population such that every organism carries **many** fitness-influencing genes evolving close-to-independently.

The polygenic limit

The **polygenic limit** describes a **large** population such that every organism carries **many** fitness-influencing genes evolving close-to-independently.

A **polygenic trait** is a measurable characteristic of an organism which is influenced by many genes.

Importance of the polygenic limit

- Interpreting the result of genomic data (Sella Barton 2019...)

Importance of the polygenic limit

- Interpreting the result of genomic data (Sella Barton 2019...)
- Understanding the genetic architecture of diseases (Koch et al 2025+)

Importance of the polygenic limit

- Interpreting the result of genomic data (Sella Barton 2019...)
- Understanding the genetic architecture of diseases (Koch et al 2025+)
- Survival of a population (Charlesworth 2013)

Importance of the polygenic limit

- Interpreting the result of genomic data (Sella Barton 2019...)
- Understanding the genetic architecture of diseases (Koch et al 2025+)
- Survival of a population (Charlesworth 2013)
- Predicting long-term response of a population to a new environment (Hayward et al 2022...).

A tale of four queens

Selection (Darwin/Wallace 1859)

- There is a well-defined measurable quantity called the **fitness** of an organism which measures how well adapted an organism is to its current environment.

Image from Wikipedia

A tale of four queens

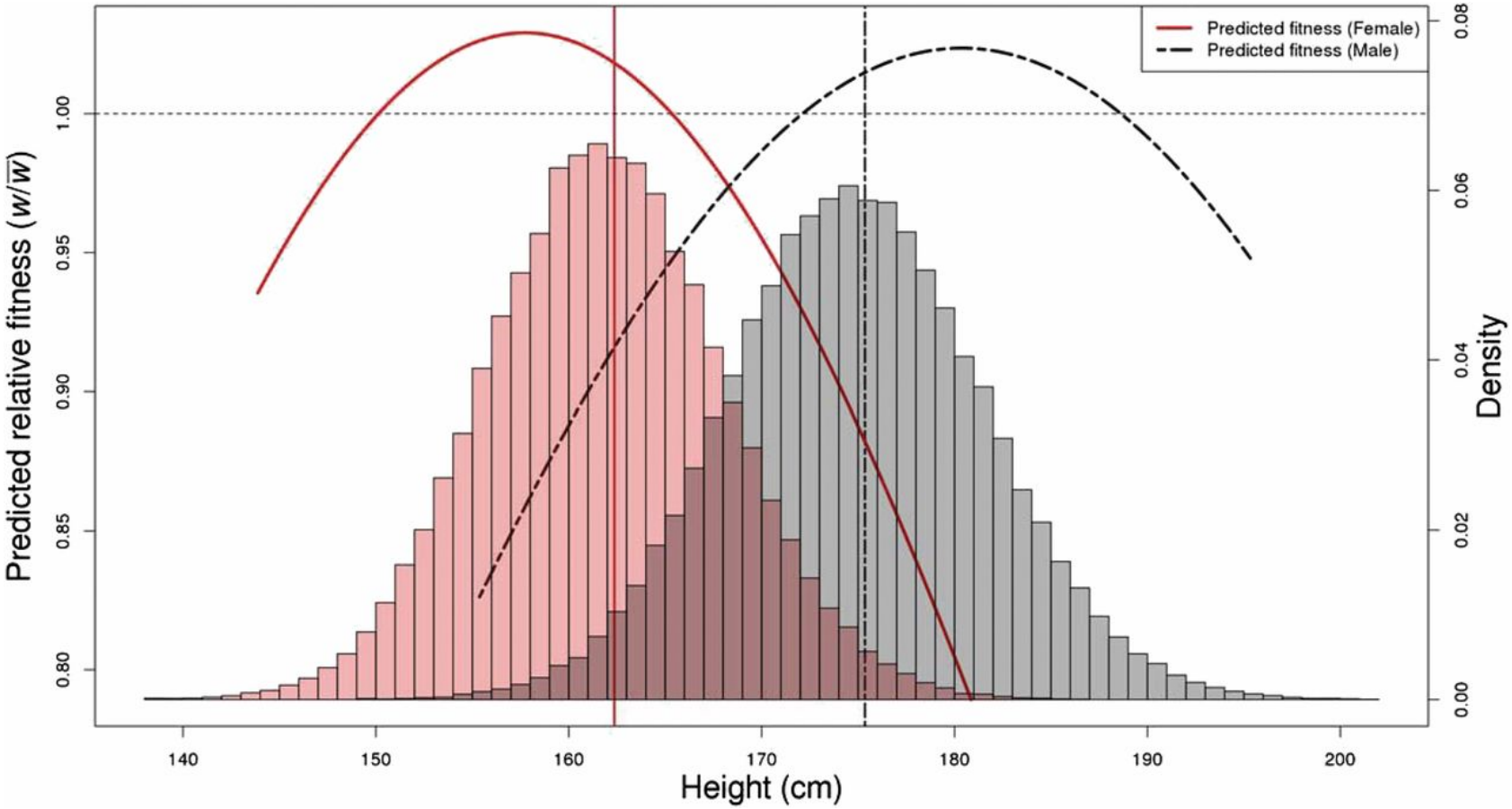
Selection (Darwin/Wallace 1859)

- There is a well-defined measurable quantity called the **fitness** of an organism which measures how well adapted an organism is to its current environment.
- The lower the fitness, the lower the chances to have healthy offspring.

Image from Wikipedia

A tale of four queens

Selection



From Sanjak et al, 2017

A tale of four queens

Mutation (Hugo de Vries 1901)

Heritable variability spontaneously appears within populations.

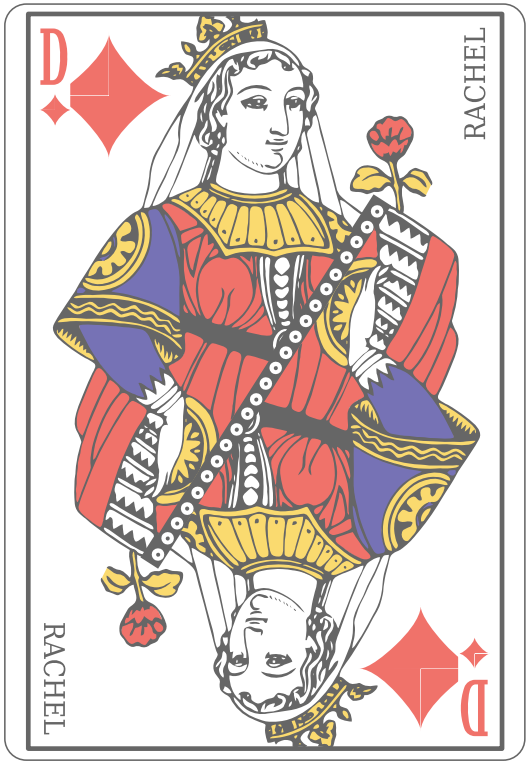


Image from Wikipedia

A tale of four queens

Recombination (Mendel 1865, Bateson, Saunders, Punnett 1905)

Offspring inherit some characteristics from each of their parents.

Image from Wikipedia

A tale of four queens

Recombination

From bartongroup.pages.ist.ac.at

A tale of four queens

Genetic drift (Hagedoorn, Hagedoorn-Vorstheuvel La Brand 1921, Wright 1929)

The evolution of a population is inherently random.

Image from Wikipedia

A tale of four queens

We will not consider

- Out-of-equilibrium dynamics (Hayward et al 2022, Höllinger et al 2023...)

A tale of four queens

We will not consider

- Out-of-equilibrium dynamics (Hayward et al 2022, Höllinger et al 2023...)
- Fluctuating environments, plasticity, bet-hedging, maternal effects, environmental effects (Bertram 2025+...)

A tale of four queens

We will not consider

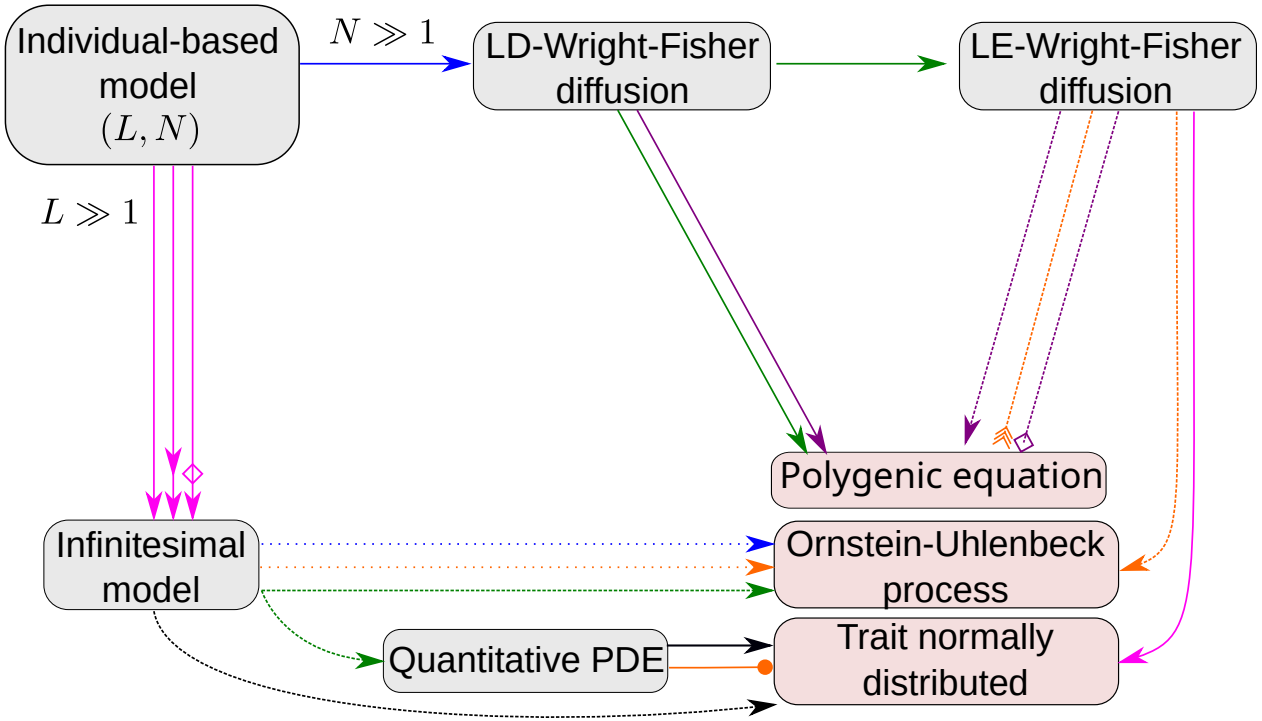
- Out-of-equilibrium dynamics (Hayward et al 2022, Höllinger et al 2023...)
- Fluctuating environments, plasticity, bet-hedging, maternal effects, environmental effects (Bertram 2025+...)
- Migration, population structure, demography (Pólechova Barton 2015, Szép et al 2021 ...).

A tale of four queens

We will not consider

- Out-of-equilibrium dynamics (Hayward et al 2022, Höllinger et al 2023...)
- Fluctuating environments, plasticity, bet-hedging, maternal effects, environmental effects (Bertram 2025+...)
- Migration, population structure, demography (Pólechova Barton 2015, Szép et al 2021 ...).
- Sexual selection, homogamy (Surendranadh and Sachdeva, 2025)

Conclusion for mathematicians



Degree of proof
- Proven
- Discussed
- Not discussed

Type of proof
- Diffusion approximation
- Mixing
- Mean-field approximation
- Separation of timescales
- Central limit theorem

Complications
- Diploidy/Dominance
- Strong selection
- Epistasis
- Population structure/Demography

The individual-based model

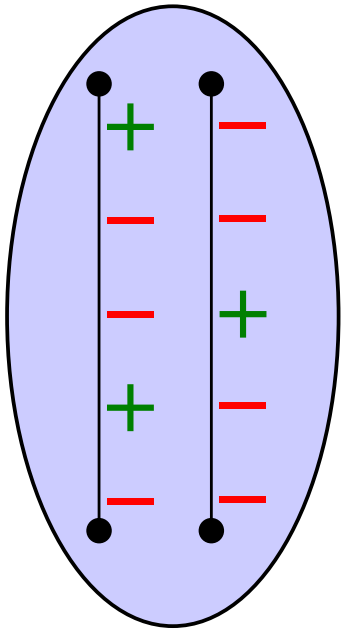
The individual-based model

The genome of an organism

The genome is structured as a pair of chromosomes with L loci.

Each locus has two alleles labelled $+$ and $-$.

A diploid genome can therefore be represented as an element of $(\{0,1\}^L)^2$.

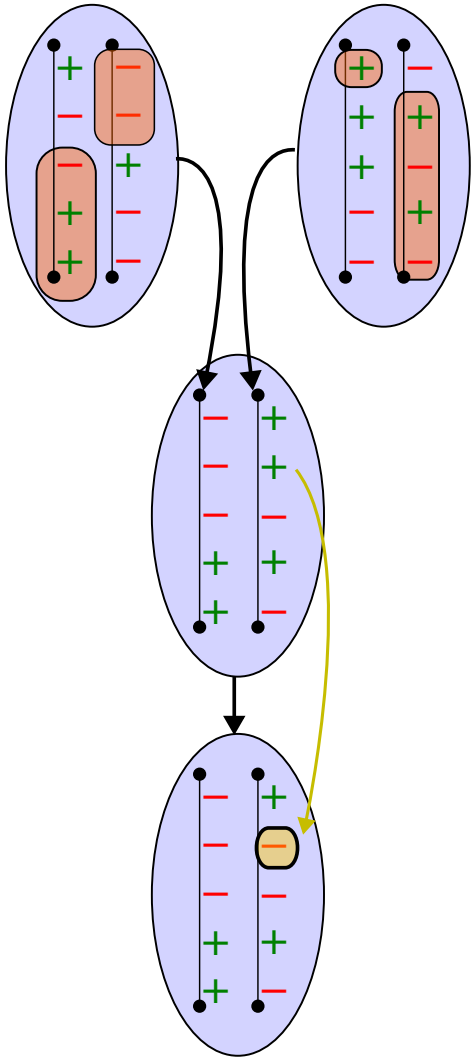


The individual-based model

Inheritance

Recombination ♥:
Single uniform crossover

Mutation ♦:
locus ℓ has mutation probabilities (μ_ℓ^+, μ_ℓ^-)



Additive model: The **trait** of an organism with genome $G = (G_{\ell,i})_{\ell\in[L],i\in[2]} \in (\{0,1\}^L)^2$ is

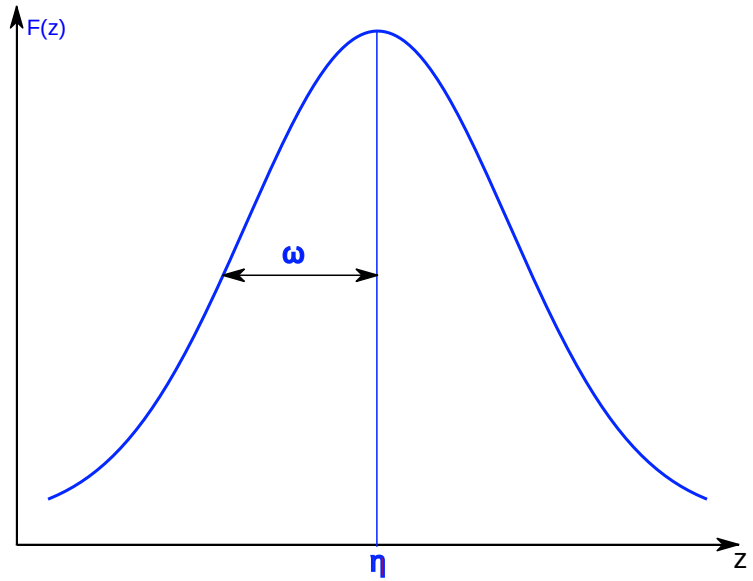
$$Z(G) := \sum_{\ell\in[L]} \alpha_\ell (G_{\ell,1} + G_{\ell,2})$$

with $\alpha_\ell \in \mathbb{R}_+$ the **additive effect on the trait at locus** ℓ.

Stabilizing selection: The **fitness** of an organism with trait value z is

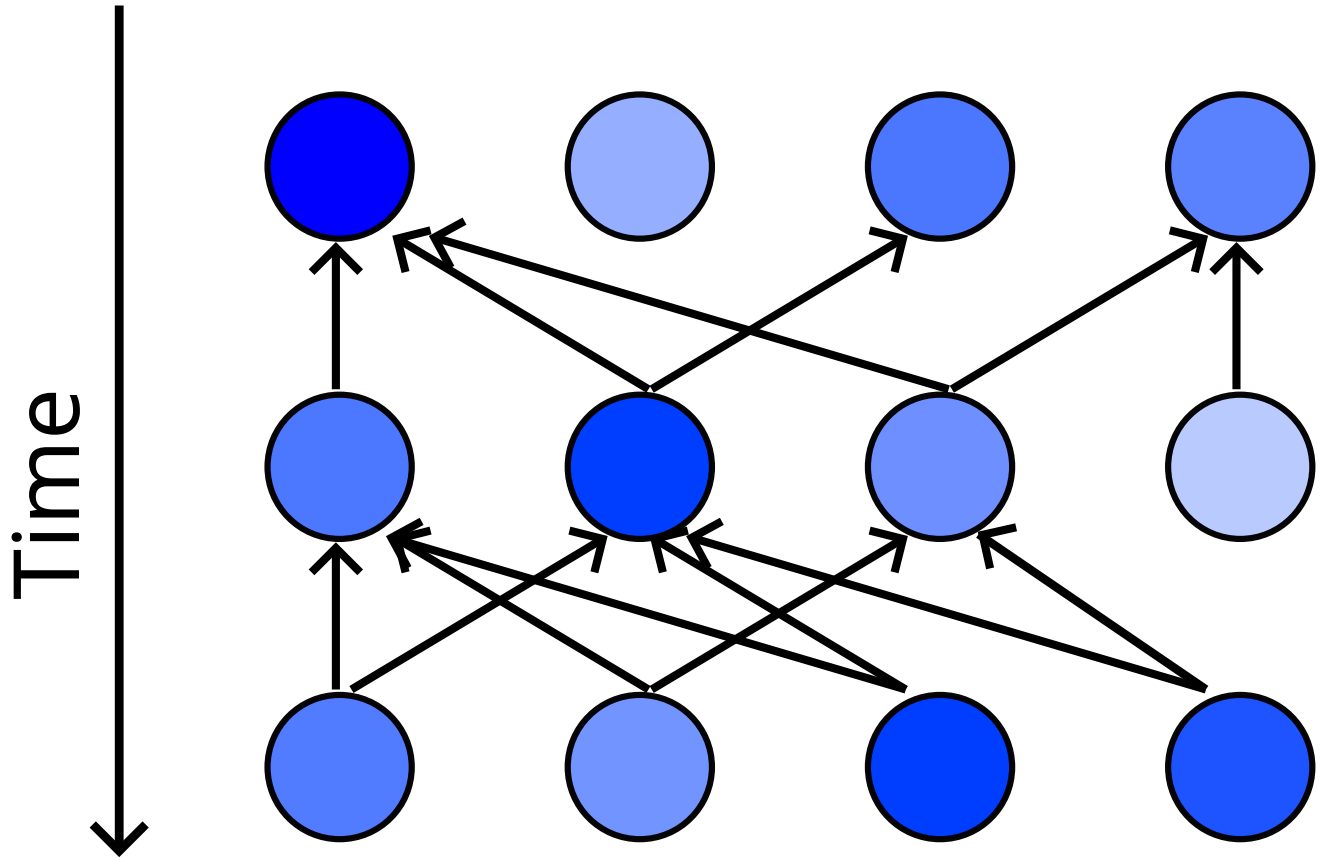
$$F(z) := \exp\left[-\frac{1}{2\omega^2}(z-\eta)^2\right]$$

where ω^{-2} is the **strength of selection** and η is the **selection optimum**.



The individual-based model

Fitness and reproduction



Every generation, organisms pick two parents at random (♣) with probability proportional to fitness (♠). The population size is fixed at N.

Summary of the parameters

N	Population size
L	Number of loci
μ_ℓ^+, μ_ℓ^-	Mutation probabilities at locus ℓ
α_ℓ	Additive effect at locus ℓ
ω^{-2}	Strength of selection
η	Selection optimum

Summary of the parameters

N	Population size
L	Number of loci
μ_ℓ^+, μ_ℓ^-	Mutation probabilities at locus ℓ
α_ℓ	Additive effect at locus ℓ
ω^{-2}	Strength of selection
η	Selection optimum

We will take $(\alpha_\ell, \mu_\ell^+, \mu_\ell^-)_{\ell \in [L]}$ to be an exchangeable vector of random variables, and assume $\sum_{\ell \in [L]} \alpha_\ell = 1$.

Summary of the parameters

N	Population size
L	Number of loci
μ_ℓ^+, μ_ℓ^-	Mutation probabilities at locus ℓ
α_ℓ	Additive effect at locus ℓ
ω^{-2}	Strength of selection
η	Selection optimum

We will take $(\alpha_\ell, \mu_\ell^+, \mu_\ell^-)_{\ell \in [L]}$ to be an exchangeable vector of random variables, and assume $\sum_{\ell \in [L]} \alpha_\ell = 1$.
This means $Z(G)$ is always in $[0, 2]$.

Three trait values

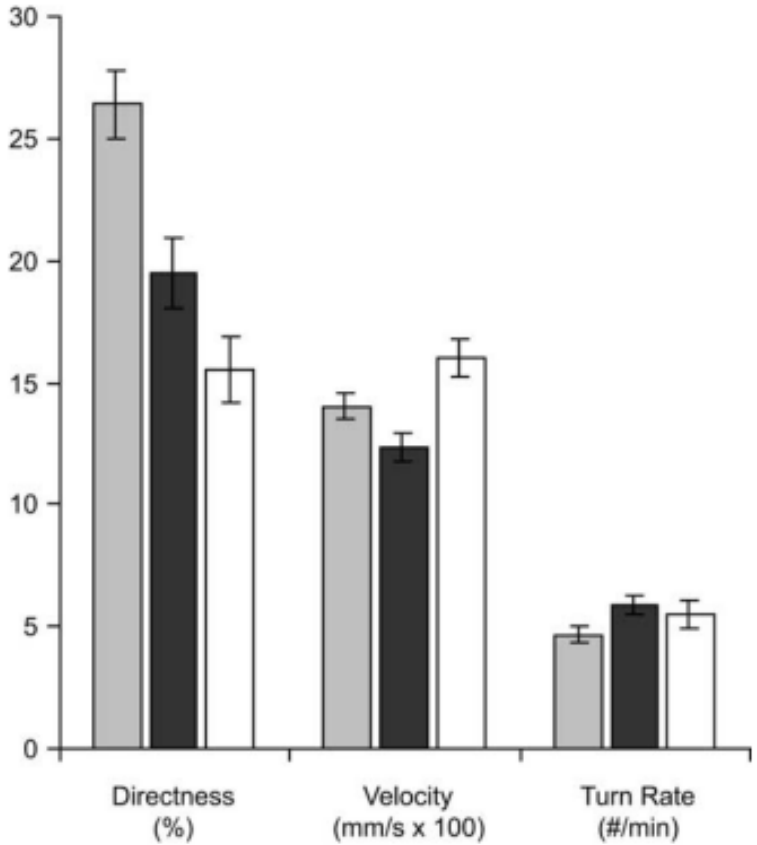
Three trait values of importance

The **selection optimum** is η.

The **heterozygote trait** is $\sum_{\ell \in [L]} \alpha_\ell$.

The **mutational optimum** is $\sum_{\ell \in [L]} 2\alpha_\ell \frac{\mu_\ell^+}{\mu_\ell^+ + \mu_\ell^-}$.

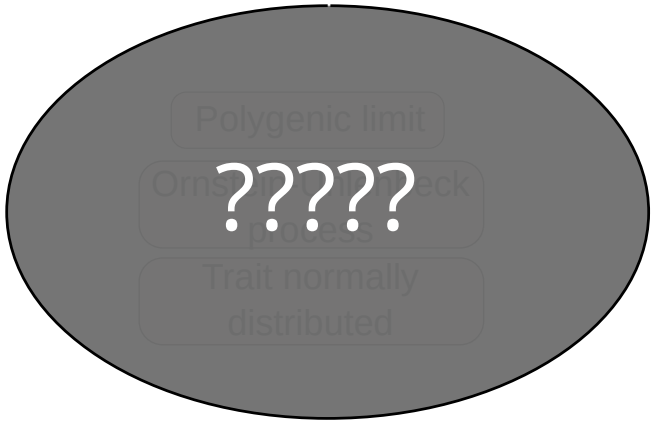
On mutational bias



From Ajie et al 2005.

The roadmap

Individual-based model
(L, N)



The polygenic limit

Features of the polygenic limit

There are three features which are expected to arise in the polygenic limit

- The trait is normally distributed (Quetelet 1835, Galton 1886, Fisher 1918), meaning if we know the population at present and sample a genome G uniformly at random, then $Z(G)$ is distributed as $\mathcal{N}(\bar{z}_t, \sigma^2)$.

Features of the polygenic limit

There are three features which are expected to arise in the polygenic limit

- The trait is normally distributed (Quetelet 1835, Galton 1886, Fisher 1918), meaning if we know the population at present and sample a genome G uniformly at random, then $Z(G)$ is distributed as $\mathcal{N}(\bar{z}_t, \sigma^2)$.
- The trait mean evolves as an Ornstein-Uhlenbeck process (Lande 1976)

$$\mathrm{d}\bar{z}_t = -\rho \bar{z}_t \mathrm{d}t + \nu \sqrt{2\rho} \mathrm{d}B_t$$

for some Brownian motion B_t and parameters ρ, ν.

Features of the polygenic limit

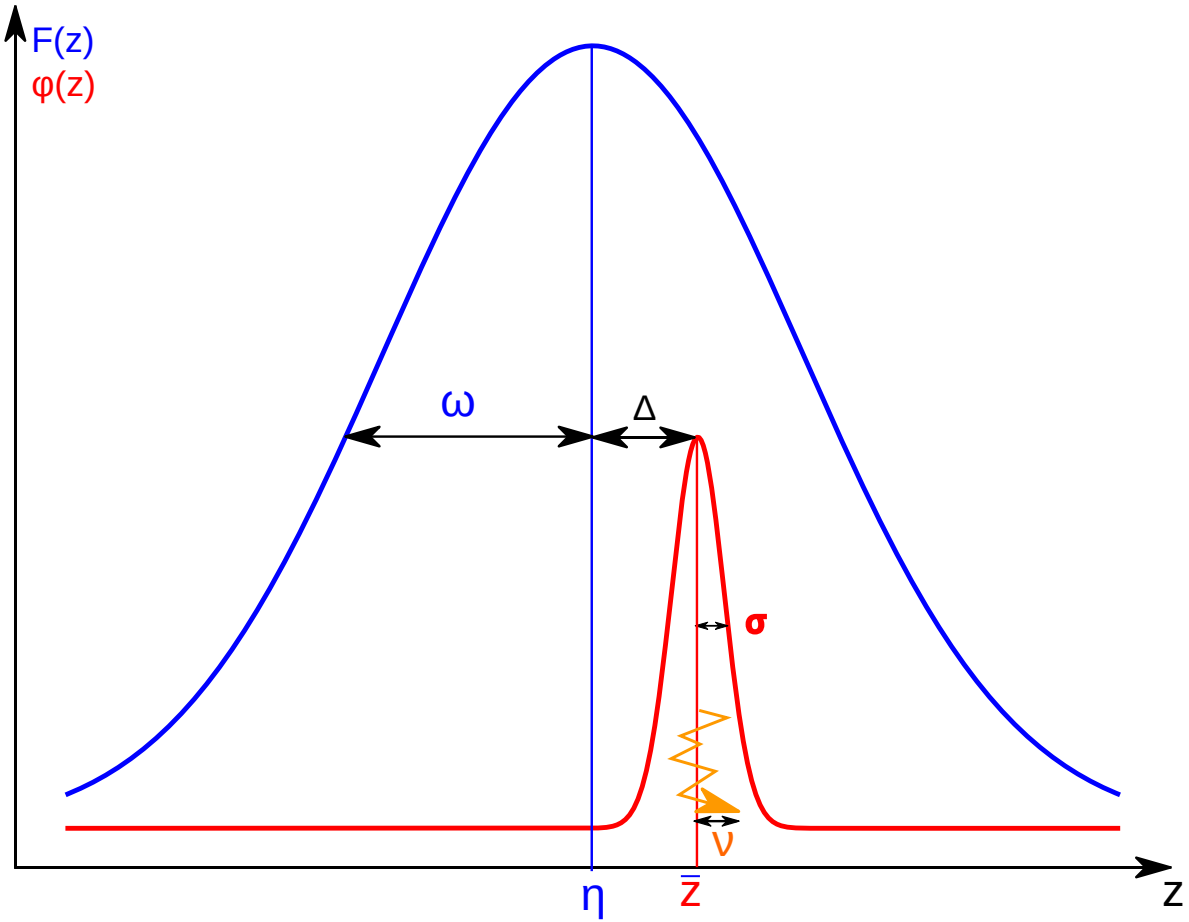
There are three features which are expected to arise in the polygenic limit

- The trait is normally distributed (Quetelet 1835, Galton 1886, Fisher 1918), meaning if we know the population at present and sample a genome G uniformly at random, then $Z(G)$ is distributed as $\mathcal{N}(\bar{z}_t, \sigma^2)$.
- The trait mean evolves as an Ornstein-Uhlenbeck process (Lande 1976)

$$\mathrm{d}\bar{z}_t = -\rho \bar{z}_t \mathrm{d}t + \nu \sqrt{2\rho} \mathrm{d}B_t$$

  for some Brownian motion B_t and parameters ρ, ν.
- The dynamics at a locus can be described with an autonomous equation.

Macroscopic observables



The roadmap

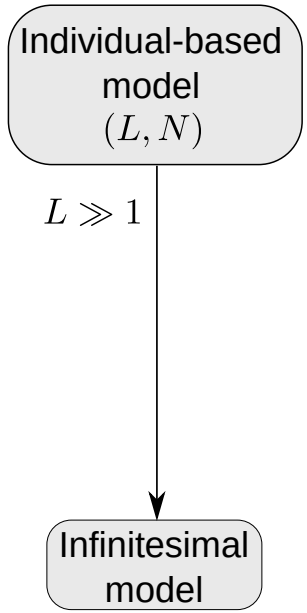
Individual-based model
(L, N)

Polygenic equation

Ornstein-Uhlenbeck process

Trait normally distributed

The roadmap



Polygenic equation

Ornstein-Uhlenbeck process

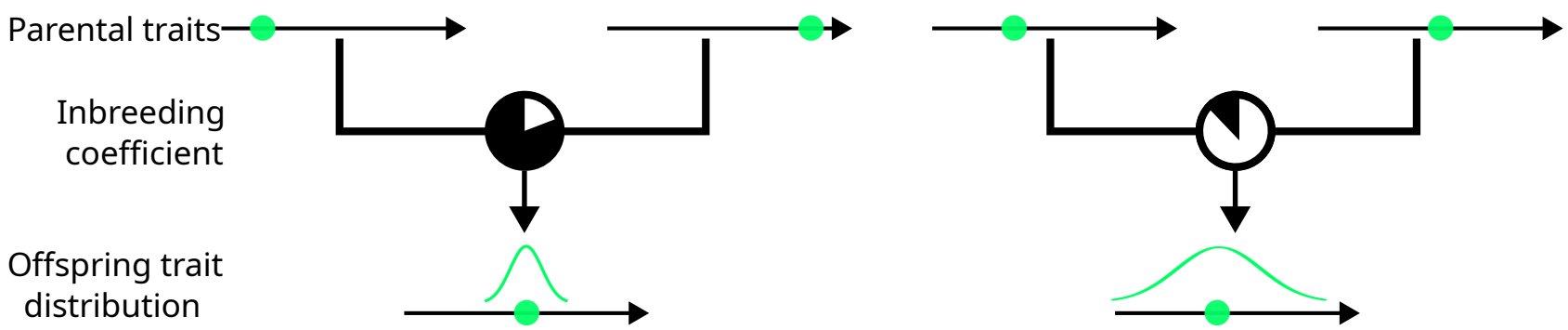
Trait normally distributed

The infinitesimal model

If we let $L \to +\infty$ in the previous model and scale the other parameters accordingly with N finite, we may obtain the infinitesimal model (Barton, Etheridge, Véber 2017).

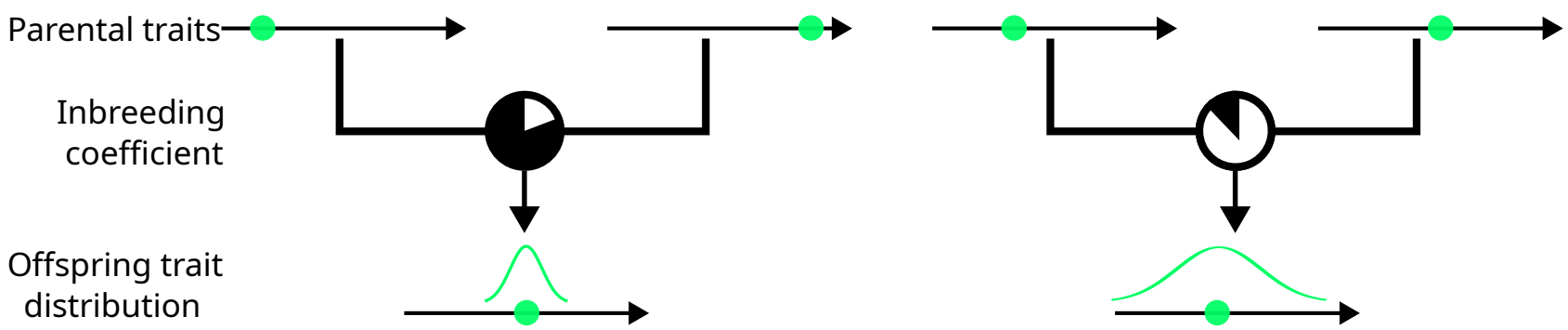
The infinitesimal model

If we let $L \to +\infty$ in the previous model and scale the other parameters accordingly with N finite, we may obtain the infinitesimal model (Barton, Etheridge, Véber 2017).



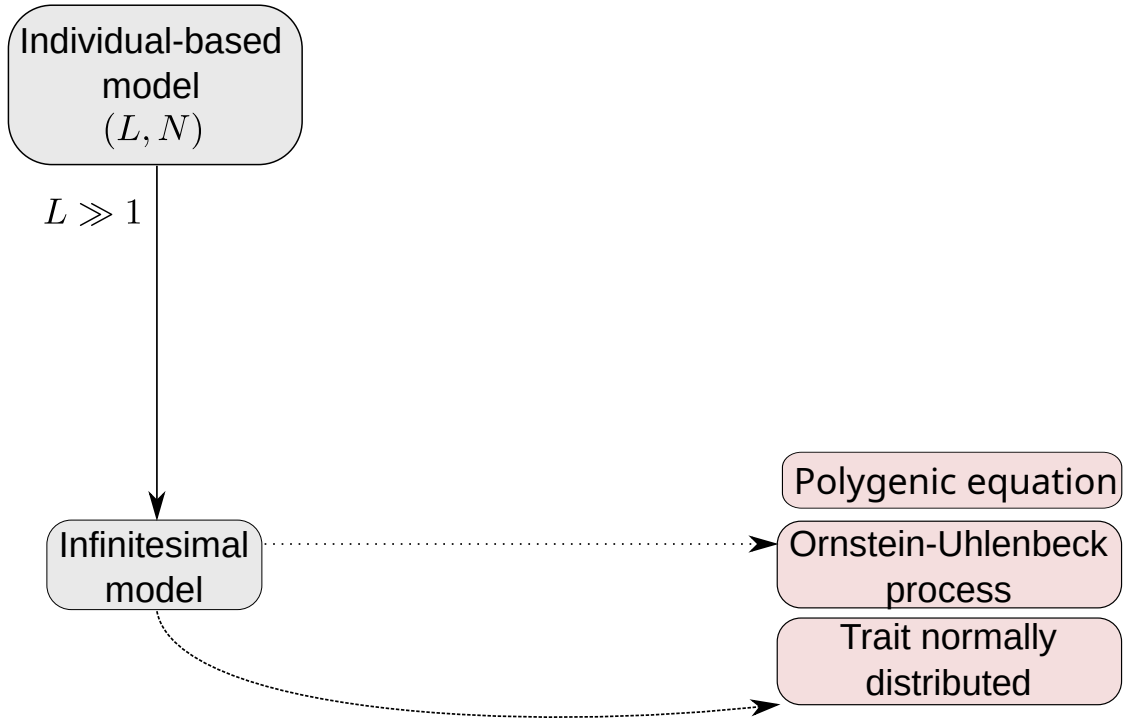
The infinitesimal model

If we let $L \to +\infty$ in the previous model and scale the other parameters accordingly with N finite, we may obtain the infinitesimal model (Barton, Etheridge, Véber 2017).

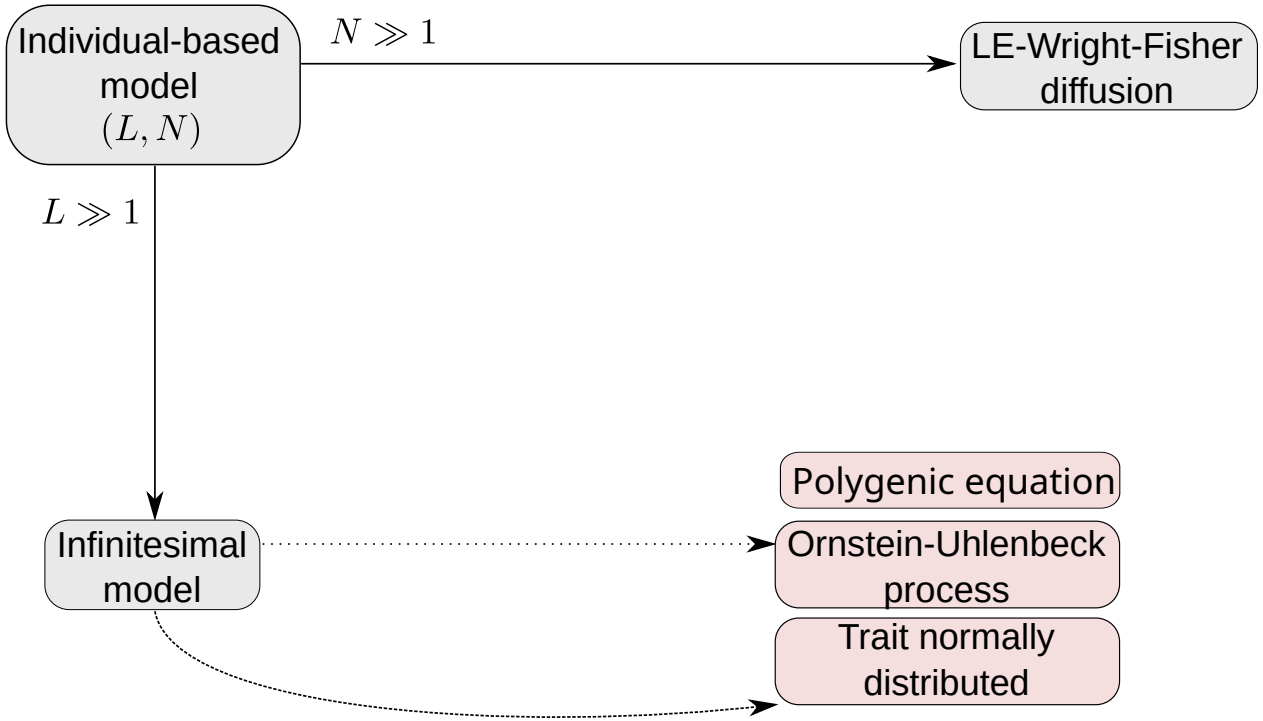


Key parameter: the segregation variance.

The roadmap

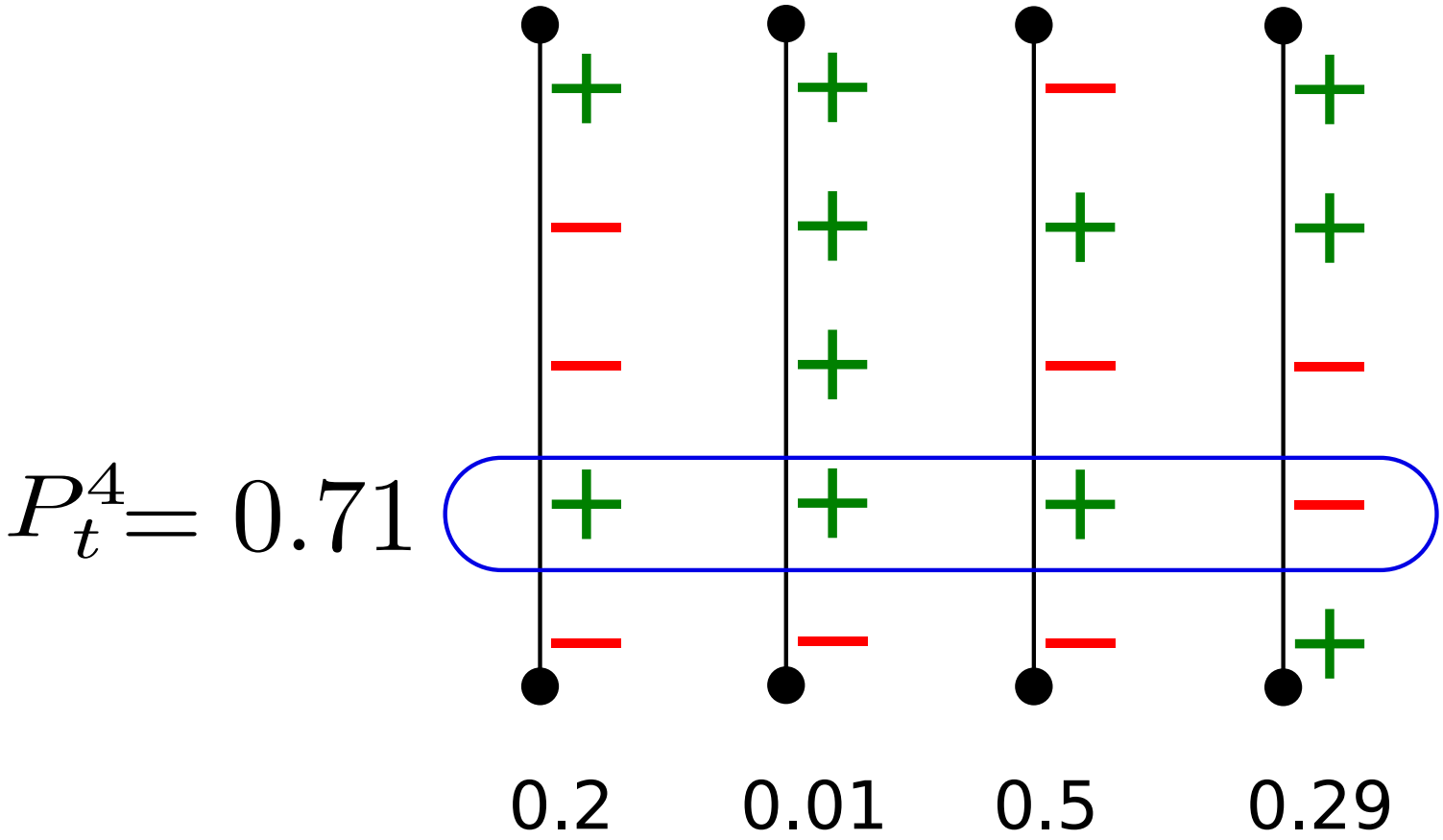

Individual-based model (L, N)
$L \gg 1$
Infinitesimal model
Polygenic equation
Ornstein-Uhlenbeck process
Trait normally distributed

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Infinitesimal model
Polygenic equation
Ornstein-Uhlenbeck process
Trait normally distributed

Taking the gene's eye-view

Let P_t^ℓ be the frequency of the $+$ allele at locus ℓ at generation $\lfloor 2Nt \rfloor$.



Taking the gene's eye-view

Step 0: rescaling

Let P_t^ℓ be the frequency of the $+$ allele at locus ℓ at generation $\lfloor 2Nt \rfloor$, with $\mathbf{P}_t = (P_t^\ell)_{\ell \in [L]}$.

We also define

$$\theta_\ell^\pm := 2N\mu_\ell^\pm \qquad\qquad \omega_e^{-2} := 2N\omega^{-2}$$

First and second moment computations suggest the dynamics of $(P_t^\ell)_{t\geq 0}$ can be approximated with the SDE

$$\mathrm{d}P_t^\ell = S^\ell(\mathbf{X}_t)\underset{\spadesuit}{P_t^\ell}(1-P_t^\ell)\mathrm{d}t + \Big((1-P_t^\ell)\underset{\diamondsuit}{\theta_\ell^+} - P_t^\ell\theta_\ell^-\Big)\mathrm{d}t + \sqrt{\underset{\clubsuit}{P_t^\ell(1-P_t^\ell)}}\mathrm{d}B_t^\ell$$

Taking the gene's eye-view

Step 1: the diffusion approximation

First and second moment computations suggest the dynamics of $(P_t^\ell)_{t\geq 0}$ can be approximated with the SDE

$$\mathrm{d}P_t^\ell = S^\ell(\mathbf{X}_t)\underset{\spadesuit}{P_t^\ell(1-P_t^\ell)}\mathrm{d}t + \Big(\underset{\blacklozenge}{(1-P_t^\ell)\theta_\ell^+} - P_t^\ell\theta_\ell^-\Big)\mathrm{d}t + \sqrt{\underset{\clubsuit}{P_t^\ell(1-P_t^\ell)}}\mathrm{d}B_t^\ell$$

where we have (Robertson 1966)

$$S^\ell(\mathbf{X}_t) := 2N\frac{\mathbf{Cov}_{\mathbf{X}_t}[W(g), g_\ell]}{\mathbf{Var}_{\mathbf{X}_t}[g_\ell]}$$

First and second moment computations suggest the dynamics of $(P_t^\ell)_{t\geq 0}$ can be approximated with the SDE

$$\mathrm{d}P_t^\ell = S^\ell(\mathbf{X}_t)P_t^\ell(1-P_t^\ell)\mathrm{d}t + \left((1-P_t^\ell)\theta_\ell^+ - P_t^\ell\theta_\ell^-\right)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell$$

♠ ♦ ♣

where we have (Robertson 1966)

$$S^\ell(\mathbf{X}_t) := 2N\frac{\mathbf{Cov}_{\mathbf{X}_t}[W(g), g_\ell]}{\mathbf{Var}_{\mathbf{X}_t}[g_\ell]}$$

where $W(g) := \ln F(Z(g))$ is the **logfitness** and under $\mathbf{X}_t$, $g = (g_\ell)_{\ell\in[L]} \in \{0,1,2\}^L$ is the unphased genome of a randomly sampled organism at generation $\lfloor 2Nt \rfloor$ (this means $g_\ell = i$ if the organism has i copies of the $+$ allele at locus ℓ for $i \in [2]$).

Taking the gene's eye-view

Step 2: Hardy-Weinberg Linkage Equilibrium (HWLE)

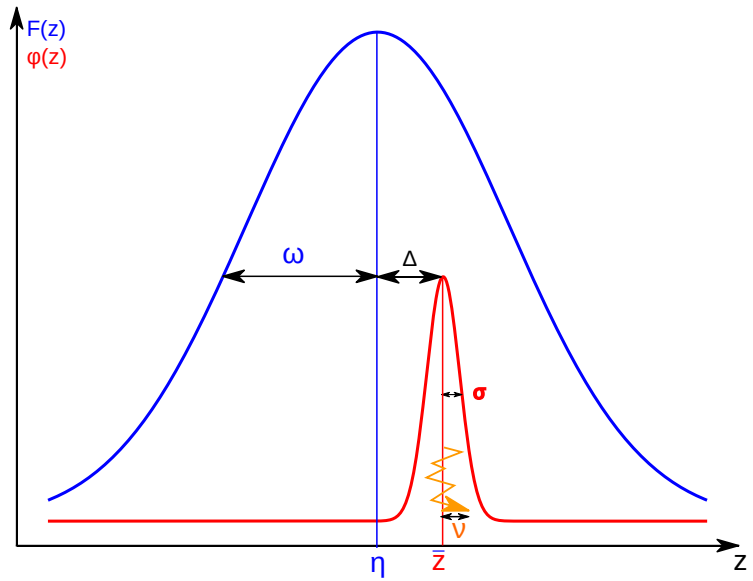
If recombination (♥) is strong then the population is close to HWLE, which means under $\mathbf{X}_t$, the $(g_\ell)_{\ell \in [L]}$ are independent $Binomial(2, P_t^\ell)$ variables.

Taking the gene's eye-view

Step 2: Hardy-Weinberg Linkage Equilibrium (HWLE)

If recombination (♥) is strong then the population is close to HWLE, which means under $\mathbf{X}_t$, the $(g_\ell)_{\ell\in[L]}$ are independent $Binomial(2, P_t^\ell)$ variables.
If we assume HWLE (Wright 1935) then we have

$$S^\ell(\mathbf{X}_t) = -\alpha_\ell \frac{\Delta_t}{\omega_e^2} + \frac{\alpha_\ell^2}{\omega_e^2}\left(P_t^\ell - \frac{1}{2}\right) =: \xi_{\alpha_\ell,\Delta_t}(P_t^\ell).$$



Taking the gene's eye-view

Step 3: LE-Wright-Fisher diffusion

We have

$$\mathrm{d}P_t^\ell = \xi_{\Delta_t,\alpha_\ell}(P_t^\ell)\underset{\spadesuit}{P_t^\ell(1-P_t^\ell)}\mathrm{d}t + \Big(\underset{\diamondsuit}{(1-P_t^\ell)\theta^+} - P_t^\ell\theta^-\Big)\mathrm{d}t + \sqrt{\underset{\clubsuit}{P_t^\ell(1-P_t^\ell)}}\mathrm{d}B_t^\ell$$

with $\Delta_t = \sum_{\ell\in[L]} 2\alpha_\ell P_t^\ell \ - \eta$.

Taking the gene's eye-view

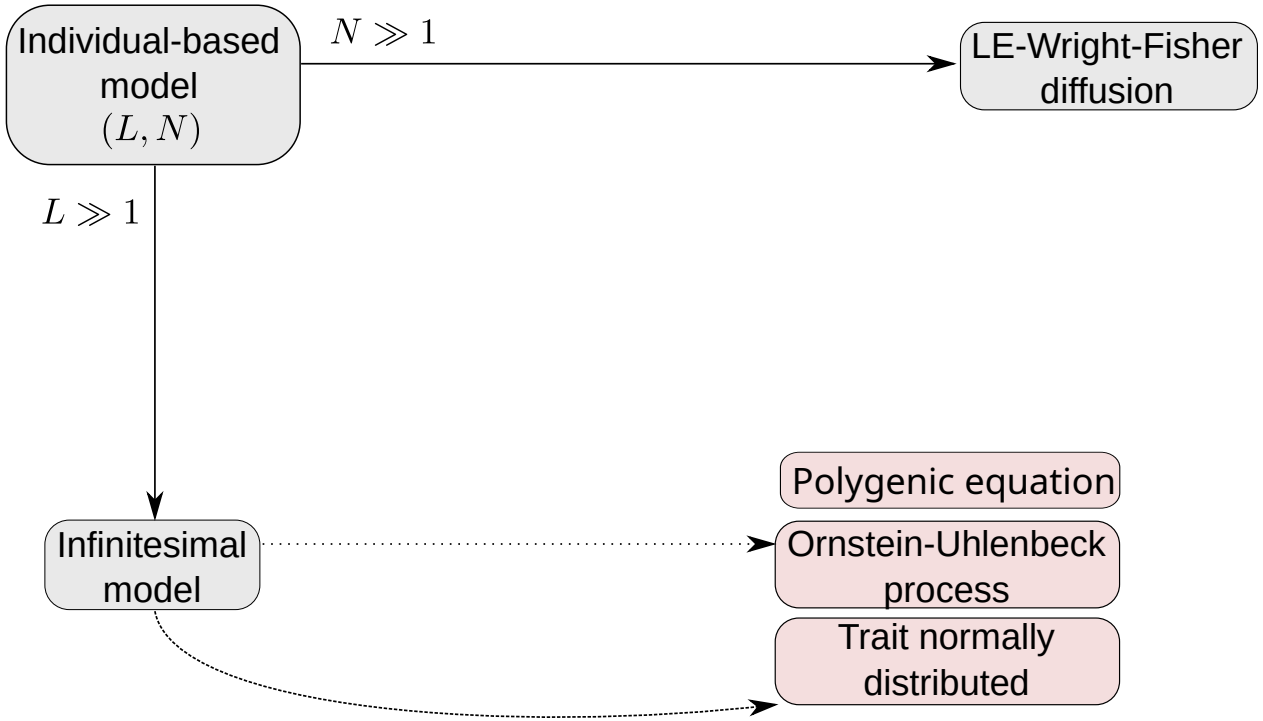
Step 3: LE-Wright-Fisher diffusion

We have

$$\mathrm{d}P_t^\ell = \xi_{\Delta_t,\alpha_\ell}(P_t^\ell)\underset{\spadesuit}{P_t^\ell(1-P_t^\ell)}\mathrm{d}t + \Big(\underset{\diamondsuit}{(1-P_t^\ell)\theta^+} - P_t^\ell\theta^-\Big)\mathrm{d}t + \sqrt{\underset{\clubsuit}{P_t^\ell(1-P_t^\ell)}}\mathrm{d}B_t^\ell$$

with $\Delta_t = \sum_{\ell\in[L]} 2\alpha_\ell P_t^\ell - \eta$.
We call this the **LE-Wright-Fisher diffusion**.

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Infinitesimal model
Polygenic equation
Ornstein-Uhlenbeck process
Trait normally distributed

Define the **typical locus** $(\vec{P}_t)_{t\geq 0} = (P_t, \alpha, \theta^{\pm})_{t\geq 0}$ is defined as $(P_t^{\ell_U}, \alpha_{\ell_U}, \theta^{\pm}_{\ell_U})$ where ℓ_U is picked uniformly at random on $[L]$.

Define the **typical locus** $(\vec{P}_t)_{t\geq 0} = (P_t, \alpha, \theta^{\pm})_{t\geq 0}$ is defined as $(P_t^{\ell_U}, \alpha_{\ell_U}, \theta^{\pm}_{\ell_U})$ where ℓ_U is picked uniformly at random on $[L]$.

We have

$$\mathrm{d}P_t = \xi_{\Delta_t,\alpha}(P_t)\underset{\spadesuit}{P_t(1-P_t)}\mathrm{d}t + \Big(\underset{\blacklozenge}{(1-P_t)\theta^+} - P_t\theta^-\Big)\mathrm{d}t + \sqrt{\underset{\clubsuit}{P_t(1-P_t)}}\mathrm{d}B_t \tag{1}$$

Define the **typical locus** $(\vec{P}_t)_{t\geq 0} = (P_t, \alpha, \theta^{\pm})_{t\geq 0}$ is defined as $(P_t^{\ell_U}, \alpha_{\ell_U}, \theta^{\pm}_{\ell_U})$ where ℓ_U is picked uniformly at random on $[L]$.

We have

$$\mathrm{d}P_t = \xi_{\Delta^*,\alpha}(P_t)\underset{\spadesuit}{P_t(1-P_t)}\mathrm{d}t + \Big(\underset{\blacklozenge}{(1-P_t)\theta^+} - P_t\theta^-\Big)\mathrm{d}t + \sqrt{\underset{\clubsuit}{P_t(1-P_t)}}\mathrm{d}B_t \tag{1}$$

If $L \gg 1$, we may replace Δ_t with

$$\Delta^* = \mathbb{E}[\Delta_t] = 2L\mathbb{E}[\alpha P_t] - \eta.$$

Taking the gene's eye-view

Step 5: fixed-point equation

If we know Δ^*, then P_t conditional on α, θ has stationary distribution $\Pi_{\Delta^*,\alpha,\theta}$ where for $\delta \in \mathbb{R}$

$$\Pi_{\delta,\alpha,\theta}(\mathrm{d}p) \quad \propto \quad p^{2\theta^+-1}(1-p)^{2\theta^--1} e^{\int_0^p \xi_{\delta,\alpha}(p')\mathrm{d}p'} \mathrm{d}p$$

Taking the gene's eye-view

Step 5: fixed-point equation

If we know Δ^*, then P_t conditional on α, θ has stationary distribution $\Pi_{\Delta^*,\alpha,\theta}$ where for $\delta \in \mathbb{R}$

$$\Pi_{\delta,\alpha,\theta}(\mathrm{d}p) \quad \propto \quad p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{\int_0^p \xi_{\delta,\alpha}(p')\mathrm{d}p'}\mathrm{d}p$$

For $\delta \in \mathbb{R}$, define

$$I(\delta) := \mathbb{E}\left[\alpha \int p\Pi_{\delta,\alpha,\theta}(\mathrm{d}p)\right]$$

where the expectation is with respect to (α, θ).

Taking the gene's eye-view

Step 5: fixed-point equation

If we know Δ^*, then P_t conditional on α, θ has stationary distribution $\Pi_{\Delta^*,\alpha,\theta}$ where for $\delta \in \mathbb{R}$

$$\Pi_{\delta,\alpha,\theta}(\mathrm{d}p) \quad \propto \quad p^{2\theta^+-1}(1-p)^{2\theta^--1} e^{\int_0^p \xi_{\delta,\alpha}(p')\mathrm{d}p'} \mathrm{d}p$$

For $\delta \in \mathbb{R}$, define

$$I(\delta) := \mathbb{E}\left[\alpha \int p \Pi_{\delta,\alpha,\theta}(\mathrm{d}p)\right]$$

where the expectation is with respect to (α, θ).
Then

$$\Delta^* = 2LI(\Delta^*) - \eta.$$

Taking the gene's eye-view

Step 5: fixed-point equation

If we know Δ^*, then P_t conditional on α, θ has stationary distribution $\Pi_{\Delta^*,\alpha,\theta}$ where for $\delta \in \mathbb{R}$

$$\Pi_{\delta,\alpha,\theta}(\mathrm{d}p) \quad \propto \quad p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{\int_0^p \xi_{\delta,\alpha}(p')\mathrm{d}p'}\mathrm{d}p$$

For $\delta \in \mathbb{R}$, define

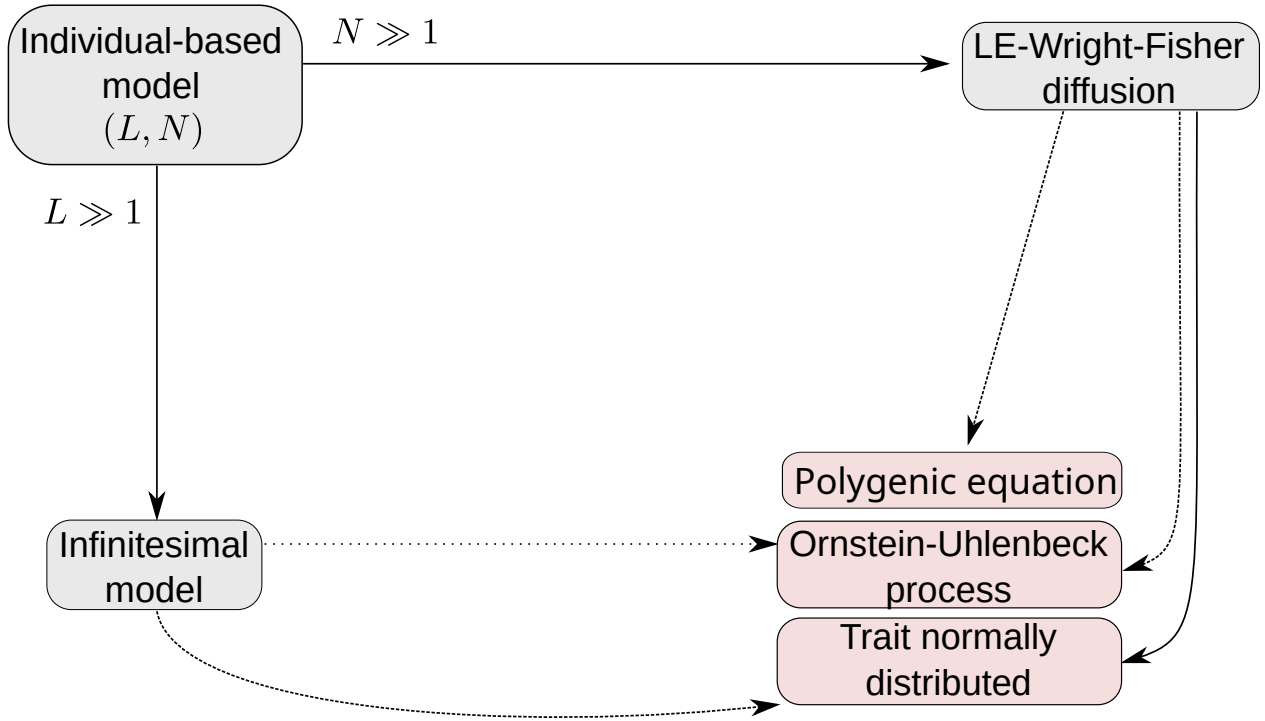
$$I(\delta) := \mathbb{E}\left[\alpha \int p\Pi_{\delta,\alpha,\theta}(\mathrm{d}p)\right]$$

where the expectation is with respect to (α, θ).
Then

$$\Delta^* = 2LI(\Delta^*) - \eta.$$

(when $\Delta^* \ll 1$ and θ^+, θ^- are constant across loci and small, this equation was written and solved by Charlesworth 2013).

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

Orders of magnitude

Reminders

L	Number of loci
$\theta_\ell^+, \theta_\ell^-$	Mutation rates at locus ℓ
α_ℓ	Additive effect at locus ℓ
ω_e^{-2}	Effective strength of selection
η	Selection optimum

Reminders

L	Number of loci
$\theta_\ell^+, \theta_\ell^-$	Mutation rates at locus ℓ
α_ℓ	Additive effect at locus ℓ
ω_e^{-2}	Effective strength of selection
η	Selection optimum

Define the mean mutation rate as $|\bar{\theta}| = \mathbb{E}[\theta^+ + \theta^-]$.

Assumptions

A 1: Uniform boundedness
There is a $C \asymp 1$ such that for any $\ell \in [L]$, $|\theta_\ell| \leq C|\bar{\theta}|$ and $\alpha_\ell \leq \frac{C}{L}$.

A 2: Mutations smaller than genetic drift $|\bar{\theta}| \lesssim 1$.

A 3: Mutational bias not too extreme
There is a constant $C \asymp 1$ such that for any ℓ, $\theta_\ell^- / C \leq \theta_\ell^+ \leq C\theta_\ell^-$.

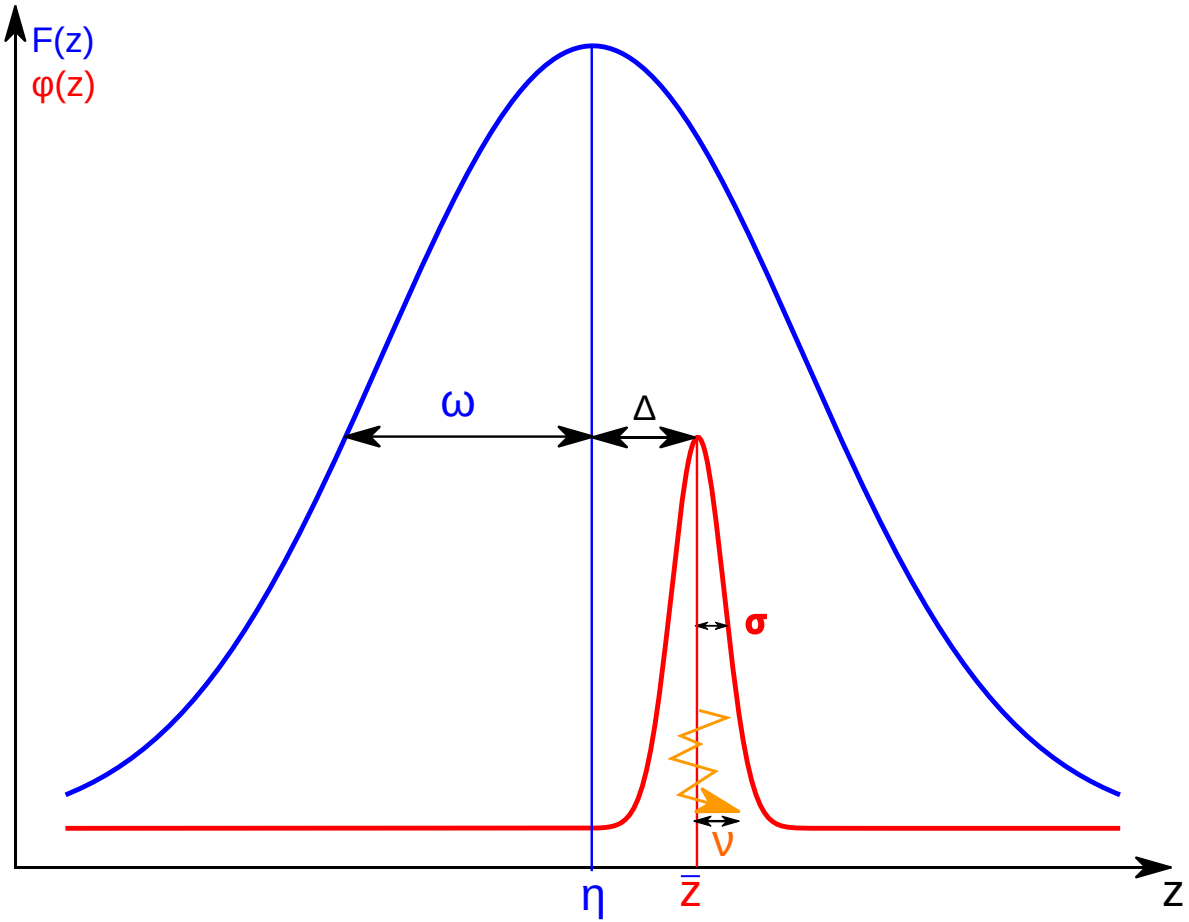
A 4: Accessibility of the selection optimum
$\eta \in (0, 2)$ satisfies $\eta(2 - \eta) \asymp 1$.

A 5: Weak/moderate/strong selection $L \lesssim \omega_e^{-2} \lesssim L^2$.

A 6: Distance between the selection and the mutation optimum
We have

$$\left|2L\ I(0) - \eta\right| \asymp 1.$$

Reminders



Orders of magnitude

In the polygenic limit, under (A1-6), we have

$$\begin{aligned} |\Delta^*| &\asymp L\omega_e^2 \\ \sigma^2 &\asymp \frac{|\bar{\theta}|}{L} \\ \nu^2 &\asymp \omega_e^2 \\ \rho &\asymp \frac{|\bar{\theta}|}{L\omega_e^2} \end{aligned}$$

The distance to the optimum is of the same order as its fluctuations

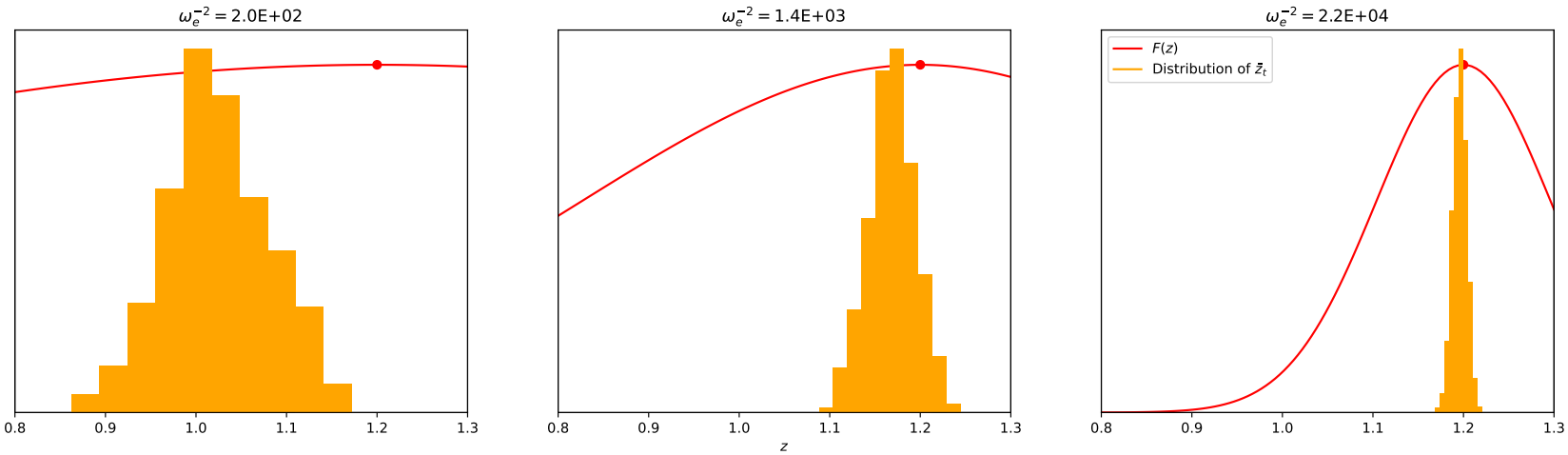


Figure: $N = 100, L = 100, \theta = (0.1, 0.2), \eta = 1.2$, and $(\alpha_\ell)_{\ell \in [L]}$ were sampled with distribution *Exponential*(L)

Bias-countering selection coefficient

Recall the selection coefficient at the typical locus is

$$\xi_{\alpha,\Delta^*}(P_t) := -\alpha\frac{\Delta^*}{\omega_e^2} + \frac{\alpha^2}{\omega_e^2}\left(P_t - \frac{1}{2}\right)$$

Bias-countering selection coefficient

Recall the selection coefficient at the typical locus is

$$\xi_{\alpha,\Delta^*}(P_t) := L\alpha s^* + \frac{\alpha^2}{\omega_e^2}\left(P_t - \frac{1}{2}\right)$$

with

$$s^* := -\frac{\Delta_t}{L\omega_e^2} \asymp 1.$$

Bias-countering selection coefficient

Recall the selection coefficient at the typical locus is

$$\xi_{\alpha,\Delta^*}(P_t) := L\alpha s^* + \frac{\alpha^2}{\omega_e^2}\left(P_t - \frac{1}{2}\right)$$

with

$$s^* := -\frac{\Delta_t}{L\omega_e^2} \asymp 1.$$

When $L \ll \omega_e^{-2} \ll L^2$ (moderate selection), the value of s^* is independent of the strength of selection ω_e^{-2}.

Weak/moderate stabilizing selection has the same effect as directional selection

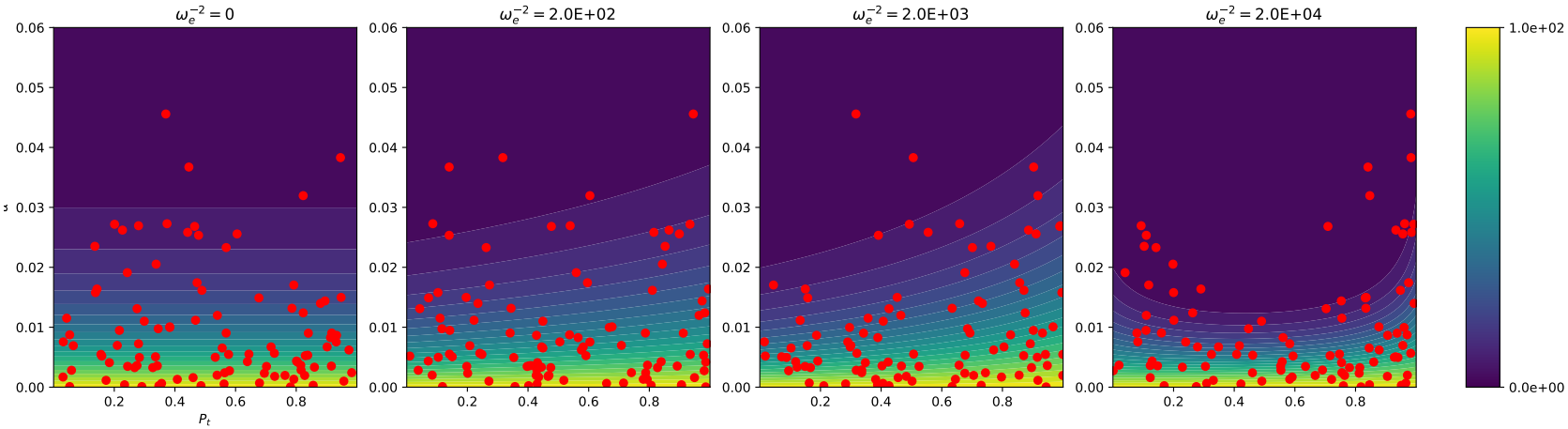


Figure: Parameters: $N = 500$, $L = 100$, $\theta = (0.5, 0.5)$, $\eta = 1.2$ and α_ℓ is distributed as *Exponential*(L).

Macroscopic observables are well predicted by the fixed point equation

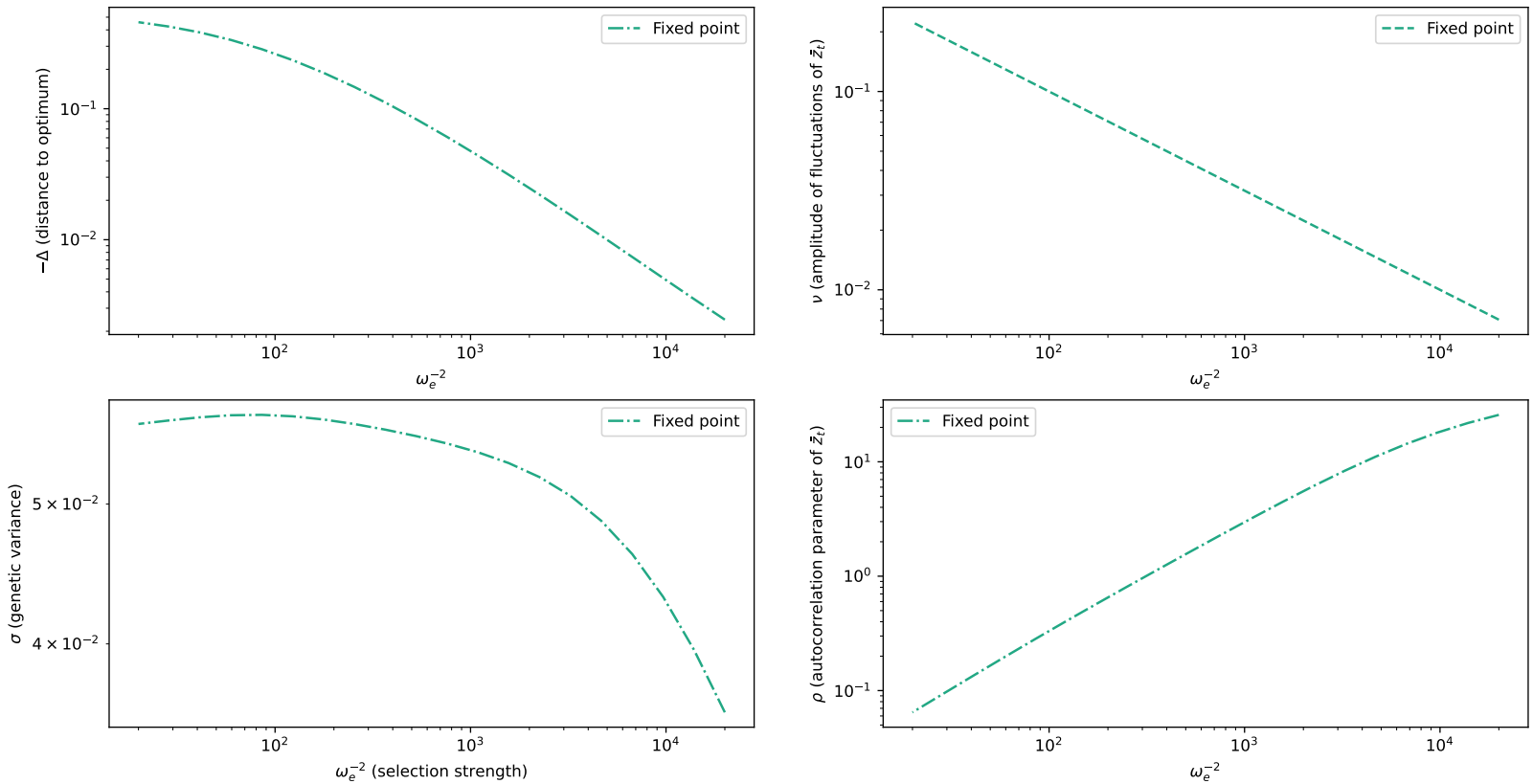


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

Macroscopic observables are well predicted by the fixed point equation

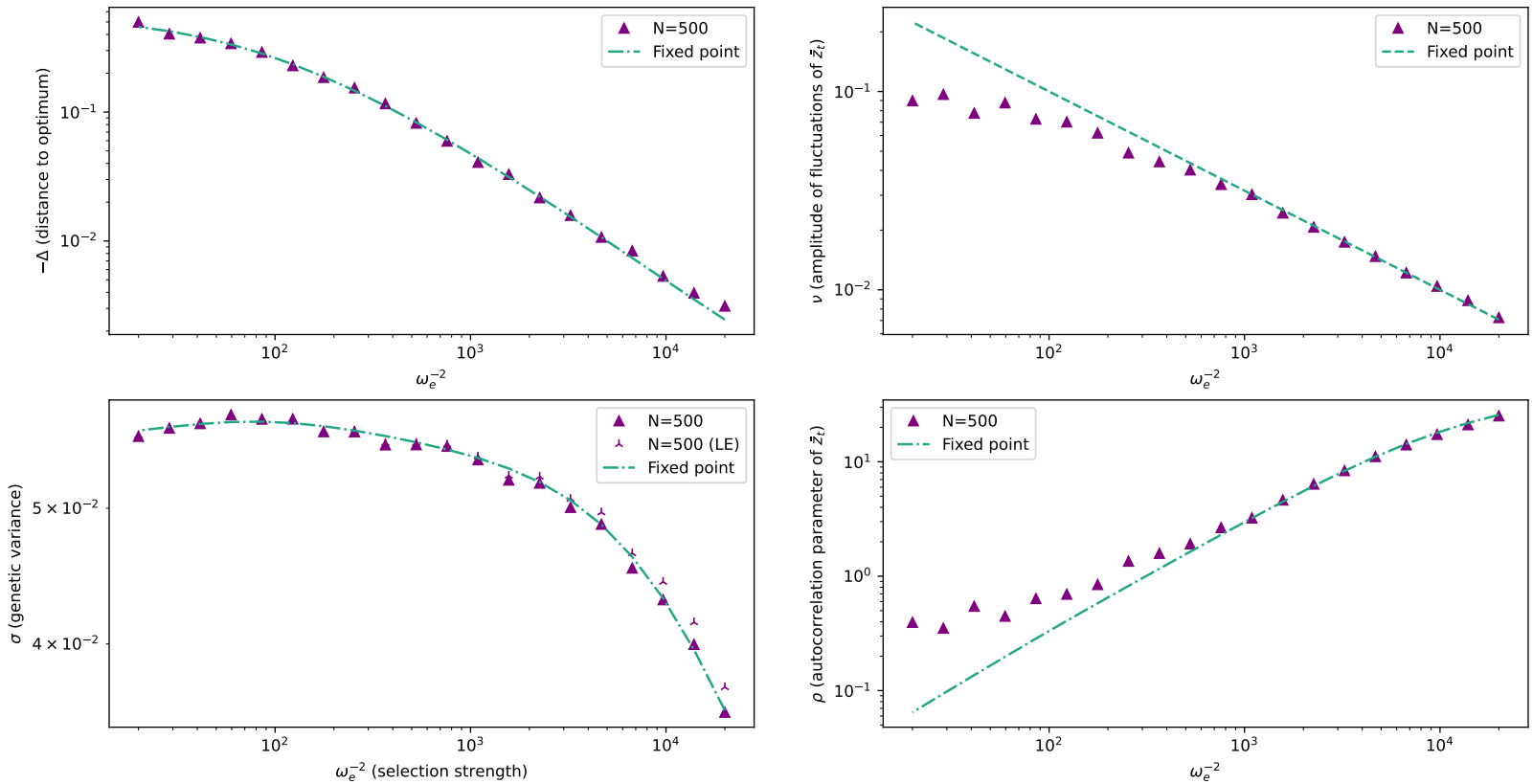


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

Macroscopic observables are well predicted by the fixed point equation

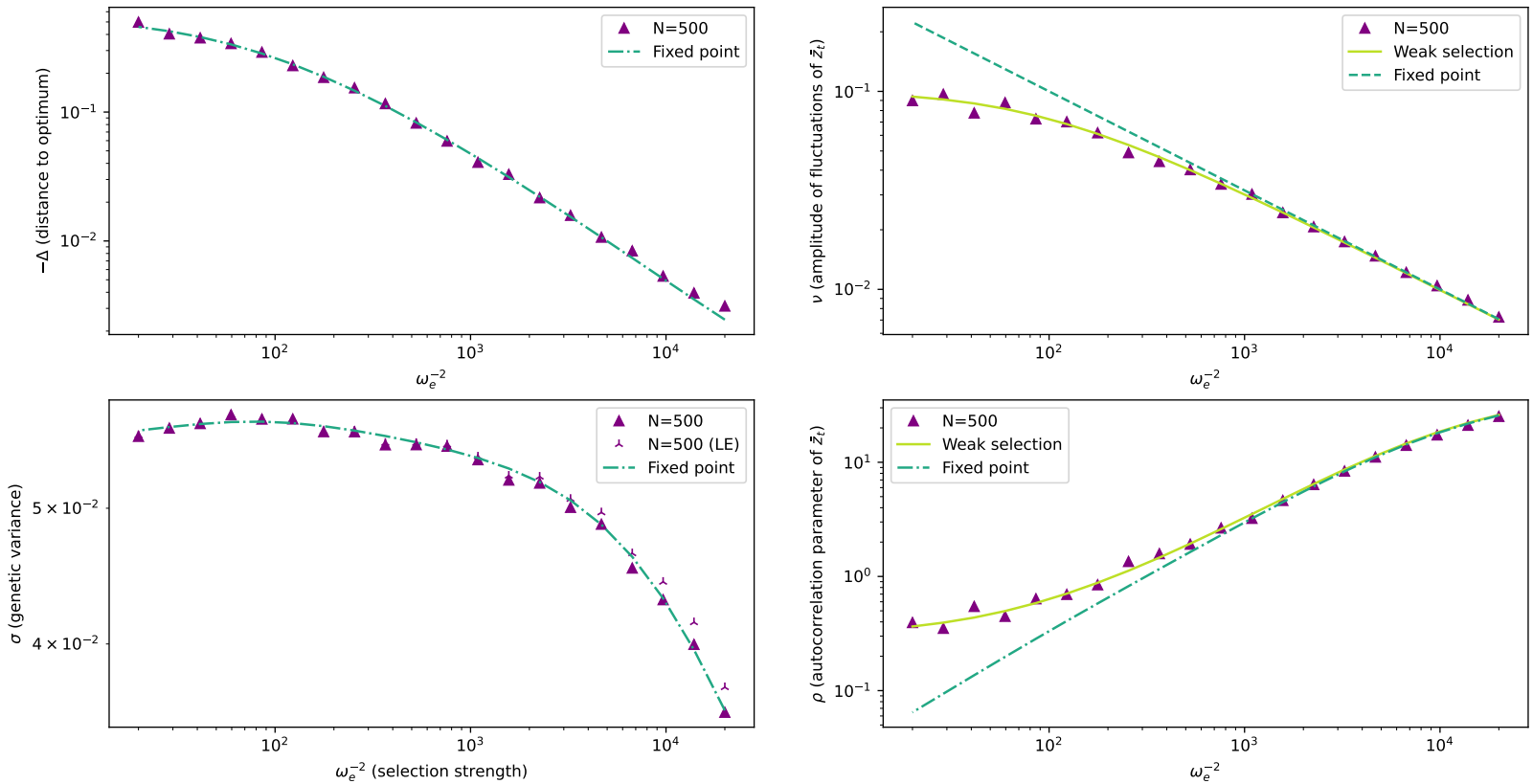


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

Macroscopic observables are well predicted by the fixed point equation

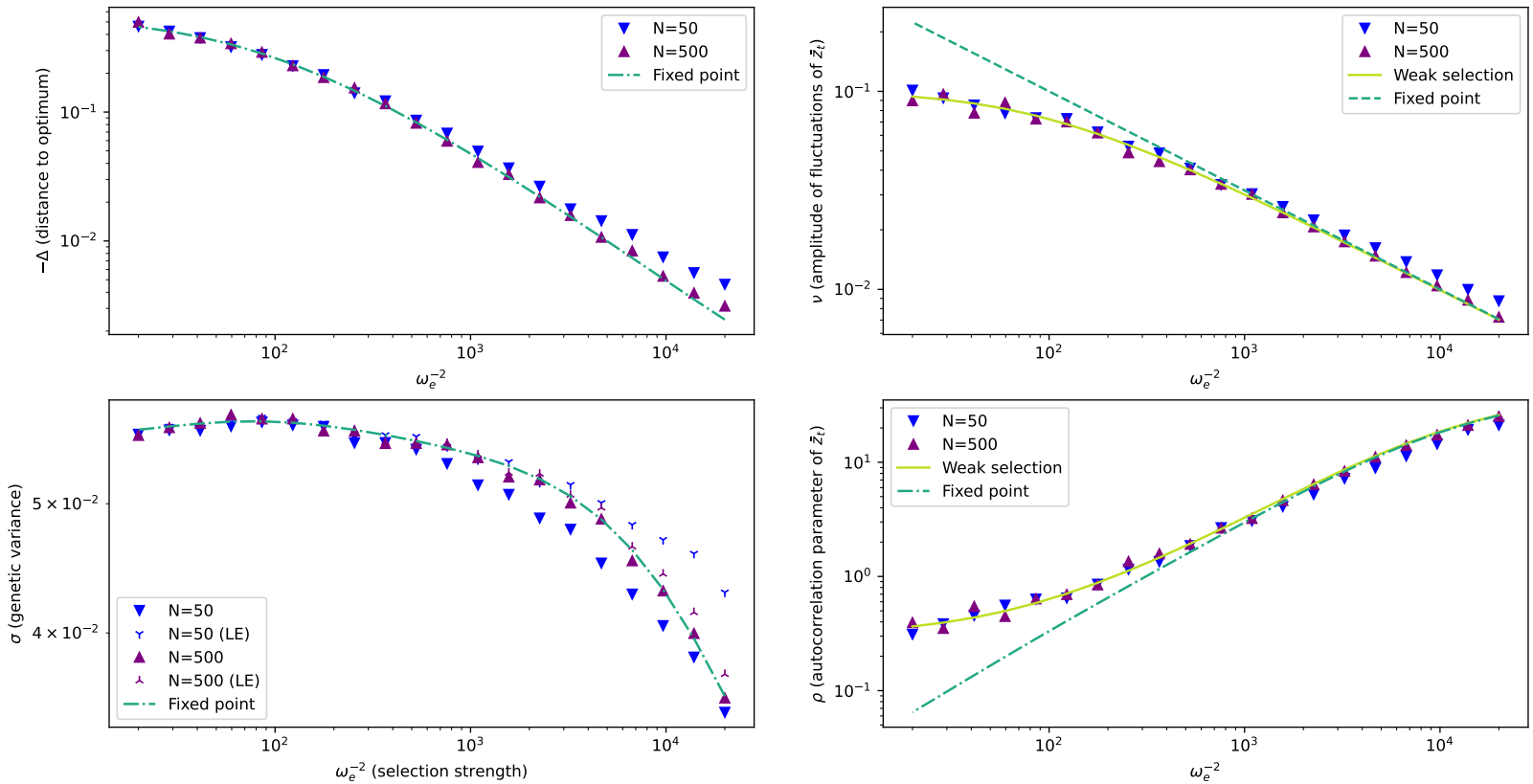
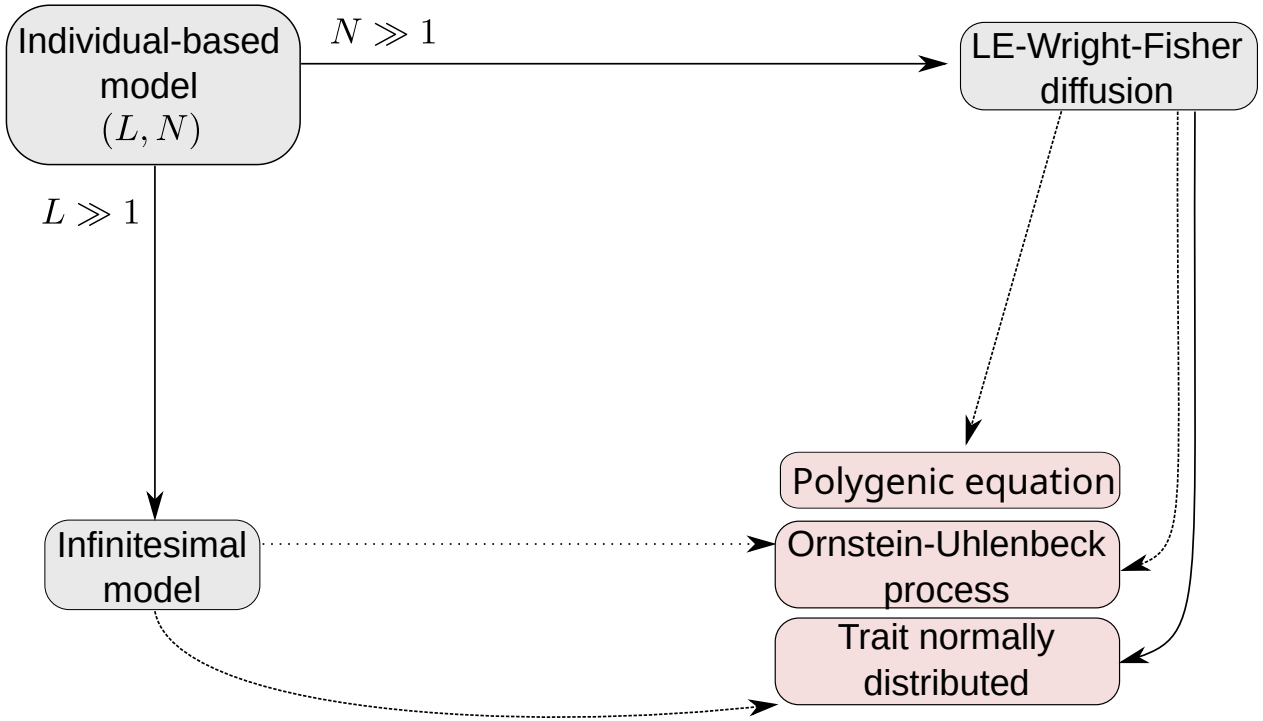
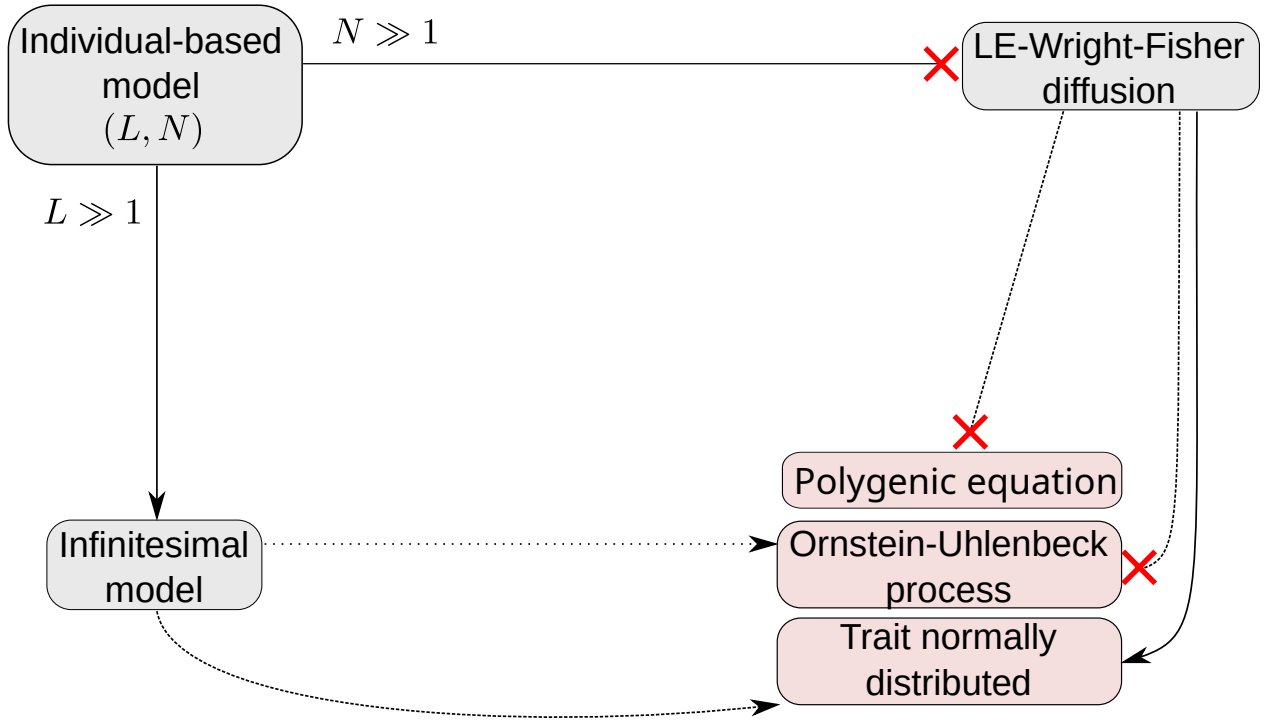


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

The roadmap

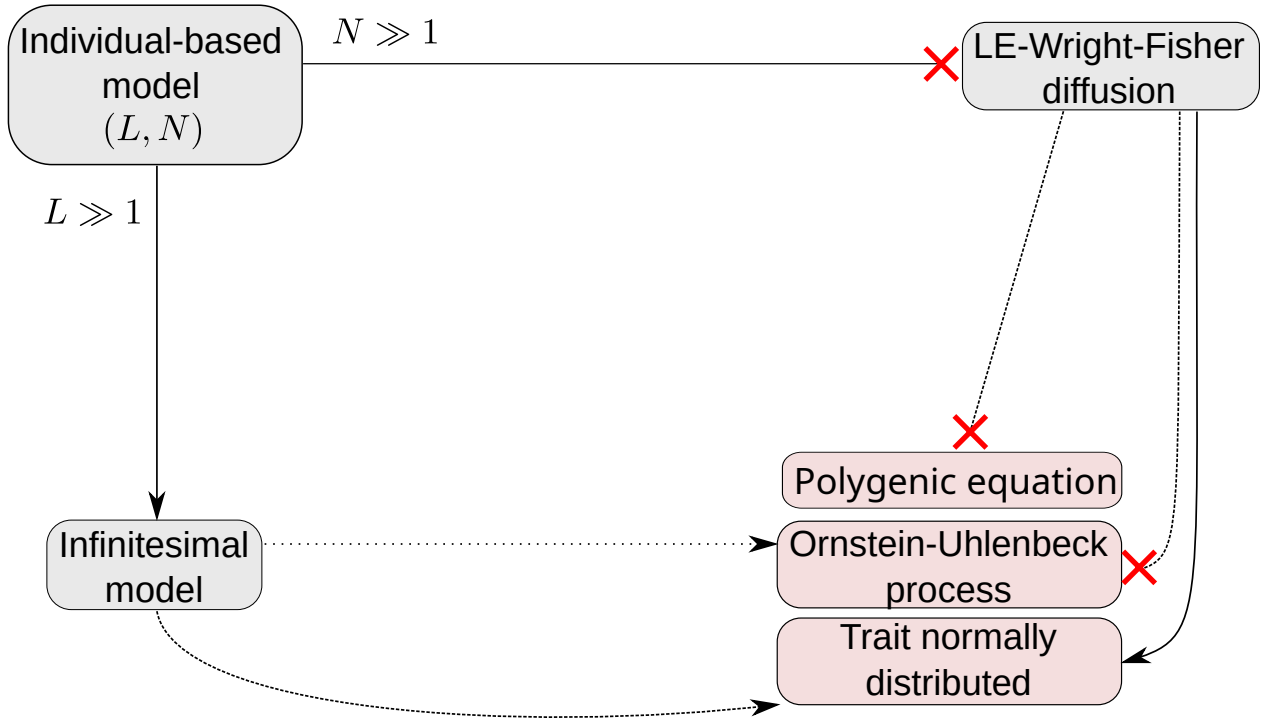

Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

Breakdown of the polygenic limit

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

The Bulmer effect (1971)

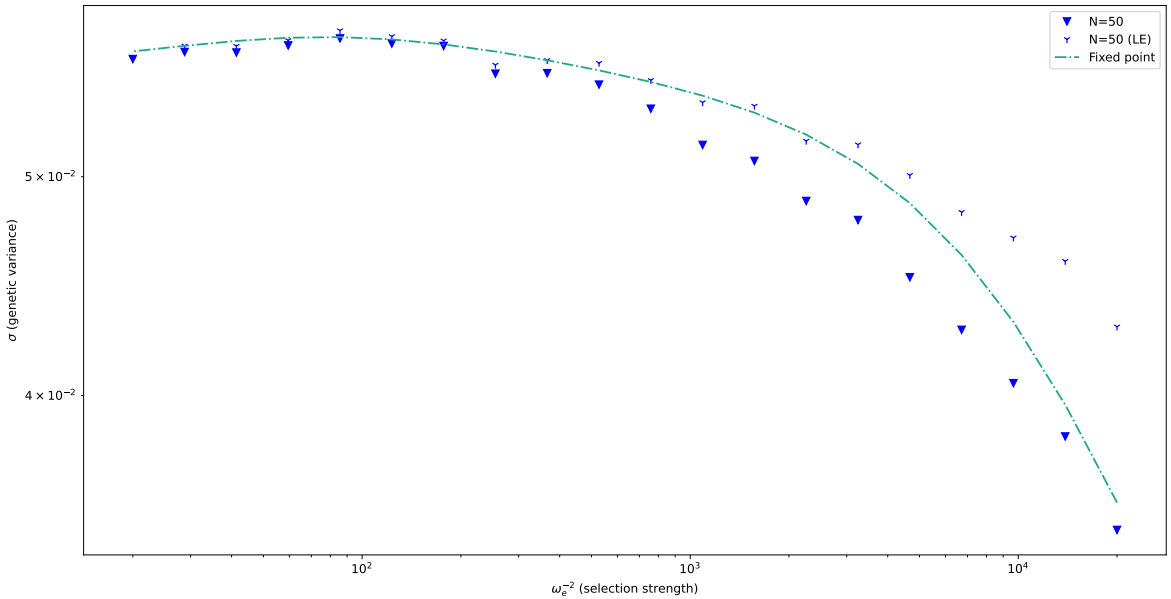


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

The Bulmer effect (1971)

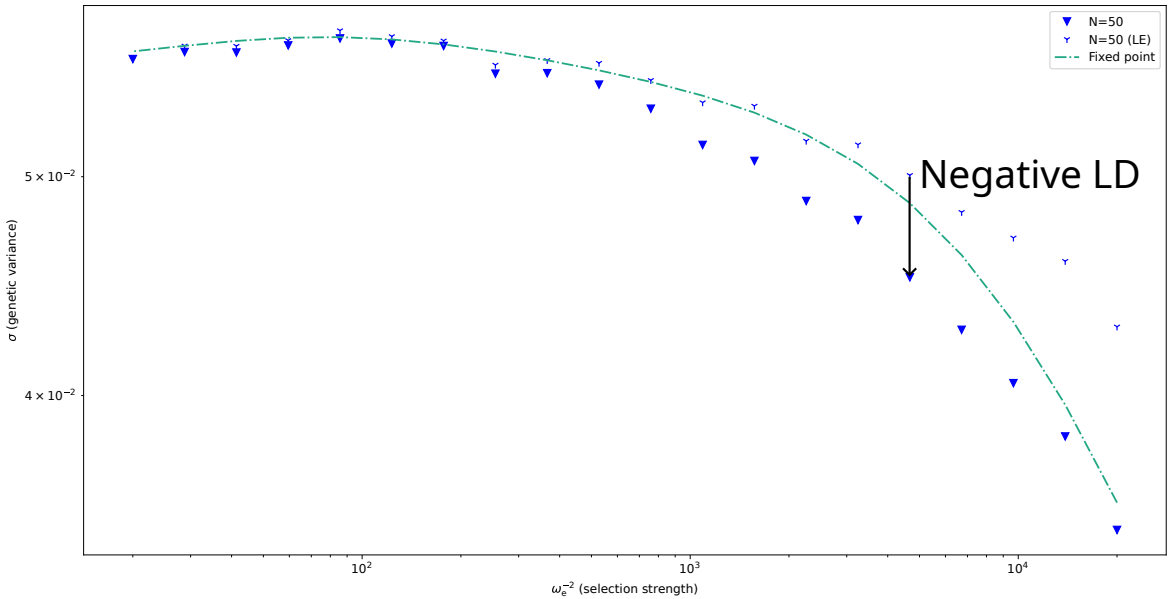


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

The Bulmer effect (1971)

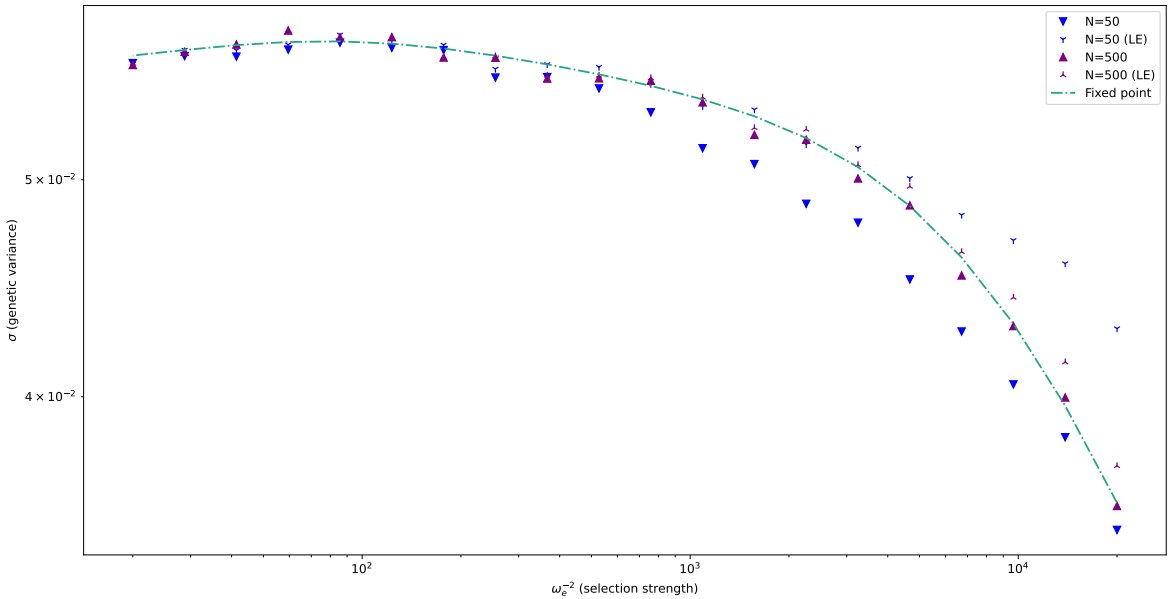


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

The Hill-Robertson effect (1966)

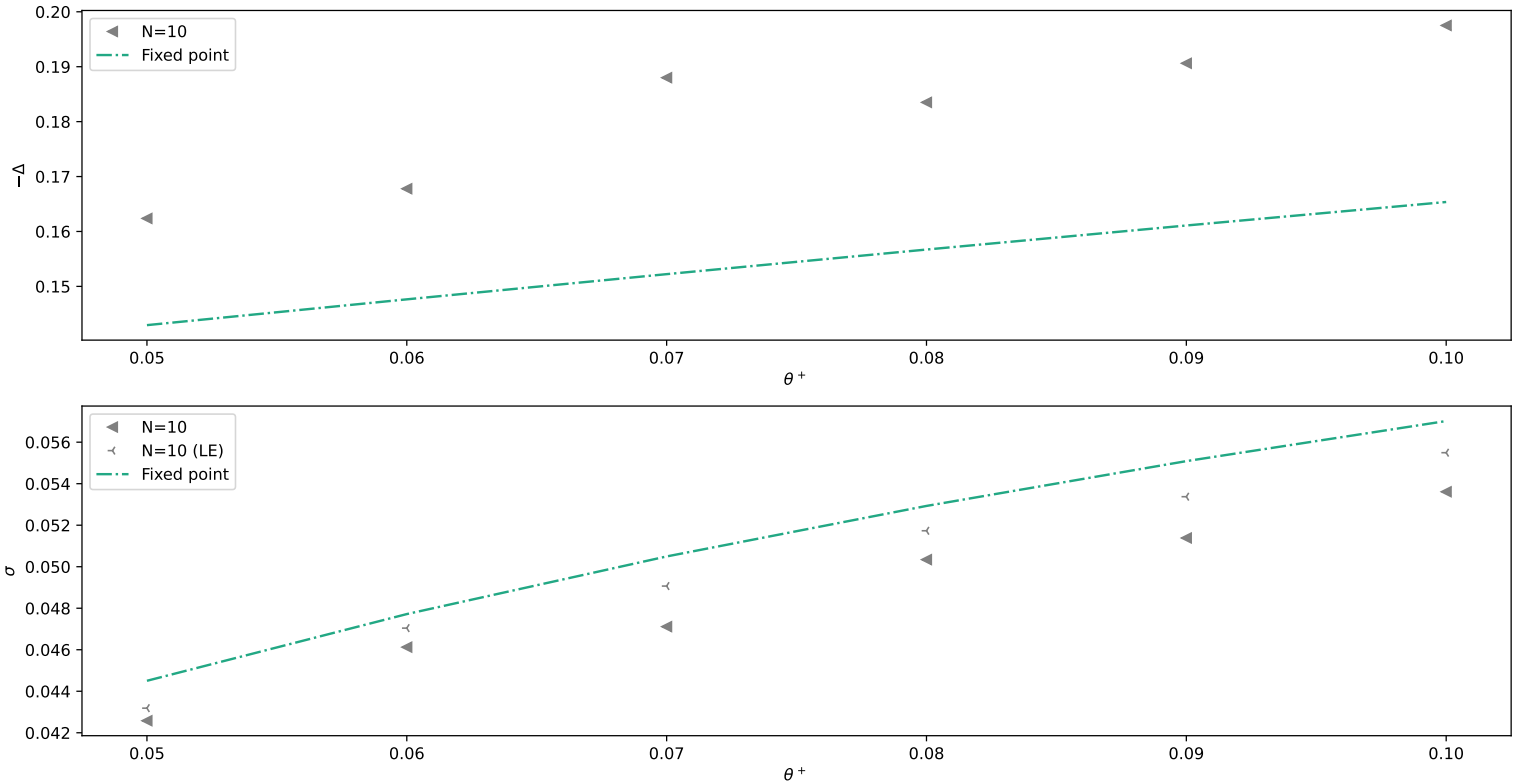


Figure: $L = 100, \theta^- = 2\theta^+, \eta = 1.2, \omega_e^{-2} = 2L, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

The Hill-Robertson effect (1966)

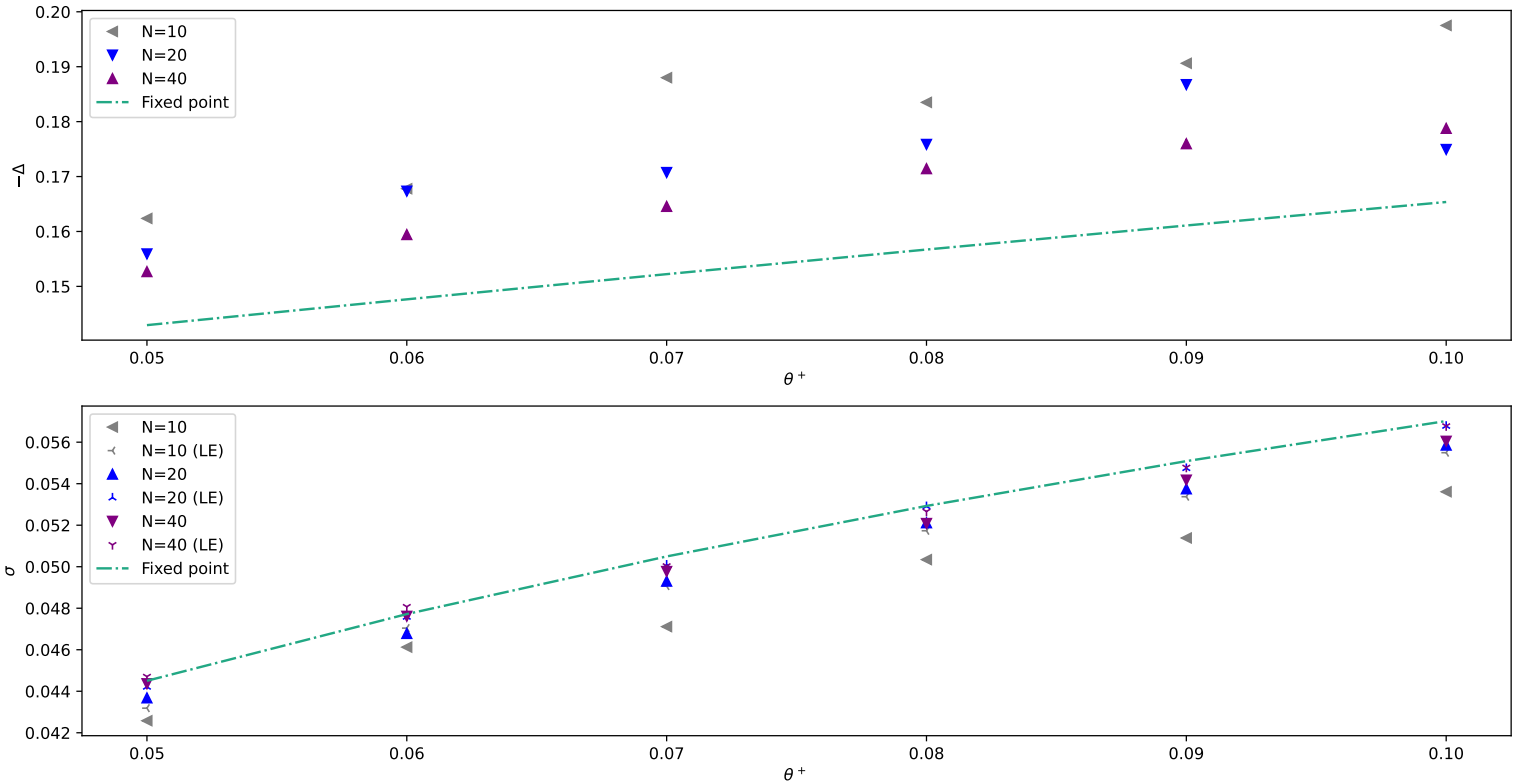


Figure: $L = 100, \theta^- = 2\theta^+, \eta = 1.2, \omega_e^{-2} = 2L, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

Assumptions

A 1: Uniform boundedness
There is a $C \asymp 1$ such that for any $\ell \in [L]$, $|\theta_\ell| \leq C|\bar{\theta}|$ and $\alpha_\ell \leq \frac{C}{L}$.

A 2: Mutations smaller than genetic drift $|\bar{\theta}| \lesssim 1$.

A 3: Mutational bias not too extreme
There is a constant $C \asymp 1$ such that for any ℓ, $\theta_\ell^- / C \leq \theta_\ell^+ \leq C\theta_\ell^-$.

A 4: Accessibility of the selection optimum
$\eta \in (0, 2)$ satisfies $\eta(2 - \eta) \asymp 1$.

A 5: Weak/moderate/strong selection $L \lesssim \omega_e^{-2} \lesssim L^2$.

A 6: Distance between the selection and the mutation optimum
We have

$$\left|2L\ I(0) - \eta\right| \asymp 1.$$

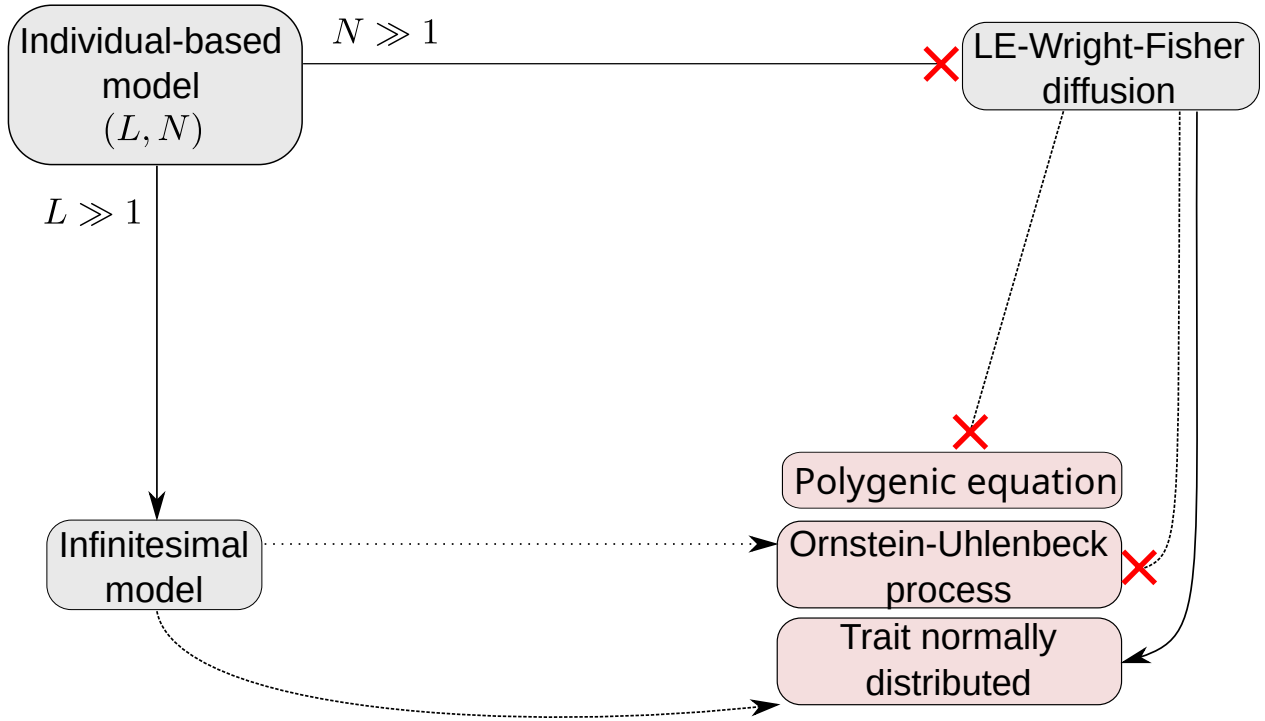
HWLE breakdown

Under (A1-6), under the Quasi-Linkage Disequilibrium approach, the breakdown occurs when the following assumption fails

N1: sufficiently large population

$$2N \gg |\bar{\theta}| \frac{\ln(L)}{L\omega_e^2} + L\sqrt{|\bar{\theta}|}$$

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

Low mutations breakdown

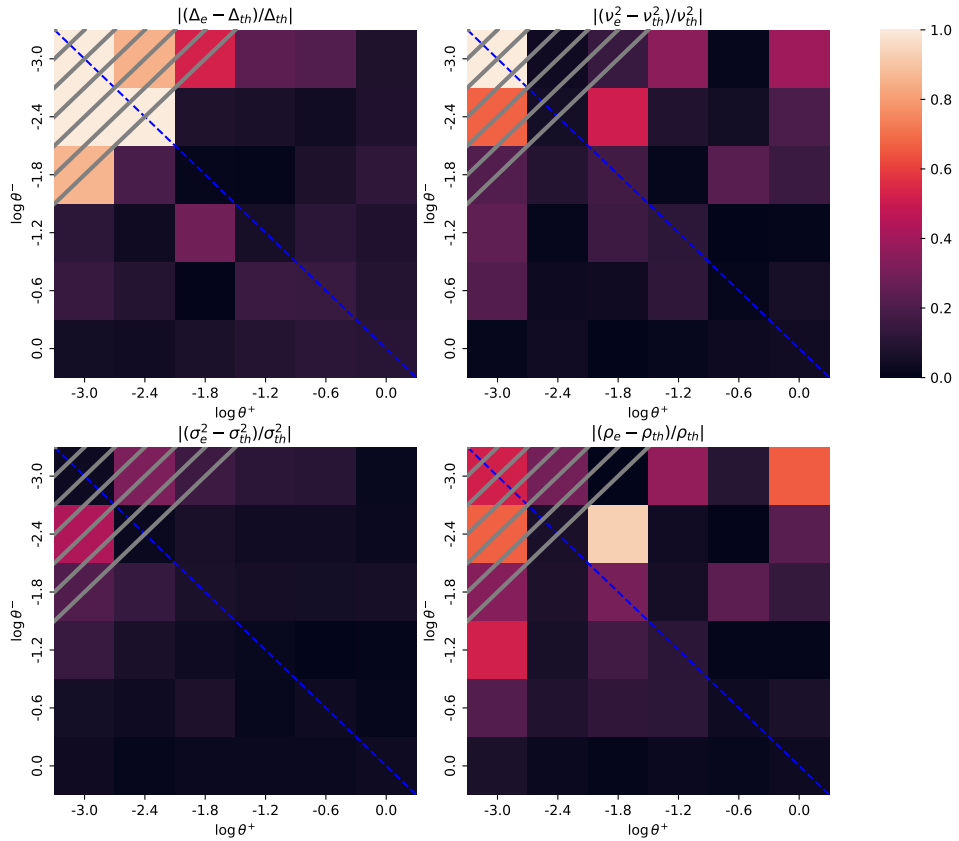


Figure: $N = 500, L = 100, \eta = 1.2, \omega_e^{-2} = 10^3, T = 500N$

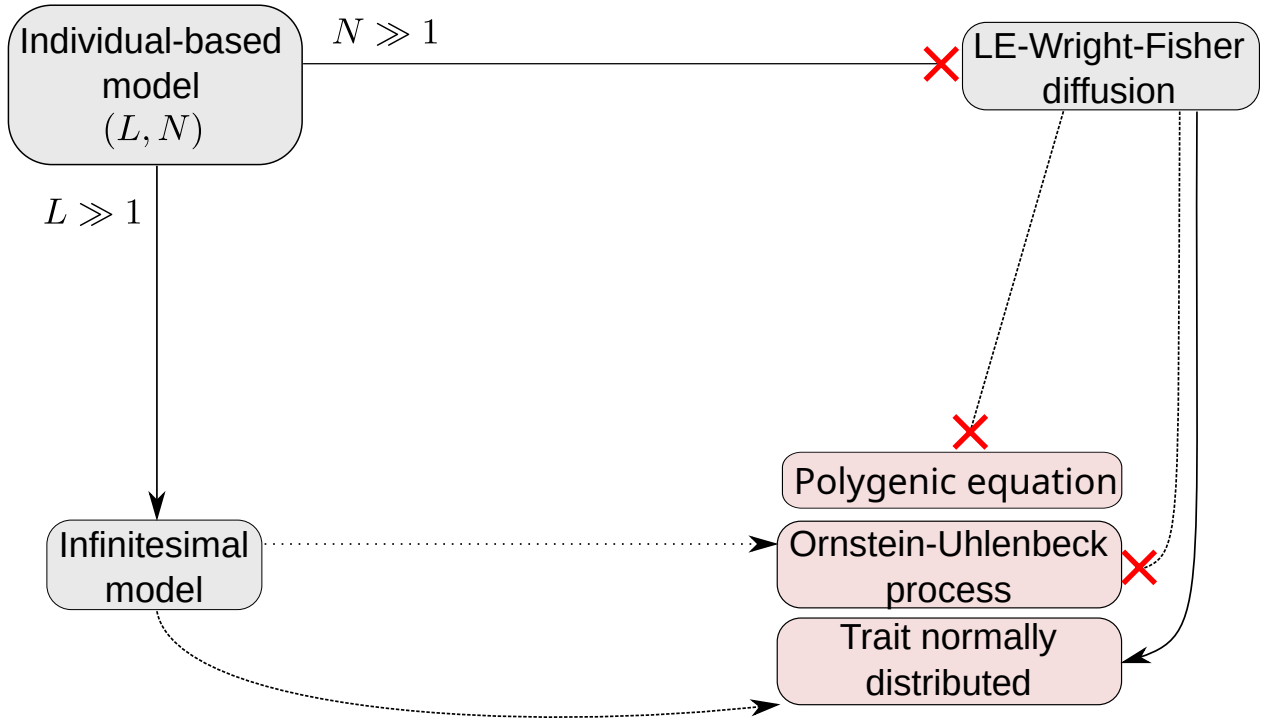
Low mutations breakdown

Under (A1-6), the breakdown for Δ^*, σ^2 is expected when the following assumption fails

N2: Minimal mutational input every generation

$|\bar{\theta}|L \gg 1$.

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

Ornstein-Uhlenbeck breakdown

The Ornstein-Uhlenbeck process is

$$\mathrm{d}\Delta_t = \rho(\Delta^* - \Delta_t)\mathrm{d}t + \nu\sqrt{2\rho}\mathrm{d}B_t$$

Ornstein-Uhlenbeck breakdown

The Ornstein-Uhlenbeck process is

$$\mathrm{d}\Delta_t = \rho(\Delta^* - \Delta_t)\mathrm{d}t + \nu\sqrt{2\rho}\mathrm{d}B_t$$

For such a process we expect

$$\ln(\mathbb{C}\mathbf{ov}[\Delta_t, \Delta_{t+u}]) = -\rho u + 2\ln(\nu)$$

Ornstein-Uhlenbeck breakdown

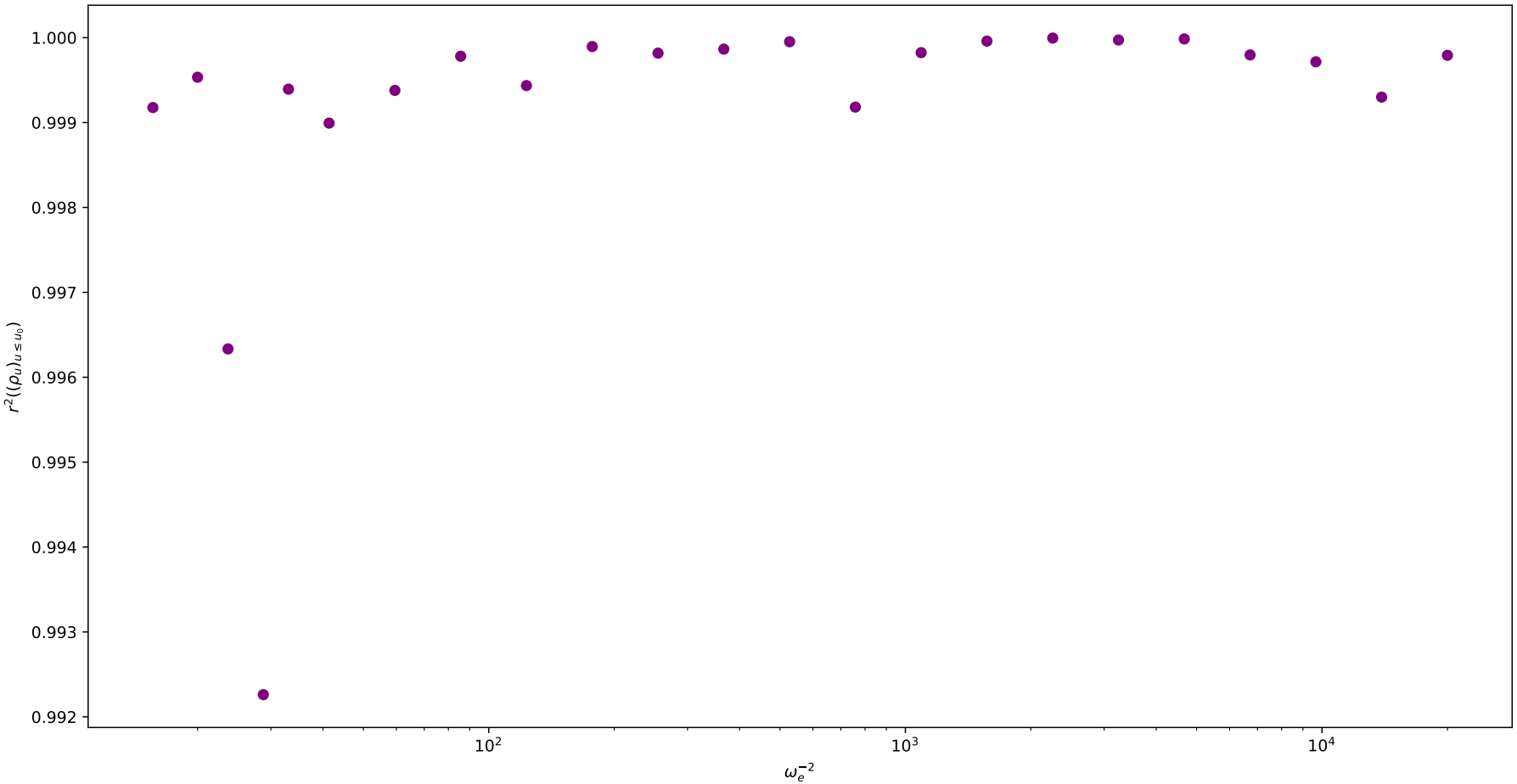


Figure: $N = 500, L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 1000N$

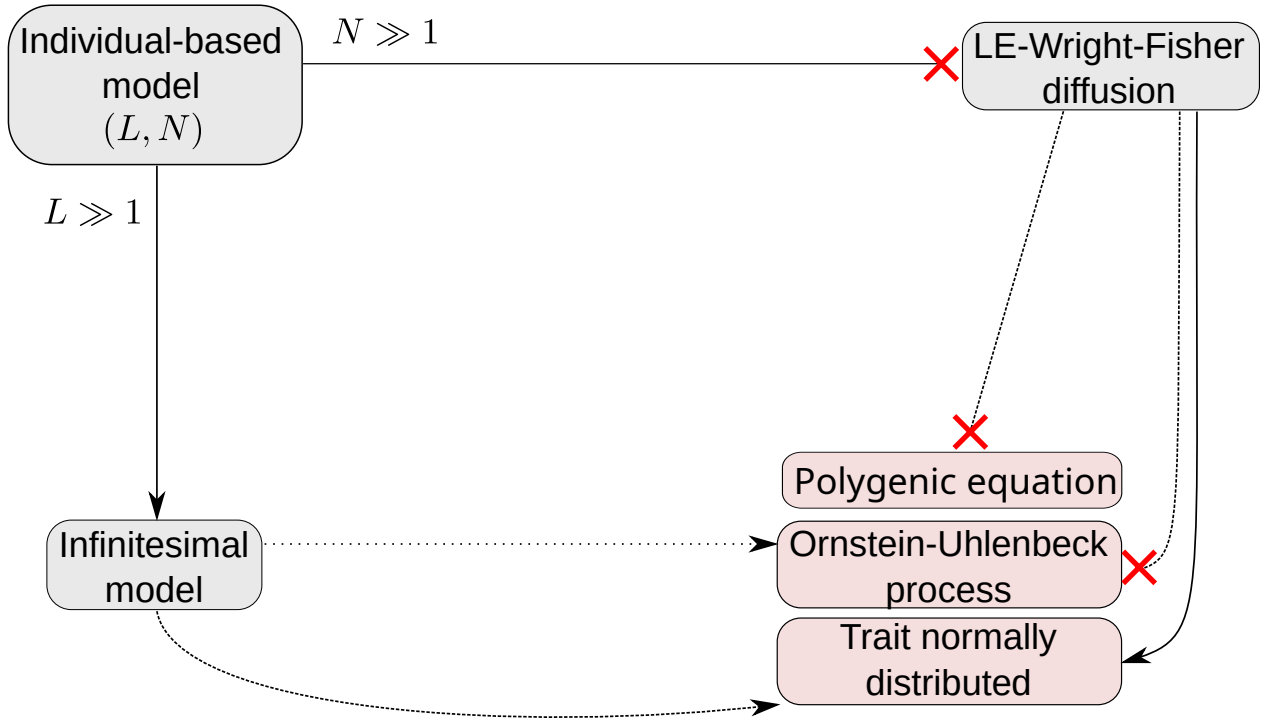
Ornstein-Uhlenbeck breakdown

Under (A1-6), the breakdown for the description of $(\Delta_t)_{t\geq 0}$ as an Ornstein-Uhlenbeck process is expected when the following assumption fails

N3: Sufficient mutational input

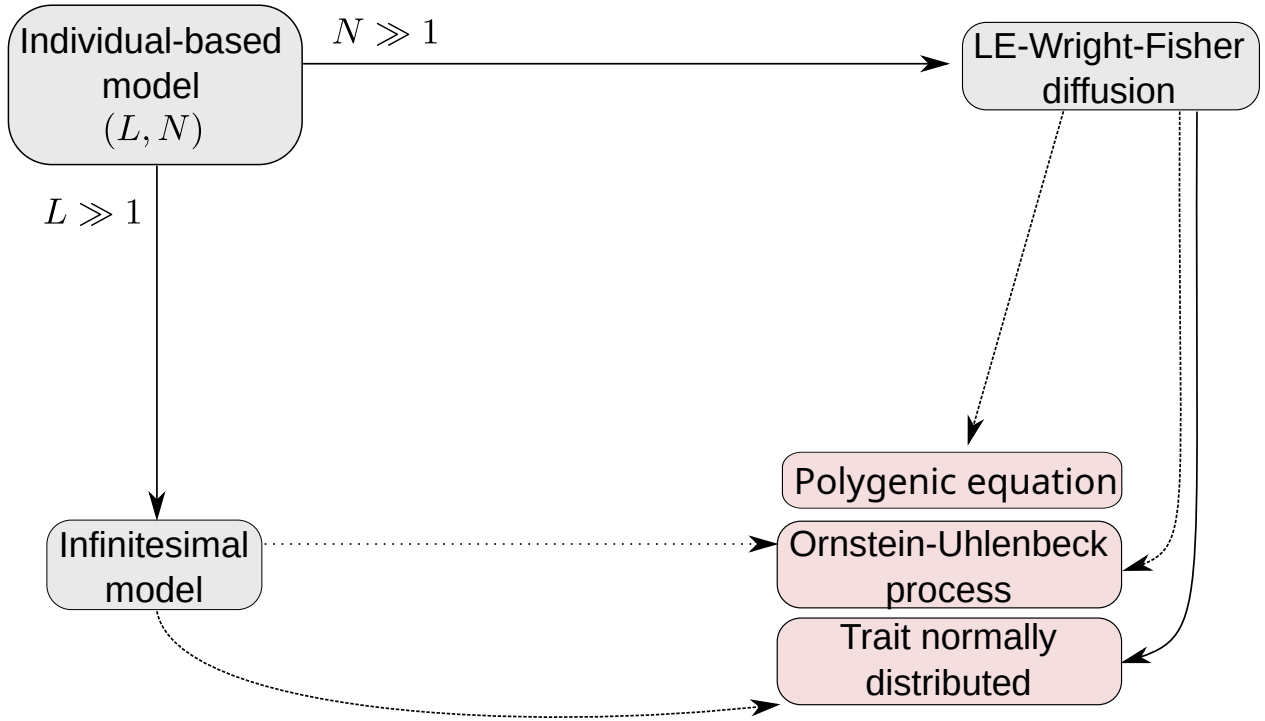
$$\frac{|\bar{\theta}|}{L\omega_e^2} \gg 1.$$

The roadmap

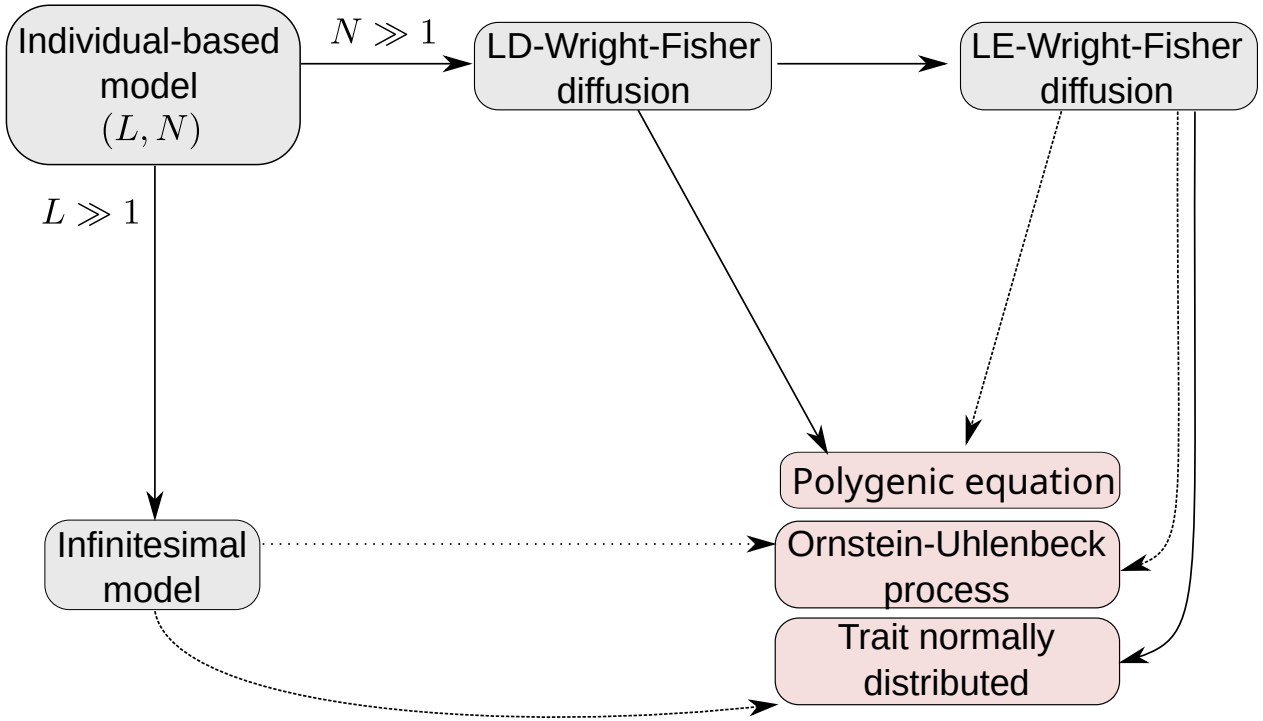

Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

A rigorous control of LD

The roadmap


Individual-based model (L, N)
$N \gg 1$
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

The roadmap


Individual-based model (L, N)
$N \gg 1$
LD-Wright-Fisher diffusion
LE-Wright-Fisher diffusion
$L \gg 1$
Polygenic equation
Infinitesimal model
Ornstein-Uhlenbeck process
Trait normally distributed

The LD-Wright-Fisher diffusion

We define $\mathbb{X}^{[L]}$ as the space of probability measures on $\{-1,+1\}^L$.

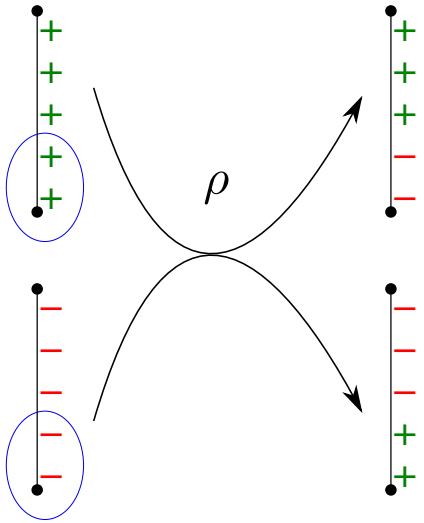
The LD-Wright-Fisher diffusion

We define $\mathbb{X}^{[L]}$ as the space of probability measures on $\{-1,+1\}^L$.
We study a diffusion on $C([0,T],\mathbb{X}^{[L]})$ which includes all four forces of interest

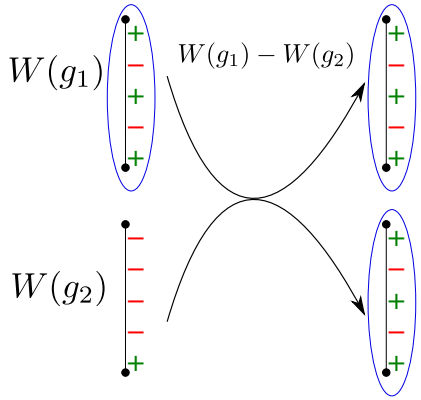
$$\mathrm{d}\mathbf{X}_t = \underset{\heartsuit}{\rho R(\mathbf{X}_t)}\mathrm{d}t + \underset{\spadesuit}{L S(\mathbf{X}_t)}\mathrm{d}t + \underset{\diamondsuit}{\Theta(\mathbf{X}_t)}\mathrm{d}t + \underset{\clubsuit}{\Sigma(\mathbf{X}_t)}\mathrm{d}\mathbf{B}_t.$$

where $\rho > 0$ is the recombination rate and L represents the strength of selection.

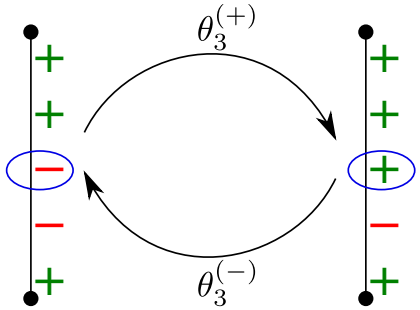
Operators



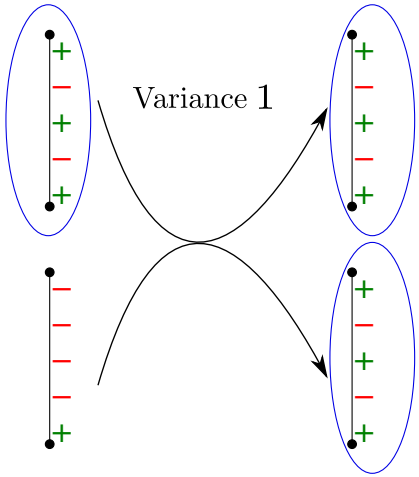
♥ Recombination



♠ Selection



♦ Mutation



♣ Genetic drift

Simplifications

- $\alpha_\ell = 1/L$

Simplifications

- $\alpha_\ell = 1/L$
- θ_ℓ^+ and θ_ℓ^- are constant across loci and of order 1.

Simplifications

- $\alpha_\ell = 1/L$
- θ_ℓ^+ and θ_ℓ^- are constant across loci and of order 1.
- ω_e^{-2} is of order L

Main theorem

Define the **local harmonic recombination rate** r^*_ℓ at locus ℓ with

$$\frac{1}{r^*_\ell} := \frac{1}{L-1} \sum_{\ell' \neq \ell} \frac{1}{r_{\{\ell,\ell'\}}}$$

where $r_{\{\ell,\ell'\}}$ is the probability of recombination between ℓ and ℓ'.

Main theorem

Define the **local harmonic recombination rate** r^*_ℓ at locus ℓ with

$$\frac{1}{r^*_\ell} := \frac{1}{L-1} \sum_{\ell' \neq \ell} \frac{1}{r_{\{\ell,\ell'\}}}$$

where $r_{\{\ell,\ell'\}}$ is the probability of recombination between ℓ and ℓ'.

Define the **global harmonic recombination rate** r^{**} as

$$\frac{1}{r^{**}} := \frac{1}{L} \sum_{\ell \in [L]} \frac{1}{r^*_\ell}$$

Main theorem

Define the **local harmonic recombination rate** r^*_ℓ at locus ℓ with

$$\frac{1}{r^*_\ell} := \frac{1}{L-1} \sum_{\ell' \neq \ell} \frac{1}{r_{\{\ell,\ell'\}}}$$

where $r_{\{\ell,\ell'\}}$ is the probability of recombination between ℓ and ℓ'.

Define the **global harmonic recombination rate** r^{**} as

$$\frac{1}{r^{**}} := \frac{1}{L} \sum_{\ell \in [L]} \frac{1}{r^*_\ell}$$

Define the allelic law process as $\mu_{\mathbf{X}_t} := \frac{1}{L} \sum_{\ell \in [L]} \delta_{P^\ell_t}$.

Main theorem (C Lambert Schertzer 25+)

Theorem 1 (part 1)

Assume $\rho r^{**} \gg L^2 \ln(\rho)$. Then $(\mu_{\mathbf{x}_t})_{t\in[0,T]}$ converges to the law of the McKean-Vlasov process (1) in the Skorokhod topology.

Main theorem (C Lambert Schertzer 25+)

Theorem 1 (part 1)

Assume $\rho r^{**} \gg L^2 \ln(\rho)$. Then $(\mu_{\mathbf{X}_t})_{t\in[0,T]}$ converges to the law of the McKean-Vlasov process (1) in the Skorokhod topology.

Theorem 1 (part 2)

Assume $\rho r^{**} \gg L^2 \ln(\rho)$. Consider n distinct loci $\ell_1 < \cdots < \ell_n$ such that $(P_0^{\ell_i})_{i\in[n]}$ are independent and

$$\min_{i\in[n]} \rho r^*_{\ell_i} \gg L^2 \ln(\rho)$$

$$\min_{i,j\in[n]} \rho r_{\{\ell_i,\ell_j\}} \gg L$$

Then $(P_t^{\ell_i})_{i\in[n],t\in[0,T]}$ converge to n independent McKean-Vlasov processes (1).

Main theorem (C Lambert Schertzer 25+)

Theorem 2

Assume $\rho r^{**} \gg L^2 \ln(\rho)$. Set

$$\varepsilon_L := \frac{1}{\sqrt{\rho r^{**}}}$$

Define the genetic variance $\sigma_t^2 := 4\mathbb{E}[f_t(1-f_t)]$. Then

$$\mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]} \left|L\,\mathbf{Var}_{\mathbf{X}_t}[Z(g)] - \sigma_t^2\right|\right] \longrightarrow 0$$

Structure of the proof

For $\mathbf{x} \in \mathbb{X}^{[L]}$, let $\pi(\mathbf{x})$ be the **LE projection** of $\mathbf{x}$

$$\pi(\mathbf{x}) = \bigotimes_{\ell \in [L]} \mathbf{x}^{\{\ell\}}.$$

Structure of the proof

For $\mathbf{x} \in \mathbb{X}^{[L]}$, let $\pi(\mathbf{x})$ be the **LE projection** of $\mathbf{x}$

$$\pi(\mathbf{x}) = \bigotimes_{\ell \in [L]} \mathbf{x}^{\{\ell\}}.$$

- We can decompose the marginal of the selector on locus ℓ, $S^\ell(\mathbf{X}_t)$, into the LE term $S^\ell(\pi(\mathbf{X}_t))$ and a sum over subsets $A \subseteq [L], \#A \leq 3$ of a term dominated by $\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)$.

Structure of the proof

For $\mathbf{x} \in \mathbb{X}^{[L]}$, let $\pi(\mathbf{x})$ be the **LE projection** of $\mathbf{x}$

$$\pi(\mathbf{x}) = \bigotimes_{\ell \in [L]} \mathbf{x}^{\{\ell\}}.$$

- We can decompose the marginal of the selector on locus ℓ, $S^\ell(\mathbf{X}_t)$, into the LE term $S^\ell(\pi(\mathbf{X}_t))$ and a sum over subsets $A \subseteq [L], \#A \leq 3$ of a term dominated by $\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)$.
- Using the eigenvalues of the marginal recombinator R^A, we can dominate $\sup_{t \in [0,T]} \|\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)\|_2$.

Structure of the proof

For $\mathbf{x} \in \mathbb{X}^{[L]}$, let $\pi(\mathbf{x})$ be the **LE projection** of $\mathbf{x}$

$$\pi(\mathbf{x}) = \bigotimes_{\ell \in [L]} \mathbf{x}^{\{\ell\}}.$$

- We can decompose the marginal of the selector on locus ℓ, $S^\ell(\mathbf{X}_t)$, into the LE term $S^\ell(\pi(\mathbf{X}_t))$ and a sum over subsets $A \subseteq [L], \#A \leq 3$ of a term dominated by $\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)$.
- Using the eigenvalues of the marginal recombinator R^A, we can dominate $\sup_{t\in[0,T]} \|\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)\|_2$.
- We can then show the convergence of the martingale problem associated with P_t^ℓ to the limit martingale problem.

Structure of the proof

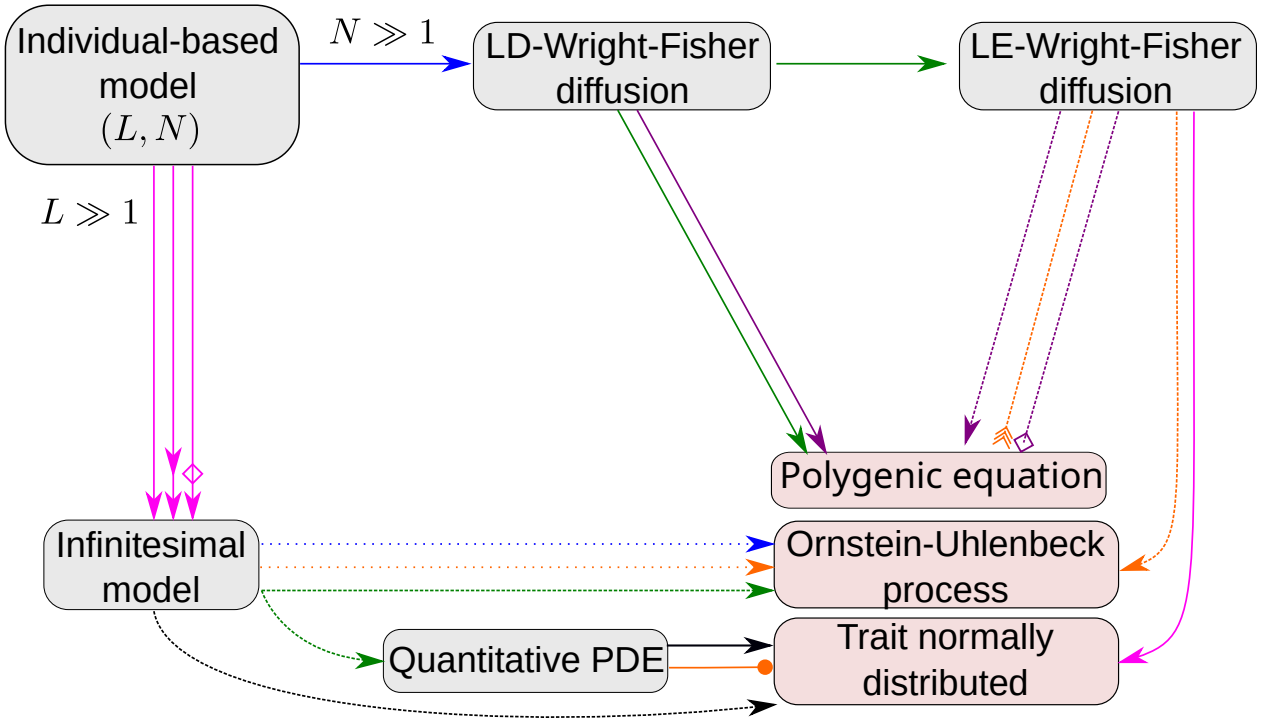
For $\mathbf{x} \in \mathbb{X}^{[L]}$, let $\pi(\mathbf{x})$ be the **LE projection** of $\mathbf{x}$

$$\pi(\mathbf{x}) = \bigotimes_{\ell \in [L]} \mathbf{x}^{\{\ell\}}.$$

- We can decompose the marginal of the selector on locus ℓ, $S^\ell(\mathbf{X}_t)$, into the LE term $S^\ell(\pi(\mathbf{X}_t))$ and a sum over subsets $A \subseteq [L], \#A \leq 3$ of a term dominated by $\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)$.
- Using the eigenvalues of the marginal recombinator R^A, we can dominate $\sup_{t\in[0,T]} \|\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)\|_2$.
- We can then show the convergence of the martingale problem associated with P_t^ℓ to the limit martingale problem.
- The well-posedness of the limit martingale problem is obtained from a Girsanov transform.

Conclusion

Conclusion for mathematicians



Degree of proof
- Proven
- Discussed
- Not discussed

Type of proof
- Diffusion approximation
- Mixing
- Mean-field approximation
- Separation of timescales
- Central limit theorem

Complications
- Diploidy/Dominance
- Strong selection
- Epistasis
- Population structure/Demography

Conclusion for mathematicians

Conclusion for mathematicians

- The subject deserves further study.

Conclusion for mathematicians

- The subject deserves further study.
- The distribution of α_ℓ should be degenerate:

Conclusion for mathematicians

- The subject deserves further study.
- The distribution of α_ℓ should be degenerate:
- The greatest task ahead is to characterize the phase transitions

Conclusion for mathematicians

- The subject deserves further study.
- The distribution of α_ℓ should be degenerate:
- The greatest task ahead is to characterize the phase transitions
  - Weak mutations $*$

Conclusion for mathematicians

- The subject deserves further study.
- The distribution of α_ℓ should be degenerate:
- The greatest task ahead is to characterize the phase transitions
  - Weak mutations $*$
  - Bulmer effect $* * * *$

Conclusion for mathematicians

- The subject deserves further study.
- The distribution of α_ℓ should be degenerate:
- The greatest task ahead is to characterize the phase transitions
  - Weak mutations $*$
  - Bulmer effect $* * * *$
  - Hill-Robertson effect $* * * * *$

Conclusion for biologists

Conclusion for biologists

- When genetic diversity is maintained by mutations, the demographic noise in trait mean should be of the same order as the distance of the trait mean to the selection optimum: $|\Delta_t| \gtrsim \nu$.

Conclusion for biologists

- When genetic diversity is maintained by mutations, the demographic noise in trait mean should be of the same order as the distance of the trait mean to the selection optimum: $|\Delta_t| \gtrsim \nu$.

- We discuss how the gene's eye-view can be extended to account for polyploidy, pleiotropy, dominance, and some forms of epistasis.

Conclusion for biologists

- When genetic diversity is maintained by mutations, the demographic noise in trait mean should be of the same order as the distance of the trait mean to the selection optimum: $|\Delta_t| \gtrsim \nu$.

- We discuss how the gene's eye-view can be extended to account for polyploidy, pleiotropy, dominance, and some forms of epistasis. We can also hope to extend it to spatial structure, demographic variability and more.

Colleague acknowledgements

Supervisors: Amaury Lambert, Emmanuel Schertzer

Smile team in Paris: Guillaume Achaz, François Blanquart, Augustin Chen, Elisa Couvert, Jasmine Gamblin, Frith Edbrooke, Thomas Forest, Thibaut Morel-Journel, Abdelmajid Omarajee, Aurore Picot, Guillaume Thomas.

And also: Louis-Pierre Chaintron, Nicolas Fournier, Rodrigue Friaud.

Lab in Vienna: Mathilde André, Florin Boenkost, Nathanaël Boutillon, Colin Desmarais, François Ged, Denis Grange, Paul de Lambert, Etienne Marecaux, Agathe Saloux, Zsófia Talyigás, Julie Tourniaire, Yannic Wenzel.

And also: Melda Akyazi, Nick Barton, Nathanaël Berestycki, Reinhard Bürger, Ariane Carrance, Archana Devi, Laura Hayward, Jitka Polechova, Himani Sachdeva, William da Silva, Lucas Teyssier, Paul Thévenin.

Appendix

First-order Bulmer correction

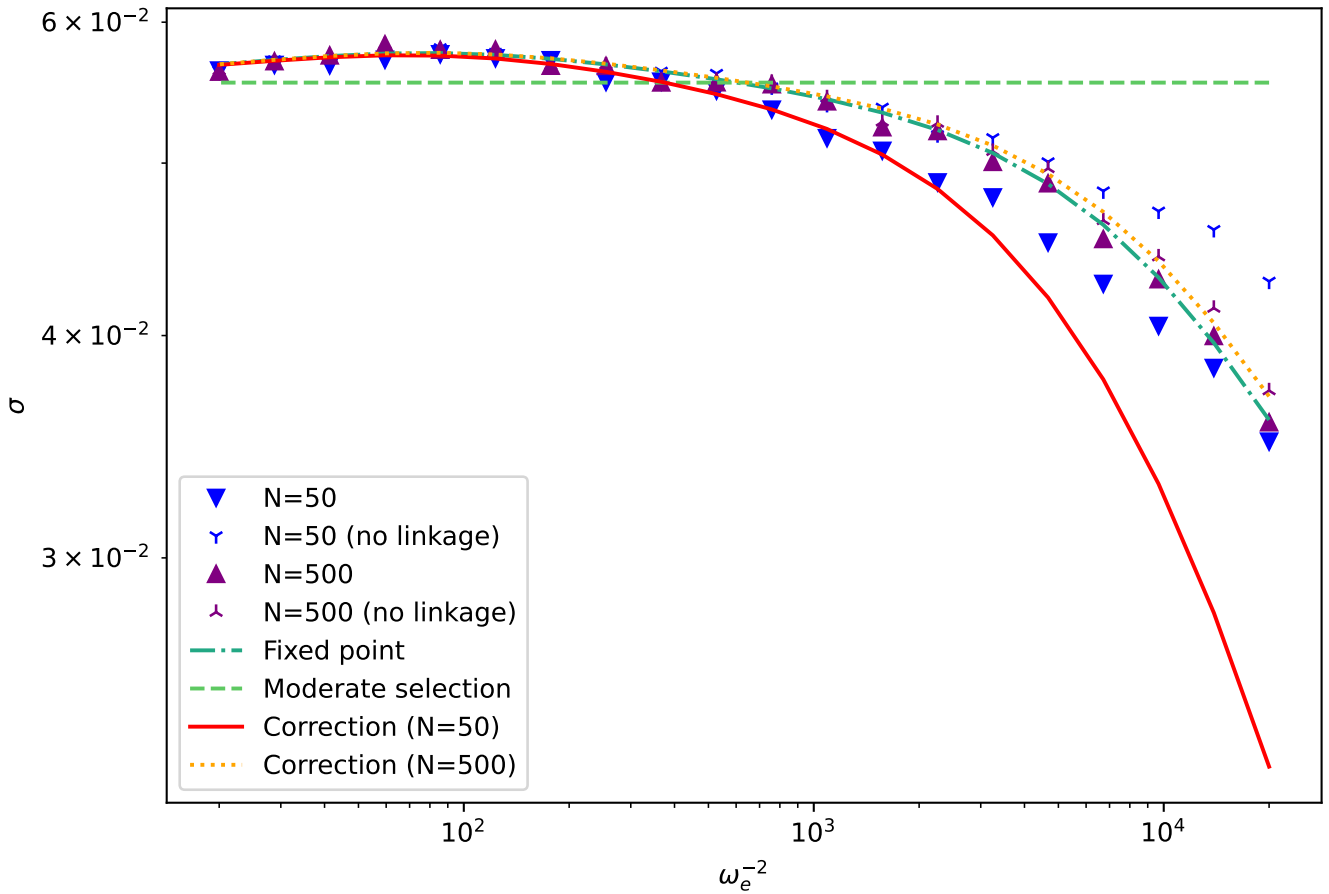
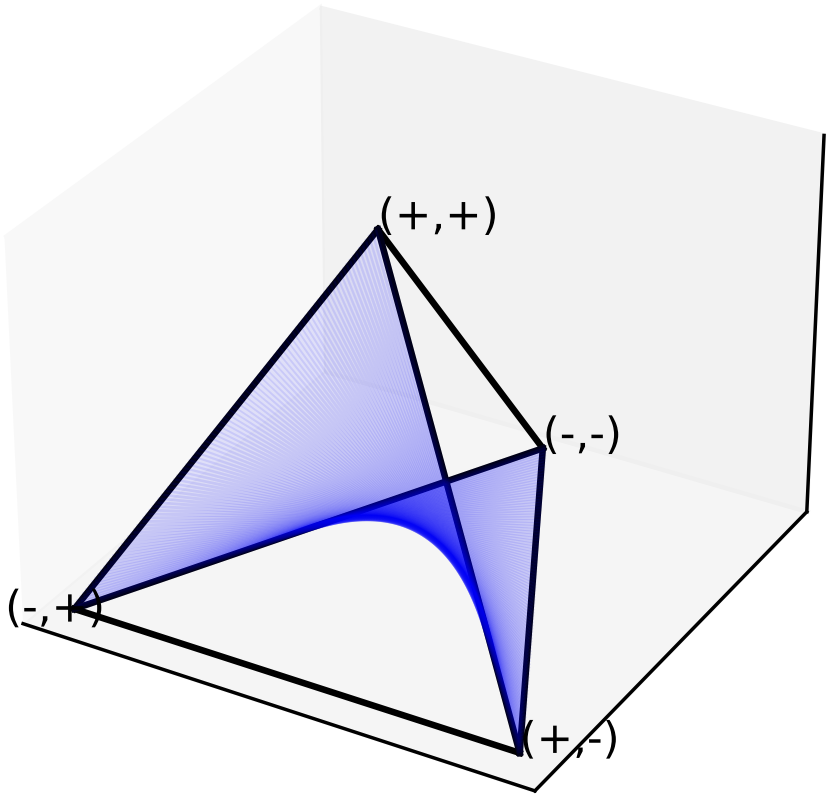


Figure: $L = 100, \theta = (0.1, 0.2), \eta = 1.2, T = 500N$ and α_ℓ is distributed as *Exponential*(L).

HWLE



Slow/fast principle for strong selection

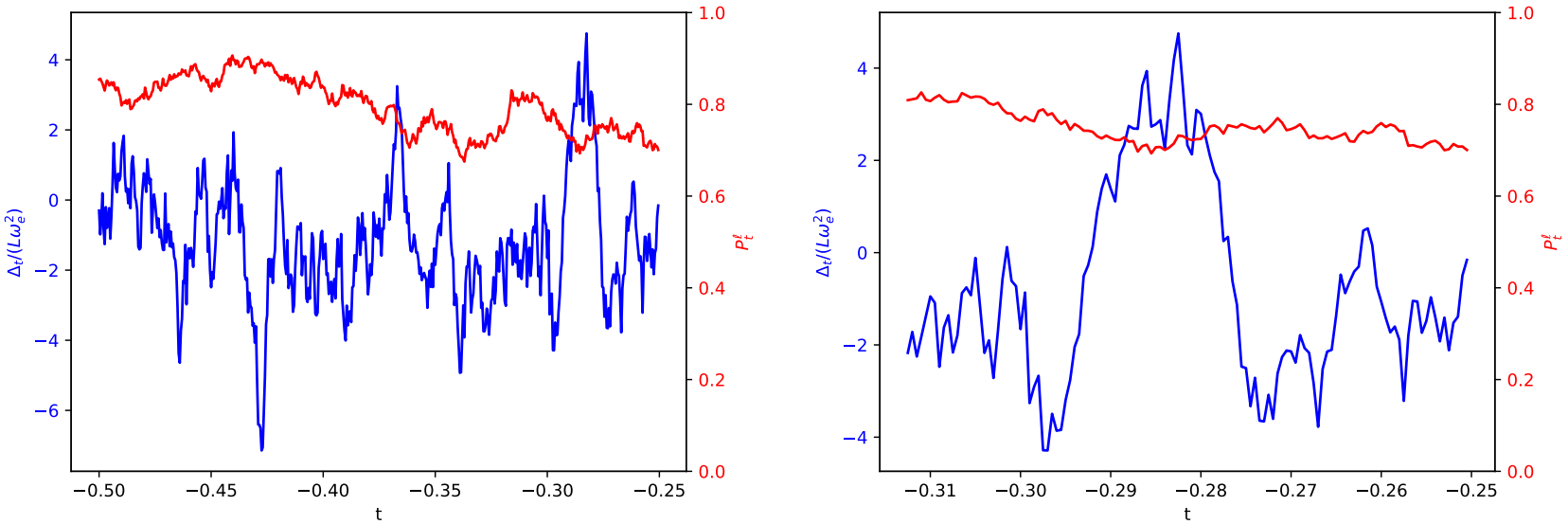


Figure: $N = 1000, L = 1000, \theta = (0.1, 0.2), \eta = 1.2, \omega_e^{-2} = L^2$ and α_ℓ has law *Exponential*(L)

Breakdown for small mutations.

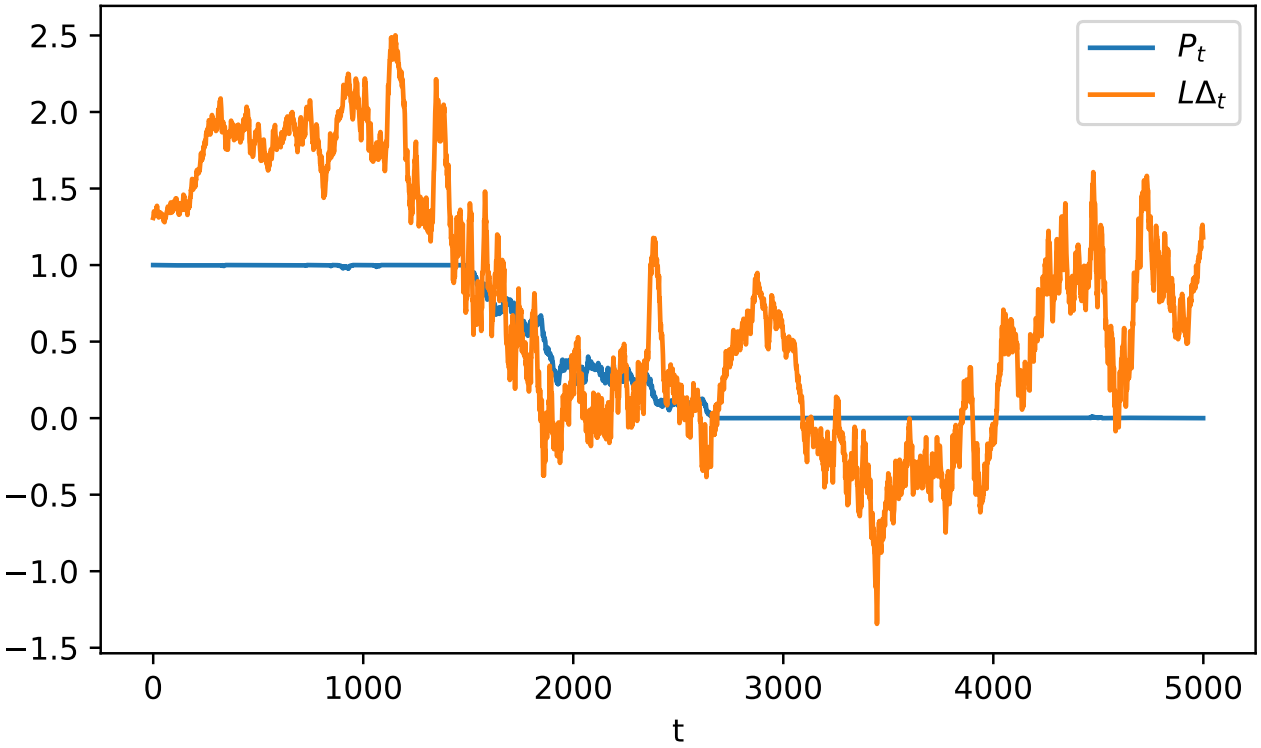
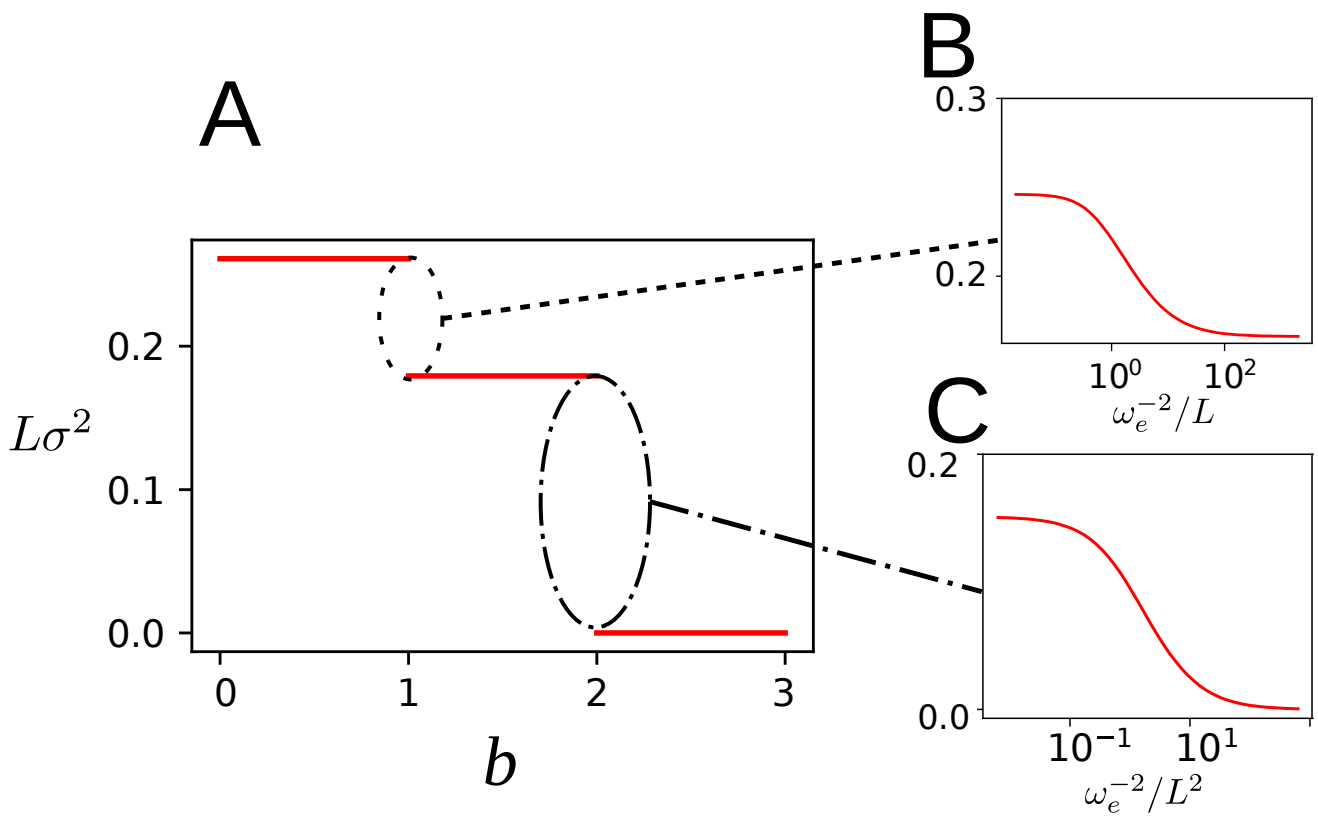


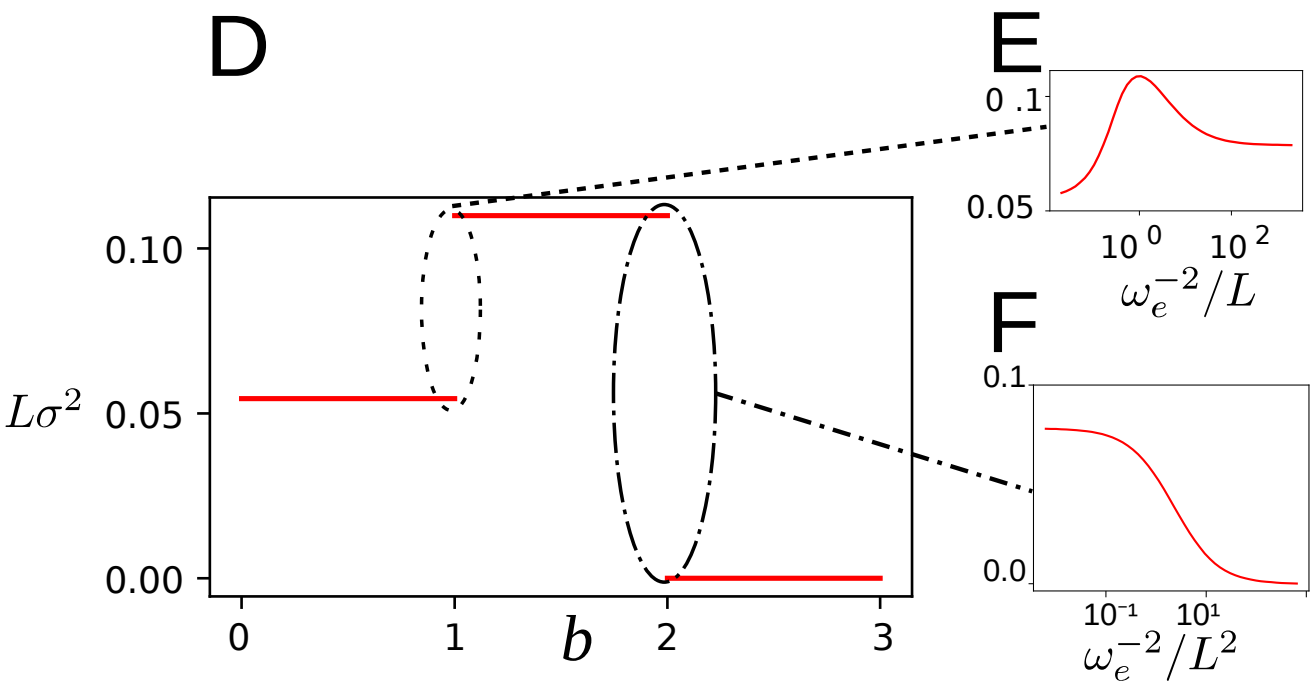
Figure: $N = 500, L = 100, \theta = (1/(2L), 1/(2L)), \omega_e^{-2} = 2000, \eta = 1.2$, and α_ℓ has law *Exponential*(L).

Scaling limit



with b such that $\omega_e^{-2} = L^b$, $L \to +\infty, N \to +\infty, \eta = 1.5$, and $(\alpha_\ell)_{\ell \in [L]}$ be exponentially distributed with parameter L.
No mutational bias ($\theta = (0.1, 0.1)$)

Scaling limit



with b such that $\omega_e^{-2} = L^b$, $L \to +\infty, N \to +\infty, \eta = 1.5$, and $(\alpha_\ell)_{\ell \in [L]}$ be exponentially distributed with parameter L.

Strong mutational bias ($\theta = (0.01, 0.1)$)

Trajectory view of the Hill-Robertson effect

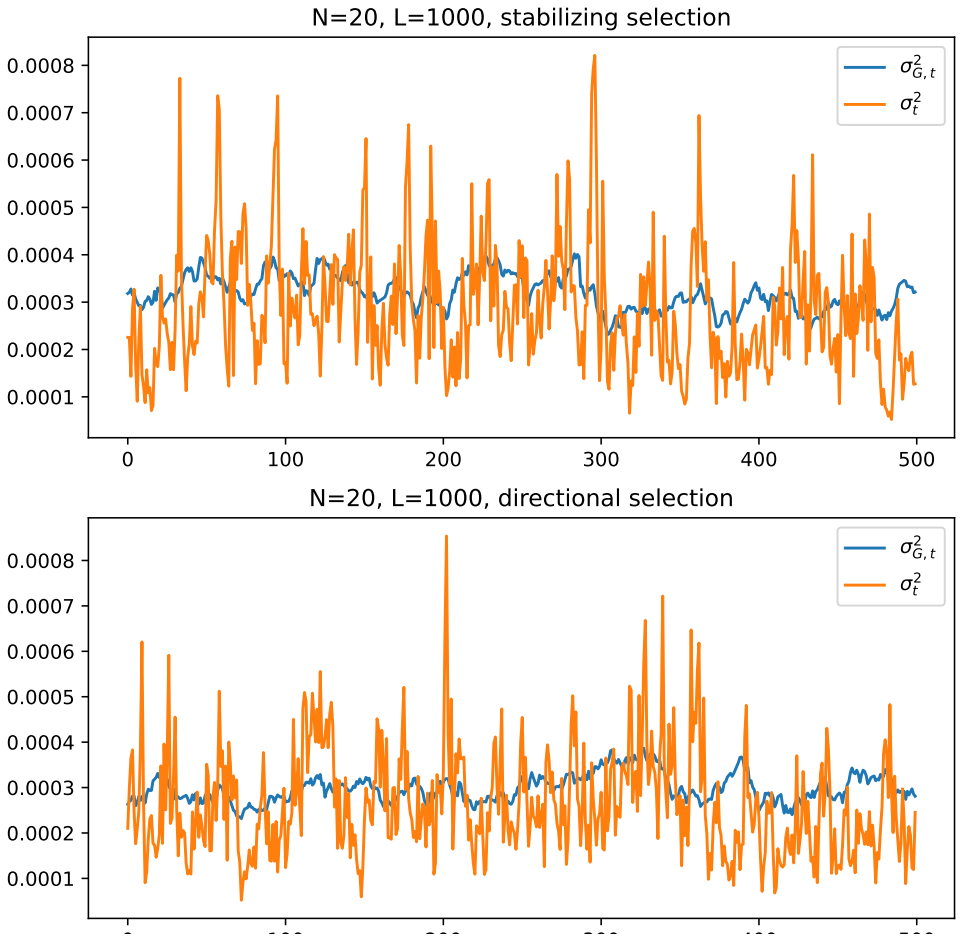


Figure: $\theta = (0.1, 0.2), \eta = 1.2, \omega_e^{-2} = 2L$.

Heavy tails of α

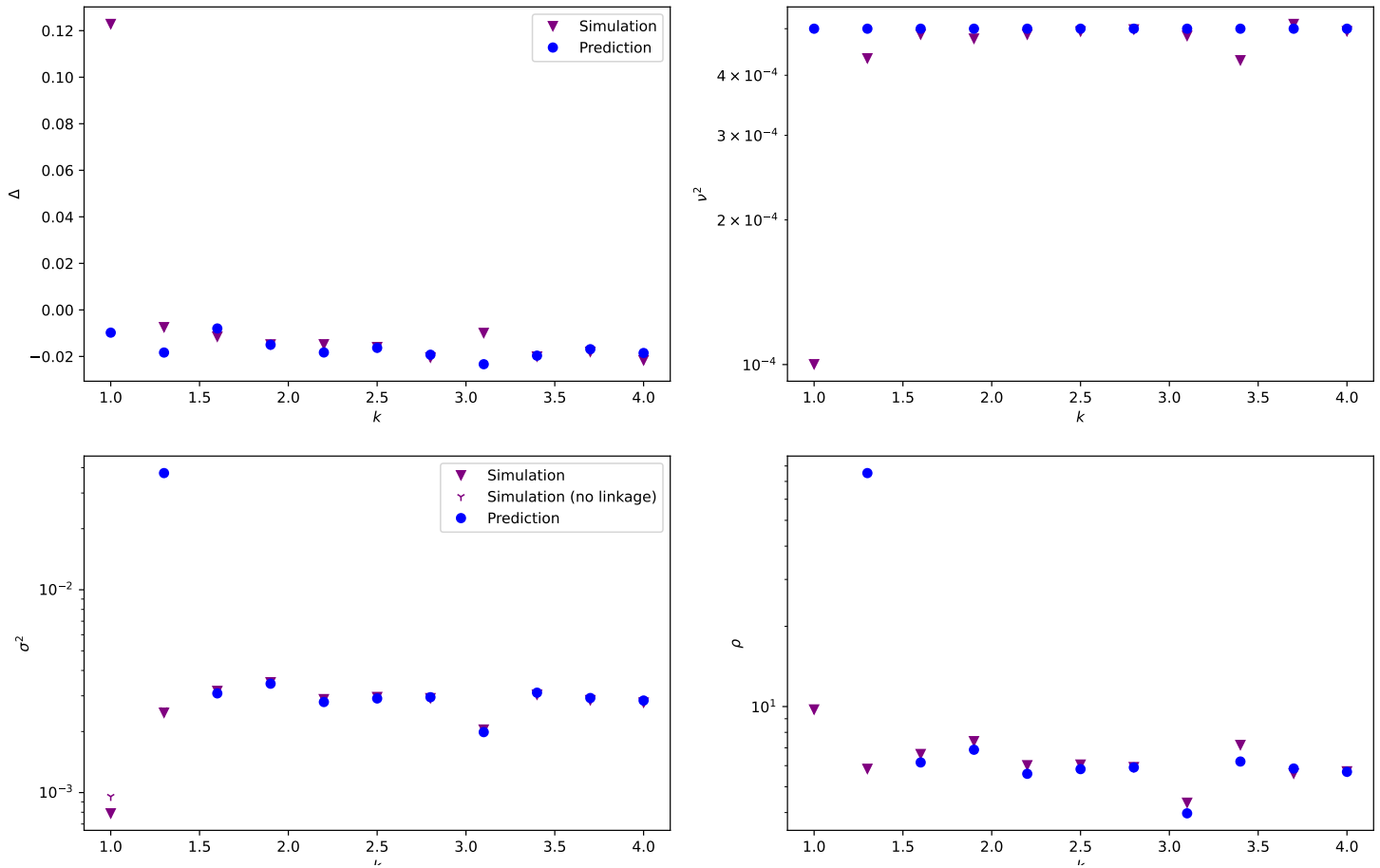
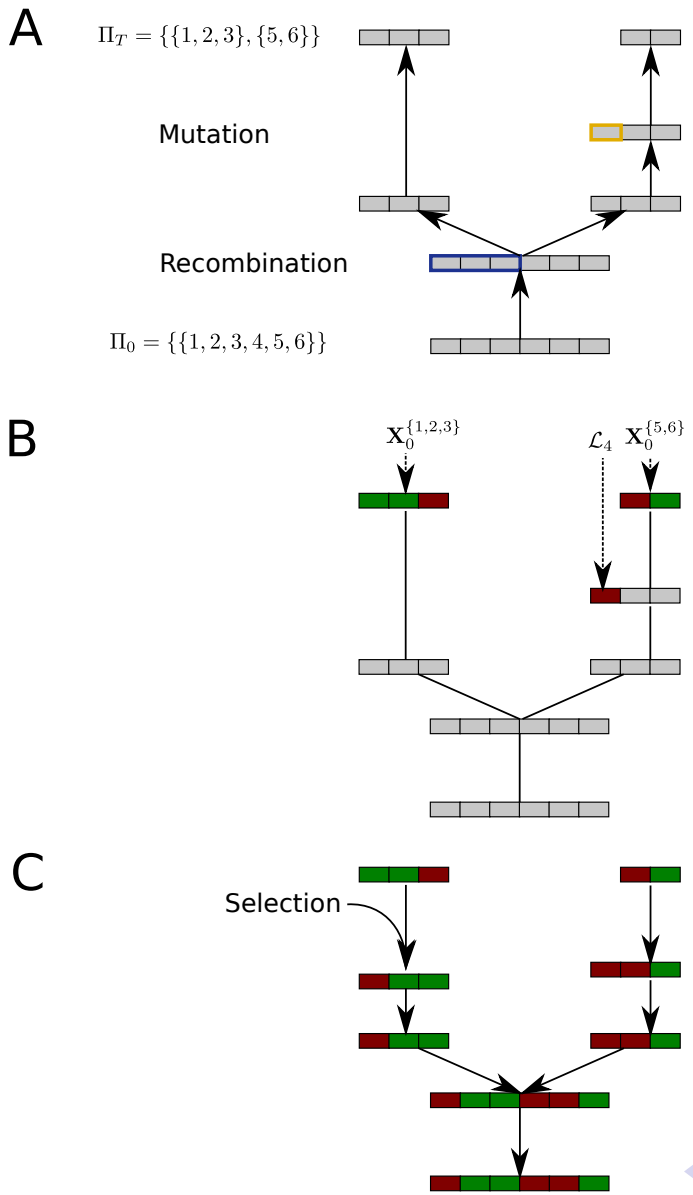


Figure: $N = 500, L = 100, \theta = (0.1, 0.2), \eta = 1.2, \omega_e^{-2} = 10^3$ and α has law $Pareto(k)$.

An ancestral process in an infinite population

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}_t = R(\mathbf{X}_t) + S(\mathbf{X}_t) + \Theta(\mathbf{X}_t)$$

An ancestral process in an infinite population



Duality formula

For $\mathbf{x} \in \mathbb{X}^{[L]}, \pi \in \mathscr{S}, \gamma \in \square_{[L]}$ define

$$f(\mathbf{x}, \pi, \gamma) := \prod_{A \in \pi} x^A(\gamma^{|A}) \quad \times \prod_{\ell \notin \pi} \mathcal{L}_\ell(\gamma^{|\{\ell\}}).$$

Duality formula

For $\mathbf{x} \in \mathbb{X}^{[L]}, \pi \in \mathscr{S}, \gamma \in \square_{[L]}$ define

$$f(\mathbf{x}, \pi, \gamma) := \prod_{A \in \pi} x^A(\gamma^{|A}) \quad \times \prod_{\ell \notin \pi} \mathcal{L}_\ell(\gamma^{|\{\ell\}}).$$

We have

$$\mathcal{A}^{\mathbf{X}_t} f(\cdot, \pi, \gamma)(\mathbf{x}) = \mathcal{A}^{\Pi_t} f(\mathbf{x}, \cdot, \gamma)(\pi) + \mathcal{A}^{\mathcal{G}_t}_{\mathbf{x}} f(\mathbf{x}, \pi, \cdot)(\gamma).$$