

STOCHASTIC NEUTRAL FRACTIONS AND THE EFFECTIVE POPULATION SIZE

joint work with R. Fric, E. Schertler and Z. Talyagas.

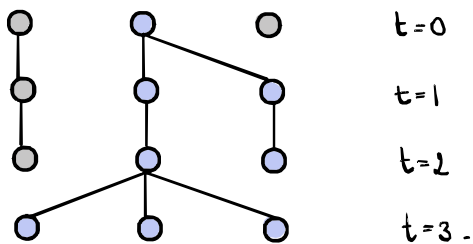
1) DEFINITIONS AND MOTIVATION

(i) Neutral fractions

Wright-Fisher model:
population model evolving in discrete time t
with fixed population size $N \in \mathbb{N}$.

at time $t=0$, N initial individuals
at $t \in \mathbb{N}$, the population is renewed
each newborn chooses a parent \perp uniformly
at random from generation $t-1$.

ex: $N=3$ $K=2$



Q1: Allelic composition?

at $t=0$, each individual is assigned an allele
alleles are indexed by $\{1, \dots, K\}$.

at $t \in \mathbb{N}$, each individual inherits the allele of its parent
alleles are neutral.

For $k \in [K]$, $t \in \mathbb{N}_0$ define

$X_t^{N,k}$ = frequency of allele k at generation t -
size of the k -th neutral fraction.

Diffusion approximation (Kimura)

Assume $(X_0^{N,k})_k \Rightarrow x_0 \in [0,1]^K$. Then, as $N \rightarrow +\infty$

$(X_{\lfloor Nt \rfloor}^{N,k}, t \geq 0) \Rightarrow (Z_t^k)$

(Z_t^k) $(K-1)$ dimensional standard Wright-Fisher diffusion

ex $K=2$ diffusion on the simplex $\{(z_1, z_2) \in [0,1]^2, z_1+z_2=1\}$

$dZ_t^1 = \sqrt{Z_t^1(1-Z_t^1)} dB_t$

random fluctuations in the allelic composition is only due
to binomial sampling \Rightarrow genetic drift.

Rk: Cannings model = generalisation of WF.
offspring distribution \perp exchangeable random vector

Assume $\mathbb{E}[\# \text{offspring}] = 1$
 $\text{Var}(\# \text{offspring}) = \sigma_N^2 \rightarrow \sigma^2 < +\infty$

Then, the diffusion approximation holds on the timescale
 N/σ^2 .

(ii) ~~Effective population size~~

Q₂: Impact of population structure on the allelic composition?

~~Structured population models~~

ex: spatial structure, fitness space ...

- each individual carries a type (\neq allele) impacting its ability to reproduce -
- reproduction is non-local (migrations / mutations)

"the standard approach" ~~the backward-in-time approach~~

BACKWARD
 structured genealogy (lineages carry types) $\xrightarrow[N \rightarrow +\infty]{\text{separation of timescale}}$ collapse of structure - Kingman coalescent on timescale $N_e \leq N$

duality.

FORWARD
 Diffusion approximation on time scale N_e

N_e = effective population size

size of the WF model that experiences the same level of genetic drift as the structured population under study -

Q₃: Can we bypass the genealogical approach to compute effective population sizes?

2) THE MODEL

GOAL: derive a general framework to answer Q₃.

(i) ~~∞ -decomposable SDEs~~

Finite type space $E = \{x_1, \dots, x_d\}$.

Ecological description of the population given by the d -dimensional SDE (= large-pop limit)

$$dU_t^N = b(U_t^N) dt + \frac{1}{\sqrt{N}} a(U_t^N) dW_t \quad (\text{SDE}_1)$$

$U_t^N \in \mathbb{R}^E$ configuration of the syst-
 $U_t^N(x)$ "number of individual of type x "

$b: \mathbb{R}^E \rightarrow \mathbb{R}^E$ drift coefficient

= mean growth of the system

$a: \mathbb{R}^E \rightarrow \mathbb{R}^{E \times E}$ noise correlation matrix

= demographic fluctuations

W_t d -dimensional Brownian motion.

a and b are smooth enough to ensure existence + uniqueness of weak solutions to (SDE₁)

ASSUMPTION 1 (co-decomposability)

1) Drift term: $b(u) = F(u)u$

with $F: \mathbb{R}^E \rightarrow \mathbb{R}^{E \times E}$

$F(u)$ "mean offspring matrix"

2) Noise term: $\sigma \sigma^*(u) = C(u)u$

with $C: \mathbb{R}^E \rightarrow \mathbb{R}^{(E \times E) \times E}$

$C(u)v = \sigma \sigma^*(u, v)$

with $\sigma: \mathbb{R}^E \times \mathbb{R}^E \rightarrow \mathbb{R}^{E \times E}$

~~Neutral fractions / allelic components~~ (K alleles)

$(X_t^{N,1}, \dots, X_t^{N,K})$ solutions to

$$dX_t^k = F\left(\sum_{i=1}^K X_t^{N,i}\right) X_t^k + \frac{1}{\sqrt{N}} \sigma\left(\sum_{i=1}^K X_t^{N,i}, X_t^k\right) dW_t^k \quad (\text{SDE}_2)$$

(W_t^k) independent d -dimensional Brownian motions

Fractions: $\sum X_t^{N,k}$ solution to (SDE₁)
drift and covariance structure "compatible" with the sum

Neutral: the fractions are exchangeable + consistent

ii) ~~Connection with IBM~~

Consider a discrete model with frequency-dependent reproduction rates

$(r(z, u, m)); z \in E, u, m \in \mathbb{N}^E$

Continuous-time Markov process with generator

$$G f(u) = \sum_{\substack{z \in E \\ m \in \mathbb{N}^E}} \mu_z r(z, u, m) (f(u + m - e_z) - f(u))$$

(e_z) canonical basis of \mathbb{R}^E

Assume the system is parametrised by a demographic parameter N

$$r^{(N)}(z, u, m) = r\left(z, \frac{u}{N}, m\right).$$

The generator of the rescaled process is given by

$$G^N f(u) = \langle F(u)u, \nabla f(u) \rangle + \frac{1}{2N} \sum_{z, y \in E} (C(u)u)_{zy} \partial_{zy} f(u) + o\left(\frac{1}{N}\right)$$

with $F(u)_{zy} = \sum_m r(z, u, m) (m_y - \delta_{zy})$

$(C(u)u)_{zy} = \sum_m r(z, u, m) (m_z - \delta_{zz}) (m_y - \delta_{zy})$

In the large- N limit, the generator "converges" with an co-decomposable SDE.

3) RESULT

ASSUMPTION 2

The (F, G) decomposition is "regular" and "positive"

ASSUMPTION 3 (stable ecological equilibrium)

The ODE $\dot{x} = b(x)$ has a locally exponentially stable equilibrium $\tilde{h} > 0$

+ Perron-Frobenius type conditions \Rightarrow the principal eigenvalue of $F(\tilde{h})$ is 0 and the associated eigenspace is spanned by \tilde{h} .

RK1: $\langle \tilde{h}, 1 \rangle \neq 1$ carrying capacity of the system.

RK2: $\exists \hat{h} > 0$ such that $t_{\hat{h}} F(\hat{h}) = 0$ and $\langle \hat{h}, \tilde{h} \rangle = 1$

PF: $\lambda_{\hat{h}}$ reproductive value of type \rightarrow individuals.

Define $\Sigma^2 := \langle \tilde{h}, a^*(\tilde{h}) \tilde{h} \rangle$

EX: $a(U) = \text{diag}((V U_{\alpha})_{\alpha \in E})$ (local branching)
 $\Sigma^2 = \sum_{\alpha} h_{\alpha}^2 \tilde{h}_{\alpha}$.

Then, under natural conditions on the I.C.,

$$(X_{\alpha}^{nk} / \Sigma^2) \Rightarrow \partial_t \otimes \tilde{h}$$

where (∂_t) is a $(K-1)$ dimensional Wright-Fisher diffusion

$$(\partial_t \otimes \tilde{h})_{\alpha k} = \partial_t^k \tilde{h}_{\alpha}$$

~~Effective population size~~

$$N_e = \frac{N}{\Sigma^2} \quad (\text{cf. Cannings model})$$

Σ^2 : reproductive variance of the structured population.

RK3: Σ^2 is defined from the ecological description of the pop. BUT

the choice of the (F, G) decomposition impacts the result

\tilde{h} does not depend on F
 \hat{h} does depend on F } same macroscopic behavior
 different fraction dynamics

F depends on the underlying individual based model

RK4: ~~Slow-fast principle~~ (= collapse of structure)

fast ecological dynamics: fractions quickly converge to multiples of \tilde{h}

slow evolutionary dynamics: proportions evolve on a slower timescale according to a WF diffusion.