The effective selective advantage of a clone within a population is well-described by the drift term of the birth-death process' diffusion approximated SDE. Typically one defines selection through an altered growth rate $\rho(1+s)$ with ρ the wildtype growth rate and s parameterizing the strength of selection. One finds the drift term of the SDE from the stochastic birth-death process by subtracting the probability of decrease of the clone size in an infinitesimal timestep from the probability of an increase. In a population of fixed size this is

$$\mathrm{d}x = \mathbb{P}\{\text{increase in } \mathrm{d}t\} - \mathbb{P}\{\text{decrease in } \mathrm{d}t\} \tag{1}$$

For the Wright-Fisher (WF) process – which describes competitive dyamics within a population of fixed size – with positive selection ($s > 0$) this becomes

$$\mathrm{d}x = \rho s\, x(1-x)\,\mathrm{d}t \tag{2}$$

which equates to a logistic growth with rate $\gamma = \rho s$. It can be shown that we obtain the same deterministic dynamics if we consider a two-type linear birth-death process where the total birth rate $\mathcal{A}$ in the population is the same as the total death rate $\mathcal{B}$. If we consider a system of V alleles with respective sizes X_i (with $i \in 0,\dots,V-1$ and $\sum_i X_i = N$) and birth and death rates α_i and β_i, the total rates are given by

$$\mathcal{A} = \sum_i \alpha_i X_i$$
$$\mathcal{B} = \sum_i \beta_i X_i$$

1 Positive selection

Defining selection through an increased birth rate, variant i has (per cell) birth and death rates

$$\begin{aligned} \alpha_i &= \alpha(1+s_i) \\ \beta_i &= \alpha\frac{\sum_j (1+s_j)X_j}{N} \end{aligned} \tag{3}$$

where we have used that $\beta_i = \mathcal{B}/N$ and $\mathcal{B} = \mathcal{A}$. The drift term of the birth-death process is then given by

$$\begin{aligned} \frac{\mathrm{d}X_i}{\mathrm{d}t} &= \alpha_i X_i - \beta_i X_i \\ &= \alpha(1+s_i)X_i - \alpha\frac{\sum_j (1+s_j)X_j}{N}X_i \end{aligned}$$

If we for simplicity assume $s_i = s\,\forall i \neq 0$, ($s_0 = 0$ describing wildtype cells), then we can easily split contributions from wildtype cells and other variants:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \alpha(1+s)X_i - \left(\alpha(1+s)\left[\frac{X_i}{N} + \frac{\sum_{k\neq i} X_k}{N}\right] + \alpha\left[1 - \left(\frac{X_i}{N} + \frac{\sum_{k\neq i} X_k}{N}\right)\right]\right)X_i \tag{4}$$

which after simplification and change of variables to $x_i = X_i/N$ becomes

$$\mathrm{d}x_i = \alpha s x_i \left(1 - x_i - \sum_{j \neq i} x_j\right) \mathrm{d}t + \sqrt{\frac{\alpha}{N}\left(2 + s\left[1 + \sum_j x_j\right]\right) x_i}\, \mathrm{d}W_t \tag{5}$$

or, allowing s to vary across variants:

$$\mathrm{d}x_i = \alpha x_i \left(s_i - \sum_{j>0} s_j x_j\right) \mathrm{d}t + \sqrt{\frac{\alpha}{N}\left(2 + s_i + \sum_{j>0} s_j x_j\right) x_i}\, \mathrm{d}W_t \tag{6}$$

The sum over all variants effectively turns this into a set of coupled differential equations:

$$\begin{cases} \mathrm{d}x_1 = \alpha s x_1 \left(1 - \sum_j x_j\right) \mathrm{d}t + \sigma(x_1)\, \mathrm{d}W_t \\ \mathrm{d}x_2 = \alpha s x_2 \left(1 - \sum_j x_j\right) \mathrm{d}t + \sigma(x_2)\, \mathrm{d}W_t \\ \quad \vdots \\ \mathrm{d}x_{V-1} = \alpha s x_{V-1} \left(1 - \sum_j x_j\right) \mathrm{d}t + \sigma(x_{V-1})\, \mathrm{d}W_t \end{cases} \tag{7}$$

However, the tricky part is that the number of variants V changes over time, according to the stochastic arrival of new variants through mutation as well as the loss of variants due to stochastic extinction. We can however consider a expected value picture, in which we drop the noise terms and assume (survival-conditioned) variants arise linearly from a Poisson process. To this end we move to a continuous picture of variants $x_i(t) \to x(v, t)$, which allows us to write the system of DE's as a single partial differential equation:

$$\frac{\partial x(v,t)}{\partial t} = \alpha s\, x(v,t) \left(1 - \int_V x(v,t)\, \partial v\right) \tag{8}$$

Because new variants arise at size $1/n$ at rate μt, we have the boundary conditions

$$x(0,0) = 1/n \tag{9}$$

$$x(\mu t, t) = 1/n \tag{10}$$

Furthermore, because at time t there are only μt existing variants, the integral boundaries can be written as $\int_0^{\mu t} x(v,t)\, \mathrm{d}v$.

2 Maximum Likelihood estimation of logistic growth curves

Denote $f(t, \beta)$ the logistic growth function bounded between 0 and 0.5, taken at time t with parameter set $\beta = \{t_0, n_0, \gamma\}$.:

$$f(t, \beta) = \frac{0.5}{1 + \frac{0.5 - n_0}{n_0} e^{-\gamma(t - t_0)}} \tag{11}$$

The probability of measuring j counts of a variant under $f(t, \beta)$ with a coverage k is given by the binomial distribution:

$$P(j|k, \beta) = \binom{k}{j} f(t, \beta)^j \left[1 - f(t, \beta)\right]^{(K-j)} \tag{12}$$

Denote $V = \{v_i\}$ and $K = \{k_i\}$ as respectively the sets of measured variant calls and coverages obtained at times $\{t_i\}$. The probability of obtaining the sample set V under a model β is thus given by

$$\mathcal{L}(\beta|V, K) = \prod_i P(v_i|k_i, \beta), \tag{13}$$

the likelihood of the measurement given the model parameterized by β. In order to obtain the best fitting set of parameters β we thus simply maximize $\mathcal{L}(\beta|V, K)$.

3 Competition-adjusted growth curve

The deterministic term of the growth function described by the full model

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \alpha x_i \left(s_i - s_i x_i - \sum_{j \neq i} s_j x_j \right) \tag{14}$$

cannot easily be used as the expected growth, as the dependence on other existing variants through the summation (the population-wide average fitness) keeps stochasticity in the expression. Let us as a simplification then assume this average fitness changes slowly, so that for a short observation period we might approximate it as a constant:

$$\sum_{j \neq i} s_j x_j \approx z \tag{15}$$

Then the above DE can be solved as

$$\frac{0.5 - {}^{z}\!/_{s}}{1 + \frac{0.5 - x_0 - {}^{z}\!/_{s}}{x_0} e^{-(s - 2z)(t - t_0)}} \tag{16}$$

4 Solving the growth curve exactly

4.1 Fixed fitness values

Taking only the deterministic part of the growth curve, we may write for a variant that originated at time u and is measured at time t:

$$\begin{cases} \dfrac{\mathrm{d}x(t,u)}{\mathrm{d}t} = \alpha s\, x(t,u)[1-y(t)] \\ y(t) = \displaystyle\int_0^t \mu[1-y(u)]\, x(t,u)\,\mathrm{d}u \end{cases} \tag{17}$$

where we have written the sum over all existing variants at t as an integral over the past times (bottom equation). Taking the derivative in time of the second expression we obtain:

$$\begin{aligned} \frac{\mathrm{d}y(t)}{\mathrm{d}t} &= \int_0^t \mu[1-y(u)]\frac{\mathrm{d}x(t,u)}{\mathrm{d}t}\,\mathrm{d}u + \mu[1-y(t)]\frac{x(t+\mathrm{d}t\,,t)}{\mathrm{d}t} \\ &= \alpha s[1-y(t)]\int_0^t \mu[1-y(u)]x(t,u)\,\mathrm{d}u + [1-y(t)]\frac{\mu}{N} \end{aligned}$$

In the second expression we identify the integral as $y(t)$, so that we may rewrite our system as

$$\begin{cases} \dfrac{\mathrm{d}x(t,u)}{\mathrm{d}t} = \alpha s\, x(t,u)[1-y(t)] \\ \dfrac{\mathrm{d}y(t)}{\mathrm{d}t} = \Big(\alpha s y(t) + \dfrac{\mu}{N}\Big)[1-y(t)] \end{cases} \tag{18}$$

The expression for $y(t)$ is closed, and has with initial condition $y(0)=0$ the solution

$$y(t) = 1 - \frac{1+\frac{\mu}{\alpha s N}}{1+\frac{\mu}{\alpha s N}e^{\left(\alpha s+\frac{\mu}{N}\right)t}} \tag{19}$$

Inserting this into the expression for $x(t,u)$ with initial condition $x(u,u)=1/N$ gives

$$x(t,u) = \frac{\frac{1}{N}+\frac{\alpha s}{\mu}e^{-\left(\alpha s+\frac{\mu}{N}\right)u}}{1+\frac{\alpha s}{\mu}Ne^{-\left(\alpha s+\frac{\mu}{N}\right)t}} \tag{20}$$

4.2 Individual fitness in average fitness landscape

Consider a variant with fitness s arriving in a system with average fitness s. We investigate approximating this as the system

$$\begin{cases} \dfrac{\mathrm{d}x(t,u)}{\mathrm{d}t} = \alpha s\, x(t,u)\Big[1-\dfrac{a}{s}y(t)\Big] \\ \dfrac{\mathrm{d}y(t)}{\mathrm{d}t} = \Big(\alpha a y(t) + \dfrac{\mu}{N}\Big)[1-y(t)] \end{cases} \tag{21}$$

4.3 Fully stochastic fitness

Let us now generalize to the case where a clone may have any positive fitness $s \in \mathbb{R}$, drawn from some distribution with pdf $p(s)$. We then have

$$\begin{cases} \dfrac{\mathrm{d}x(t,u,s)}{\mathrm{d}t} = \alpha s\, x(t,u,s) \left[1 - \dfrac{z(t)}{s}\right] \\ z(t) = \displaystyle\int_0^t \mathrm{d}u\, \tilde{\mu}(u) \int_0^\infty \mathrm{d}s\, p(s) s\, x(t,u,s) \end{cases} \tag{22}$$

Again deriving the second expression with respect to t, and denoting $\int_0^\infty \mathrm{d}s\, p(s) f(s) = \langle f(s) \rangle$ we obtain

$$\begin{cases} \dfrac{\mathrm{d}x(t,u,s)}{\mathrm{d}t} = \alpha s\, x(t,u,s) \left[1 - \dfrac{z(t)}{s}\right] \\ \dfrac{\mathrm{d}z(t)}{\mathrm{d}t} = \alpha \displaystyle\int_0^t \mathrm{d}u\, \tilde{\mu}(u) \big[\big\langle s^2 x(t,u,s) \big\rangle - \langle x(t,u,s) z(t) \rangle\big] + \dfrac{\tilde{\mu}(t)}{N} \end{cases} \tag{23}$$

We can unfortunately no longer perform the same 'trick' as before, since the expression for $z(t) = \int_0^t \mathrm{d}u\, \tilde{\mu}(u) \, \langle s x(t,u,s) \rangle$ does not appear directly in the expression above.