

1 VAF spectrum of the pure-birth process

In a pure-birth process every cell divides independently with exponentially distributed times between divisions at rate γ . If the population size is 1 at time $t = 0$, the expected size of the population at time t is $N(t) = e^{\gamma t}$. Allowing mutations to occur during division at rate μ per daughter, we investigate the expected number of variants at each prevalence in the population. Denote $V_k(t)$ the expected number of variants with size k (i.e. found in k cells) at time t . In an infinitesimal time step, the expected number of variants moving from prevalence $k > 1$ to $k+1$ is given by $k/N(t) \cdot \gamma N(t) dt$. Furthermore, the expected number of variants entering the system at state $k = 1$ due to mutations is $2\mu\gamma N(t) dt$. Thus the expected variation of $V_k(t)$ is given by the master equation

$$\begin{cases} \frac{dV_1(t)}{dt} = -\gamma V_1(t) + 2\mu\gamma N(t) \\ \frac{dV_k(t)}{dt} = -k\gamma V_k(t) + (k-1)\gamma V_{k-1}(t) \end{cases} \quad (1)$$

The first state can be immediately solved to obtain

$$V_1(t) = \mu(e^{\gamma t} - e^{-\gamma t}) \quad (2)$$

On the other hand, for any state k we have

$$V_k(t) = (k-1)\gamma e^{-k\gamma t} \int_0^t e^{k\gamma t'} V_{k-1}(t') dt' \quad (3)$$

Thus, integrating iteratively gives

$$\begin{aligned} V_k(t) &= \mu(k-1)! \gamma^{k-1} e^{-k\gamma t} \\ &\times \int_0^{t_{k-1}} dt_{k-1} \int_0^{t_{k-2}} dt_{k-2} \cdots \int_0^{t_1} dt_1 e^{\gamma t_{k-1}} e^{\gamma t_{k-2}} \cdots e^{\gamma t_1} (e^{2\gamma t_1} - 1) \end{aligned} \quad (4)$$

It can be shown (At this point I just read it off Mathematica... :-p) that this results in

$$V_k(t) = \frac{2\mu}{k^2 + k} (N(t) - 1)^k \left(\frac{k}{N(t)^k} + \frac{N(t)}{N(t)^k} \right) \quad (5)$$

In the limit of large k this becomes

$$V_k(t) \approx \frac{2\mu}{k^2 + k} (N(t) + k) \quad (6)$$

We can check to see if this result is properly normalized by checking whether

$$\sum_{k=1}^N V_k = \int_0^t 2\mu\gamma N(t) dt = 2\mu N(t) \quad (7)$$

Mathematica tells me this is true in the limit of large N but I haven't quite been able to show this yet...

2 Normalization with respect to total number of variants

For a population subject to pure birth dynamics, we have that the VAP spectrum $V_k(t)$ (number of variants at each prevalence) is in equilibrium if normalized by the number of variants with prevalence 1:

$$\frac{V_k(t)}{V_1(t)} = \frac{2}{k^2 + k} \quad (8)$$

We can find the non-normalized $V_k(t)$ by finding the value of $V_1(t)$. To this end we use the fact that the sum over all states k must equal the total number of mutations that are in the system at time t

$$V_1(t) \sum_{k=1}^N \frac{2}{k^2 + k} = \int_0^t 2\mu\gamma N(t) dt \quad (9)$$

To simplify the LHS we move from prevalences to frequency space (VAF spectrum) $v_{k/N} = V_k$, and note that we are working in the limit of large N

$$V_1(t) \sum_{k=1}^N \frac{2}{k^2 + k} = v_{1/N} \sum_{f=1/N}^1 \frac{1}{N} \frac{2}{Nf^2 + f} \quad (10)$$

$$\approx v_{1/N} \int_{1/N}^1 df \frac{2}{Nf^2 + f} \quad (11)$$

$$= v_{1/N} 2 \log \left[\frac{2N}{1 + N} \right] \quad (12)$$

$$\approx v_{1/N} 2 \log(2) \quad (13)$$

Meanwhile, performing the integration on the RHS gives

$$\int_0^t 2\mu\gamma N(t) dt = 2\mu(e^{\gamma t} - 1) \quad (14)$$

$$\approx 2\mu N(t) \quad (15)$$

So that we obtain

$$v_{1/N(t)}(t) = \frac{\mu N(t)}{\log(2)} \quad (16)$$

And thus for any state f

$$v_f(t) = \frac{2\mu}{N(t)f^2 + f} \frac{1}{\log(2)} \quad (17)$$