Fano Factor for Mutator Phenotypic State Switching

November 7, 2024

We aim to calculate the mean $E[M(t)]$ and variance $Var[M(t)]$ of the number of mutations $M(t)$ after t generations in the switching process, using the expressions for the fraction of high rate mutator cells, $f_H(t)$.

Definitions and Assumptions:

1. States:
   - State 1: Low mutator cells with mutation rate μ_L
   - State 2: High mutator cells with mutation rate μ_H, where $\mu_H \gg \mu_L$.
2. Switching rates:
   - r_{LH} : Rate of switching from State 1 to State 2
   - r_{HL} : Rate of switching from State 2 to State 1
   - Equilibrium fraction of high mutator cells: $\hat{f}_H = \frac{r_{LH}}{r_{LH}+r_{HL}}$
   - Equilibrium fraction of low mutator cells: $\hat{f}_L = 1 - \hat{f}_H = \frac{r_{HL}}{r_{LH}+r_{HL}}$
3. Define $\mu = \hat{f}_H \mu_H + \hat{f}_L \mu_L$
4. If we draw a random cell from the equilibrium colony, and let it grow over time, the fraction of high mutator cells in the growing colony at time t can be solved:

$$f_H(t) = \frac{x_H(t)}{x(t)} = \begin{cases} \hat{f_H}\left(1 - e^{-(r_{LH}+r_{HL})t}\right) & p = 1 - \hat{f_H} \\ \hat{f_H} + \left(1 - \hat{f_H}\right) e^{-(r_{LH}+r_{HL})t} & p = \hat{f_H} \end{cases}$$

where $x_H(t)$ is the number of cells in high mutator state, and $x(t)$ is the total number of cells at time t.

Luria-Delbruck vs. State Switching Hypothesis

Say there are 2 hypothesis for how a culture of mutating organisms mutate:

1. State switching as defined above, with mutation rates μ_L, μ_H

2. All the population mutates with rate μ

Can we distinguish between the 2 states somehow? We shall see the Fano factor(variance over the mean) for the number of mutations(the genetic richness of the population) distinguishes between both. For the LD case this is easy - the number of mutations is a regular Poisson process with Fano factor(FF) 1.

We shall derive the FF for the switching hypothesis.

Derivation

- Population size(given growth rate λ that does not change between mutants) :

$$N(t) = N_0 e^{\lambda t}$$

- Mean and Variance of $f_H(t)$:

$$\begin{aligned} E[f_H(t)] &= \hat{f}_H \left(\hat{f}_H + \left(1 - \hat{f}_H\right) e^{-(r_{LH}+r_{HL})t} \right) + \left(1 - \hat{f}_H\right) \hat{f}_H \left(1 - e^{-(r_{LH}+r_{HL})t}\right) \\ &= \hat{f}_H \end{aligned}$$

$$\begin{aligned} E[f_H^2(t)] &= \hat{f}_H \left(\hat{f}_H + \left(1 - \hat{f}_H\right) e^{-(r_{LH}+r_{HL})t} \right)^2 + \left(1 - \hat{f}_H\right) \hat{f}_H^2 \left(1 - e^{-(r_{LH}+r_{HL})t}\right)^2 \\ &= \hat{f}_H^2 + \hat{f}_H \left(1 - \hat{f}_H\right) e^{-2(r_{LH}+r_{HL})t} \end{aligned}$$

$$Var\left[f_H(t)\right] = E[f_H^2(t)] - E[f_H(t)]^2 = \hat{f}_H \left(1 - \hat{f}_H\right) e^{-2(r_{LH}+r_{HL})t}$$

- Mean Number of Mutations $E[M(t)]$:

$$\begin{aligned} E[M(t)] &= E[M_L(t)] + E[M_H(t)] \\ &= E\left[\int_0^t N_L(t')\mu_L dt'\right] + E\left[\int_0^t N_H(t')\mu_H dt'\right] \\ &= N_0\mu_L \int_0^t e^{\lambda t'} E\left[f_L(t')\right] dt' + N_0\mu_H \int_0^t e^{\lambda t'} E\left[f_H(t')\right] dt' \\ &\approx N_0 \frac{\mu_L \hat{f}_L + \mu_H \hat{f}_H}{\lambda} e^{\lambda t} = N(t)\frac{\mu}{\lambda} \end{aligned}$$

where:

$$N_i(t) = N(t) f_i(t)$$

- Variance of Number of Mutations $\mathrm{Var}[M(t)]$, using the law of total variance:

$$\mathrm{Var}[M(t)] = E\left[\mathrm{Var}[M(t) \mid f_H(t)]\right] + \mathrm{Var}\left[E[M(t) \mid f_H(t)]\right]$$

First term("unexplained component"), mutation is a Poisson process:

$$\begin{aligned}
\mathrm{Var}[M(t) \mid f_H(t)] &= \mathrm{Var}[M_L(t) + M_H(t) \mid f_H(t)] \\
&= \mathrm{Var}[M_L(t) \mid f_L(t)] + \mathrm{Var}[M_H(t) \mid f_H(t)] \\
&= \int_0^t \left(N_L(t')\mu_L(1-\mu_L) + N_H(t')\mu_H(1-\mu_H)\right) dt'
\end{aligned}$$

Where we treat $M_L(t), M_H(t)$ as independent RVs with no covariance.
Taking $\mu_L, \mu_H \ll 1$, we can approximate:

$$\begin{aligned}
\mathrm{Var}[M(t) \mid f_H(t)] &\approx \int_0^t \left(N_L(t')\mu_L + N_H(t')\mu_H\right) dt' \\
&= N_0 \int_0^t \left(f_L(t')\mu_L + f_H(t')\mu_H\right) e^{\lambda t'} dt'
\end{aligned}$$

$$\begin{aligned}
E\left[\mathrm{Var}[M(t) \mid f_H(t)]\right] &= N_0 \int_0^t \left(E\left[f_L(t')\right]\mu_L + E\left[f_H(t')\right]\mu_H\right) e^{\lambda t'} dt' \\
&= N(t)\frac{\mu}{\lambda}
\end{aligned}$$

This term in the law of total variance is what is called the "unexplained component". It stems from the variance of the Poisson process even for a given IC of $f_H(t)$.

Now for the second term("explained component"):

$$\begin{aligned}
E[M(t) \mid f_H(t)] &= \int_0^t \left(N_L(t')\mu_L + N_H(t')\mu_H\right) dt' \\
&= \int_0^t N(t')\left(\mu_L\left(1 - f_H(t')\right) + \mu_H f_H(t')\right) dt' \\
&= N_0 \int_0^t \left(\mu_L + \left(\mu_H - \mu_L\right) f_H(t)\right) e^{\lambda t'} dt' \\
&\approx N_0 \int_0^t \left(\mu_L + \mu_H f_H(t')\right) e^{\lambda t'} dt'
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\left[E[M(t)\mid f_H(t)]\right] &= N_0^2\mu_H^2\mathrm{Var}\left[\int_0^t \left(\begin{cases}\hat{f_H}\left(1-e^{-(r_{LH}+r_{HL})t'}\right) & p=1-\hat{f_H}\\ \hat{f_H}+\left(1-\hat{f_H}\right)e^{-(r_{LH}+r_{HL})t'} & p=\hat{f_H}\end{cases}\right)\times e^{\lambda t'}dt'\right]\\
&\approx N_0^2\mu_H^2e^{2\lambda t}\mathrm{Var}\left[\begin{cases}\hat{f_H}\left(\frac{1}{\lambda}-\frac{1}{\lambda-(r_{LH}+r_{HL})}e^{-(r_{LH}+r_{HL})t}\right) & p=1-\hat{f_H}\\ \hat{f_H}\frac{1}{\lambda}+\left(1-\hat{f_H}\right)\frac{1}{\lambda-(r_{LH}+r_{HL})}e^{-(r_{LH}+r_{HL})t} & p=\hat{f_H}\end{cases}\right]\\
&:= N_0^2\mu_H^2e^{2\lambda t}\mathrm{Var}\left[I\right]
\end{aligned}$$

Compute the variance of I:

$$E\left[I\right]=\frac{\hat{f_H}}{\lambda}$$

$$\begin{aligned}
E\left[I^2\right] &= \left(1-\hat{f_H}\right)\hat{f_H}^2\left(\frac{1}{\lambda}-\frac{1}{\lambda-(r_{LH}+r_{HL})}e^{-(r_{LH}+r_{HL})t}\right)^2\\
&+\hat{f_H}\left(\hat{f_H}\frac{1}{\lambda}+\left(1-\hat{f_H}\right)\frac{1}{\lambda-(r_{LH}+r_{HL})}e^{-(r_{LH}+r_{HL})t}\right)^2\\
&=\frac{\hat{f_H}^2}{\lambda^2}+\hat{f_H}\left(1-\hat{f_H}\right)\frac{e^{-2(r_{LH}+r_{HL})t}}{\left(\lambda-(r_{LH}+r_{HL})\right)^2}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\left[I\right] &= E\left[I^2\right]-E\left[I\right]^2\\
&=\hat{f_H}\left(1-\hat{f_H}\right)\frac{e^{-2(r_{LH}+r_{HL})t}}{\left(\lambda-(r_{LH}+r_{HL})\right)^2}
\end{aligned}$$

So in total:

$$\begin{aligned}
\frac{\mathrm{Var}\left[M(t)\right]}{\mathrm{E}\left[M(t)\right]} &= \frac{N(t)\frac{\mu}{\lambda}+N_0^2\mu_H^2e^{2\lambda t}\hat{f_H}\left(1-\hat{f_H}\right)\frac{e^{-2(r_{LH}+r_{HL})t}}{\left(\lambda-(r_{LH}+r_{HL})\right)^2}}{N(t)\frac{\mu}{\lambda}}\\
&=1+N_0\hat{f_H}\left(1-\hat{f_H}\right)\frac{\lambda}{\mu}\frac{\mu_H^2}{\left(\lambda-(r_{LH}+r_{HL})\right)^2}e^{(\lambda-2(r_{LH}+r_{HL}))t}\\
&=1+N(t)\mathrm{Var}\left[f_H(t)\right]\frac{\mu_H}{\mu}\frac{\lambda\mu_H}{\left(\lambda-(r_{LH}+r_{HL})\right)^2}
\end{aligned}$$

Where $\mathrm{Var}\left[f_H(t)\right]$ is what we derived as the variance of the fraction of high mutators in single colony that originates from a randomly drawn cell from the equilibrium distribution. It is obvious this can differ vastly from 1, which would be the value of the FF for the uniform mutation-rate hypothesis.