

The number of mutations carried by a single cell is given by the sum over all mutations occurring during each of its past divisions:

$$m = \sum_i^l u_i \quad (1)$$

Without making any assumptions about the distributions of the number of past divisions l and the number of mutations per division u , The expected value of m can easily be found from the law of total expectation

$$E(m) = E(l) E(u) \quad (2)$$

and the variance of m can be written through the law of total variance

$$\text{Var}(m) = E(l) \text{Var}(u) + E(u)^2 \text{Var}(l) \quad (3)$$

Now we introduce the fact that the number of mutations per division are Poisson distributed with parameter μ (i.e. $E(u) = \text{Var}(u) = \mu$) and subtract the mean from the variance:

$$\begin{aligned} \text{Var}(m) - E(m) &= \mu^2 \text{Var}(l) \\ \Rightarrow \frac{\text{Var}(m) - E(m)}{E(m)} &= \mu \frac{\text{Var}(l)}{E(l)} \end{aligned}$$

so finally we can find the mutation rate from

$$\mu = \left(\frac{\text{Var}(m)}{E(m)} - 1 \right) \frac{E(l)}{\text{Var}(l)} \quad (4)$$

If the number of divisions is Poisson distributed, then $E(l) = \text{Var}(l)$ so that $\mu = \text{Var}(m)/E(m) - 1$. This is the compound Poisson estimate which we will denote as $\tilde{\mu}$. If we take the $\mu = 1.2$ as true, then the CPD estimate $\tilde{\mu}$ gives an indication of how the system dynamics differ from the Poisson model. In particular, if $\tilde{\mu} < \mu$ then we must have $\text{Var}(l) < E(l)$, which could be an indication of the existence memory in the system, for example in the form of divisions occurring according to a timed cell cycle. Conversely, if $\tilde{\mu} > \mu$, we have $\text{Var}(l) > E(l)$, which implies that various lineages actually divided far more or far less than what would be appropriate for a memoryless system. Interestingly, the latter observation need not be at odds with current knowledge of stem cell dynamics, as the possibility of a quiescence state – which cells enter and exist according to some complex dynamics – could potentially lead to overdispersion compared to the Poisson model.