

A Bayesian approach for semi-automatic cell segmentation manual correction

Chang Chang @ Cluzel lab
 changchang@fas.harvard.edu
 MCB, Harvard
 Sep 7, 2018

Within one cell cycle, there is no cell division event. In this model, we assume the parameters are the underlying “true” cell length \vec{l} , and the growth rate λ . The posterior probability of observing these parameters given the observed images \vec{im} , is

$$P(\vec{l}, \lambda | \vec{im}) \propto P(\vec{im} | \vec{l}, \lambda) P(\vec{l}, \lambda). \quad (1)$$

The likelihood $P(\vec{im} | \vec{l}, \lambda)$ could be defined as the probability of observing cell lengths inferred from images directly (I name them as observed cell lengths)

$$\begin{aligned} P(\vec{im} | \vec{l}, \lambda) &\stackrel{\text{def}}{=} P(\vec{l}_{obs} | \vec{l}, \lambda) \\ &= P(\vec{l}_{obs} | \vec{l}) \\ &= \prod_k P(l_{obs,k} | l_k). \end{aligned} \quad (2)$$

where $l_{obs,k}$ and l_k is \vec{l}_{obs} and \vec{l} at time t . It would be reasonable to assume $l_{obs,k} \sim N(l_k, \sigma_{obs,k}^2)$. We analytically have the expression for the likelihood part.

For the second term in Eq. (1),

$$P(\vec{l}, \lambda) = P(\vec{l} | \lambda) P(\lambda). \quad (3)$$

Let's assume an uninformative prior for $P(\lambda)$. Thus,

$$\begin{aligned} P(\vec{l}, \lambda) &= P(\vec{l} | \lambda) \\ &= P(l_1) \prod_{k=2}^n P(l_k | l_{k-1}, \lambda) \end{aligned} \quad (4)$$

For $P(l_1)$, we could define it as $\delta(l_1, l_{obs,1})$, and no longer update it in the iteration. For the second term, we could assume $l_k \sim N(El_k, \sigma_\lambda^2)$, where El_k is the expected length at time k , based on the cell length at $k-1$ and growth rate λ . We assume an exponential growth in this model.

Overall, the posterior probability is given by

$$\ln P(\vec{l}, \lambda | \vec{im}) \propto \ln P(\vec{im} | \vec{l}, \lambda) + \ln P(\vec{l}, \lambda)$$

$$= \sum_{k=2}^n [\ln P(l_{obs,k}|l_k) + \ln P(l_k|l_{k-1}, \lambda)], \quad (5)$$

which is a sum of many normal distributions.

To analytically have the equations for update, we write down the partial derivative of the posterior probability to each parameter, and we get

$$l_k = \alpha_k l_{obs,k} + (1 - \alpha_k) El_k, \quad (6)$$

$$\lambda = \frac{\ln \frac{\sum_k l_k l_{k-1}}{\sum_k l_{k-1} l_{k-1}}}{\Delta t}, \quad (7)$$

$$El_k = l_{k-1} \exp(\lambda \Delta t), \quad (8)$$

$$\alpha_k = \frac{\sigma_\lambda^2}{(l_{obs,k} - l_k)^2 + \sigma_\lambda^2}. \quad (9)$$

The physical meaning of Eq. (6) is, the underlying “true” cell length is an average based on the information from the image and the information from the cell growth. Whether the first part of more information is more reliable or the second part, depending on the α_k , *i.e.*, Eq. (9). In the limiting case of $(l_{obs,k} - l_k)^2 \ll \sigma_\lambda^2$, *i.e.*, the observed cell length is close enough to the inferred cell length (and then must close to the one calculated from cell growth), $\alpha_k \rightarrow 1$, we would like to trust the observed cell length. On the other hand, in the limiting case of $(l_{obs,k} - l_k)^2 \gg \sigma_\lambda^2$, the observed cell length is far away from the inferred cell length, $\alpha_k \rightarrow 0$, we could more trust the information from cell growth part.