Estimating half-life

Peter Swain

We use a Bayesian approach to estimate the half-life of Pma1 from the data of Fig. 1.

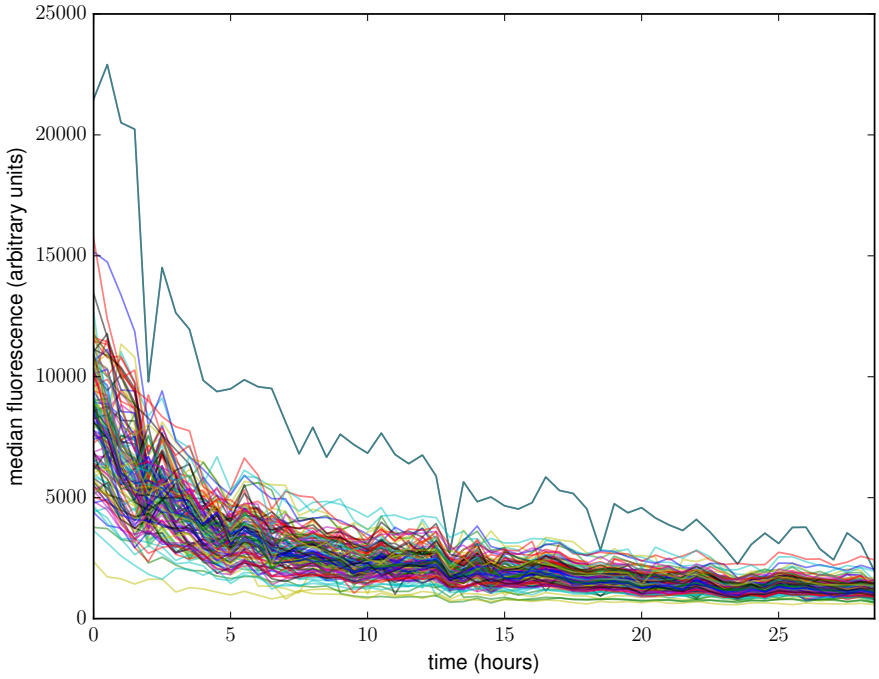

Figure 1. The median fluorescence of single cells falls over time. Each cell has a different initial and final value. 120 cells were followed over 58 time points with intervals of 30 minutes.

Considering first a single cell, the observed fluorescence will have three components: from GFP bound to Pma1 at the plasma membrane; from GFP in the cytosol; and from autofluorescence. At the membrane, GFP decays at a rate determined by the half-life of Pma1 because Pma1 is not transported into the daughter cell and GFP itself hardly decays; in the cytoplasm, GFP principally decays through entering and remaining in the daughter cell. We assume that autofluorescence does not decay. Mathematically, then, the observed fluorescence, f, obeys

$$f(t) = m_0 \mathrm{e}^{-d_m t} + c_0 \mathrm{e}^{-d_c t} + a_0 \tag{1}$$

where the initial GFP at the membrane, m_0, decays exponentially with a decay rate of d_m and the initial GFP in the cytoplasm, c_0, decays exponentially with a decay rate of d_c. The autofluorescence, a_0, does not change with time t.

The decay rates d_m and d_c are our focus, and m_0, c_0, and a_0 are 'nuisance' parameters, which ideally we would integrate away. For a cell indexed by j with data $D^{(j)}$, the posterior probability of d_m and d_c is

$$\begin{aligned} P^{(j)}\big(d_m, d_c|D^{(j)}\big) &= \int dm_0 dc_0 da_0 \, P^{(j)}\big(m_0, c_0, a_0, d_m, d_c|D^{(j)}\big) \\ &\sim \int dm_0 dc_0 da_0 \, P^{(j)}\big(D^{(j)}, m_0, c_0, a_0|d_m, d_c\big) \end{aligned} \tag{2}$$

assuming uniform, bounded prior probabilities for d_m and d_c.

We further assume that the errors in measuring fluorescence are identically and independently distributed with a Gaussian distribution of zero mean, but with a variance that changes with time. Assuming the measurement error is dominated by shot noise, we empirically estimate this variance as the mean fluorescence taken across all cells at each time point.

The likelihood is then

$$\begin{aligned} P^{(j)}\big(D^{(j)}|d_m,d_c\big) &\sim \int dm_0 dc_0 da_0 \prod_i \exp\left[\frac{-(d_i - m_0 \mathrm{e}^{-d_m t_i} - c_0 \mathrm{e}^{-d_c t_i} - a_0)^2}{2\sigma_i^2}\right] \\ &= \int dm_0 dc_0 da_0 \exp\left[\sum_i \frac{-(d_i - m_0 \mathrm{e}^{-d_m t_i} - c_0 \mathrm{e}^{-d_c t_i} - a_0)^2}{2\sigma_i^2}\right] \end{aligned} \tag{3}$$

where i runs over all time points and σ_i is known.

By making c_0 and m_0 be independent variables in Eq. 3, we have assumed that the amount of GFP bound to Pma1 at the membrane, m_0, and the amount of GFP in the cytoplasm, c_0, are independent. With the Spycatcher tag, GFP irreversibly binds to Pma1, and our assumption of independence is strengthened if all the Pma1 at the membrane is bound by GFP. The integral in Eq. 3 should peak at positive values of m_0, c_0, and a_0, and therefore we can extend the range of integration to be over all real numbers, enabling an exact integration if an approximate calculation.

Integrating Eq. 3 and defining $M_i = \mathrm{e}^{-d_m t_i}$ and $C_i = \mathrm{e}^{-d_c t_i}$, we find ten sufficient statistics:

$$\begin{array}{ccc} T_1 = \sum_i C_i^2/\sigma_i^2 & ; & T_2 = \sum_i d_i C_i/\sigma_i^2 \\ T_3 = \sum_i C_i M_i/\sigma_i^2 & ; & T_4 = \sum_i d_i^2/\sigma_i^2 \\ T_5 = \sum_i d_i M_i/\sigma_i^2 & ; & T_6 = \sum_i M_i^2/\sigma_i^2 \\ T_7 = \sum_i C_i/\sigma_i^2 & ; & T_8 = \sum_i \sigma_i^{-2} \\ T_9 = \sum_i d_i/\sigma_i^2 & ; & T_{10} = \sum_i M_i/\sigma_i^2 \end{array} \tag{4}$$

and that

$$\begin{aligned} P^{(j)}\big(D^{(j)}|d_m,d_c\big) \sim \exp\left[\frac{-T_4}{2} + \frac{T_1 T_5^2 + T_2^2 T_6 - 2T_2 T_3 T_5}{2u_T}\right] \times \\ \exp\left[\frac{(T_2 T_3 T_{10} - T_1 T_5 T_{10} + T_3 T_5 T_7 - T_2 T_6 T_7 + T_9 u_T)^2}{2u_T v_T}\right] / \sqrt{v_T} \end{aligned} \tag{5}$$

with

$$u_T = T_1 T_6 - T_3^2 \tag{6}$$

and

$$v_T = T_8 u_T - T_6 T_7^2 + 2T_3 T_7 T_{10} - T_1 T_{10}^2. \tag{7}$$

To extend our analysis to more than one cell, we assume that m_0, c_0, and a_0 for each cell are independent of their values in other cells so that the posterior probability satisfies

$$\begin{aligned} P(d_m,d_c|D) &\sim P(D|d_m,d_c) \\ &= \prod_j P^{(j)}\big(D^{(j)}|d_m,d_c\big) \end{aligned} \tag{8}$$

for j running over all cells and $P^{(j)}\big(D^{(j)}|d_m,d_c\big)$ being given by Eq. 5 with the sufficient statistics of Eq. 4 evaluated using the data from cell j.

For our data comprising 120 cells and 58 time-points per cell, plotting the negative logarithm of the posterior probability (Fig. 2) shows a well-defined minimum (corresponding to a maximum of the probability). We are interested in the half-life of Pma1, τ_m, which is determined by d_m. Integrating the posterior probability over d_c and using $\tau_m = \log 2/d_m$ and that

$$P(\tau_m) = \frac{\tau_m^2}{\log 2} P\Big(d_m = \frac{\log 2}{\tau_m}\Big) \tag{9}$$

through changing variables, we find that the marginal posterior probability for τ_m is sharply peaked at $\simeq 11.5$ hours with a 90% credible interval of $11.3 < \tau_m < 11.7$ hours (Fig. 3).

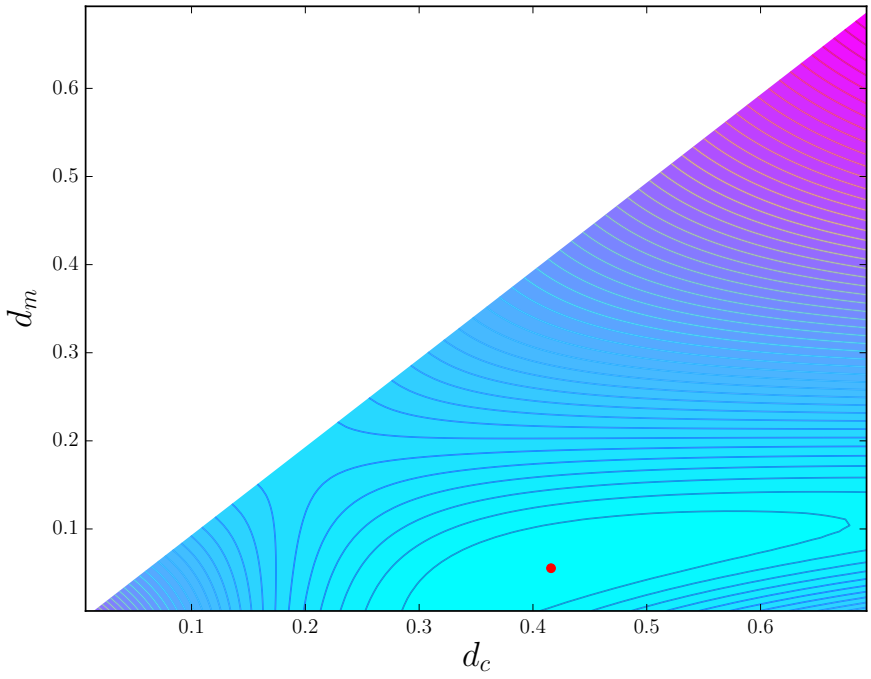

Figure 2. The posterior probability of d_m and d_c has a maximum at $d_c \simeq 0.4$ hr^{-1} and $d_m \simeq 0.06$ hr^{-1} (red dot). We plot the negative logarithm of the posterior probability, which has a minimum (light blue) at the most probable values of d_m and d_c. Darker blue and violet shading correspond to higher values. The probability is symmetric in d_m and d_c, and we plot only for $d_m < d_c$ because decay at the membrane is assumed to be slower than decay in the cytoplasm.

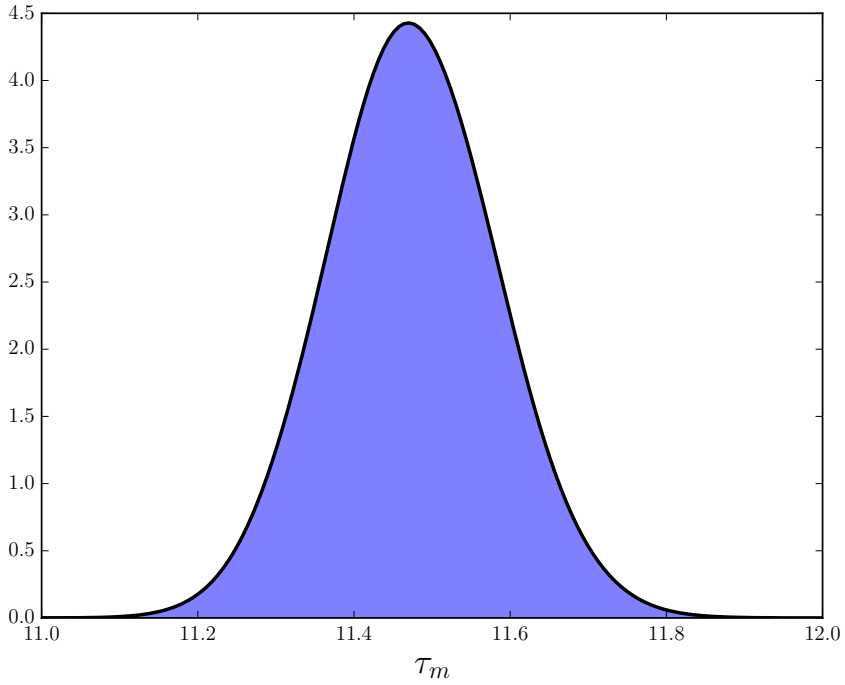

Figure 3. The posterior probability for the half-life of Pma1, τ_m, found by integrating the probability corresponding to the surface in Fig. 2 over d_c and changing variables from d_m to τ_m.