Supporting Information for "On the Dielectric Dependence of Single-Molecule Photoluminescence Intermittency: Nile Red in Poly(vinylidene fluoride)"

Chelsea M. Hess, Erin A. Riley, and Philip J Reid*

Box 351700, Dept. of Chemistry, University of Washington, Seattle, WA 98195

Description of Bayesian detection of intensity traces (BDIC)

The general assumption that molecule emission is Poisson distributed is employed. A Bayes factor is computed by comparing the probability for the hypothesis (H_1) that there is no change point in the data (D) and the probability for the alternate hypothesis (H_2) that there are two emitting states separated by a change point at time t_s. Assuming no prior knowledge of which hypothesis is likely to be true, the odds that there is a change point is simply the ratio of the two probabilities. The probabilities of the two hypotheses given the data are computed as the likelihoods, with the ratio being the Bayes factor (B):

$$B = \frac{P(D|H_2)}{P(D|H_1)}$$

After incorporating the appropriate details for the Poisson case and prior probabilities, the Bayes factor is [20]

$$B = \frac{2}{\pi} \frac{N^C}{(C-1)!\ (N-1)} \sum_{t_s} \frac{C_1!\, C_2!}{N_1^{C_1+1} N_2^{C_2+1}} \left[\left(\frac{C_1}{N_1}\right)^2 + \left(\frac{C_2}{N_2}\right)^2 \right]^{-1}$$

In the above expression t_s is the possible location of a change point, C_1 is the number of photons before the change point and C_2 is the number of photon counts after the change point, and C is the total number of photon counts in the trace. Analogously, N_1 is the number of time points before t_s, N_2 is the time points after the change point, and N is the total number of time points. The probability naturally sums over all the possible change points in the trace. The Bayes factor must be sufficiently large to contain substantial evidence for which hypothesis is appropriate. If there are only the two outcomes, then a Bayes factor of 4 is interpreted as 4:1 odds in favor of H_2, in terms of probabilities there is a 4/(4+1) x 100% = 80% probability that H_2 is correct. We chose the authors recommendation for B = 10, as an appropriate level of skepticism at 91% confidence.

Once it is determined that a change point is probable, then the location must be estimated. This is the maximum of the posterior probability distribution of t_s:

$$P(t_s|D,H_2) \propto \frac{C_1!\,C_2!}{N_1^{C_1+1}N_2^{C_2+1}}\left[\left(\frac{C_1}{N_1}\right)^2+\left(\frac{C_2}{N_2}\right)^2\right]^{-1}$$

Once the change point location is determined, a recursive algorithm is employed to find the remaining change points. Once all the change points in the trace have been located, a clean-up algorithm is employed to check the change points by calculating the Bayes parameters for the change point (i) in the trace segments between change point(i-1) and (i+1), and spurious change points are eliminated. One can go further and group the intensity states; however, we were not interested in this particular aspect at this time.

In our algorithm we explicitly calculate B and $P(t_s|D, H_2)$ by simplifying the factorials in the above equation with the following form of Stirling's approximation:

$$\ln(n!) \cong \left(n+\frac{1}{2}\right)\ln(n) - n + \frac{\ln(2\pi)}{2}$$

This allows for the calculation of very large B values and makes possible the calculation of $P(t_s|D, H_2)$ which can also be very large. The following expressions were used in our program:

$$B = \frac{2}{\pi}\sum_{t_s} \exp\left[C_1 \ln\left(\frac{C_1}{N_1}\right) + C_2 \ln\left(\frac{C_2}{N_2}\right) - C \ln\left(\frac{C}{N}\right) + \ln\left(\frac{\sqrt{C_1 C_2 (C-1)}}{N_1 N_2 (N-1)}\right)\right]\left[\left(\frac{C_1}{N_1}\right)^2 + \left(\frac{C_2}{N_2}\right)^2\right]^{-1}$$

$$P(t_s|D, H_2) = \frac{2}{\pi}\exp\left[C_1 \ln\left(\frac{C_1}{N_1}\right) + C_2 \ln\left(\frac{C_2}{N_2}\right) - C \ln\left(\frac{C}{N}\right) + \ln\left(\frac{\sqrt{C_1 C_2 (C-1)}}{N_1 N_2 (N-1)}\right)\right]\left[\left(\frac{C_1}{N_1}\right)^2 + \left(\frac{C_2}{N_2}\right)^2\right]^{-1}$$

Where the constant pre-factor from the Bayes parameter is used in the calculation of $P(t_s|D, H_2)$ to keep the calculation constrained.