A Bayesian Approach for Extracting Fluorescence Lifetimes from Sparse Data Sets and Its Significance for Imaging Experiments

Kalyan Santra
_Iowa State University and Ames Laboratory_, kalyans@iastate.edu

Emily A. Smith
_Iowa State University and Ames Laboratory_, esmith1@iastate.edu

Xueyu Song
_Iowa State University and Ames Laboratory_, xsong@iastate.edu

Jacob W. Petrich
_Iowa State University and Ames Laboratory_, jwp@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/ameslab_manuscripts

Part of the Chemistry Commons, and the Mathematics Commons

Recommended Citation
Santra, Kalyan; Smith, Emily A.; Song, Xueyu; and Petrich, Jacob W., "A Bayesian Approach for Extracting Fluorescence Lifetimes from Sparse Data Sets and Its Significance for Imaging Experiments" (2019). Ames Laboratory Accepted Manuscripts. 305. https://lib.dr.iastate.edu/ameslab_manuscripts/305

This Article is brought to you for free and open access by the Ames Laboratory at Iowa State University Digital Repository. It has been accepted for inclusion in Ames Laboratory Accepted Manuscripts by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
A Bayesian Approach for Extracting Fluorescence Lifetimes from Sparse Data Sets and Its Significance for Imaging Experiments

Abstract
The measurement of fluorescence lifetimes, especially in small sample volumes, presents the dual challenge of probing a small number of fluorophores and fitting the concomitant sparse data set to the appropriate excited-state decay function. A common method of analysis, such as the maximum likelihood (ML) technique, assumes a uniform probability distribution of the parameters describing the fluorescence decay function. An improvement is thus suggested by implementing a suitable nonuniform distribution, as is provided by a Bayesian framework, where the distribution of parameters is obtained from both their prior knowledge and the evidence-based likelihood of an event for a given set of parameters. We have also considered the Dirichlet prior distribution, whose mathematical form enables analytical solutions of the fitting parameters to be rapidly obtained. If Gaussian and exponential prior distributions are judiciously chosen, they reproduce the experimental target lifetime to within 20% with as few as 20 total photon counts for the data set, as does the Dirichlet prior distribution. But because of the analytical solutions afforded by the Dirichlet prior distribution, it is proposed to employ a Dirichlet prior to search parameter space rapidly to provide, if necessary, appropriate parameters for subsequent employment of a Gaussian or exponential prior distribution.

Disciplines
Chemistry | Mathematics
A Bayesian Approach for Extracting Fluorescence Lifetimes from Sparse Data Sets and Its Significance for Imaging Experiments

Kalyan Santra\(^1\)\(^2\), Emily A. Smith\(^1\)\(^2\), Xueyu Song\(^1\)\(^2\) and Jacob W. Petrich\(^*1\)\(^2\)

\(^1\)Department of Chemistry, Iowa State University, Ames, IA
\(^2\)Ames Laboratory, U.S. Department of Energy, Ames, IA

Received 26 June 2018, accepted 6 November 2018, DOI: 10.1111/php.13057

ABSTRACT

The measurement of fluorescence lifetimes, especially in small sample volumes, presents the dual challenge of probing a small number of fluorophores and fitting the concomitant sparse data set to the appropriate excited-state decay function. A common method of analysis, such as the maximum likelihood (ML) technique, assumes a uniform probability distribution of the parameters describing the fluorescence decay function. An improvement is thus suggested by implementing a suitable nonuniform distribution, as is provided by a Bayesian framework, where the distribution of parameters is obtained from both their prior knowledge and the evidence-based likelihood of an event for a given set of parameters. We have also considered the Dirichlet prior distribution, whose mathematical form enables analytical solutions of the fitting parameters to be rapidly obtained. If Gaussian and exponential prior distributions are judiciously chosen, they reproduce the experimental target lifetime to within 20% with as few as 20 total photon counts for the data set, as does the Dirichlet prior distribution. But because of the analytical solutions afforded by the Dirichlet prior distribution, it is proposed to employ a Dirichlet prior to search parameter space rapidly to provide, if necessary, appropriate parameters for subsequent employment of a Gaussian or exponential prior distribution.

INTRODUCTION

Time-correlated, single-photon counting has become an integral part of techniques such as fluorescence-lifetime imaging microscopy (1–5), Förster resonance energy transfer (6–8) and fluorescence correlation spectroscopy (9–11). The technique records the time difference between the arrival times of an excitation pulse and a pulse resulting from a photon detected from fluorescence emission. A histogram of arrival-time differences is accumulated and fit to a model function for the fluorescence decay. The most frequently used fitting method (residual minimization, RM) minimizes the weighted squares of the residuals of the experimental data and the continuously optimized fitting function. RM requires a histogram of very high quality to extract the mean lifetime with high accuracy, and such a histogram is only obtained with a large number of total photon counts (~20 000 for rose bengal).

In super-resolution microscopies, however, such as stimulated emission depletion microscopy (12–14), high spatial resolution is only obtained at the expense of the fluorescence signal, as the latter decreases with decreasing detection volume. Additional factors such as a low intrinsic fluorescence quantum yield or photodegradation of the sample contribute to reducing the magnitude of the total photon counts, thus making it more difficult to generate a histogram of high quality. Unless there is a certain number of total counts, RM yields a poor estimate of the mean lifetime (15,16). In these cases (15–17), probability-based methods, such as maximum likelihood (ML), provide considerable improvement over RM. One of the limitations of ML, however, is that it assumes a uniform probability distribution of the parameters describing the fluorescence decay function. Thus, ML can be further improved by implementing a suitable nonuniform distribution. In this context, we note that alternative approaches such as phasor-based methods (18–21) have been adopted in fluorescence-lifetime imaging microscopy, especially, for biological samples. These methods performed well compared to RM methods in terms of computational efficiency and extracting contributions of different components.

Here, we consider photon-counting data analysis using a Bayesian framework, where the distribution of parameters is obtained from both their prior knowledge and the evidence-based likelihood of an event for a given set of parameters. If \( \beta \) and \( E \) represent the parameter space and the evidence (i.e. experimental observations), respectively, then the posterior distribution of the parameters for \( E, P(\beta|E) \), is given by the Bayes’ theorem (22–24):

\[
P(\beta|E) = \frac{P(E|\beta)P(\beta)}{P(E)}
\]

\( P(E|\beta) \) is the likelihood of evidence given the set of parameters \( \beta \); and \( P(\beta) \) is the prior distribution of the parameters, which is obtained from the prior knowledge of the parameters. As the evidence is collected, the prior knowledge can be updated for the prediction of the parameters. \( P(E) \) is the total likelihood (also known as the marginal likelihood) of the evidence at all possible points in the parameter space and acts as a normalization.

The Bayesian method is employed (23,25–38) to estimate parameters where there are insufficient evidences. It has been used in fluorescence-lifetime imaging (30,32,37), Förster resonance energy transfer (30) and fluorescence correlation spectroscopy (35,36) experiments. The choices, however, of the parameters to which priors are assigned and the functional form of the priors themselves varied widely. In some cases, an
exponential prior has been assigned to “the relaxation time of the photon-generating emission process” based on the argument that it has the maximum entropy within the allowed parameter range (37). In other cases, a uniform prior was assigned for the fraction of mean lifetime components; but this defeats the point of implementing the Bayesian approach because it becomes reduced to the ML method (30). Thus, one of the major challenges in implementing a Bayesian analysis of photon-counting data is determining the choice of the prior distribution and updating it as more evidence is successively acquired. In this work, we compare Gaussian and exponential prior distributions, where the lifetime parameter is directly incorporated in the posterior that is to be optimized, as well as Dirichlet prior distribution, where the lifetime parameter is indirectly calculated using the estimated probability of the bins. For the Gaussian and exponential prior, two analysis schemes were employed. In one, an identical prior was used for every data trace collected for a fixed number of counts. In the other, the prior is calculated and updated using the statistics of the results obtained from a data trace having a similar number of counts. As the latter method is preferable, the discussion of the former is given in the Supporting Information.

These prior distributions and the utility of the Bayesian approach were tested by analyzing photon-counting data obtained from the very well-characterized fluorophore, rose bengal. Rose bengal in methanol has an excited-state lifetime of 0.49 ± 0.01 ns at room temperature (15). Three sets of data were collected, each consisting of 50 individual traces, with a total number of counts of approximately 20, 200 and 20 000, respectively. In all the analyses, we incorporated the real instrument response function (IRF), a very narrow ∼20-ps time channel (to avoid the limitations incurred from binning time channels (15)) and a shift parameter.

MATERIALS AND METHODS

Rose bengal (Sigma-Aldrich, St. Louis) was purified by thin-layer chromatography (15). 550 nm was the excitation wavelength. Time-resolved data were collected using a homemade instrument (15). The full width at half-chromatography (15) was 200 ns. Time-gating also decreases the number of total photons. In all the analyses, we incorporated the real instrument response function (IRF), a very narrow ∼20-ps time channel (to avoid the limitations incurred from binning time channels (15)) and a shift parameter.

The likelihood of observing a sequence of counts \( \{c_1, c_2, \ldots, c_K\} \) with probability \( \{p_1, p_2, \ldots, p_K\} \) for a given set of parameters \( \{\tau, b\} \) and subject to the condition, \( \tau = \sum c_j \), is given by the multinomial form (15-17):

\[
P(c_1, c_2, \ldots, c_K | \tau, b) \propto \frac{\prod_i p_i^{c_i}}{c_1! c_2! \cdots c_K!} = \frac{\prod_j (p_j)^{c_j}}{c_1! c_2! \cdots c_K!} \tag{4}
\]

Using Eq. (3) for the probability, \( p_j = \frac{c_j}{\tau C_T} \), we obtain:

\[
P(c_1, c_2, \ldots, c_K | \tau, b) = \frac{\prod_{j=1}^{K} (\tau/C_T)^{c_j}}{c_1! c_2! \cdots c_K!} \tag{5}
\]

Note that both the probability of a photon being detected in the \( j \)th channel, \( p_j \), and the predicted counts, \( \hat{c}_j \) in that channel are functions of the parameters \( \tau \) and \( b \). The experimental data, the "evidence" of photon-counting events for a given parameter space \( \beta \equiv (\tau, b) \), are the observed counts. Therefore, we have \( E \equiv \mathcal{C} = (c_1, c_2, \ldots, c_K) \) and Eq. (5) can be rewritten as:

\[
P(E | \beta) = C_T^{K} \frac{\prod_{j=1}^{K} (\hat{c}_j/C_T)^{c_j}}{c_1! c_2! \cdots c_K!} \tag{6}
\]

Gaussian and exponential prior distributions. The critical part of the Bayesian analysis is identifying and selecting a suitable prior distribution for the parameters. Since our analysis includes two independent parameters, \( \tau \) and \( b \), the prior distribution is:

\[
P(\beta) = P(\tau)P(b). \tag{7}
\]

We have shown that the estimated mean lifetime of a fluorophore approximately follows a normal distribution (15,16). This conclusion is also obtained from the central limit theorem (41,42), which states that with a sufficiently large number of samples or of observations the distribution will converge to a normal distribution. Therefore, a Gaussian function with a preselected mean and variance is arguably a good choice for a prior distribution:

\[
P(\tau) = \left(2\pi \sigma_{\tau}^2\right)^{-1/2} e^{-\frac{(\tau-\mu_{\tau})^2}{2\sigma_{\tau}^2}}. \tag{8}
\]

The "hyperparameters" \( \mu_{\tau} \) and \( \sigma_{\tau} \) are the mean and the standard deviation of the prior distribution for \( \tau \). We also have tested the exponential prior distribution for the mean lifetime with known hyperparameter, \( \log_{10} \tau \).
\[ P(\tau) = \lambda_0 e^{-\lambda_0 \tau}, \quad (9) \]

where \( \langle \tau \rangle = 1/\lambda_0 \) is the mean of the prior distribution. For the shift parameters, \( b \), since we limit ourselves to a small range, \(-0.1\) to \(0.1\), it is convenient to assume that their distribution is uniform. Therefore, we take \( P(b) = 1/(b_{\text{max}} - b_{\text{min}}) \), which is a constant and does not affect the overall prior distribution, \( P(b) \). From Eq. (1), we write:

\[ P(\tau|\beta) = \frac{P(\tau)P(b)P(\beta|\tau)b}{P(\beta)} \]

The marginal likelihood, \( P(\beta) \), is given its name from the process of “marginalization,” which is an integration over all the parameters (27):

\[ P(\beta) = \frac{P(\tau|\beta)P(b)P(\beta|\tau)b}{\int d\tau db P(\tau|\beta)P(b)P(\beta|\tau)b} \]

Since \( P(b) \) is a constant, it can be eliminated from Eq. (11). Substituting \( P(\tau) \) from Eq. (8) and the expression for \( P(\beta|\tau)b \) from Eq. (6), the logarithm of the posterior for the Gaussian prior distribution can be written as:

\[ \ln P(\beta|\tau) = \gamma_1 - \frac{\tau - \mu_0}{2\sigma_0^2} + \frac{1}{2} \sum_{j=1}^{K} \ln \tilde{c}_j \]

(12)

where all the terms that are independent of the parameters \( \tau \) and \( b \) are condensed into the constant \( \gamma_2 \). Similarly, if we choose the exponential prior distribution for the mean lifetime given in Eq. (9), we have:

\[ \ln P(\beta|\tau) = \gamma_2 - \lambda_0 \tau + \sum_{j=1}^{K} \ln \tilde{c}_j, \]

(13)

where \( \gamma_2 \) is another constant, independent of the parameters \( \tau \) and \( b \). Maximization, therefore, of the logarithm of the posterior probability distribution in Eqs. (12) and (13) provides the optimum values of the parameters.

**Dirichlet prior distribution.** Since the joint probability distribution given in Eq. (4) is in multinomial form, the Dirichlet prior (43–49) distribution is a natural choice for estimating the probability of the channels because it forms a conjugate prior (48,49) with the multinomial distribution insofar as it combines with the likelihood function to form a posterior distribution that belongs to same Dirichlet family. Thus, analytical solutions for the parameters can be easily formulated. The process for extracting the lifetime from the estimated probabilities of the channels is the following. We rewrite the likelihood distribution function from Eq. (4) as \( P(c|x) \) in the following by considering the probabilities of the channels as unknown parameters given by \( \pi = (\pi_1, \pi_2, \ldots, \pi_K) \), where \( \sum_j \pi_j = 1 \).

\[ P(c|\pi) = C_T^{\sum_j} \prod_{j=1}^{K} \left( \frac{\pi_j}{c_j} \right)^{\pi_j}, \]

(14)

Let \( x = (x_1, x_2, \ldots, x_K) \) be the “precounts” (virtual counts (48) or pseudo counts (49)) of the channels with probabilities \( (\pi_1, \pi_2, \ldots, \pi_K) \) before the evidence is collected; and let the sum of all “precounts” be \( \Sigma_j x_j = A_T \). Then, the Dirichlet prior distribution is:

\[ P(\pi|x) \propto P(\pi|x)P(c|\pi) \propto N(x_1 + c_1, x_2 + c_2, \ldots, x_K + c_K) \]

\[ = \frac{\Gamma(A_T + C_T)}{\prod_{j=1}^{K} \Gamma(x_j + c_j)} \prod_{j=1}^{K} \pi_j^{x_j+c_j-1} \]

(16)

and the posterior mean is given by

\[ E(\pi|x) = \frac{\sum_j x_j + c_j}{A_T + C_T} = \frac{1}{A_T + C_T} \theta_j + \frac{C_T}{A_T + C_T} \phi_j \]

(17)

where \( \theta_j = E(\pi_j) = x_j/A_T \) and \( \phi_j = c_j/C_T \). The posterior mean is thus the weighted average of the prior mean \( \theta_j \) and the sample mean \( \phi_j \) with respect to the total “precounts” and the total experimental counts, respectively (43,45). The most important aspect of the Dirichlet prior is that, unlike the Gaussian and exponential priors, it does not combine the prior distribution of the lifetime parameter (\( \tau \)) directly in the estimation. Rather, the method of employing a Dirichlet prior evaluates the expected probability of the channels given the experimental counts.

In order to find the lifetime parameter, the bin-averaged time of the photon counts data was evaluated from the posterior and then compared with the same quantity calculated from the convoluted model. In order to do this, first, one needs to estimate the “precounts” of the channels. For a given set of initial parameters \((\tau_{\text{int}}, b_{\text{init}})\), we propose to distribute the total number of experimental counts \( C_T \) into the \( K \) bins to estimate \( x = (x_1, x_2, \ldots, x_K) \) as follows, using Eq. (3):

\[ x_j = C_T P_j(\tau_{\text{int}}, b_{\text{init}}) = C_T \sum_{j=1}^{K} \frac{s_{1j} - b_{ij}}{C_0} \frac{1}{C_1} \]

(18)

where \( \Sigma_j x_j = A_T = C_T \). The expectation values of the bin probabilities \( \pi_j \) are then calculated using Eq. (17) for all the channels. Let \( \tau_{av} \) represent the bin-averaged time calculated from the expectation value of the bin probability. Therefore,

\[ \tau_{av} = \sum_{j=1}^{K} \tau_{av}(\pi_j|x) = \sum_{j=1}^{K} \theta_j + \frac{C_T}{A_T + C_T} \phi_j \]

(19)

Similarly, for a given set of values of the parameters \( (\tau, b) \), we can define another bin-averaged time \( \tau_{av} \) for the convoluted model using the form of the probability \( p_T \) given in Eq. (3):

\[ \tau_{av} = \sum_{j=1}^{K} \tau_{av}(\pi_j|x) = \sum_{j=1}^{K} \theta_j + \frac{C_T}{A_T + C_T} \phi_j \]

(20)

Theoretically, the values of these two bin-averaged times \( \tau_{av} \) and \( \tau_{av} \) should be equal for the ideal data without any noise. Therefore, for experimental data we can minimize the absolute difference \( (\Delta_{\text{abs}}) \) between \( \tau_{av} \) and \( \tau_{av} \) as shown in Eq. (21) to obtain the optimum values of the parameters \((\tau_{\text{opt}}, b_{\text{opt}})\):

\[ \Delta_{\text{abs}} = |\tau_{av} - \tau_{av}| \]

(21)

The obtained optimal values are set as the new initial parameters, \((\tau_{\text{init}}, b_{\text{init}}) = (\tau_{\text{opt}}, b_{\text{opt}})\) and the entire procedure is repeated for several iterations until the results converge to a preset tolerance.

**Computational methods.**

The optimizations of the posterior distributions given in Eqs. (12) and (13) are performed using codes written in MATLAB. (These codes are provided in the SI.) The GlobalSearch toolbox in MATLAB uses the “fmincon” solver to minimize the objective function with respect to the parameters; and in each calculation, a global minimum is reached. The ranges of the parameters \( \tau \) and \( b \) are assigned as 0.01 to 1.5 ns and \(-0.1\) to \(0.1\) ns.
to 0.1 ns, respectively. Within the specified ranges, we run our in-house routine with different initial values of the parameters and always retrieve the same results through the third decimal place.

Gaussian and exponential priors. Both Eqs. (12) and (13) depend on the initial values of the hyperparameters. We employ two schemes to assign the values of the hyperparameters. In the first, we use identical prior hyperparameters (i.e., fixed \( \mu_0 \) and \( \sigma_0 \) for the Gaussian prior or fixed \( \lambda_0 \) for the exponential prior) for all the fifty decay traces in a set. In the second, we update the prior hyperparameters for the analysis of \( N^b \) decay trace using the calculated statistics of the results obtained from all the analyzed \( N-1 \) decay traces of that set according to the Eq. (22) given below. In the second scheme, we update the mean and the standard deviation after the analysis of 1 and 5 decay traces, respectively, to obtain sufficient statistics:

\[
(\mu_0)_N = \frac{1}{N-1} \sum_{i=1}^{N-1} \tau_i; \quad \text{for all } N > 1
\]

\[
(\sigma_0)_N = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N-1} (\tau_i - (\mu_0)_N)^2}; \quad \text{for all } N > 5
\]  

(22)

\[
(\lambda_0)_N = \frac{1}{(\mu_0)_N}; \quad \text{for all } N > 1
\]

In both schemes, different combinations of the initial values of the hyperparameters are assigned.

Dirichlet prior. As shown in Eq. (21), the absolute difference between \( \sum \mu_i \) and \( \sum \mu_i^c \) is minimized to obtain the new set of initial parameters \( \tau_{\text{new}, \text{best}} = (\sum \mu_{i, \text{best}}) \). The change of the value of the lifetime parameter is monitored; and convergence is obtained if the change between two successive iterations, \( \delta \tau \), is less than a preset tolerance value, which we set to \( \delta_{\text{tol}} = 10^{-7} \) ns. If \( b < 10^{-4} \) in an iteration, then \( b \) is set to zero. (We find \( b \approx 0 \) using the maximum likelihood estimation and other Bayesian analyses considered here for our data sets. Setting \( b = 0 \) simplifies the computation.) All the calculations converged in \( \leq 50 \) iterations. To test the influence of the initial conditions, the parameter space for the lifetime has been expanded (0.001 to 15 ns), and various initial values of \( \tau_{\text{init}} \) are chosen within that range. In all cases, the results converge to the same lifetime value.

RESULTS AND DISCUSSION

Gaussian and exponential priors

We assign the initial values of the hyperparameters for a decay trace and those values are mentioned in the corresponding figures and tables. After obtaining the results from a certain number of traces, we calculate the statistics of the results for all the decay traces considered up to that point using Eq. (22). The calculated statistics provide the hyperparameters for the subsequent analysis of the remaining decay traces. After each step, a new set of hyperparameters is obtained. Estimated lifetimes using this scheme are presented in Fig. 1 for all 50 decay traces for each set of data having a total number of 20, 200 and 20 000 counts, respectively. Each panel is labeled with the initial values of \( \mu_0 \) and \( \sigma_0 \). The histograms of the lifetimes obtained by using Gaussian and exponential priors with different sets of initial hyperparameters are given in Fig. 2. Statistics are summarized in Table 1.

For a Gaussian prior, where \( \sigma_0 = 0.5 \) ns, the results converge to the correct mean value as more and more decay traces are analyzed for a data set. As a result, the distribution of the estimated lifetimes becomes very narrow, with standard deviations of 8%, 2% and less than 1% of the mean lifetime for the data sets with total number of 20, 200 and 20 000 counts, respectively, as shown in Fig. 2a and Table 1. The identical-prior counterpart (see Supporting Information) has much wider distributions, as noted in the previous section. As in the identical-prior counterpart, however, the estimated lifetime is not very sensitive to the initial values of the prior mean, \( \mu_0 \), when the initial value of the prior standard deviation \( \sigma_0 \) is wide. For an exponential prior, the convergence is not as rapid as in the case of the Gaussian prior using this strategy (Fig. 2b). Again, the initial choice of \( \lambda_0 \) has no influence on the estimated lifetime for the three data sets. These results also suggest that the Gaussian prior is preferable to the exponential prior.

This example of updating the prior distribution using the results from data sets with the same number of total counts is purely illustrative. The point is that the fitting results can be improved by employing data collected using similar experimental conditions and choosing the prior hyperparameter accordingly. Once obtained, higher-quality data (e.g. from a decay trace having 20 000 total counts) can be used to extract the hyperparameters for the prior when analyzing lesser-quality data (e.g. from a decay trace having 20 total photon counts). Further updating of the prior might even be unnecessary, since it is possible that one data set of sufficiently high quality can provide a suitable prior. Such higher-quality data sets may be obtained from bulk solutions or from imaging data from STED experiments, for example, using pixels of higher intensity where the experimental conditions and fluorophore environment are similar.
Dirichlet prior

Using the strategy of minimization of the absolute difference between two bin-average time ($t_{av}$ and $t_0$) given in Eq. (21) for the Dirichlet prior, each of the decay traces was analyzed; and convergence was obtained when $\delta t<10^{-4}$ ns. The lifetimes of the individual traces are given in Figure S3 for all the decay traces. The histograms of the lifetimes obtained from an analysis employing the Dirichlet prior are given in Fig. 3 for all the data traces for a given initial condition. The statistics of the results for the fifty decay traces of each set are summarized in Table 2.

The value of the initial lifetime ($\tau_{int}$), which has been used to estimate the “precounts” ($\alpha$) for the Dirichlet prior at the beginning of the iteration, is 0.4 ns; and it is given in Fig. 3, Figure S3 and Table 2.

To test the influence of the initial value, the parameter space for the lifetime was expanded ($0.001 \, \text{ns} \leq \tau_{int} \leq 15 \, \text{ns}$), and various initial values of $\tau_{int}$ are chosen within that range. In all cases, the results converge to the same lifetime value. An example is shown in Figure S4, where the convergence is tested with various initial conditions ($\tau_{int}$) for a representative data trace randomly chosen from each of the data set with total number of counts 20, 200 and 20 000, respectively.

It can be seen (Tables 2 and S1) that the mean and the standard deviation of the lifetimes obtained for all the fifty decay traces in a set using a Dirichlet prior are comparable to those obtained using a Gaussian prior when $\sigma_0$ is 0.5 ns in the case of the three values (0.2 ns, 0.5 ns and 1.0 ns) of $\mu_0$. When $\sigma_0$ is 0.3 ns, the statistical results of the lifetimes are comparable to those we obtained from the Dirichlet prior for all cases except that where $\mu_0$ is 1.0 ns and the data set has 20 total counts. Here, the Gaussian prior yields $0.6 \pm 0.1$ ns and the Dirichlet prior yields $0.5 \pm 0.1$ ns. As mentioned above, the statistical results of the lifetimes for the Gaussian prior analysis depend on the value of $\mu_0$ when $\sigma_0$ is 0.1 ns for the data set with a total number of counts of 200 or less. On the other hand, the statistical results of the lifetimes obtained from the Dirichlet prior analysis are comparable to those we obtained using an exponential prior for all cases except that where the exponential prior parameter ($\lambda_0$) is 5.0 ns$^{-1}$ and the data set has 20 total counts. Here, the exponential prior yields $0.4 \pm 0.1$ ns.

Thus, the advantage of employing a Dirichlet prior is not so much for the result it yields but rather because its use does not require any \textit{a priori} knowledge of the lifetime of the sample. The change of the value of the lifetime parameter between two successive iterations, $\delta \tau$, should converge to yield the optimized results from any given starting point (initial value) for all the three data sets we have considered with total number of counts 20, 200 and 20 000, respectively. The Gaussian prior, on the other hand, can yield much smaller standard deviations, but its use requires prior knowledge of the parameters. In the case of the exponential prior, the initial condition (the value of the hyperparameter, $\lambda_0$) has little influence on the estimated lifetimes for the data set with a total count number of 20. The Dirichlet prior, being a natural conjugate prior for the multinomial distribution, combines with the joint probability of the data obtained in the photon-counting experiments to estimate the posterior of channel probability parameters analytically. It also differs significantly from the Gaussian and exponential prior cases in how the parameters are evaluated.

CONCLUSIONS

We have formulated and demonstrated the usefulness of a Bayesian approach for analyzing time-correlated, single-photon-counting data to estimate the mean fluorescence lifetime of a well-characterized fluorophore, rose bengal. Although the exponential
and the estimation of the value of the prior is less sensitive to the initial values of the hyperparameters, the type of prior and the initial values of the hyperparameters are given in the shaded rows.

### Table 1. Fitting results for three sets of 50 decay traces employing a Bayesian analysis using updated prior distributions.

<table>
<thead>
<tr>
<th>Total number of counts</th>
<th>Mean lifetime ± one standard deviation (ns)*</th>
<th>Mean lifetime ± one standard deviation (ns)*</th>
<th>Mean lifetime ± one standard deviation (ns)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian prior</td>
<td>( \mu_0 = 0.2 \text{ ns, } \sigma_0 = 0.5 \text{ ns} )</td>
<td>( \mu_0 = 0.5 \text{ ns, } \sigma_0 = 0.5 \text{ ns} )</td>
<td>( \mu_0 = 1.0 \text{ ns, } \sigma_0 = 0.5 \text{ ns} )</td>
</tr>
<tr>
<td>20</td>
<td>( 0.48 \pm 0.04 )</td>
<td>( 0.48 \pm 0.04 )</td>
<td>( 0.49 \pm 0.04 )</td>
</tr>
<tr>
<td>200</td>
<td>( 0.49 \pm 0.01 )</td>
<td>( 0.49 \pm 0.01 )</td>
<td>( 0.49 \pm 0.01 )</td>
</tr>
<tr>
<td>20 000</td>
<td>( 0.489 \pm 0.001 )</td>
<td>( 0.489 \pm 0.001 )</td>
<td>( 0.489 \pm 0.001 )</td>
</tr>
<tr>
<td>Exponential prior</td>
<td>( \lambda_0 = 0.5 \text{ ns}^{-1} )</td>
<td>( \lambda_0 = 2.0 \text{ ns}^{-1} )</td>
<td>( \lambda_0 = 10 \text{ ns}^{-1} )</td>
</tr>
<tr>
<td>20</td>
<td>( 0.5 \pm 0.1 )</td>
<td>( 0.5 \pm 0.1 )</td>
<td>( 0.5 \pm 0.1 )</td>
</tr>
<tr>
<td>200</td>
<td>( 0.5 \pm 0.03 )</td>
<td>( 0.49 \pm 0.03 )</td>
<td>( 0.49 \pm 0.03 )</td>
</tr>
<tr>
<td>20 000</td>
<td>( 0.490 \pm 0.004 )</td>
<td>( 0.490 \pm 0.004 )</td>
<td>( 0.490 \pm 0.004 )</td>
</tr>
</tbody>
</table>

*Mean lifetime ± one standard deviation (ns) of fifty decay traces calculated using a Bayesian analysis for three data sets with a total number of counts of 20, 200, and 20 000, respectively. The priors for a data trace in a set are updated using the statistics of the results of all the analyzed decay traces of that set, as given in Eq. (22). The type of prior and the initial values of the hyperparameters are given in the shaded rows.

Figure 3. Histograms of the estimated lifetimes of all fifty decay traces obtained by the Bayesian analysis using a Dirichlet prior distribution. The initial values of the lifetime, \( \tau_{\text{init}} \), and the estimation of the “pre-counts” are given at the top of the panel. The mean and the standard deviation of the estimated lifetimes are given in each histogram.

### Table 2. Fitting results for three sets of 50 decay traces employing a Bayesian analysis using Dirichlet prior distributions.

<table>
<thead>
<tr>
<th>Total number of counts</th>
<th>Mean lifetime ± one standard deviation (ns)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirichlet prior</td>
<td>( \tau_{\text{init}} = 0.4 \text{ ns, } \alpha_i = C_i P_i )</td>
</tr>
<tr>
<td>20</td>
<td>( 0.5 \pm 0.1 )</td>
</tr>
<tr>
<td>200</td>
<td>( 0.50 \pm 0.04 )</td>
</tr>
<tr>
<td>20 000</td>
<td>( 0.489 \pm 0.004 )</td>
</tr>
</tbody>
</table>

*Mean lifetime ± one standard deviation (ns) of fifty decay traces calculated using a Bayesian analysis for three data sets with a total number of counts of 20, 200, and 20 000, respectively. The absolute difference between \( \tau_{\text{init}} \) and \( \tau_{\text{fit}} \) shown in Eq. (21) is minimized to obtain optimum values of the lifetime, and the convergence is obtained if the change between two successive iteration, \( \delta \tau < 10^{-5} \text{ ns} \). The initial parameter, \( \tau_{\text{init}} \), and the estimation of “pre-counts” are given in the shaded row.

prior is less sensitive to the initial values of the hyperparameters, the Gaussian prior yields a much narrower distribution of the estimated lifetime and, thus, a more precise value of the retrieved value of the fluorescence lifetime. The greatest advantage, however, of the Dirichlet prior is that for the cases we investigated, the same optimized results are obtained regardless of the initial conditions for the prior parameters. Thus, an analysis strategy is suggested in which parameter space can rapidly be searched with the Dirichlet prior, and a subsequent, more refined search may be carried out with a Gaussian or exponential prior, if necessary.

Acknowledgements—Support was provided by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through the Ames Laboratory, under Contract No. DE-AC02-07CH11358. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University. We thank Dr. Hyung Jun Woo for encouraging the consideration of the Dirichlet prior and for his helpful comments.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:

**Figure S1.** Estimated lifetimes of all decay traces obtained by the Bayesian analysis where identical priors are used.

**Figure S2.** Histograms of the estimated lifetimes of all the decay traces obtained by the Bayesian analysis where identical priors are used.

**Figure S3.** Estimated lifetimes of all fifty decay traces obtained by the Bayesian analysis using a Dirichlet prior distribution.

**Figure S4.** A representative trace from each data set is analyzed using a Dirichlet prior with various initial values (\( \tau_{\text{init}} \)).

**Table S1.** Fitting results for three sets of 50 decay traces employing a Bayesian analysis using identical prior distributions.

REFERENCES

8. Van Der Meer, B. W., G. Coker and S.-Y. S. Chen (1994) Reso-
nance Energy Transfer: Theory and Data. VCH, New York.
Springer, Boston, MA.
10. Elson, E. L. and D. Magde (1974) Fluorescence correlation spec-
lation spectroscopy. II. An experimental realization. Biopolymers 13,
29–61.
12. Lesoine, M. D., U. Bhattacharjee, Y. Guo, J. Vela, J. W. Petrich and
E. A. Smith (2013) Subdiffraction, luminescence-depletion imaging
Supercontinuum stimulated emission depletion fluorescence lifetime
Smith (2014) The number of accumulated photons and the quality of
90, 767–772.
15. Santra, K., J. Zhan, X. Song, E. A. Smith, N. Vaswani and J. W.
Petrich (2016) What is the best method to fit time-resolved data? A
comparison of the residual minimization and the maximum likel-
hood techniques as applied to experimental time-correlated, single-
counting data analysis: Application of the maximum likelihood and
related methods for the determination of lifetimes in mixtures of rose
chi-square and likelihood functions in fits to histograms. Nucl. Instr.
18. Fereidouni, F., A. Esposito, G. Blab and H. Gerritsen (2011) A mod-
ified phasor approach for analyzing time-gated fluorescence lifetime
images. J. Microsc. 244, 248–258.
19. Schirmpf, W., A. Barth, J. Hendrix and D. C. Lamb (2018) Pam: A
framework for integrated analysis of imaging, single-molecule, and
The phasor approach to fluorescence lifetime imaging analysis. Biophys.
J. 94, L14–L16.
lifetime imaging system that optimizes photon efficiency. Microsc.
Res. Techniq. 71, 201–213.
Springer, New York, NY.
don, UK.
25. Bretthorst, G. L. and U. V. Toussaint. The maximum entropy
using Bayesian Probability Theory. Springer, Dordrecht, Nether-
lands.
Transient multiplexponential signals analysis using Bayesian deconvo-
29. Barber, P., S. Ameer-Beg, S. Pathmanathan, M. Rowley and A.
Coolsen (2010) A Bayesian method for single molecule, fluorescence
Developing and testing a Bayesian analysis of fluorescence lifetime
31. Van Dyk, D. A., A. Conners, V. L. Kashyap and A. Siemiginowska
(2001) Analysis of energy spectra with low photon counts via Baye-
Robust Bayesian fluorescence lifetime estimation, decay model selec-
tion and instrument response determination for low-intensity FLIM
in a mixed-order model of BOD decay. Water Res. 34, 1830–1836.
35. Guo, S.-M., J. He, N. Monnier, G. Sun, T. Wohland and M. Bathe
(2012) Bayesian approach to the analysis of fluorescence correlation
spectroscopy data II: Application to simulated and in vitro data. Anal.
Chem. 84, 3880–3888.
36. He, J., S.-M. Guo and M. Bathe (2012) Bayesian approach to the
Chem. 84, 3871–3879.
37. Rowley, M. L., P. R. Barber, A. C. Coolen and B. Vojnovic. Baye-
790325–790325. International Society for Optics and Photonics, Pro-
cedings of the SPIE BioS011.
38. Lubrano, M. Bayesian analysis of nonlinear time series models with a
improves the spatial resolution of STED microscopy. Opt. Express
19, 4242–4254.
40. Vicedomini, G., G. Moneron, K. Y. Han, V. Westphal, H. Ta, M.
Reuss, J. Engelhardt, C. Eggeling and S. W. Hell (2011) Sharper
low-power STED nanoscopy by time gating. Nat. Methods 8, 571.
41. Dobrushin, R. L. (1956) Central limit theorem for nonstationary mark-
York-Sydney-Toronto.
& Sons, Hoboken, NJ.
report, MIT.
45. Agresti, A. and D. B. Hitchcock (2005) Bayesian inference for cate-
categorical data: The problem of nonresponse. J. Stat. Plann. Infer-
ence 6, 345–362.
47. Huang, J. (2005) Maximum likelihood estimation of Dirichlet distri-
bution parameters. CMU Technique Report.
mial distribution. Online tutorial.