An MCMC approach for estimating a fluorescence lifetime with pile-up distortion

Tabea Rebafka
ENST, Département TSI, 46 rue Barrault, 75634 Paris Cedex 13, France
CEA, LIST, Laboratoire d'Electronique et de Traitement du Signal, 91191 Gif sur Yvette Cedex, France
rebafka@tsi.enst.fr

Résumé − Ce travail présente un nouvel estimateur de la distribution de la durée de vie en fluorescence. Un échantillonneur de Gibbs est développé pour estimer les paramètres quand le minimum d'un nombre aléatoire de variables distribuées selon un mélange exponentiel est observé. L'algorithme est testé avec des données simulées, et une comparaison avec des méthodes utilisées en pratique est faite. Nos résultats indiquent que la méthode proposée requiert moins d'observations que des méthodes classiques pour obtenir la même qualité statistique de l'estimation.

Abstract − This work presents a new estimator of the fluorescence lifetime distribution. A Gibbs sampler is developed for estimating the parameters when the minimum of a random number of variables following an exponential mixture distribution is observed. The algorithm is tested on simulated data and compared to methods used in practice. It turns out that the new method requires less measurements than standard methods to obtain the same statistical quality of estimation.

1 Model

In time-resolved fluorescence the technique of Time Correlated Single Photon Counting (TCSPC) measures the fluorescence lifetime, i.e. the time from a laser pulse that excites a random number of fluorescent molecules until the emission of photons (see [8], [4]). For technical reasons only the lifetime of the first arriving photon can be measured. The current practice to estimate the probability distribution of the lifetime is to diminish the laser intensity such that there is (almost) never more than one photon per light pulse. Then the recorded arrival times can be considered as i.i.d. observations of the lifetime distribution, but as a result for more than 90% of the laser pulses there is no photon at all and the experiment takes long. From a statistical point of view it is a waste to carry out measurements where most of the observations do not contain any information about the distribution of interest. In this work we propose an MCMC algorithm to estimate the lifetime distribution for more intensive laser pulses, so the probability of several photons per excitation pulse is no longer negligible. Then one says that 'pile-up' occurs, i.e. the distribution of the measurements is a nonlinear transformation of the lifetime distribution since only the minimum is observed. Fluorescence measurements are highly sensitive to unstable experimental conditions, so whenever a large number of lifetimes have to be estimated methods may be useful that require less observations and so allow faster experiments. Those mehtods can be obtained by dealing with pile-up affected data.

Actually, in 1968 Coates proposed a correction formulae for histograms of TCSPC data for the pile-up case [1]. Although this method works well, it has rarely been used in practice, maybe for the reason that it can only be an intermediate step on the way to parameter estimation of the lifetime distribution. That means that after correcting the data by Coates, another method like the least squares method or an EM algorithm has to be applied to obtain parameter estimates, whereas the MCMC method that is proposed here provides parameter estimates directly. Section 4 includes a comparison of the MCMC method to an estimation using the method of Coates.

The lifetime of a single molecule is known to be exponentially distributed, and so the distribution of a mixture of molecules represents an exponential mixture whose density is given by

$$f(y) = \sum_{k=1}^{K} \alpha_k \nu_k e^{-\nu_k y} \tag{1}$$

where $\nu_k > 0$, $\sum_{k=1}^{K} \alpha_k = 1$ and the number of components K is supposed to be known. We assume to observe the following variable

$$Z = \begin{cases} T & \text{if } N = 0 \\ \min\{Y_1, \dots, Y_N, T\} & \text{otherwise.} \end{cases} \tag{2}$$

where

- the sequence $(Y_k)_{k\geq 1}$ consists of i.i.d. variables from the exponential mixture distribution (1),
- the random variable N follows a Poisson distribution Poi(λ) and is independent from $(Y_k)_{k\geq 1}$,
- T denotes the observation time, i.e. photons arriving after T cannot be observed and are censored.

For the variable Z we have for all $z \in (0, T)$

$$\begin{aligned} \mathbb{P}(Z > z) &= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{P}(\min\{Y_1, \dots, Y_n, T\} > z) \\ &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} (\overline{F}(z))^n \\ &= e^{-\lambda F(z)} \end{aligned}$$

where $F(.)$ denotes the cdf of the exponential mixture distribution given by (1) and $\overline{F}(.) = 1 - F(.)$. Thus, Z admits

a density $p(.)$ w.r.t. $\mathcal{L}_{(0,T)} \otimes \delta_T$ where $\mathcal{L}_{(0,T)}$ denotes the Lebesgue measure on $(0,T)$ and δ_T the measure having mass one in T. The density is given by

$$p(z) = \begin{cases} \lambda f(z) e^{-\lambda F(z)} & \text{if } z \in (0,T) \\ e^{-\lambda F(T)} & \text{if } z = T. \end{cases}$$

2 Statistical Inference

Let $Z_1, \ldots, Z_M$ be M i.i.d. random variables having the same distribution as Z defined in (2). Then, for $T = \infty$, the distribution of Z can be interpreted as an exponential mixture with an infinte number of components where the set of exponential parameters is given by $\nu_1\mathbb{N} + \nu_2\mathbb{N} + \ldots + \nu_K\mathbb{N}$ and the weights depend on $\lambda, \nu_1, \ldots, \nu_K, \alpha_1, \ldots, \alpha_K$. Existing methods for estimating the mixing distribution of a finite or infinite exponential mixture are the EM algorithm as proposed in [3], the penalized dual method developed in [7] that searches the nonparametric maximum likelihood estimator (NPMLE) and an MCMC approach proposed in [2]. In our context the EM algorithm is hard to handle since the maximisation step does not provide explicit solutions for all parameters. The NPMLE method seems to be irrelevant unless one takes into account the particular structure of the parameter set and the weights. In contrast the MCMC approach which is more precisely a Gibbs sampler is easier to adapt to our context, although, due to the more involved structure of the model, the data augmentation scheme in [2] cannot be applied. Instead, we propose to use the missing data $(B_1, \ldots, B_K)$ where B_k is the number of photons generated from the k-th mixture component. Due to our choice of the missing data the posterior distributions turn out to be different from the ones in [2].

3 Algorithm

Following the approach of [2] we obtain identifiability of the model by ordering the exponential parameters

$$\nu_1 > \nu_2 > \ldots > \nu_K.$$

In order to use improper priors as in [6], the following reparametrization of the exponential parameters is used

$$\varphi = \nu_1, \quad \tau_k = \frac{\nu_k}{\nu_{k-1}} \quad \text{for } k = 2, \ldots, K.$$

In the Gibbs sampler new parameter values are generated one by one depending on the other parameter values. So with the initial parametrization of the weights α_k and the constraint $\sum_{k=1}^K \alpha_k = 1$, the set of possible values of $(\alpha_1, \ldots, \alpha_K)$ at one step of the algorithm is not the entire set $\{(\alpha_1, \ldots, \alpha_K) \in [0,1]^K : \sum_{k=1}^K \alpha_k = 1\}$. The following reparametrization is necessary to elude that problem

$$q_k = \frac{\alpha_k}{1 - \sum_{j=1}^{k-1} \alpha_j} \quad \text{for } k = 1, \ldots, K-1.$$

Now we assume the noninformative prior distribution that leads to proper posterior distributions as shown in [2]

$$\pi(\lambda, q_1, \ldots, q_{K-1}, \varphi, \tau_2, \ldots, \tau_K) = \frac{1}{\lambda\varphi} \prod_{k=1}^{K-1} \mathbb{1}\{q_k \in [0,1]\} \prod_{k=2}^{K} \mathbb{1}\{\tau_k \in [0,1]\}.$$

The Gibbs sampling algorithm consists of two steps repeated many times. The first step is a data augmentation and the second is the generation of new parameter values drawn from the posterior distributions. For each observation Z we propose the data augmentation of the missing data $(B_1, \ldots, B_K)$ where B_k is the number of photons generated from the k-th mixture component. Denoting by Y_z a random variable following the Poisson distribution $\mathrm{Poi}(\lambda\overline{F}(z))$, the missing data can be generated by first drawing the total number of photons N from the distribution

$$N|Z \overset{\mathcal{D}}{=} \begin{cases} Y_Z + 1, & \text{if } Z \in (0,T) \\ Y_T & \text{if } Z = T, \end{cases}$$

then by drawing a vector $(B_1, \ldots, B_K)$ from the distribution

$$\mathbb{P}(B_1 = b_1, \ldots, B_K = b_K | N = n, Z) = \begin{cases} \frac{(n-1)! \prod_k \alpha_k^{b_k} \sum_k b_k\nu_k}{f(Z)(\overline{F}(Z))^{n-1} \prod_k b_k!} e^{-Z\sum_k b_k\nu_k}, \\ \qquad\qquad \text{for } n \geq 1 \text{ if } Z \in (0,T) \\ \frac{n! \prod_k \alpha_k^{b_k}}{(\overline{F}(T))^n \prod_k b_k!} e^{-T\sum_k b_k\nu_k}, \quad \text{for } n \geq 0 \text{ if } Z = T, \end{cases}$$

where $b_k \in \mathbb{N}$ and $\sum_k b_k = n$.

We introduce the notation $\theta = (\lambda, \phi, \tau_2, \ldots, \tau_K, q_1, \ldots, q_{K-1})$ for the parameter vector and $\theta^{(-\lambda)}$ for the vector θ where the element λ is missing. The notations $\theta^{(-\phi)}, \theta^{(-\tau_l)}$ and $\theta^{(-q)}$ are defined the same way. Further, we put $\tau_1 = 1$. One can show that the posterior distributions for the different parameters given the data $\mathbf{Z} = (Z_1, \ldots, Z_M)$, the missing data $\mathbf{B} = (B_{ik}, i = 1, \ldots, M,\ k = 1, \ldots, K)$ and the other parameters turn out to be

$$\lambda|(\mathbf{Z},\ \mathbf{B}, \theta^{(-\lambda)}) \sim \Gamma\left(\sum_{i=1}^M \sum_{k=1}^K B_{ik}, M\right)$$

$$\phi|(\mathbf{Z},\ \mathbf{B}, \theta^{(-\phi)}) \sim \Gamma\left(\sum_{i=1}^M \mathbb{1}\{Z_i < T\}, \sum_{k=1}^K \tau_1 \ldots \tau_k \sum_{i=1}^M B_{ik} Z_i\}\right)$$

$$q_l|(\mathbf{Z},\ \mathbf{B}, \theta^{(-q_l)}) \sim \mathrm{Beta}\left(\sum_{i=1}^M B_{il} + 1, \sum_{i=1}^M \sum_{k=l+1}^K B_{ik} + 1\right)$$

The posterior distribution of τ_l is a mixture of Gamma distributions restricted on the interval $[0,1]$ with $\sum_{i=1}^M \mathbb{1}\{Z_i < T\}$ components. Since the computation of all parameters of this distribution is costly, it is more convenient to use a slice sampler or the accept-reject method for simulation from the distribution where the density function is given by

$$p(\tau_l|\mathbf{Z},\ \mathbf{B}, \theta^{(-\tau_l)}) \propto \exp\left\{-\sum_{k=1}^K \phi\tau_1 \ldots \tau_k \sum_{i=1}^M B_{ik} Z_i\right\} \times \prod_{i:Z_i<T} \left(\sum_{k=1}^K \phi\tau_1 \ldots \tau_k B_{ik}\right) \mathbb{1}\{\tau_l \in [0,1]\}.$$

Typically the parameter values from the first 5.000 steps are thrown away to ensure convergence of the algorithm. For a more sophisticated control of convergence an adaption of the Rao-Blackwellized control variate presented in [2] can be used. Parameter estimators are obtained by taking the mean values of the generated values. Moreover,

the algorithm is robust to different starting values. We recommend to use uniformly distributed values for the weights α_k and any distinct values for the exponential parameters ν_k (like $1, 2, 3, \ldots$). For the intensity λ the initial value

$$-\log\left(\frac{M}{\sum_{i=1}^{M}\mathbb{1}\{Z_i = T\}}\right),$$

may be used since $\mathbb{P}(N = 0) = e^{-\lambda}$.

4 Experimental Results

In this section simulations are carried out to test the performance of the new Gibbs sampler. Table 1 presents results obtained for data simulated from models with different numbers of components K. Note that all estimated values are close to the true parameter values and, of course, by increasing the number of observations one may obtain even higher precision. Especially in the monoexponential case ($K = 1$) very good estimates are provided for relatively intense laser pulses ($\lambda = 3$). For illustration the quantile-quantile plot for the two-components model is presented in Figure 1. It shows that the estimated distribution is close to the true one, which is also confirmed by a high p-value of the Kolmogorov-Smirnov test (0.984).

TAB. 1 – Simulation results for $K = 1, 2, 3$ components and 3000, 3000, 10.000 observations, resp.

$K = 1$	true values			estimated values		
λ	3			3.05		
ν	.1			.101		
$K = 2$						
λ	1			.994		
α_1, α_2	.75	.25		.742	.258	
ν_1, ν_2	6	.5		6.032	.518	
$K = 3$						
λ	1			1.01		
$\alpha_1, \alpha_2, \alpha_3$	.33	.33	.33	.347	.329	.323
ν_1, ν_2, ν_3	5	1	.1	4.88	.970	.101

From Table 2 we derive the typical reasons that worsen the estimation for any mixture of distributions (see [5]) : first, increasing the number of components which means having more unknown parameters ; second, component densities that are close to each other and hence difficult to distinguish. In both cases one can show a heavy decline of the Fisher information. For the problem where the minimum of a mixture of distributions is considered there is a third reason that is related to the intensity λ that may worsen the estimation. Increasing λ means increasing the mean number of emitted photons. So the observed minimum is taken over more and more arrival times and measurements are mostly close to zero and information gets lost. Hence, we conclude that our MCMC method provides reliable estimates when the model parameter values are such that the Fisher information is not too small.

Recall that our main goal is to obtain a method that handles with pile-up distorsion (i.e. $\lambda > 0.1$) in order to reduce the required number of measurements M. Let us now compare the MCMC method to the following estimation practice. Data from model (2) are obtained at a small laser intensity ($\lambda = 0.05$) so that the probability for 2 or more photons per laser pulse becomes negligible. Then the observed arrival times, i.e. measurements smaller than T, are considered as i.i.d. observations from the exponential mixture distribution described by (1) and a classical EM algorithm for this distribution is applied. Repeated simulations provide estimates of the bias and the variance of the estimators for a model with two components and various numbers of observations. For the same two-components model we simulated data with a more intense laser ($\lambda = 1$) and applied the MCMC method. From the results shown in Table 3 we see first that for both methods bias and variances decrease by increasing the number of observations and, second and more notable, that any bias and any variance from the standard method exceeds the corresponding value of the MCMC method. Actually, comparing the results for 10.000 observations of

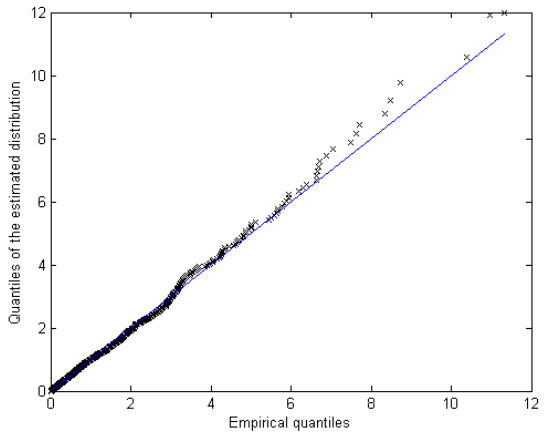

FIG. 1 – Quantile-quantile plot for the two-component model with the true parameter values $\lambda = 1$, $(\nu_1, \nu_2) = (6, .5)$, $(\alpha_1, \alpha_2) = (.75, .25)$.

TAB. 2 – Simulation results for $K = 5, 2, 1$ components and 10.000, 3000, 3000 observations, resp.

$K = 5$	true values			estimated values		
λ	1			1.23		
$\alpha_1, \alpha_2, \alpha_3$	.2	.2	.2	.476	.089	.105
α_4, α_5	.2	.2		.160	.171	
ν_1, ν_2, ν_3	10	7	5	7.25	1.29	1.14
ν_4, ν_5	1	.1		.104	.001	
$K = 2$						
λ	1			.972		
α_1, α_2	.066	0.33		.811	.189	
ν_1, ν_2	6	6.1		5.243	6.34	
$K = 1$						
λ	8			8.46		
ν	.1			.092		

the first method to the results for 3.000 observations of the new method shows that all values except one from the MCMC method are smaller than the first. Hence, we conclude that for this choice of parameters for the MCMC method less than a third of the observations of the standard method are necessary to obtain estimates having a comparable statistical quality.

TAB. 3 – Comparison of a standard estimation method and the MCMC method for a two-components model with true parameter values $\alpha_1 = .33, \alpha_2 = .66, \nu_1 = 0.5, \nu_2 = 6$.

Standard estimation method ($\lambda = 0.05$)					
nb of obs		1000	3000	5000	10000
α_1	bias	.0357	.0135	0.0116	.0098
	var	.0231	.0064	.0036	.0019
ν_1	bias	.0147	.0061	.0010	.0030
	var	.0173	.0042	.0026	.0014
ν_2	bias	1.939	.4197	.2015	.1331
	var	88.67	4.633	1.8343	.8316
MCMC method ($\lambda = 1$)					
nb of obs		1000	3000	5000	10000
α_1	bias	.0138	.0027	.4333e-3	.0026
	var	.0012	.4599e-3	.3550e-3	.1769e-3
ν_1	bias	.0089	.0044	.0006	.0029
	var	.0142	.0003	.0001	.0001
ν_2	bias	.4984	.0518	.0023	.0465
	var	.3619	.1574	.0753	.0555

Finally, the MCMC method is compared to the approach of Coates another method dealing with the pile-up case. In [1] Coates proposes a correction formulae for histograms that works for any distribution F of photon arrival times where the number of independent photons per excitation N follows a Poisson distribution and only the first arriving photon per laser pulse is observed. Denote m_i the counts in the i-th bin of the histogram based on M laser excitations. Then the 'corrected' histogram that estimates the density of the distibution of the arrival times F is given by the new counts

$$-\log\left(1 - \frac{m_i}{M - \sum_{j=1}^{i-1} m_j}\right).$$

As Coates' method is a correction of histograms we conducted the following simulation study : a histogram from the distribution of Z defined in (2) is simulated with a bin width in the range of the resolution of the TCSPC technique, then the Coates correction is applied and finally an EM algorithm for the exponential mixture distribution provides parameter estimates. On the same data set the MCMC method is applied. By repeated simulations we obtained estimates of the bias and the standard deviation of the estimators. The results for various model parameters are displayed in Table 4 and we note that the performance of the MCMC method is superior to the method based on Coates' correction when there are one or two components, i.e. bias and variances of all estimators are smaller. In the case of three components both methods may be considered as comparable.

TAB. 4 – Comparison of the bias and standard deviation for the Coates correction and the MCMC method, for models with $K = 1, 2, 3$ components.

True values	Coates		MCMC	
$\nu_1 = 3$	.414	(.138)	.344	(.114)
$\lambda = 3$	-	-	.386	(.060)
True values				
$\nu_1 = 6$	.655	(.219)	.616	(.203)
$\nu_2 = 0.5$	.016	(.031)	.003	(.025)
$\alpha_1 = 0.66$	.013	(.022)	.003	(.020)
$\lambda = 1$	-	-	.953	(.047)
True values				
$\nu_1 = 3$	.128	(.269)	.204	(.157)
$\nu_2 = 1$	.151	(.345)	.059	(.367)
$\nu_3 = 0.1$	.106	(.010)	.003	(.015)
$\alpha_1 = 0.50$	.004	(.107)	.035	(.081)
$\alpha_2 = 0.25$	.034	(.099)	.036	(.070)
$\alpha_3 = 0.25$	.038	(.024)	.049	(.031)
$\lambda = 3$	-	-	.404	(.043)

5 Conclusion

We conclude within the limits of these simulations that the proposed Gibbs sampling algorithm requires less observations than the standard method based on less intense lasers. Moreover, the method provides estimators of at least the same statistical quality as a method using the Coates correction. Thus, faster TCSPC measurements to estimate the fluorescence lifetime distribution are possible.

Références

[1] P. B. Coates. The correction of photon 'pile-up' in the measurement of radiative lifetimes. *Journal of Physics E : Scientific Instruments*, 1 :878–879, 1968.

[2] Marie-Anne Gruet, Anne Philippe, and Christian P. Robert. MCMC control spreadsheets for exponential mixture estimation. *Journal of Computational and Graphical Statistics*, 8(2) :298–317, 1999.

[3] Nicholas P. Jewell. Mixtures of exponential distributions. *Annals of Statistics*, 10(2) :479–484, 1982.

[4] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Academic/Plenum, New York, 1999.

[5] Bruce G. Lindsay. *Mixture Models : Theory, Geometry and Applications*. NFS-CBMS Regional Conference Series in Probability and Statistics. Hayward CA, Institute of Mathematical Statistics, 1995.

[6] Jean-Michel Marin, Kerrie Mengersen, and Christian P. Robert. Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics 25*, D. Dey and C.R. Rao (eds). Elsevier-Sciences, 2005.

[7] Ramani Pilla and Bruce P. Lindsay. Alternative EM methods for nonparametric finite mixture models. *Biometrika*, 8(2) :535–550, 2001.

[8] Bernard Valeur. *Molecular Fluorescence*. WILEY-VCH, Weinheim, 2002.