How many photons are necessary for fluorescence-lifetime measurements?

Malte Köllner and Jürgen Wolfrum
Physikalisch-Chemisches Institut, Im Neuenheimer Feld 253, W-6900 Heidelberg, Germany

Received 15 June 1992; in final form 12 August 1992

On planning time-correlated single photon counting experiments for fluorescence-lifetime measurements with minute samples, the question of ultimate sensitivity arises. To answer this question, we will find out how many detected photons one needs to achieve a certain accuracy in the lifetime estimate. Using least-squares or maximum-likelihood procedures and the multinomial distribution, the accuracy of the lifetime estimation can be expressed explicitly as a function of the signal strength. This allows one to determine the required number of detected photons.

1. Introduction

Time-correlated single photon counting [1] is an important technique for measuring lifetimes of fluorescing species with applications in kinetics, energy transfer studies, local distance measurements, rotational diffusion measurements, microviscosity measurements and other fields.

In the last few years, interest in lifetime measurements on minute samples has been growing [2–4]. Lifetime measurements are now performed on single cells [5,6] or even single molecules [7], and the question arises as to how far the limit of sensitivity can ultimately be pushed. In this work we give for the first time the lower bounds for the signal strength that is needed in these experiments. We calculate with analytical methods how many detected photons one needs to achieve a certain accuracy in the lifetime estimates. Thereby, the validity of the model function is not questioned and no systematic errors are taken into account.

Our general approach will be

(a) to find an optimum estimation procedure,

(b) to determine the error of the lifetime estimate as a function of the signal strength,

(c) to calculate the signal strength for the desired accuracy.

Correspondence to: M. Köllner, Physikalisch-Chemisches Institut, Im Neuenheimer Feld 253, W-6900 Heidelberg, Germany.

2. Efficient estimators

In time-correlated single-photon-counting experiments, we are confronted with the problem of estimating a parameter – the lifetime – from a given data set. This is essentially a statistical problem. Many methods of analysis have been proposed that have specific advantages like simplicity, exactness or short calculation times on a computer. In this work we have to focus our attention on estimation procedures that give the highest accuracy, in other words, those procedures that need little data to obtain fairly exact results. Only this approach will allow us to predict the very limit of sensitivity.

In the statistical literature [8,9] the Rao–Cramér theorem can be found, which states that in general, the standard deviation or variance of an estimate cannot be smaller than a well-defined limit. Estimators reaching this limit are called efficient. More precisely, the theorem states that the distribution of the parameter estimates for efficient estimators is asymptotically, i.e. for sufficiently large signals, normal with covariance matrix equal to the inverse Fisher-information matrix **F**, defined in our case by

0009-2614/92/$ 05.00

$$F_{hj} = \sum_i \frac{1}{y_i} \frac{\partial y_i}{\partial \alpha_h} \frac{\partial y_i}{\partial \alpha_j}, \quad (1)$$

where y_i is the average number of photons in channel i, as predicted by the model function y. The function y depends on the parameters α_h and α_j that are to be estimated, and h and j run from l to M, where M is the number of parameters to be estimated. The sum is performed over all channels. For small signals the variance of the estimate differs from the lower limit, but if the signal is large enough to estimate the parameters with a relative error of 10% – which can be calculated with eq. (6) of this paper – the variance coincides almost completely with the lower limit.

The weighted least-squares method and the maximum-likelihood method [9–13] for example are efficient [14–16], while linear regression is not efficient due to improper weighting.

The weighted least-squares method consists of finding the set of parameters that minimizes the expression

$$\mathrm{LS}(\alpha_1, \alpha_2, \ldots) = \sum_i \frac{(n_i - y_i)^2}{\sigma_i^2}, \quad (2)$$

where n_i is the number of observed photons or counts in channel i and σ_i is the standard deviation of n_i.

The maximum likelihood method, on the other hand, proceeds as follows: If one observes $n = \{n_i, i = 1, 2, \ldots, k\}$ counts in channels 1 through k, and the true lifetime is τ_{true}, one can assign a probability $P(n, \tau_{\mathrm{true}})$ to this outcome of the experiment. The lifetime is, of course, not known. The maximum likelihood estimator τ_{true} is the value that maximizes the expression $P(n, \tau)$. Therefore, we need an explicit expression for $P(n, \tau)$.

3. Multinomial distribution

Since single photon counting involves distributing N photons, $N = \sum n_i$, over k channels, this probability is the multinomial probability

$$P(n, \alpha_1, \alpha_2, \ldots) = \frac{N!}{n_1! \ldots n_k!} p_1^{n_1} \times \ldots \times p_k^{n_k}, \quad (3)$$

where p_i is the probability that a photon will fall into channel i and $\sum p_i = 1$. The p_i are functions of the parameters α. The multinomial distribution is a generalization of the binomial distribution from two channels to k channels.

Since the expected number of counts y_i in channel i is now Np_i, we have as Fisher-information matrix for the multinomial approach

$$(\mathbf{F}^{\mathrm{m}})_{hj} = \sum_i \frac{1}{Np_i} \frac{\partial Np_i}{\partial \alpha_h} \frac{\partial Np_i}{\partial \alpha_j}$$

$$= N \sum_i \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha_h} \frac{\partial p_i}{\partial \alpha_j}, \quad (4)$$

where the superscript m refers to the multinomial distribution. $\mathbf{F}^{\mathrm{m}}$ is linear in the signal strength N. Therefore, the variance of an estimate of τ is

$$\mathrm{var}_N(\tau) = (\mathbf{F}^{\mathrm{m}})^{-1}_{\tau,\tau} = \frac{1}{N} [\mathbf{F}^{\mathrm{m}}(N=1)]^{-1}_{\tau\tau}. \quad (5)$$

If we then denote by $\mathrm{var}_1(\tau)$ the variance of τ when the signal is only one photon, the required number of photons for a desired variance is

$$N \geq \frac{\mathrm{var}_1(\tau)}{\text{desired variance}(\tau)}, \quad (6)$$

which is what we have been looking for.

4. Poisson distribution

Usually the counts are not described by the multinomial distribution (3) but by a Poisson distribution with the probability for n_i counts in channel i being

$$P(n_i) = \exp(-y_i) \frac{y_i^{n_i}}{n_i!}. \quad (7)$$

The σ_i of eq. (2) are then

$$\sigma_i = \sqrt{y_i}. \quad (8)$$

A clarification of the difference between the multinomial and a Poisson description of the counts n_i is given in ref. [10]. We will only mention here that the n_i are Poisson distributed only if N is Poisson distributed, which is not always the case [17]. On the other hand, under the condition that N is known – which it is once the measurement is finished – the n_i follow rigorously a multinomial distribution. Thus, the Poisson approach does not make use of all the

information, it introduces an insecurity in N that is actually not present.

Nevertheless, both approaches give the same parameter estimates when using maximum likelihood fitting [10]. Furthermore, the estimates have the same variances for parameters that do not depend on the signal strength N, especially the lifetimes. This can be seen as follows: If we perform a substitution of the parameters so that y_i is written as Np_i and the p_i are functions of the same parameter set as in the multinomial approach but with the intensity N as additional parameter, the diagonal element with index NN of $\mathbf{F}$ becomes

$$F_{NN}=\sum_i \frac{1}{Np_i}\frac{\partial Np_i}{\partial N}\frac{\partial Np_i}{\partial N}=\frac{1}{N}\sum_i p_i=\frac{1}{N}, \tag{9}$$

and the off-diagonal elements with index $N\alpha$ vanish,

$$F_{N\alpha}=\sum_i \frac{1}{Np_i}\frac{\partial Np_i}{\partial N}\frac{\partial Np_i}{\partial \alpha}=\sum_i \frac{1}{Np_i}p_i N\frac{\partial p_i}{\partial \alpha}$$

$$=\sum_i \frac{\partial p_i}{\partial \alpha}=\frac{\partial}{\partial \alpha}\sum_i p_i=\frac{\partial}{\partial \alpha}1=0\,, \tag{10}$$

so that the Fisher-information $\mathbf{F}^{\mathrm{p}}$ for the Poisson estimator becomes

$$\mathbf{F}^{\mathrm{p}}=\begin{pmatrix}1/N & 0\\ 0 & \mathbf{F}^{\mathrm{m}}\end{pmatrix}, \tag{11}$$

where the first row and column are associated with the intensity N. And for the covariance matrix we get

$$\mathrm{cov}^{\mathrm{p}}=\begin{pmatrix}N & 0\\ 0 & \mathrm{cov}^{\mathrm{m}}\end{pmatrix}. \tag{12}$$

So the Poisson approach introduces an uncertainty in N of the square root of N. Parameters depending on N will thus have an increased variance, while parameters not depending on N will have the same variance as in the multinomial approach. However, since the relative error in N is $1/\sqrt{N}$ – which is generally small – this effect can be neglected in most cases and two approaches are equivalent. Therefore the lower bounds calculated for the multinomial approach are also valid for the usual least-squares approach with Poisson statistics.

In the following we will work with the multinomial distribution because of its mathematical rigour and because it leads in a straightforward way to eq. (6), which allows one to answer the posed questions.

5. Monoexponential decays

Let us consider a background-free monoexponential decay with a lifetime τ. The probabilities p_i are defined by

$$p_i=\int_{\Delta_i} \mathrm{d}t\, d(t)\,, \tag{13}$$

where Δ_i is the temporal interval corresponding to channel i and $d(t)$ is the probability density of the monoexponential decay. It has the form

$$d(\tau,T,t)=\frac{1}{\tau}\exp(-t/\tau)\frac{1}{1-\exp(-T/\tau)}\,, \tag{14}$$

and is normalized to the finite width T of the measurement window. The probabilities p_i are

$$p_i(\tau,T,k)=\int_{(i-1)T/k}^{iT/k}\mathrm{d}t\, d(\tau,T,t)$$

$$=\exp(-ir/k)\frac{\exp(r/k)-1}{1-\exp(-r)}\,, \tag{15}$$

with the channels numbered from 1 to k and $r=T/\tau$ is the number of lifetimes in the measurement window.

The Fisher-information matrix $\mathbf{F}$ for this case is a number and the variance is [11,18]

$$\mathrm{var}_N(\tau,T,k)=\frac{1}{N}\tau^2\frac{k^2}{r^2}[1-\exp(-r)]$$

$$\times\left(\frac{\exp(r/k)[1-\exp(-r)]}{[\exp(r/k)-1]^2}-\frac{k^2}{\exp(r)-1}\right)^{-1}. \tag{16}$$

This can be separated and rewritten as

$$\mathrm{var}_N(\tau,T,k)=\frac{1}{N}\mathrm{var}_1(\tau,T,k)$$

$$=\frac{1}{N}\tau^2\,\mathrm{var}_1(r,k)\,. \tag{17}$$

Let us consider an example: To measure with 10% accuracy a lifetime of $\tau=2.5$ ns with a measurement window of $T=8$ ns, or $r=3.2$, and a resolution of 8 bits or $k=2.56$ one would need only 185 photons ac-

cording to eqs. (6) and (16). This astonishingly low number of photons makes monoexponential lifetime measurements on single molecules in solution feasible [7].

On the other hand, the equations can be used to optimize the experimental boundary conditions, such as the width of the measurement window T or the number of channels k, by simply minimizing $\mathrm{var}_1(\tau, T, k)$ or better $\mathrm{var}_1(\tau, T, k)/\tau^2 = \mathrm{var}_1(r, k)$. In fig. 1 we see that small measurement windows are detrimental, but if T is too large, all the signal falls into the first channel, making estimates rather difficult. A measurement window of ten or more lifetimes is generally best, and little is to be gained beyond two or three bits of resolution. Note that under ideal conditions (infinite, continuous measurement window) $\mathrm{var}_1(r, k)$ is unity.

6. Monoexponential decays with constant background

If there is a constant background underlying the fluorescence decay, one can write the multonomial probabilities as

$$p_i(\tau, b, T, k) = \frac{b}{k} + (1-b)p_i(\tau, T, k)\,, \qquad (18)$$

where b is the portion of the signal that is due to background. (The background need not be constant.) In order not to add noise to the data, the background should never be subtracted. Even if a constant background is subtracted the relative noise increases and the counts cannot be described by a Poisson distribution anymore.

Two cases have to be distinguished: (I) the background has to be fitted from the data and (II) the background has been determined by a separate experiment. Case (I) is a usual fitting situation. The best way to treat case (II) would be a global analysis [19] but usually the amount of information about the background is large in the preliminary experiment (pe) and small in the experiment used for the lifetime estimate (le). Therefore, one might simply take the estimate of the background component (b) from the preliminary experiment, calculate the Fisher-information of the lifetime experiment at the point b and take its error into account by adding the information matrices, since the information is an additive quantity [20]:

$$\mathbf{F}^{\mathrm{m}} = \mathbf{F}^{\mathrm{m}}_{\mathrm{le}} + \mathbf{F}^{\mathrm{m}}_{\mathrm{pe}} = \mathbf{F}^{\mathrm{m}}_{\mathrm{le}} + \begin{pmatrix} 0 & 0 \\ 0 & 1/\sigma^2 \end{pmatrix}, \qquad (19)$$

where σ is the standard deviation of the background estimate from the preliminary experiment.

Fig. 2 shows var_1/τ^2 for (I) the background-free case ($b=0$), (II) the case with b fixed to 20% and (III) for the case where b is 20% but has to be estimated (all curves are drawn for 256 channels). In

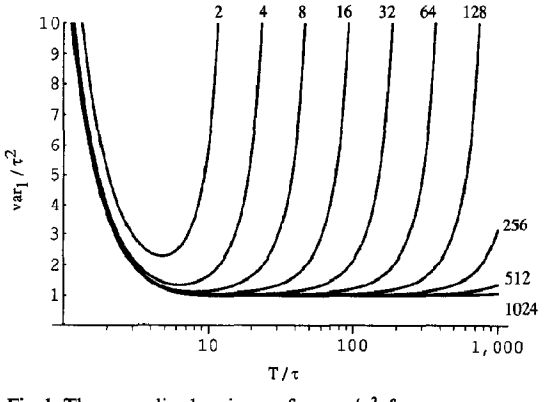

Fig. 1. The normalized variance of τ, var_1/τ^2, for monoexponential decays as a function of T/τ for different numbers of channels. T is the total width of the measurement window. The normalized variance is proportional to the number of photons needed for a certain accuracy.

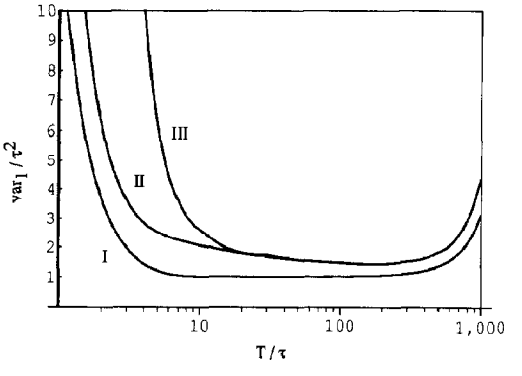

Fig. 2. Monoexponential decays with background: The normalized variance of τ, var_1/τ^2, as a function of T/τ for (I) the background free case, (II) the case with the background known to be 20% and (III) for the case where the background is 20% but has to be estimated. T is the total width of the measurement window. All curves are drawn for 256 channels.

the presence of background, it is better to choose a larger measurement window to increase the significance of the last bins, here of about 30 lifetimes. The loss in accuracy or sensitivity is then less than a factor of two in this example and it does not make a big difference whether the background has to be estimated or is known.

Another example showing the number of fluorescence photons needed for a certain accuracy as a function of the background counts is given in ref. [14].

7. Biexponential decays

As a third example let us consider a biexponential decay with constant background. The probabilities p_i can be written as

$$p_i = \frac{b}{k} + (1-b) \times [a p_i(\tau_1, T, k) + (1-a) p_i(\tau_2, T, k)] , \qquad (20)$$

where a is the portion of the fluorescence coming from the first lifetime component.

Besides the questions of the influence of the background and the appropriate choice of the number of channels or the width of the temporal window – that can be treated for the biexponential case much in the same way as they were for the monoexponential case – let us look as a third example at the index of separability S [13] defined by

$$S = \frac{|\tau_1 - \tau_2|}{\sqrt{\mathrm{var}(\tau_1) + \mathrm{var}(\tau_2)}} . \qquad (21)$$

We can extract the intensity N:

$$S = \sqrt{N} \frac{|\tau_1 - \tau_2|}{\sqrt{\mathrm{var}_1(\tau_1) + \mathrm{var}_1(\tau_2)}} \qquad (22)$$

and find the signal strength N necessary for a desired separability S:

$$N \geq S^2 \frac{\mathrm{var}_1(\tau_1) + \mathrm{var}_1(\tau_2)}{(\tau_1 - \tau_2)^2} . \qquad (23)$$

For instance, in a biexponential decay with $\tau_1 = 2$ ns and $\tau_2 = 4$ ns and a temporal window of $T = 10$ ns, $a = 10\%$ and 256 channels, one would need 400000 photons for a separability of $S = 2$. Again the multinomial approach has led in a straightforward way to an analytical expression for the required signal strength.

8. Conclusion

An analytic expression is given for the number of detected photons that are needed to achieve a certain accuracy of a parameter estimate in time-correlated single photon counting experiments. This was achieved by a multinomial approach which, although equivalent in performance to the officially recommended Poisson-based procedure [21], has the advantage of clarifying the structure of the fitting problem and allowing the calculation of suitable signal strength. The influence of dark counts on the required signal strength was calculated.

The techniques developed here are not limited to time-correlated single photon counting experiments and lifetime estimates, but can be applied to different kinds of single photon counting experiments involving, for instance, Lorentzian or Gaussian lineshape distributions in fluorescence spectra. An important requirement, however, is that for the applicability of multinomial statistics. This is normally the case if photon counting is combined with multichannel analysis.

Future work will try to incorporate the convolution of the decay curve with the experimentally determined excitation pulse shape. Furthermore, the techniques developed here should allow one to calculate the signal strength one needs to distinguish different models, for instance to tell whether a decay is bi- or tri-exponential. This will be investigated in the near future.

Acknowledgement

This work was supported by a grant from the University of Heidelberg.

References

[1] D.V. O'Connor and D. Phillips, Time-correlated single photon counting (Academic Press, New York, 1984).

[2] L.A. Sklar, Laser Focus World, August (1991) 61.
[3] F.J. Jett, R.A. Keller, J.C. Martin, B.L. Marrone, R.K. Moyzis, R.L. Ratliff, N.K. Seitzinger, E.B. Shera and C.C. Stewart, J. Biomol. Struct. Dynam. 7 (1989) 301.
[4] C. Seidel, K. Rittinger, J. Cortés, R.S. Goody, M. Köllner, J. Wolfrum and K.O. Greulich, in: Bimolecular spectroscopy II, Proc. SPIE 1432, eds. R.R. Birge and L.A. Nafie, p. 105.
[5] T. Minami, M. Kawahigashi, Y. Sakai, K. Shimamoto and S. Hirayama, J. Luminescence 35 (1986) 247.
[6] S.M. Keating and R.G. Wensel, Biophys. J. 59 (1991) 186.
[7] S.A. Soper, L.M. Davis and B.E. Shera, J. Opt. Soc. Am. B (1992), submitted for publication.
[8] S. Zacks, The theory of statistical inference (Wiley, New York, 1971).
[9] J. Orear, Notes on statistics for physicists, University of California, Radiation Laboratory, Berkeley, UCRL-8417 (1958).
[10] S. Baker and R.D. Cousins, Nucl. Instr. Methods 221 (1984) 437.
[11] P. Hall and B. Selinger, J. Phys. Chem. 85 (1981) 2941.
[12] P. Hall and B. Selinger, Z. Physik. Chem. NF 141 (1984) 77.
[13] Z. Bajzer, T.M. Therneau, J.C. Sharp and F.G. Prendergast, Eur. Biophys. J. 20 (1991) 247.
[14] M. Köllner, in: Ultrafast processes in spectroscopy 1991, eds. A. Laubereau and A. Seilmeier (The Institute of Physics, Bristol and Philadelphia, 1992) p. 151.
[15] C.R. Rao, Sankhya A 25 (1963) 189.
[16] R. Ponnapalli, Commun. Stat. Theoret. Methods 12 (1983) 47.
[17] M. Köllner, Appl. Optics (1992), in press.
[18] H.P. Good, A.J. Kallir and U.P. Wild, J. Phys. Chem. 88 (1984) 5435.
[19] J.R. Knutson, J.M. Beechem and L. Brand, Chem. Phys. Letters 102 (1983) 501.
[20] S. Kullback, Information theory and statistics (Wiley, New York, 1953).
[21] D.F. Eaton, Pure Appl. Chem. 62 (1990) 1631.