

Supplementary information on estimating the true emission count for “Photon counting, censor corrections, and lifetime imaging for improved detection in two-photon microscopy”

Jonathan D. Driscoll, Andy Y. Shih, Satish Iyengar, Jeffrey J. Field, G. Allen White, Jeffrey A. Squier, Gert Cauwenberghs and David Kleinfeld

For each experimental run, we assume that the number of photocurrent events, i.e., detected fluorescent emission events, N , is a Poisson random variable with mean α . If $N = n \geq 1$, then let W_1, \dots, W_n be the times of the N emissions; we assume that each W_i is exponentially distributed and that the emission times are independent of the number of emissions. For all calculations below, our clock is in units of the fluorescent lifetime, τ , of the emission times; thus, each W_i has mean 1. Next, sort the emission times in increasing order to get the order statistics $W_{(1)} < \dots < W_{(n)}$. In general, the joint distribution is a bit involved, but there is a convenient representation for order statistics from exponential random variables that facilitates many computations. In short, when $N = n$, start with independent and identically distributed (i.i.d.) exponential random variables $U = (U_1, \dots, U_n)$, each with mean 1; then the joint distribution of $(W_{(1)}, W_{(2)}, \dots, W_{(n)})$ is the same as that of $V = (V_1, \dots, V_n)$, where

$$V_k = \left[\frac{U_1}{n} + \frac{U_2}{n-1} + \dots + \frac{U_k}{n-k+1} \right], \text{ for } k = 1, \dots, n, \quad (1)$$

(Cox and Hinkley [1]). Thus, calculations involving the order statistics can be recast as those involving V , which in turn are weighted linear combinations of the i.i.d. exponential random variables, U .

The distribution of an order statistic or the joint distribution of several order statistics is easy in principle, but the expressions are a bit involved. For example, suppose that X_1, \dots, X_n are i.i.d. random variables with continuous cumulative distribution function (cdf) F . Let $X_{(1)} < \dots < X_{(n)}$ be the corresponding order statistics. Then, writing $C(n, j) = n!/[j!(n-j)!]$, we have

$$[X_{(k)} \leq x] = [\text{at least } k \text{ of the } X_i \text{ are at most } x],$$

so that

$$P(X_{(k)} \leq x) = \sum_{j=k}^n C(n, j) F(x)^j (1 - F(x))^{n-j},$$

and the probability density function (pdf) of $X_{(k)}$ is

$$f_{X_{(k)}}(t) = kC(n, k) f(x) F(x)^{k-1} (1 - F(x))^{n-k}. \quad (2)$$

For each experimental run, the observation period is normalized by the fluorescent lifetime, so that $T \equiv 1/(\tau f_{pixel})$. If an emission occurs after T , it is not observed in that run. If an emission occurs before T , there is a normalized dead period of length $\delta \equiv \Delta/\tau$ during which any emissions are censored. As a practical matter, we can observe at most T/δ emissions during an experimental run. For the current hardware the observation period is 13.2 ns and the dead period is 3.1 ns, so we can essentially observe at most four emissions. In order to estimate α from a series of such experiments, we need the probabilities of observing i emissions (or dead periods) for $i = 1, \dots, 4$. Let D denote the number of emissions observed in an experimental run. We now evaluate $p_i = P(D = i)$ for $i = 0, 1$ (we suppress the dependence of $p_i(\alpha)$ on T and δ for simpler notation). Our main results are that

$$p_0(\alpha) = e^{-\alpha(1-e^{-T})} \quad (3)$$

and

$$p_1(\alpha) = \frac{e^{-\alpha[e^{-\delta}-e^{-T}]} - e^{-\delta}e^{-\alpha[1-e^{-(T-\delta)}]} - (1-e^{-\delta})e^{-\alpha[1-e^{-T}]}}{(1-e^{-\delta})}. \quad (4)$$

The computations of $p_i(\alpha)$ for $i = 2, 3, 4$ are considerably more cumbersome, have progressively smaller contributions, and are thus omitted.

The proof of equation (3) is elementary. First, note that the generating function of a Poisson random variable is

$$\sum_{n=0}^{\infty} P(N = n)\zeta^n = e^{\alpha(\zeta-1)}.$$

Then note that the event $[D = 0]$ occurs if and only if either $[N = 0]$ or $[N \geq 1, W_{(1)} > T]$. Conditioning on $N = n$, we know from (1) that $W_{(1)}$ has an exponential distribution with pdf ne^{-nx} ; thus,

$$\begin{aligned} p_0(\alpha) &= P(N = 0) + \sum_{n=1}^{\infty} P(N = n)e^{-nT} = e^{-\alpha} + \sum_{n=1}^{\infty} e^{-\alpha} \frac{\alpha^n}{n!} e^{-nT} \\ &= \sum_{n=0}^{\infty} e^{-\alpha} \frac{\alpha^n}{n!} e^{-nT} = e^{-\alpha(1-e^{-T})}. \end{aligned} \quad (5)$$

The proof of equation (4) is more involved. We consider the cases $N = 1$, $N = 2$, and $N \geq 3$ separately. When $N = 1$, we observe one emission only if $W_1 < T$, so

$$\begin{aligned} P(D = 1, N = 1) &= P(N = 1, W_1 < T) = P(N = 1)(1 - e^{-T}) \\ &= P(N = 1)[(1 - e^{-(T-\delta)}) + e^{-T}(e^{\delta} - 1)]. \end{aligned} \quad (6)$$

Next, when $N = 2$, we observe only one emission if one of the following conditions holds: (a) $[T - \delta < W_{(1)} < T]$, (b) $[W_{(1)} \leq T - \delta, W_{(2)} \leq W_{(1)} + \delta]$, or (c) $[W_{(1)} \leq T - \delta, W_{(2)} > T]$.

The probabilities of these three events are

$$\begin{aligned}
P(T - \delta < W_{(1)} < T) &= e^{-2(T-\delta)} - e^{-2T}, \\
P(W_{(1)} \leq T - \delta, W_{(2)} \leq W_{(1)} + \delta) &= (1 - e^{-\delta})(1 - e^{-2(T-\delta)}), \text{ and} \\
P(W_{(1)} \leq T - \delta, W_{(2)} > T) &= 2e^{-T}(1 - e^{-(T-\delta)}).
\end{aligned} \tag{7}$$

These calculations come from the representation in equation (1). Thus, $P(D = 1, N = 2)$ is $P(N = 2)$ times the sum of these three terms.

When $N = n \geq 3$, a similar listing of the configurations which allow us to observe one emission give the following probabilities. Once again, from the representation in equation (1), we have

$$P(T - \delta < W_{(1)} < T) = e^{-n(T-\delta)} - e^{-nT} = e^{-nT}(e^{n\delta} - 1).$$

Next,

$$\begin{aligned}
P(W_{(1)} \leq T - \delta, W_{(n)} \leq W_{(1)} + \delta) &= P\left(\frac{U_1}{n} \leq T - \delta, \sum_{j=1}^n \frac{U_j}{n-j+1} \leq \frac{U_1}{n} + \delta\right) \\
&= P[U_1 \leq n(T - \delta)]P\left(\sum_{j=2}^n \frac{U_j}{n-j+1} \leq \delta\right) \\
&= (1 - e^{-n(T-\delta)})(1 - e^{-\delta})^{n-1},
\end{aligned}$$

and

$$\begin{aligned}
P(W_{(1)} \leq T - \delta, W_{(2)} > T) &= P\left[\frac{U_1}{n} \leq T - \delta, \frac{U_1}{n} + \frac{U_2}{n-1} > T\right] \\
&= \int_0^{n(T-\delta)} e^{-(n-1)(T-x/n)} e^{-x} dx \\
&= ne^{-(n-1)T}(1 - e^{-(T-\delta)}).
\end{aligned}$$

Next, for $j = 2, \dots, n-1$,

$$P(W_{(1)} \leq T - \delta, W_{(j)} \leq W_{(1)} + \delta, W_{(j+1)} > T) = P(A_1 \leq T - \delta, A_2 < \delta, \sum_{i=1}^3 A_i > T),$$

where $A_1 = U_1/n$ has pdf ne^{-nx} ,

$$A_2 = \sum_{i=2}^j \frac{U_i}{n-i+1}$$

has pdf

$$g_{A_2}(y) = (j-1)C(n-1, j-1)(1 - e^{-y})^{j-2}e^{-(n-j+1)y}$$

because it has the same distribution as that of the $(j - 1)$ st order statistic from a sample of size $(n - 1)$, and $A_3 = U_{j+1}/(n - j)$ has pdf $(n - j)e^{-(n-j)z}$. Thus, conditioning on A_1 and A_2 , we get

$$\begin{aligned} P[W_{(1)} \leq T - \delta, W_{(j)} \leq W_{(1)} + \delta, W_{(j+1)} > T] &= \int_0^{T-\delta} \int_0^\delta e^{-(n-j)(T-x-y)} g_{A_2}(y) n e^{-nx} dy dx \\ &= C(n, j) e^{-(n-j)T} [1 - e^{-j(T-\delta)}] [1 - e^{-\delta}]^{j-1}. \end{aligned}$$

Combining these terms, we get

$$p_1(\alpha) = \sum_{n=1}^{\infty} P(N = n) B_n = \sum_{n=0}^{\infty} P(N = n) B_n, \quad (8)$$

where

$$\begin{aligned} B_n &= e^{-nT} (e^{n\delta} - 1) + \sum_{j=1}^n C(n, j) (1 - e^{-\delta})^j e^{-(n-j)T} (1 - e^{-j(T-\delta)}) \\ &= e^{-nT} (e^{n\delta} - 1) + \sum_{j=0}^n C(n, j) (1 - e^{-\delta})^j e^{-(n-j)T} (1 - e^{-j(T-\delta)}) \\ &= e^{-nT} (e^{n\delta} - 1) + \frac{(1 - e^{-\delta} + e^{-T})^n - (e^{-(T-\delta)}(1 - e^{-\delta}) + e^{-T})^n}{(1 - e^{-\delta})}. \end{aligned}$$

Note that $B_0 = 0$, as is the zeroth term in the binomial expansion contained in B_n . Finally, we use the generating function of a Poisson random variable to sum the series in equation (9) to get the expression in equation (4).

Estimation of α

We now turn to the problem of estimating the Poisson parameter α . Suppose that we have r experimental runs, with r_i cases having i observed emissions. Below we deal with two cases: use only r_0 or use both (r_0, r_1) . In both cases we seek the maximum likelihood estimate (MLE).

We first summarize the key facts about MLEs in general (Lehmann and Casella [2]). Suppose that $X^{(n)} = (X_1, \dots, X_n)$ are independent and identically distributed observations with pdf $f(x|\theta)$, where θ is unknown. The likelihood, up to a constant that does not depend on θ , is

$$L(\theta|X^{(n)}) = \prod_{i=1}^n f(X_i|\theta).$$

More convenient is the log-likelihood $l(\theta|X^{(n)}) = \ln L(\theta|X^{(n)})$. The MLE is the value $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ which maximizes $l(\theta|X^{(n)})$. When $l(\theta|X^{(n)})$ is sufficiently smooth, the

MLE is a solution to $l'(\theta|X^{(n)}) = 0$. Under certain regularity conditions, which are met for both cases below, $\hat{\theta}_n$ is unique and for large n , is approximately normal with mean θ and standard deviation $\hat{s} = [-l''(\hat{\theta}_n|X^{(n)})]^{-1/2}$. Thus, an approximate 95% confidence interval for θ is $\hat{\theta}_n \pm 1.96\hat{s}$. With these general results, we now turn to the details of the two cases.

1. USE ONLY r_0 .

Suppose that there are r experimental runs, with r_0 having no emissions and the remaining $(r - r_0)$ having at least one emission. The likelihood and log-likelihood functions for α are, respectively,

$$L(\alpha|r, r_0) \propto p_0(\alpha)^{r_0}(1 - p_0(\alpha))^{r-r_0}$$

and

$$l(\alpha|r, r_0) = r_0 \ln(p_0(\alpha)) + (r - r_0) \ln(1 - p_0(\alpha)).$$

In the expression for L , we omit the combinatorial coefficient $C(r, r_0)$ because it does not depend on α . In this case, maximum likelihood estimate (MLE) of $p_0(\alpha)$ is easy to compute: $\hat{p}_0(\alpha) = r_0/r$; thus, the MLE of α is

$$\hat{\alpha}_{0r} = -\frac{\ln(\hat{p}_0(\alpha))}{(1 - e^{-T})} = -\frac{\ln(r_0/r)}{(1 - e^{-T})}.$$

Using the second derivative of the log-likelihood at the MLE, the approximate standard deviation is given by

$$\hat{\sigma}_{0r} = [-l''(\hat{\alpha}_{0r}|r, r_0)]^{-1/2} = (1 - e^{-T})^{-1} \sqrt{\frac{(r - r_0)}{rr_0}}. \quad (9)$$

2. USE ONLY (r_0, r_1) .

This time, suppose that there are r experimental runs, with r_0 having no emissions, r_1 having one emission, and the remaining $(r - r_0 - r_1)$ having at least two emissions. Writing $q(\alpha) = 1 - p_0(\alpha) - p_1(\alpha)$, the log-likelihood function is

$$l(\alpha|r, r_0, r_1) = r_0 \ln(p_0(\alpha)) + r_1 \ln(p_1(\alpha)) + (r - r_0 - r_1) \ln(q(\alpha)).$$

An explicit solution for the MLE $\hat{\alpha}_{1r}$ is not available, but a search using Newton-Raphson iterations with an analytic expression for the first derivatives will yield $\hat{\alpha}_{1r}$. To do these calculations, we need the following expressions:

$$l'(\alpha|r, r_0, r_1) = r_0 \frac{p'_0(\alpha)}{p_0(\alpha)} + r_1 \frac{p'_1(\alpha)}{p_1(\alpha)} + (r - r_0 - r_1) \frac{q'(\alpha)}{q(\alpha)},$$

where

$$\frac{p'_0(\alpha)}{p_0(\alpha)} = -(1 - e^{-T}) \quad \text{or} \quad p'_0(\alpha) = -(1 - e^{-T})p_0(\alpha)$$

and

$$(1 - e^{-\delta})p_1'(\alpha) = -(e^{-\delta} - e^{-T})e^{-\alpha[e^{-\delta} - e^{-T}]} + e^{-\delta}(1 - e^{-(T-\delta)})e^{-\alpha[1 - e^{-(T-\delta)}]} \\ + (1 - e^{-\delta})(1 - e^{-T})e^{-\alpha[1 - e^{-T}]}$$

allow for the calculation of $l'(\alpha|r, r_0, r_1)$. Next, because p_0'/p_0 does not depend on α , the second derivative is

$$l''(\alpha|r, r_0, r_1) = r_1 \frac{p_1(\alpha)p_1''(\alpha) - [p_1'(\alpha)]^2}{p_1(\alpha)^2} + (r - r_0 - r_1) \frac{q(\alpha)q''(\alpha) - [q'(\alpha)]^2}{q(\alpha)^2},$$

for which we need the expressions

$$p_0''(\alpha) = (1 - e^{-T})^2 p_0(\alpha)$$

and

$$(1 - e^{-\delta})p_1''(\alpha) = (e^{-\delta} - e^{-T})^2 e^{-\alpha[e^{-\delta} - e^{-T}]} - e^{-\delta}(1 - e^{-(T-\delta)})^2 e^{-\alpha[1 - e^{-(T-\delta)}]} \\ - (1 - e^{-\delta})(1 - e^{-T})^2 e^{-\alpha[1 - e^{-T}]}.$$

The approximate standard deviation of $\hat{\alpha}_{1r}$ is then $\hat{\sigma}_{1r} = [-l''(\hat{\alpha}_{1r}|r, r_0, r_1)]^{-1/2}$; although it does not have an elementary expression, it is easy to compute.

References

1. Cox DR, Hinkley DV. (1974) *Theoretical Statistics*. Chapman and Hall.
2. Lehmann EL, Casella G. (1998) *Theory of Point Estimation*. Springer.