LLNL-JRNL-420247

Efficient Levenberg-Marquardt Minimization of the Maximum Likelihood Estimator for Poisson Deviates

T. A. Laurence and B. Chromy

November 13, 2009

Nature Methods

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Efficient Levenberg-Marquardt minimization of the maximum likelihood estimator for Poisson deviates

Ted A. Laurence, Brett Chromy

Abstract

Histograms of counted events are Poisson distributed, but are typically fitted without justification using nonlinear least squares fitting. The more appropriate maximum likelihood estimator (MLE) for Poisson distributed data is seldom used. We extend the use of the Levenberg-Marquardt algorithm commonly used for nonlinear least squares minimization for use with the MLE for Poisson distributed data. In so doing, we remove any excuse for not using this more appropriate MLE. We demonstrate the use of the algorithm and the superior performance of the MLE using simulations and experiments in the context of fluorescence lifetime imaging.

Scientists commonly form histograms of counted events from their data, and extract parameters by fitting to a specified model. Assuming that the probability of occurrence for each bin is small, event counts in the histogram bins will be distributed according to the Poisson distribution [1]. We develop here an efficient algorithm for fitting event counting histograms using the maximum likelihood estimator (MLE) for Poisson distributed data, rather than the non-linear least squares measure. This algorithm is a simple extension of the common Levenberg-Marquardt (L-M) algorithm [2], is simple to implement, quick and robust.

Fitting using a least squares measure is most common, but it is the maximum likelihood estimator only for Gaussian-distributed data. Non-linear least squares methods may be applied to event counting histograms in cases where the number of events is very large, so that the Poisson distribution is well approximated by a Gaussian. However, it is not easy to satisfy this criterion in practice – which requires a large number of events.

It has been well-known for years that least squares procedures lead to biased results when applied to Poisson-distributed data; a recent paper providing extensive characterization of these biases in exponential fitting is given in [3]. The more appropriate measure based on the maximum likelihood estimator (MLE) for the Poisson distribution is also well known, but has not become generally used. This is primarily because, in contrast to non-linear least squares fitting, there has been no quick, robust, and general fitting method. In the field of fluorescence lifetime spectroscopy and imaging, there have been some efforts to use this estimator through minimization routines such as Nelder-Mead optimization [4, 5], exhaustive line searches [6], and Gauss-Newton minimization [7]. Minimization based on specific one- or multi-exponential models [8, 9] has been used to obtain quick results, but this procedure does not allow the incorporation of the instrument response, and is not generally applicable to models found in other fields. Methods for using the MLE for Poisson-distributed data have been published by the wider spectroscopic community [10, 11], including iterative minimization schemes based on Gauss-Newton minimization[12].

The slow acceptance of these procedures for fitting event counting histograms may also be explained by the use of the ubiquitous, fast Levenberg-Marquardt (L-M) fitting procedure for fitting non-linear models using least squares fitting [2] (simple searches obtain ~10000 references – this doesn't include those who use it, but don't know they are using it). The benefits of L-M include a seamless transition between Gauss-Newton minimization and downward gradient minimization through the use of a regularization parameter. This transition is desirable because Gauss-Newton methods converge quickly, but only within a limited domain of convergence; on the other hand the downward gradient methods have a much wider domain of convergence, but converge extremely slowly nearer the minimum. L-M has the advantages of both procedures: relative insensitivity to initial parameters and rapid convergence. Scientists, when wanting an answer quickly, will fit data using L-M, get an answer, and move on. Only those that are aware of the bias issues will bother to fit using the more appropriate MLE for Poisson deviates. However, since there is a simple, analytical formula for the appropriate MLE measure for Poisson deviates, it is inexcusable that least squares estimators are used almost exclusively when fitting event counting histograms. There have been ways found to use successive non-linear least squares fitting to obtain similarly unbiased results [3], but this procedure is justified by simulation, must be re-tested when conditions change significantly, and requires two successive fits. There is a great need for a fitting routine for the MLE estimator for Poisson deviates that has convergence domains and rates comparable to the non-linear least squares L-M fitting.

We show here that a simple way to achieve that goal is to use the L-M fitting procedure not to minimize the least squares measure, but the MLE for Poisson deviates. We replace the least squares calculation in the L-M routine with the MLE estimator, make corresponding changes in the gradients calculated, and achieve the desired procedure.

Consider a data set $\mathbf{y}=(y_1,y_2,\ldots,y_N)$ where each y_i is the number of independent events in a histogram bin i defined as a range in variables $\mathbf{x}=(x_1,x_2,\ldots)$. N is large enough that the probability that an event is within any bin is small (Poisson distributed rather than multinomial). The model function to be fitted is $\mathbf{f}=(f_1,f_2,\ldots,f_N)$ and depends on parameters $\mathbf{a}=(a_1,a_2,\ldots,a_M)$. The MLE for Poisson deviates is

$$\chi^2_{\text{mle}} = 2\sum_{i=1}^{N} f_i - y_i - 2\sum_{i=1\ y_i\neq 0}^{N} y_i \ln\left(f_i/y_i\right) \tag{1}$$

This function is minimized to find the best fit. Note that each f_i is required to be positive; any fitting procedure must adjust **a** in such a way that **f** is positive for every element. Otherwise, the minimization will run into problems. For example, if y_i is positive, then the last term in Eq. (1) is undefined. Eq. (1) may be compared to the more familiar least-squares χ^2 statistic,

$$\chi^2 = \sum_{i=1}^{N}\left(f_i - y_i\right)^2 \Big/ \sigma_i^2. \tag{2}$$

In this case, σ_i is the uncertainty for each data point y_i.

As we extend the L-M algorithm for use with the MLE for Poisson-distributed histogram bins [Eq. (1)], we follow many of the arguments made in Marquardt's original paper and in the book Numerical Recipes[13].

Iterative minimization schemes use information about the function near current parameters $\mathbf{a}_{\text{curr}}$ to find improved estimates of the parameters that minimize that function. In the neighborhood of the current parameters $\mathbf{a}_{\text{curr}}$, the function χ^2_{MLE} may be approximated by its Taylor expansion,

$$\chi^2_{\text{mle}}\left(\mathbf{a}_{\text{curr}}+\Delta\mathbf{a}\right)\approx\chi^2_{\text{mle,curr}}+\nabla_{\mathbf{a}}\chi^2_{\text{mle}}\cdot\Delta\mathbf{a}+\tfrac{1}{2}\Delta\mathbf{a}^{\text{T}}\cdot\nabla_{\mathbf{a}}\nabla_{\mathbf{a}}\chi^2_{\text{mle}}\cdot\Delta\mathbf{a} \tag{3}$$

The gradient of χ^2_{mle} with respect to the vector of model parameters $\boldsymbol{a}$ is

$$\nabla_{\mathbf{a}}\chi^2_{\text{mle}}=\frac{\partial\chi^2_{\text{mle}}}{\partial a_k}=2\sum_{i=1}^{\text{N}}\left(1-\frac{y_i}{f_i}\right)\frac{\partial f_i}{\partial a_k}. \tag{4}$$

The gradient of the gradient is

$$\nabla_{\mathbf{a}}\nabla_{\mathbf{a}}\chi^2_{\text{mle}}=\frac{\partial^2\chi^2_{\text{mle}}}{\partial a_k\partial a_l}=2\sum_{i=1}^{\text{N}}\frac{\partial f_i}{\partial a_k}\frac{\partial f_i}{\partial a_l}\frac{y_i}{f_i^2}+\left(1-\frac{y_i}{f_i}\right)\frac{\partial^2 f_i}{\partial a_k\partial a_l}. \tag{5}$$

If $\mathbf{a}_{\text{curr}}$ is sufficiently near the minimum, then by differentiating Eq. (3) and setting to 0, we determine $\mathbf{a}_{\text{min}}$,

$$\nabla_{\mathbf{a}}\nabla_{\mathbf{a}}\chi^2\left(\mathbf{a}_{\text{curr}}\right)\cdot\left(\mathbf{a}_{\text{min}}-\mathbf{a}_{\text{curr}}\right)=-\nabla_{\mathbf{a}}\chi^2\left(\mathbf{a}_{\text{curr}}\right) \tag{6}$$

In Newton and Gauss-Newton iterative schemes, Eq. (6) is used to find successively improved estimates of $\mathbf{a}_{\text{min}}$. We will now imitate the procedure for the L-M algorithm for least squares, and ignore the term with the second derivatives in Eq. (5) when using Eq. (6). Gauss-Newton minimization schemes ignore the corresponding term for least squares analysis, both because of the frequent difficulty in computing the second derivatives and for their potential destabilizing influence [13]. We will discuss this issue further at the end of the letter. We define the following matrix $\boldsymbol{\alpha}$ and vector $\boldsymbol{\beta}$,

$$\boldsymbol{\alpha}=\alpha_{kl}=\sum_{i=1}^{\text{N}}\frac{\partial f_i}{\partial a_k}\frac{\partial f_i}{\partial a_l}\frac{y_i}{f_i^2};\boldsymbol{\beta}=\beta_k=-\sum_{i=1}^{\text{N}}\left(1-\frac{y_i}{f_i}\right)\frac{\partial f_i}{\partial a_k}. \tag{7}$$

We can then write Eq. (6) as

$$\alpha_k\Delta a_l=\beta_k. \tag{8}$$

Just as in the least squares case, the Gauss-Newton minimization scheme of Eq. (8) runs into trouble when Eq. (3) is a poor approximation, leading to increases in χ^2_{mle} after an iteration step. We solve this problem in precisely the same way as in the L-M least squares algorithm: we augment the diagonal elements of the matrix by a multiplicative factor $\left(1+\lambda\right)$,

$$\alpha'_k=\alpha_k\left(1+\lambda\delta_{kl}\right) \tag{9}$$

We then have the iteration step

$$\alpha'_k \Delta q = \beta_k . \tag{10}$$

In the L-M algorithm, if an iteration step in Eq. (10) fails to decrease χ^2_{mle}, we increase λ (typically by a factor of 10), and try Eq. (10) again. As λ increases, the solution to Eq. (10) rotates toward the downhill gradient of χ^2_{mle}. For sufficiently high λ, this will guarantee a decrease in χ^2_{mle}, although convergence may be slow in certain cases. Marquardt provides three theorems to justify his algorithm [2]. These can be seen to apply in this case by replacing the expressions for the least squares estimator with the corresponding expressions for the MLE for the Poisson distribution. It is particularly important to note that α_{kl} (which replaces A in Marquardt's paper), is positive definite – as long as f_i is strictly positive.

We are able to easily modify the Numerical Recipes code (2nd edition) for the L-M algorithm to perform the minimization for χ^2_{mle}. The only changes needed are in the function `mrqcof` (no changes in `mrqmin`). We simply modify the calculations for `alpha` and `beta` to correspond to Eq. (7), and the calculation of `chisq` to correspond to Eq. (1). `sig2i` is no longer used.

We now test our implementation on simulated and experimental data. In each case, we show the benefits of MLE over least squares fitting, and show that convergence rates and domains are similar. We compare our results with those obtained using non-linear least squares fitting of Eq. (2). To do this, we choose values for σ_i as typically implemented for event counting data. Since σ_i depends on the expectation $\langle y_i \rangle$ for Poisson-distributed data, an estimate for σ_i must be made based on the available information (**y** and **f**). Since $\langle y_i \rangle = \sigma_i^2$, Neyman defined χ^2_{N} by estimating σ_i^2 with y_i,

$$\chi^2_{\text{N}} = \sum_{i=1}^{N} (f_i - y_i)^2 / \max(y_i, 1) \tag{11}$$

The max function is there to prevent division by 0. Pearson instead estimates σ_i^2 with the function value f_i, yielding

$$\chi^2_{\text{P}} = \sum_{i=1}^{N} (f_i - y_i)^2 / f_i \tag{12}$$

Several other modified χ^2 statistics have been proposed; however, problems in biases are encountered with each [3, 10, 11]. In our tests, we implement both Neyman's χ^2_{N} and an unweighted χ^2_1 where we just weight all data points equally, σ_i=1. We attempted to use Eq. (12), but the final fitted values varied erratically.

We test our method on simulated and measured data of one- or two-exponential lifetime decays, as encountered in fluorescence lifetime measurements using time-correlated single photon counting (TCSPC). In such measurements, a high repetition rate laser (e.g. 40 MHz) excites fluorescence. Fluorescence is detected by single photon

detectors. The time delay between each detected photon and the exciting laser pulse is measured and recorded. Histograms of the measured photon time delays are fit to exponentials in order to extract the fluorescence lifetime. In order to accurately extract lifetimes, the finite bin widths, the finite measurement window available between laser pulses, and the measured instrument response must be taken into account. We take T as the time between laser pulses. As above, N is the number of bins in the histogram. τ is the lifetime of the decay model, and IRF_i is the measured instrument response (i=1…N). We define $r = \mathrm{T}/\tau$. A is the amplitude of the fitted lifetime curve (equal to the number of photons). The lifetime decay is modeled as in Zander et al. [6],

$$f_i(\tau) = A\left[e^{-ir/\mathrm{N}}\left(\frac{e^{r/\mathrm{N}}-1}{1-e^{-r}}\right)\right] * \mathrm{IRF}_i \qquad (13)$$

The symbol $*$ denotes convolution, which is typically calculated using fast Fourier transform algorithms. Eq. (13) is readily extended to more than one exponential. The TCSPC methodology is often used in single molecule experiments and in fluorescence lifetime imaging. These applications typically do not measure a number of photons large enough to justify use of χ^2_{N} or χ^2_1, and are ideal for use with χ^2_{mle}.

We first test our method on simulated, single exponential data with a 3 ns lifetime convolved with a measured instrument response (Figure 1). Figure 1A shows examples of the simulated data with 100 photons and the fit. We compared the performance of the L-M algorithm for χ^2_{mle}, χ^2_{N} and χ^2_1 in fitting 10000 realizations of the data with 100 photons each. The number of iterations required to reach convergence for χ^2_{mle}, χ^2_{N} and χ^2_1 is shown in Fig. 1B. The L-M algorithm quickly finds the minimum for χ^2_{mle} (4-5 iterations) and χ^2_{N} (2-3 iterations) when supplied with the amplitude and lifetime used in the simulation. However, the L-M algorithm has more difficulty in finding the minimum of χ^2_1, requiring significantly more iterations.

Minimizing χ^2_{mle} rather than χ^2_{N} or χ^2_1 more accurately finds the values of the amplitude A and the lifetime τ used in the simulation (Figs. 1C-D). Exactly 100 photons were produced in each simulation. Minimizing χ^2_{N} and χ^2_1 resulted in fitted A that deviate strongly from the number of photons in the data, whereas minimizing χ^2_{mle} resulted in precisely 100 photons for the fitted amplitude (Fig. 1C). The distribution of lifetimes fitted using χ^2_{mle} (2.99±0.31 ns) is narrower than the others (2.95±0.46 ns for χ^2_1; 3.29±0.50 ns for χ^2_{N}), and is centered around the correct value of 3 ns (Fig. 1D).

The benefit of using L-M over Gauss-Newton methods in minimizing χ^2_{mle} is revealed when changing initial fitting parameters. If one starts with the model τ=3ns, and varies the initial fitting value A, Gauss-Newton minimization (setting λ=0) fails to converge when $A > 225$ (actual value is 100), whereas the L-M implementation still converged for $A>10^8$. This is critical when doing large number of repeated fits that may have large variations in final fitting parameters; this is usually the situation of interest for lifetime imaging applications, and other similar statistical analyses.

In applications with more the one lifetime, the benefits of χ^2_{mle} over χ^2_N and χ^2_1 are even more dramatic (Fig. 2). The data in Fig. 2 were simulated using a two exponential model with equal amplitude for two lifetimes τ_1=1 ns and τ_2=3 ns. There are 1000 photons in each of the 10000 realizations of each fit. We show distributions for the two fitted lifetimes τ_1 and τ_2, and the ratio of $A_1/(A_1+A_2)$; the distribution of this ratio should ideally be centered around 0.5 (equal amplitudes for both lifetimes). Fig. 2A shows a 3D scatter plot for 10000 fits found by minimizing χ^2_{mle}, χ^2_N and χ^2_1. Figs. 2B-D show the same data, but represented as histograms over only one parameter. The results for χ^2_{mle} (black points and lines) are centered on the correct values. The results for χ^2_N (red) are offset from the correct values, not even intersecting the results using χ^2_{mle} in the 3D space in Fig. 2A. The results for χ^2_1 are centered better on the correct values, but with far more spread in the values than found with χ^2_{mle}. As seen in Fig. 2B, the fitted values often degenerate to a single exponential fit using χ^2_1, as indicated by peaks in the ratio $A_1/(A_1+A_2)$ near 0 and 1. The use of χ^2_{mle} clearly provides superior results, and the use of the L-M algorithm as developed here obtains those results quickly and robustly.

We now test the use of the L-M algorithm with χ^2_{mle} in an experimental situation with a single exponential lifetime. We perform lifetime imaging measurements on *Yersinia pestis* bacteria with constitutive expression of EGFP. We use sample scanning confocal microscopy combined with TCSPC. We excite the fluorescence using a 470 nm pulsed diode laser (LDH-P-C-470B , Picoquant, Berlin), detect fluorescence with an avalanche photodiode (PDM series, Micro Photon Devices), and measure the time delay using the PicoHarp 300 (Picoquant, Berlin). Figure 3A shows an image of several *Y. pestis* bacteria adhered to the glass surface. Figure 3B shows the fit for the lifetime (2.53±0.01 ns) for all of the photons contained in the image. We fit the lifetime histogram for every pixel with more than 20 photons, and the fitted lifetimes for each pixel are represented in Figure 3C. The color white represents the value of the lifetime extracted in Figure 3B, and deviations from that value are shown in red and green. Regions with fewer photons have a larger scatter, but the mean values of the fitted lifetime are the same throughout the image. Figures 3D and 3D compare the fitted lifetimes for, respectively, pixels with between 20 and 70 photons and pixels with between 150 and 300 photons. The histograms obtained for χ^2_{mle} are narrower than for the other cases just as we saw in Figure 1. Note than in Figure 3E that the fitted values found for χ^2_N are centered at a higher, incorrect value of 2.8 ns. Also, Figure 3F shows that convergence was easier to obtain for χ^2_{mle} and χ^2_N than for χ^2_1. The results from experiments match those obtained from simulation.

We have extended the L-M algorithm to be used with the MLE for Poisson distributed data, and have shown that it works for lifetime imaging applications both by simulation and experiment. This extension is not limited to fluorescence lifetime imaging applications. Indeed, it can be applied anytime one forms an event counting histogram in any application. We have performed simulations showing that the methodology works with the sum of two Gaussians, for instance (not shown).

In developing the algorithm, we ignored the second-derivative terms in the Hessian, Eq. (5). Minimization algorithms that build on the L-M algorithm for non-linear least squares [14, 15] estimate this ignored term when it is advantageous to do so, and are more robust than the original L-M algorithm in large residual problems. An extension of these algorithms to more general nonlinear regression models was provided later [16, 17]. Although the latter algorithm was explicitly shown to be applicable to a form of Eq. (1), we have not found any examples of this algorithm being used for this application. We suspect this has to do with the real or perceived complexity of the algorithm. There are many benefits and drawbacks of various non-linear minimization schemes. We have chosen to implement a simple change to a widely used version of the L-M algorithm to facilitate use of the MLE for Poisson-distributed data. We hope that this will inspire other researchers to take advantage of the benefits of the MLE for Poisson-distributed data, reducing biases and errors in parametric results from histogram-based analysis. There are enough uncertainties and pitfalls in science. If we can simply eliminate one variable in a large range of event counting applications, we should do so.

1. Ross, S.M., *Introduction to probability and statistics for engineers and scientists*. 1987, New York, N.Y.: Wiley. 492.
2. Marquardt, D.W., *An algorithm for least-squares estimation of nonlinear parameters.* J. Soc. Indust. Appl. Math., 1963. **11**(2): p. 431-441.
3. Turton, D.A., G.D. Reid, and G.S. Beddard, *Accurate analysis of fluorescence decays from single molecules in photon counting experiments.* Analytical Chemistry, 2003. **75**(16): p. 4182-4187.
4. Bajzer, Z. and F.G. Prendergast, *Maximum-Likelihood Analysis Of Fluorescence Data.* Methods In Enzymology, 1992. **210**: p. 200-237.
5. Bajzer, Z., et al., *Maximum-Likelihood Method For The Analysis Of Time-Resolved Fluorescence Decay Curves.* European Biophysics Journal With Biophysics Letters, 1991. **20**(5): p. 247-262.
6. Zander, C., et al., *Detection and characterization of single molecules in aqueous solution.* Applied Physics B (Lasers and Optics), 1996. **B63**(5): p. 517-23.
7. Sandor, T. and G.D. Wilson, *Maximum Likelihood Estimation Of Parameters In Multi Exponential Fits When Data Follow Poisson Distribution.* Computer Programs in Biomedicine, 1972. **2**(2): p. 111-117.
8. Hall, P. and B. Selinger, *Better Estimates Of Multiexponential Decay Parameters.* Zeitschrift Fur Physikalische Chemie Neue Folge, 1984. **141**: p. 77-89.
9. Hall, P. and B. Selinger, *Better Estimates Of Exponential Decay Parameters.* Journal Of Physical Chemistry, 1981. **85**(20): p. 2941-2946.
10. Hauschild, T. and M. Jentschel, *Comparison of maximum likelihood estimation and chi-square statistics applied to counting experiments.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2001. **457**(1-2): p. 384.
11. Stoneking, M.R. and D.J. Den Hartog. *Maximum-likelihood fitting of data dominated by Poisson statistical uncertainties*. in *Proceedings of the eleventh*

topical conference on high temperature plasma diagnostics. 1997. Monterey, California (USA): AIP.
12. Muravsky, V.A., S.A. Tolstov, and A.L. Kholmetskii, *Comparison of the least squares and the maximum likelihood estimators for gamma-spectrometry.* Nuclear Instruments & Methods In Physics Research Section B-Beam Interactions With Materials And Atoms, 1998. **145**(4): p. 573-577.
13. Press, W.H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C: the art of scientific computing*. 2nd ed. 1992, Cambridge, U.K.: Cambridge University Press. xxvi, 994.
14. Dennis, J.E., D.M. Gay, and R.E. Welsch, *An Adaptive Non-Linear Least-Squares Algorithm.* Acm Transactions On Mathematical Software, 1981. **7**(3): p. 348-368.
15. Dennis, J.E., D.M. Gay, and R.E. Welsch, *Algorithm 573 - Nl2sol - An Adaptive Non-Linear Least-Squares Algorithm [E4].* Acm Transactions On Mathematical Software, 1981. **7**(3): p. 369-383.
16. Bunch, D.S., D.M. Gay, and R.E. Welsch, *Algorithm-717 Subroutines For Maximum-Likelihood And Quasi-Likelihood Estimation Of Parameters In Nonlinear-Regression Models.* Acm Transactions On Mathematical Software, 1993. **19**(1): p. 109-130.
17. Gay, D.M. and R.E. Welsch, *Maximum-Likelihood And Quasi-Likelihood For Nonlinear Exponential Family Regression-Models.* Journal Of The American Statistical Association, 1988. **83**(404): p. 990-998.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

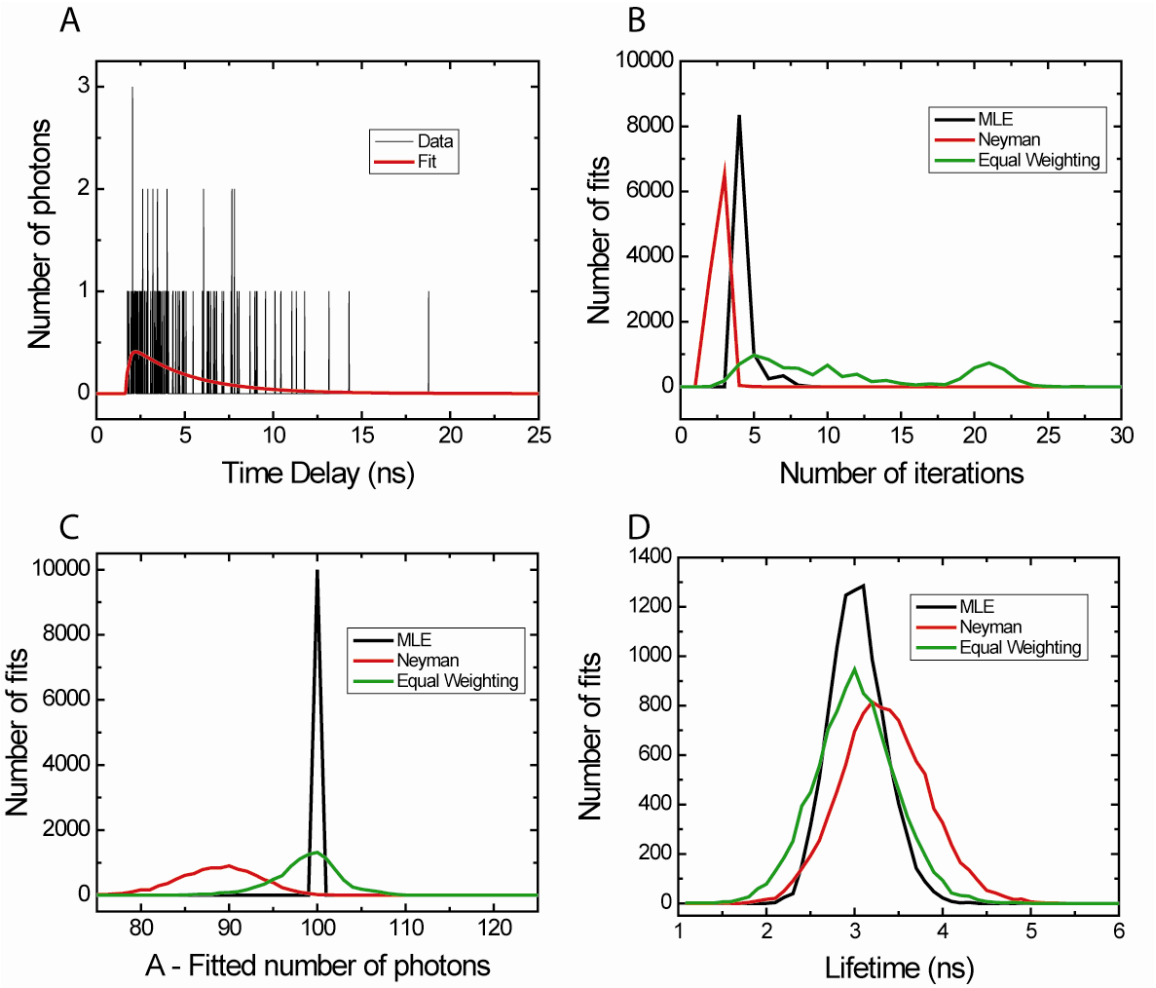

Figure 1: Fitting results for data simulated from a single exponential model with 3ns lifetime and exactly 100 photons per fit. **A**. Example data set (black) fitted using single exponential model (red). **B.** Histogram of the number of iterations required to reach the convergence criterion (a change in χ^2_{mle}, χ^2_{N} or χ^2_{l} of less than 10^{-6}) for the MLE, Neyman, and Equal Weighting figures of merit. **C**. Histogram of the fitted ampltidues A for each of the three methods. **D**. Histogram of fitted lifetimes τ.

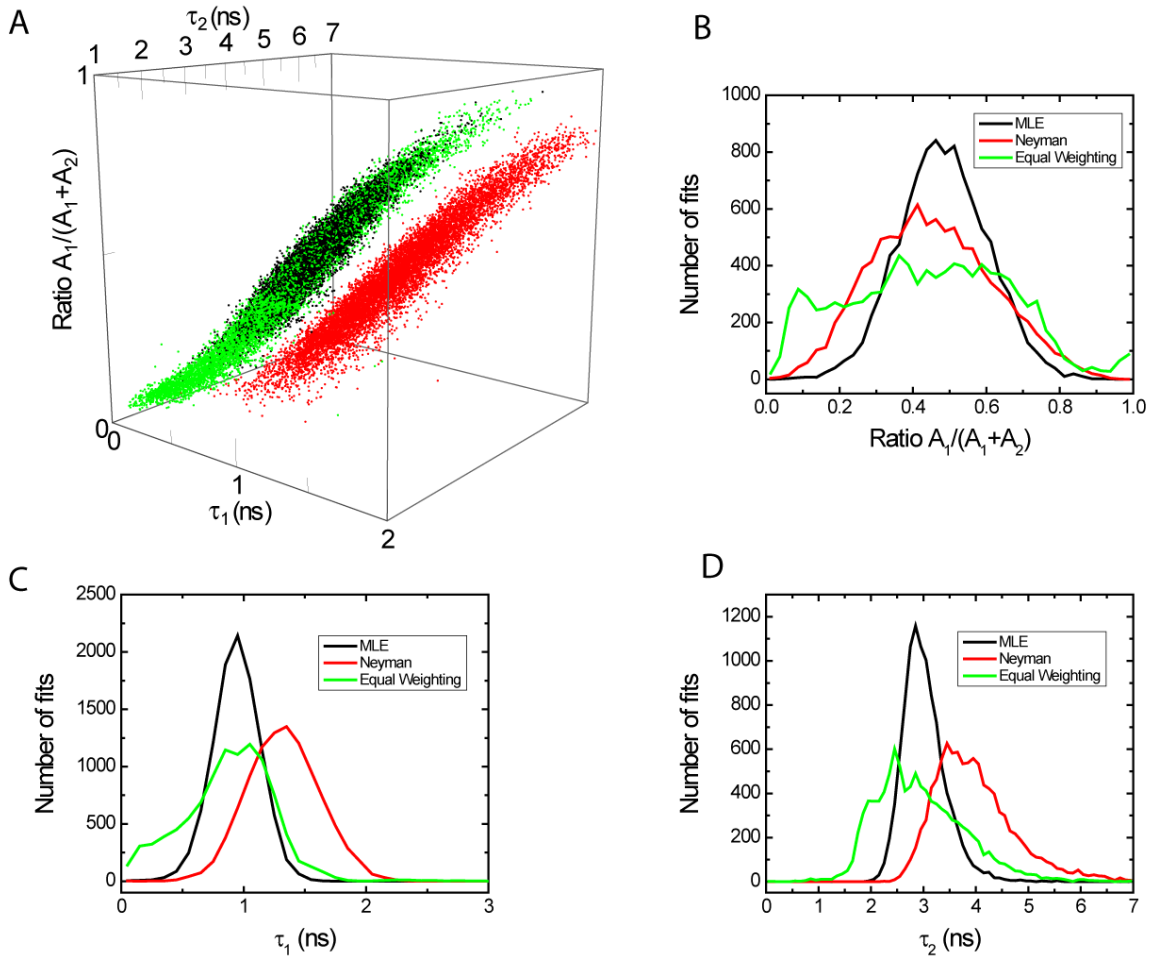

Figure 2: Fitting results for data simulated from a two exponential model with 1 ns and 3 ns components with equal weight, and exactly 1000 photons per fit. The average number of iterations for each method to converge was 4.4 for MLE, 6.5 for Neyman, and 30 for least squares with equal weighting. **A**. 3D scatter plot of fitted lifetimes τ_1 and τ_2, and the ratio of fitted amplitudes $A_1/(A_1+A_2)$ (ideally 0.5) for 10000 realizations of the data for χ^2_{mle}, χ^2_{N} and χ^2_{l}. **B**. Histograms of results from A over the ratio of amplitudes $A_1/(A_1+A_2)$. **C**. Histograms of fitted lifetime τ_1. **D**. Histograms of fitted lifetime τ_2.

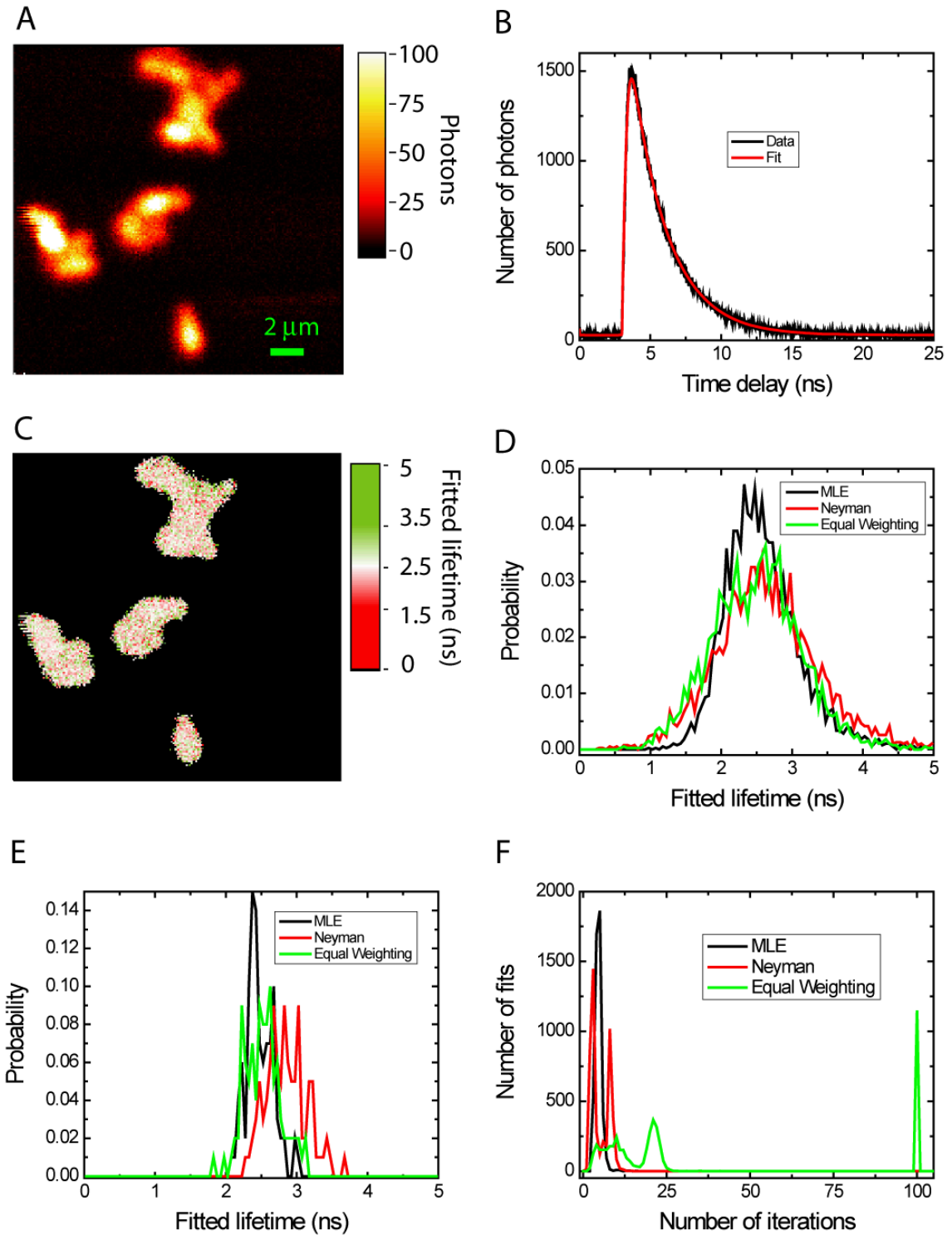

Figure 3: Lifetime imaging data for *Yersinia pestis* with constitutive expression of EGFP. **A**: Image of bacteria adhered to glass surface. **B**: Lifetime Data (all photons in image) and fit, including instrument response; fitted lifetime of 2.53 ns agrees with literature values for EGFP. **C**. Lifetime Map using the MLE procedure. Fits are shown only for pixels with 20 or more photons. **D**. Distribution of fitted

values for the different procedures for pixels with between 20 and 70 photons. **E**. Distribution of fitted values for the different procedures for pixels with between 150 and 300 photons. **F**. Histograms of the number of iterations required to obtain convergence.