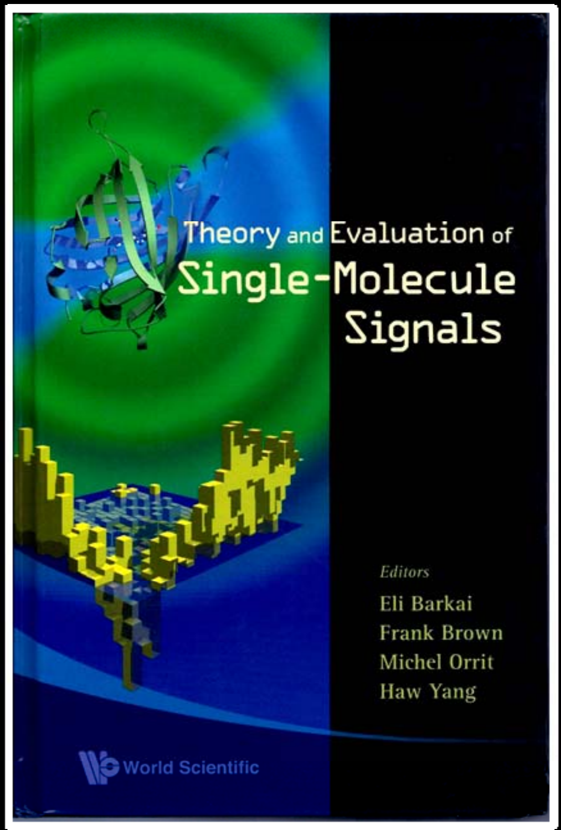


Topics

Model-Free Statistical Reduction of Single-Molecule Time Series

Testing Hypothesis with Single Molecules: Bayesian Approach

Generating Functions for Single-Molecule Statistics

Multipoint Correlation Functions for Photon Statistics in Single-Molecule Spectroscopy

Thermodynamics and Kinetics from Single-Molecule Force Spectroscopy

Theory of Photon Counting in Single-Molecule Spectroscopy

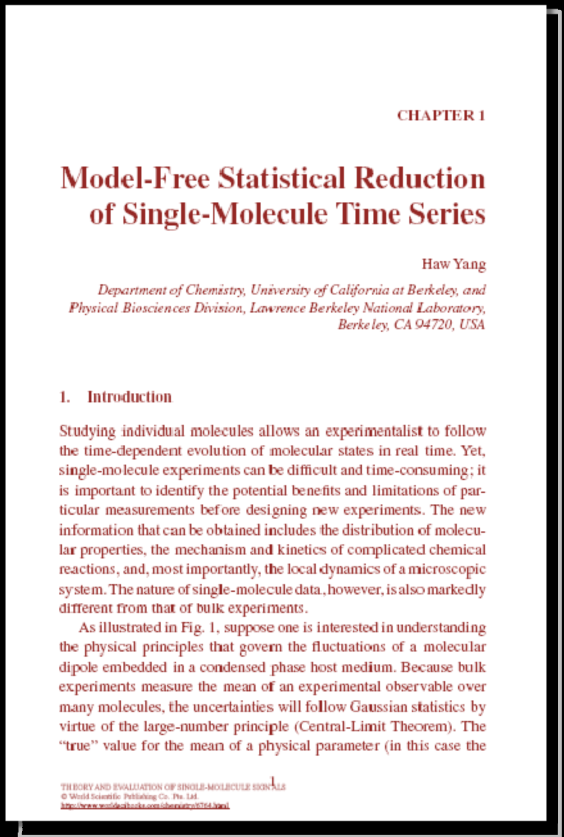
Memory Effects in Single-Molecule Time Series

Analysis of Experimental Observables and Oscillations in Single-Molecule Kinetic

Discrete Stochastic Models of Single-Molecule Motor Proteins Dynamics

Unique Mechanisms From Finite Two-State Trajectories

Weak Ergodicity Breaking in Single-Particle Dynamics



CHAPTER 1

Model-Free Statistical Reduction of Single-Molecule Time Series

Haw Yang

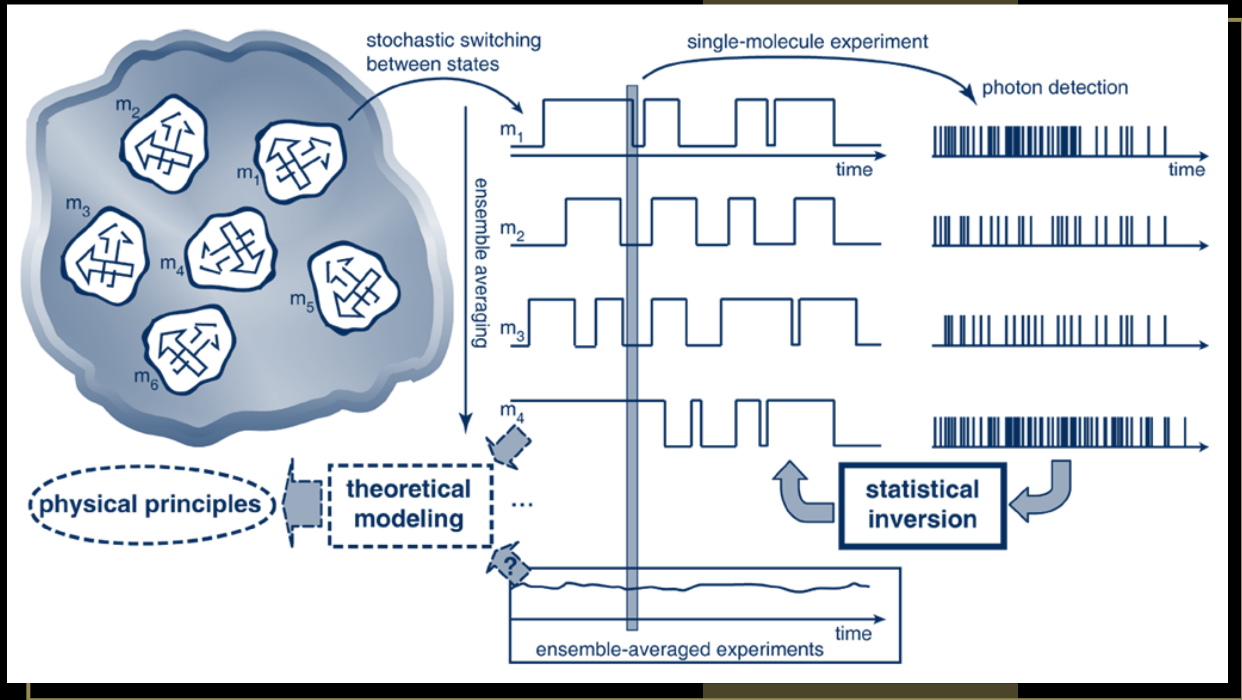
Department of Chemistry, University of California at Berkeley, and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

1. Introduction

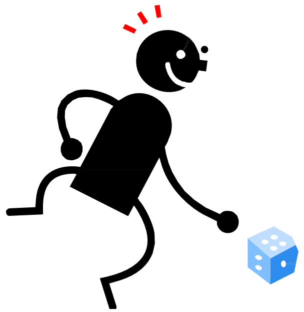
Studying individual molecules allows an experimentalist to follow the time-dependent evolution of molecular states in real time. Yet, single-molecule experiments can be difficult and time-consuming; it is important to identify the potential benefits and limitations of particular measurements before designing new experiments. The new information that can be obtained includes the distribution of molecular properties, the mechanism and kinetics of complicated chemical reactions, and, most importantly, the local dynamics of a microscopic system. The nature of single-molecule data, however, is also markedly different from that of bulk experiments.

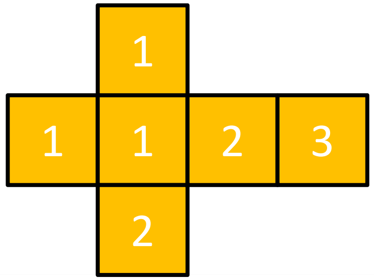
As illustrated in Fig. 1, suppose one is interested in understanding the physical principles that govern the fluctuations of a molecular dipole embedded in a condensed phase host medium. Because bulk experiments measure the mean of an experimental observable over many molecules, the uncertainties will follow Gaussian statistics by virtue of the large-number principle (Central-Limit Theorem). The "true" value for the mean of a physical parameter (in this case the

THEORY AND EVALUATION OF SINGLE-MOLECULE SIGNALS



Probability and Statistics





Probability

$$P(1) = \frac{\#\text{ of 1s}}{\text{total \# of possible outcomes}} = \frac{3}{6}$$

$$Mean = \frac{1+2+1+2+1+3+....}{\#\text{ of observations}} = 1\times P(1) + 2\times P(2) + 3\times P(3)$$

$$= 1\times\frac{3}{6} + 2\times\frac{2}{6} + 3\times\frac{1}{6} = \frac{4}{3}$$

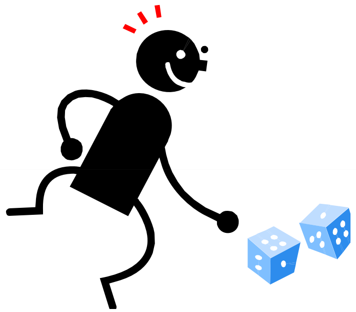
Mean (=expectation, average)

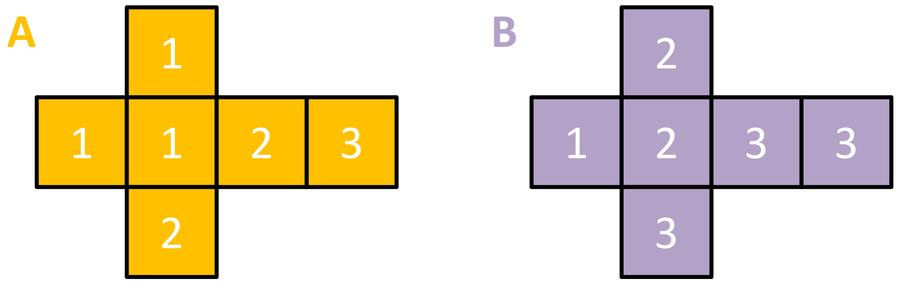
$$E[X] = \sum xP(x), \qquad \text{where X} = \{x_1, \ldots, x_N\}$$

Variance

$$Var(X) = E\left[(X - \langle X \rangle)^2\right] = E\left[X^2\right] - (E[X])^2$$

Probability and Statistics





Probability

$$P(1) = \frac{\text{\# of 1s}}{\text{total \# of possible outcomes}} = \frac{4}{12}$$

Conditional Probability

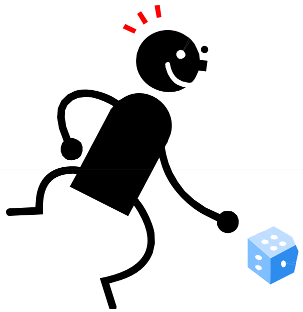
$$P(1 \mid B) = \frac{P(1 \cap B)}{P(B)} = \frac{1/12}{1/2} = \frac{1}{6}$$

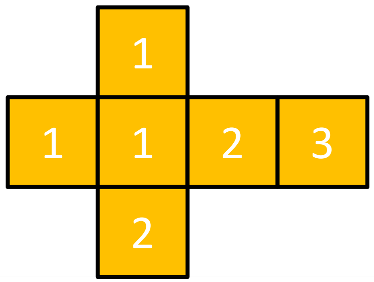
$$P(1 \mid A) = \frac{P(1 \cap A)}{P(A)} = \frac{3/12}{1/2} = \frac{1}{2}$$

Conditional Probability

$$P(X \mid Y) = \frac{P(X \cap Y)}{P(Y)}$$

Probability and Statistics





Series of observations…

Jack said he got ➔ {1, 3, 2, 2, 1, 2, 1, 1, 2, 3}

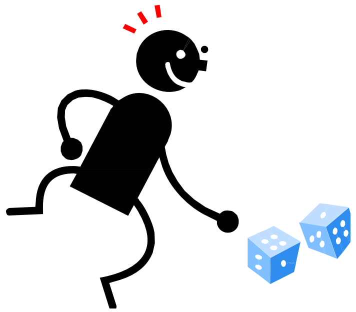
John said he got ➔ {3, 1, 2, 2, 3, 3, 2, 2, 3, 3}

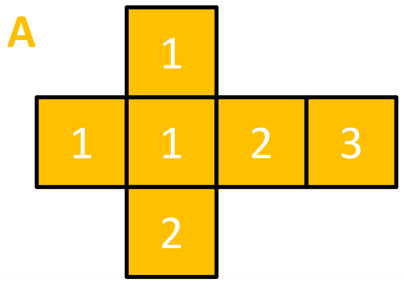
Likelihood function

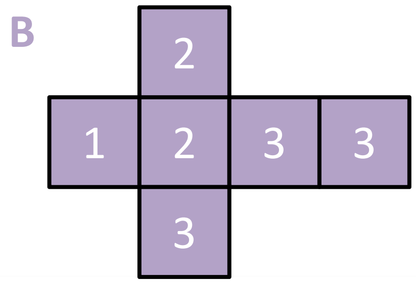
$$L_N(x_1,\ldots,x_N) = P(x_1)\times\cdots\times P(x_N)$$

$$= \prod_{i=1}^{N} P(x_i)$$

Probability and Statistics







Say, for the first k rolling, he used dice A. And he used dice B for remains.
Here's the result.

1, 3, 2, 2, 1, 2, 1, 1, 2, 3, 3, 2, 1, 3, 3, 2, 3,……

Can you tell when did he change the dice by looking at the observations?

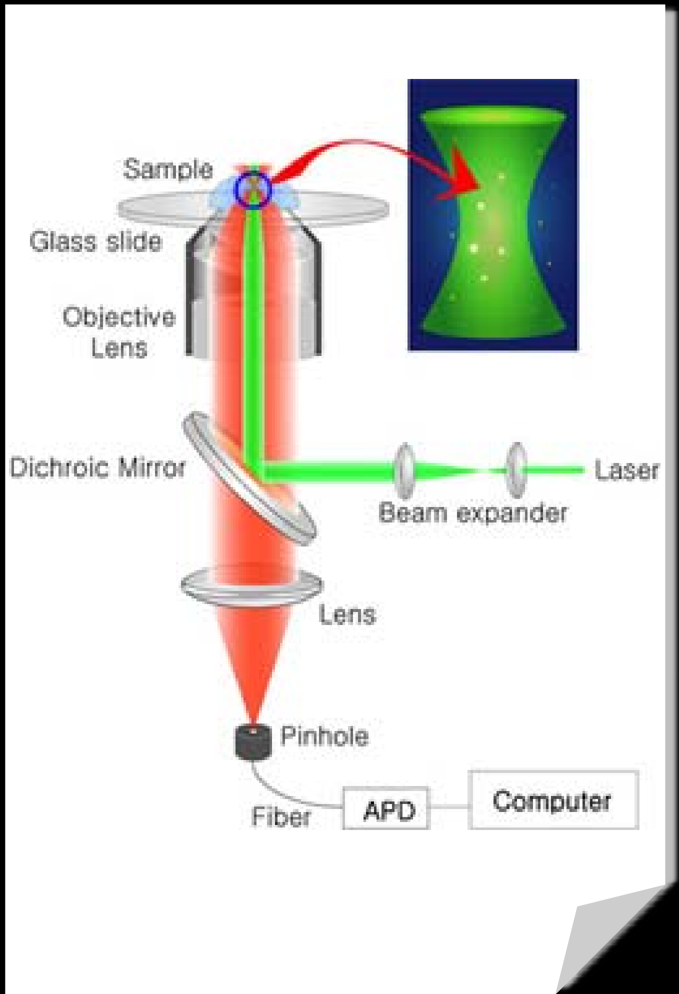
Likelihood function

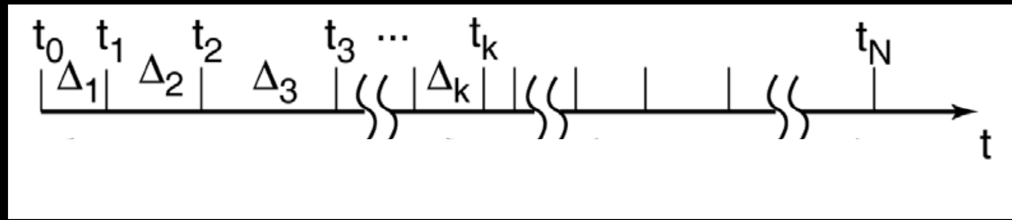
$$L_N\left(x_1,\ldots,x_N\right)=\prod_{i=1}^{N}P\left(x_i\right)$$

$$L_N=\prod_{i=1}^{k}P\left(\Delta_i\mid A\right)\times\prod_{i=k}^{N}P\left(\Delta_i\mid B\right)$$

Likelihood ratio test

$$\lambda\left(N\right)=\ln\left[\frac{L_N\left(P\left(x\mid \text{chagned the dice at k}\right)\right)}{L_N\left(P\left(x\mid \text{didnot changed the dice}\right)\right)}\right]>\lambda_C$$


Sample
Glass slide
Objective
Lens
Dichroic Mirror
Laser
Beam expander
Lens
Pinhole
Fiber
APD
Computer


t_0 t_1 t_2 t_3 ··· t_k t_N
Δ_1 Δ_2 Δ_3 Δ_k
t

The probability having Δ given I is (Poisson statistics)

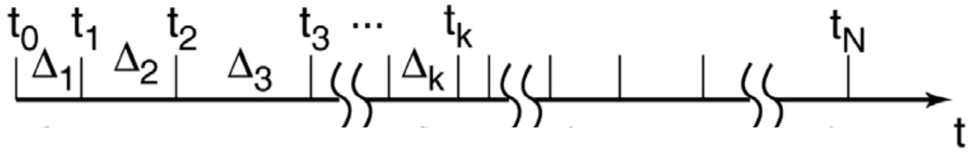
$$f\left(\Delta \mid I\right) = I \exp\left[-I \cdot \Delta\right]$$

And the number of photon detected with in a time interval, T, is

$$g\left(n \mid T, I\right) = \frac{\left(IT\right)^{n} \exp\left[-IT\right]}{n!}$$

To describe many detected photons, we need likelihood function

$$L_N\left(\Delta_1, \ldots, \Delta_N \mid I\right)$$
$$= f\left(\Delta_1 \mid I\right) \times \cdots \times f\left(\Delta_N \mid I\right)$$
$$= \prod_{i=1}^{N} f\left(\Delta_i \mid I\right)$$



The Question is

what is the most likely value of **I** that gives rise to the observed inter-photon duration sequence $\{\Delta_1, \ldots, \Delta_N\}$?

➔ Finding **I** that maximize L_N

$$\frac{\partial}{\partial I} \ln L = 0$$

Like wise, any physical parameters, θ, can be estimated by

$$\frac{\partial}{\partial \theta} \ln L = 0$$

The probability having Δ given I is (Poisson statistics)

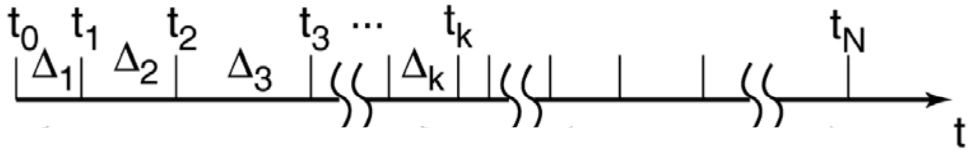
$$f(\Delta \mid I) = I \exp[-I \cdot \Delta]$$

And the number of photon detected with in a time interval, T, is

$$g(n \mid T, I) = \frac{(IT)^n \exp[-IT]}{n!}$$

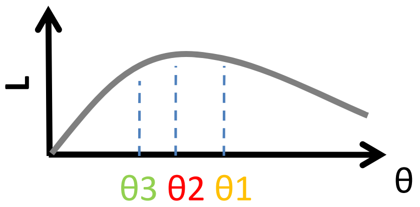
To describe many detected photons, we need likelihood function

$$L_N(\Delta_1, \ldots, \Delta_N \mid I)$$
$$= f(\Delta_1 \mid I) \times \cdots \times f(\Delta_N \mid I)$$
$$= \prod_{i=1}^{N} f(\Delta_i \mid I)$$



➔ Finding θ that maximize L_N

$$L_N = \prod_{i=1}^{N} f(\Delta_i \mid \theta)$$



In may cases, it is convenient to take a logarithm.
Then the Score function is defined by

$$S(X, \theta) \equiv \frac{\partial}{\partial \theta} \ln L \qquad \text{where, } X = \{\Delta_1, \ldots, \Delta_N\}$$

Note that $S(X,\theta_{ML})=0$

The probability having Δ given I is (Poisson statistics)

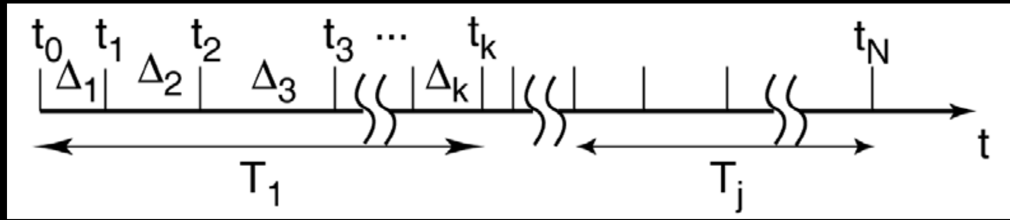
$$f\left(\Delta \mid I\right) = I \exp\left[-I \cdot \Delta\right]$$

And the number of photon detected with in a time interval, T, is

$$g\left(n \mid T, I\right) = \frac{\left(IT\right)^n \exp\left[-IT\right]}{n!}$$

To describe many detected photons, we need likelihood function

$$L_N\left(\Delta_1, \ldots, \Delta_N \mid I\right)$$
$$= f\left(\Delta_1 \mid I\right) \times \cdots \times f\left(\Delta_N \mid I\right)$$
$$= \prod_{i=1}^{N} f\left(\Delta_i \mid I\right)$$



Now, let's say there was **an intensity change at t_k** .

Then the probability having Δ for T_j is,

$$f\left(\Delta \mid I_j\right) = I_j \exp\left[-I_j \cdot \Delta\right]$$

Then, **likelihood function** is given by

$$L_N = \prod_{i=1}^{k} f\left(\Delta_i \mid I_1\right) \times \prod_{i=k}^{N} f\left(\Delta_i \mid I_2\right)$$

The problem is that
we don't know where k locates in an observed data.

The probability having Δ given I is (Poisson statistics)

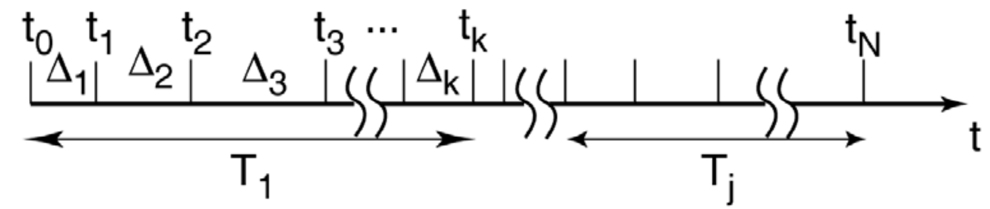
$$f(\Delta \mid I) = I \exp[-I \cdot \Delta]$$

And the number of photon detected with in a time interval, T, is

$$g(n \mid T, I) = \frac{(IT)^n \exp[-IT]}{n!}$$

To describe many detected photons, we need likelihood function

$$L_N(\Delta_1, \ldots, \Delta_N \mid I) = f(\Delta_1 \mid I) \times \cdots \times f(\Delta_N \mid I) = \prod_{i=1}^{N} f(\Delta_i \mid I)$$

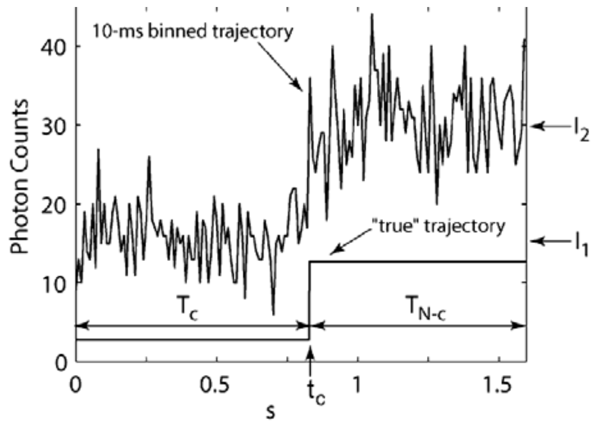


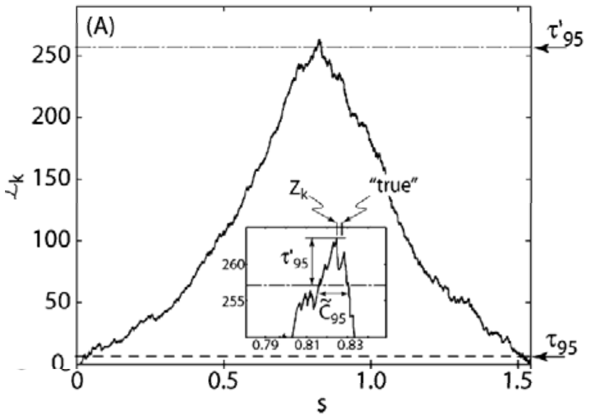
If you have two model,

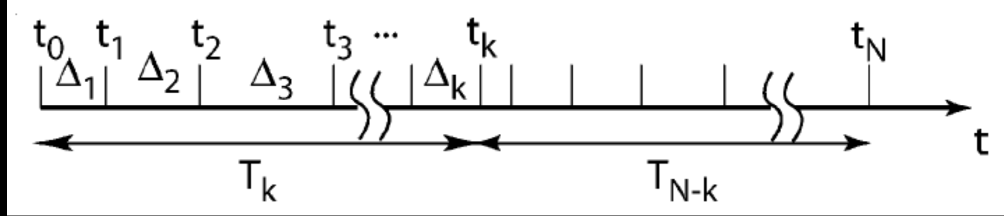
Statistical test
the likelihood ratio

$$\lambda(N) = \ln\left[\frac{L_N\left(f_{\text{model 1}}(\Delta \mid \theta_1)\right)}{L_N\left(f_{\text{model 2}}(\Delta \mid \theta_2)\right)}\right] > \lambda_C(\alpha, N)$$

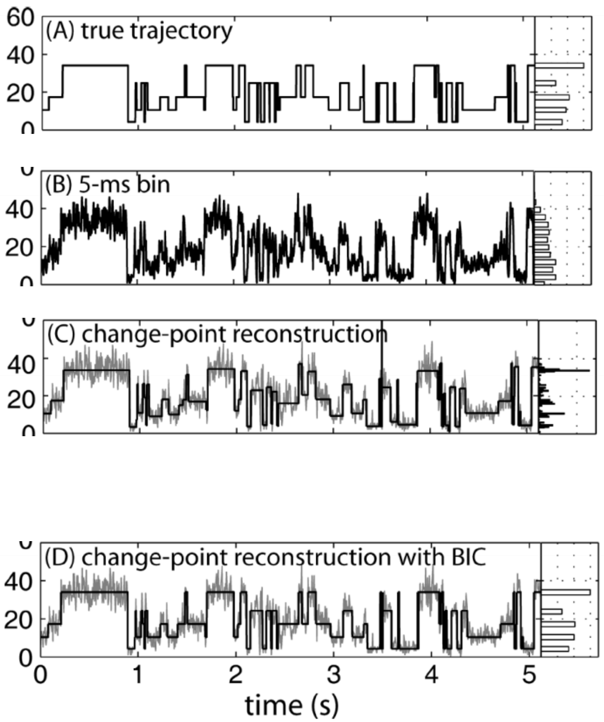
the critical value $\lambda_C(N,\alpha)$ with N observables and a confidence interval α.


10-ms binned trajectory
"true" trajectory
Photon Counts
40
30
20
10
0
0
0.5
1
1.5
s
t_c
T_c
T_{N-c}
I_2
I_1


(A)
τ'_{95}
τ_{95}
$\mathcal{L}_k$
250
200
150
100
50
0
0
0.5
1.0
1.5
s
Z_k
"true"
260
255
τ'_{95}
C_{95}
0.79
0.81
0.83

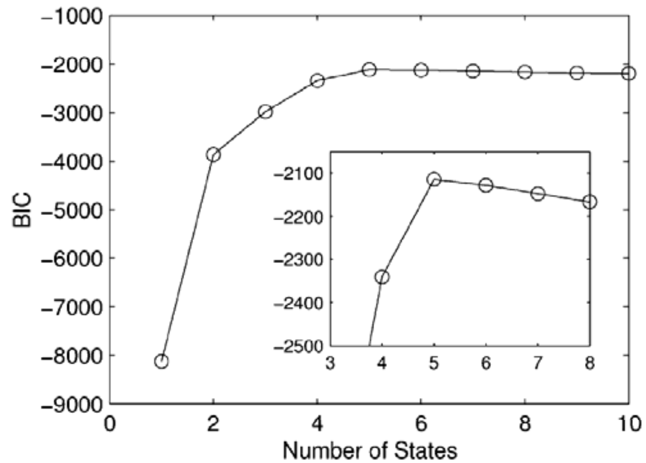

t_0 t_1 t_2 t_3 ... t_k t_N
Δ_1 Δ_2 Δ_3 Δ_k
t
T_k
T_{N-k}

To determine the # states



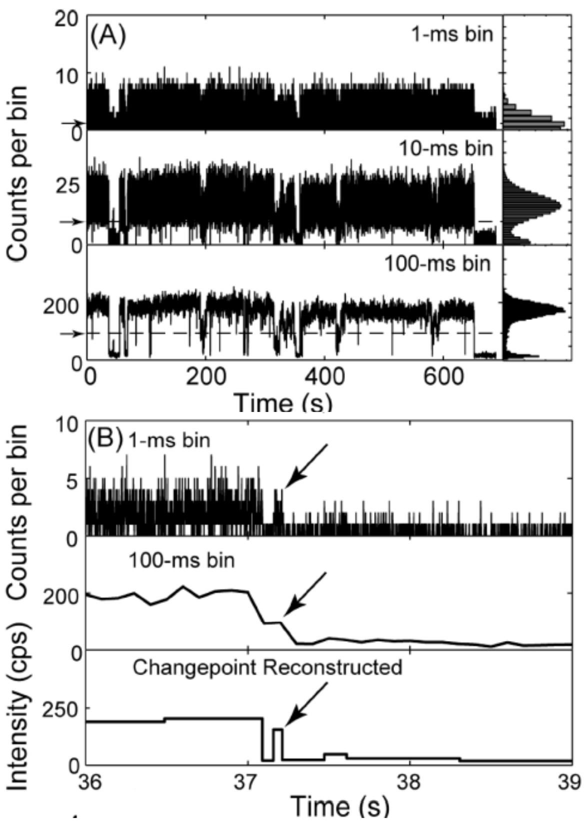
Bayesian Information Criterion

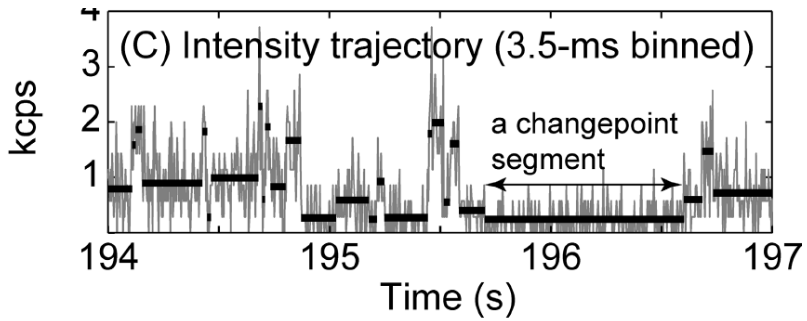
$$BIC = n \cdot \ln\left(\hat{\sigma}_{\varepsilon}^{2}\right) + k \cdot \ln(n)$$



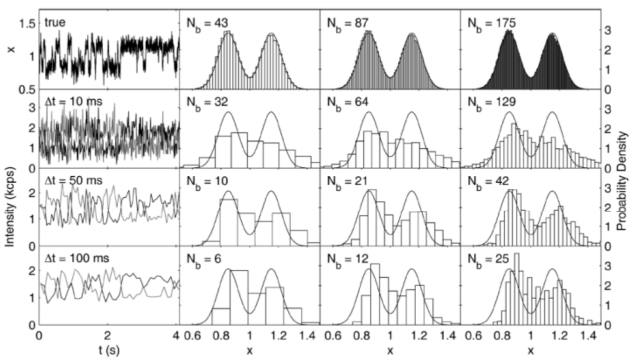
Application to the Real single-molecule data

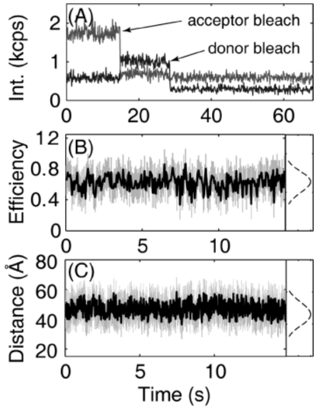
1. Quantum Dot



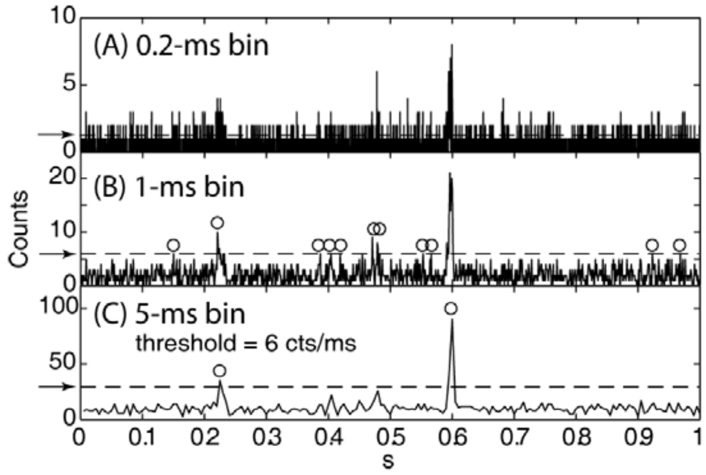


single molecule FRET





photon burst data



[redacted]

[redacted]

The probability having Δ given I is (Poisson statistics)

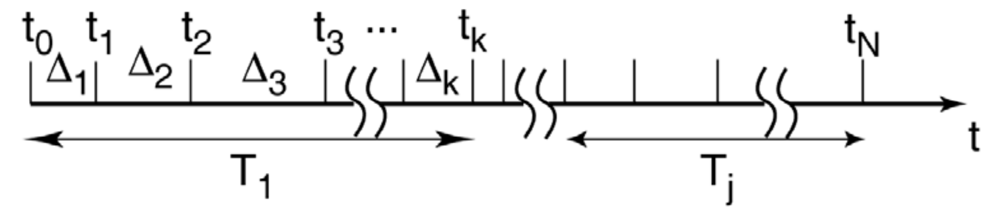
$$f\left(\Delta \mid I\right) = I \exp\left[-I \cdot \Delta\right]$$

And the number of photon detected with in a time interval, T, is

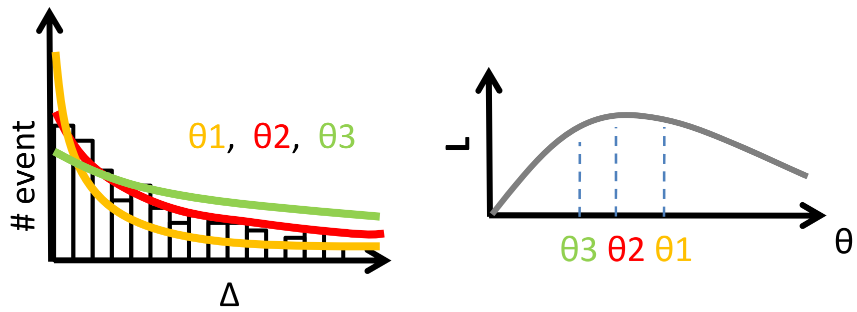
$$g\left(n \mid T, I\right) = \frac{\left(IT\right)^{n} \exp\left[-IT\right]}{n!}$$

To describe many detected photons, we need likelihood function

$$L_N\left(\Delta_1, \ldots, \Delta_N \mid I\right)$$
$$= f\left(\Delta_1 \mid I\right) \times \cdots \times f\left(\Delta_N \mid I\right)$$
$$= \prod_{i=1}^{N} f\left(\Delta_i \mid I\right)$$



➔ Finding θ that maximize L_N



$$\frac{\partial}{\partial \theta} \ln L \equiv S\left(\theta \mid X\right)$$

The probability having Δ given I is (Poisson statistics)

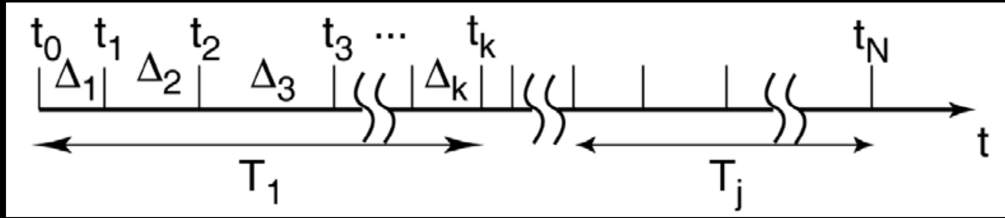
$$f\left(\Delta \mid I\right) = I \exp\left[-I \cdot \Delta\right]$$

And the number of photon detected with in a time interval, T, is

$$g\left(n \mid T, I\right) = \frac{\left(IT\right)^n \exp\left[-IT\right]}{n!}$$

To describe many detected photons, we need likelihood function

$$L_N\left(\Delta_1, \ldots, \Delta_N \mid I\right)$$
$$= f\left(\Delta_1 \mid I\right) \times \cdots \times f\left(\Delta_N \mid I\right)$$
$$= \prod_{i=1}^{N} f\left(\Delta_i \mid I\right)$$



The **Fisher information** gives the amount of information contained in a data set.

$$J\left(x\right) = \left\langle \left(\frac{\partial}{\partial x} \ln L_N\left(\left\{\Delta_1, \ldots, \Delta_N\right\} \mid x\right) \right)^2 \right\rangle_\Delta$$

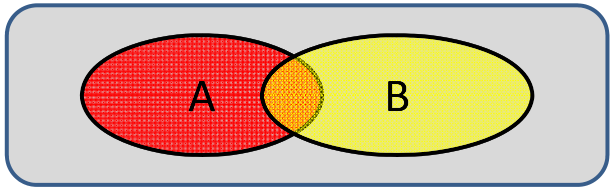
There is a relation between var(x) and J(x).

$$\operatorname{var}\left(x\right) \geq J\left(x\right)^{-1}$$

Where the equality is hold when x is calculated using MLE.

Bayes Law

$$p(B \mid A) = \frac{p(B) \times p(A \mid B)}{p(A)}$$



$$p(B \mid A) = \frac{p(A,B)}{p(A)} \qquad p(A \mid B) = \frac{p(A,B)}{p(B)}$$

$$p(A) \times p(B \mid A) = p(A,B) = p(B) \times p(A \mid B)$$

$$p(B \mid A) = \frac{p(B) \times p(A \mid B)}{p(A)}$$