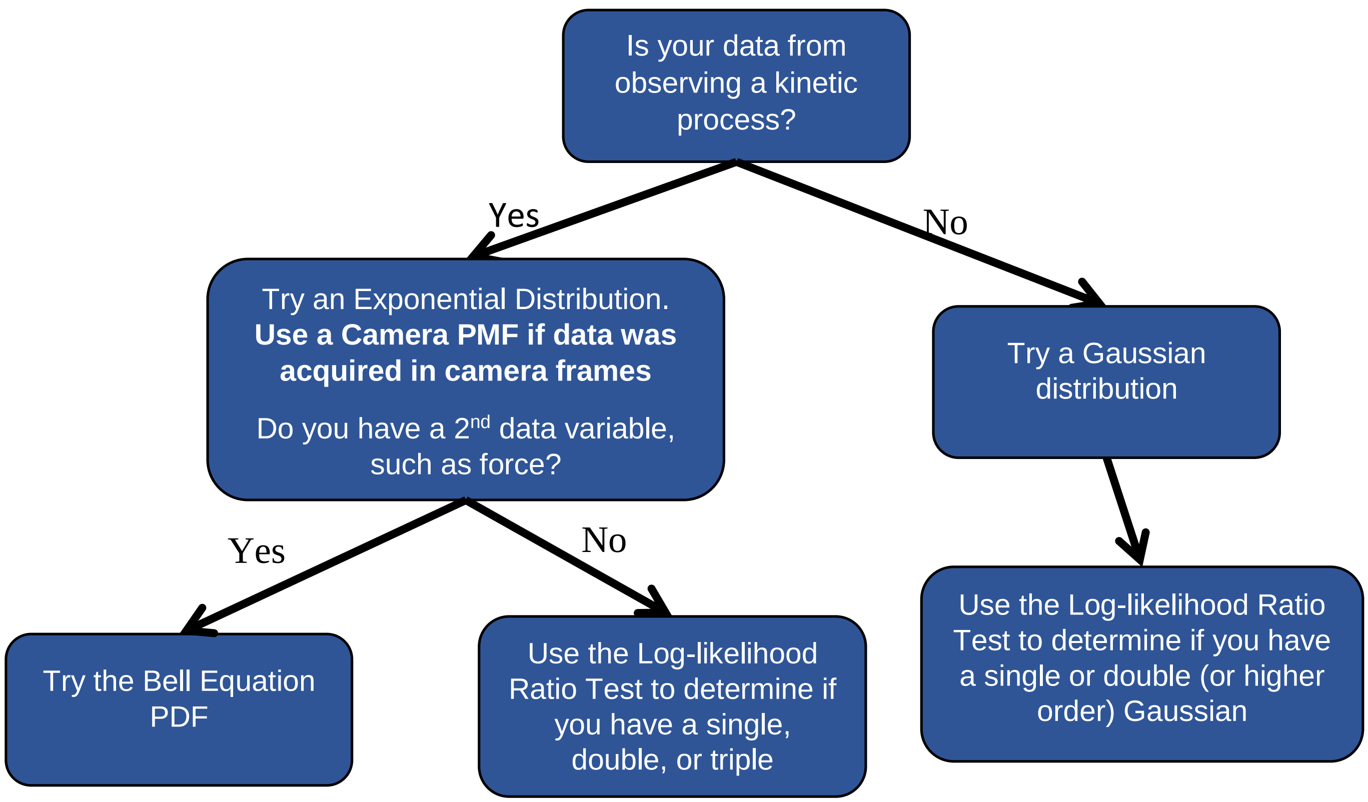
Contents

General Guidelines

To determine what kind of probability density function (PDF) you should use for your data, it is important to understand your data. Presented here are general guidelines to get you started, but since every experiment and data set is unique, it's important to not rely exclusively on these guidelines.

The flow chart below shows a general process for choosing the right model PDF. The first step is determining whether your data comes from studying the rate or lifetime of a kinetic process (including processivity data, which is related to the lifetime of attachment) or another type of data. Then you need to determine the number of components you suspect are part of your distribution. The log-likelihood ratio test of MEMLET can assist in this (see the MEMLET User's Guide for more information). If the data consists of multiple dependent variables (i.e. forces and durations) then MEMLET can fit a multidimensional model such as the Bell Equation. Multiple datasets from similar experiments can be simultaneously fit with a global model. See the User's Guide for further discussion.



A more detailed description of the probability distributions included with MEMLET is given below.

List of PDFs included in MEMLET

Gaussian

Single Gaussian
Double Gaussian

Exponential

Single Exponential
Double Exponential Independent
*Double Exponential Hidden Step **
*Gamma Function **
*Triple Exponential **
Bell Equation

Camera PDFs

*Single Exponential Camera **
*Double Exponential Independent Camera **
*Double Exponential Hidden Step Camera **

Others

*Poison **
*Chi-squared **
*Beta **

** PDFs marked with asterisks can be accessed by using the “Show PDFs” menu at the top of MEMLET next to “File”.*

General Types of Probability Distribution

Normal or Gaussian Distributions

The normal distribution (or Gaussian) distribution is widely applicable to various types of data. It can describe many types of data from single molecule experiments where experimental observations have a mean value and deviate from that value equally in the positive and negative direction. Some examples of data from single molecule experiments that often should be fit by Normal distributions include (but are not limited to):

- Fluorescence Intensities
- FRET values
- Step Sizes
- Velocities

There are two varieties of Normal distributions provided in MEMLET: Single Gaussian and Double Gaussian, but custom PDF's can easily be written with more components if desired. For information on choosing the appropriate type of normal distribution, see below.

Exponential Distributions

Exponential distributions are often useful when fitting kinetic processes in single molecule experiments. They often involve rates (often represented by the variable k) of a process, such a biochemical reaction. Combinations of multiple exponential distributions can be used for analyzing processes with multiple transitions or steps. Examples of data from single molecule experiments that often should be analyzed using exponential distributions include (but are not limited to).

- Motor Run Lengths/Processivity
- Dwell times
- Attachment Lifetimes
- Bleaching Times
- Time between mechanical steps

There are many varieties of distributions based on the exponential distribution included in MEMLET, including a set of discrete distributions that is designed to be used with data obtained with cameras or other methods where data is binned when acquired (See below for details). All the exponential distributions included with MEMLET are capable of taking into account a minimum detectable event time, which is important for achieving accurate results (as described in Woody et. al, Biophysical Journal, 2017)

Other Distributions

There are many other probability distributions that are compatible with MEMLET (Beta, Chi-squared, -log-normal, etc… nearly all continuous distributions) but are less useful for single molecule and biophysical experiments and can be accessed by the user providing their own functions.

One possible distribution of interest if the Poison distribution, which is most often applied when observing relatively rare events. This distribution can be used with data such as:

- Number of beads/particles interacting in a single molecule titration experiment
- Photon counting (when the number of photons is small)
- Number of Bleaching steps

Probability Distributions in MEMLET

Gaussian Distributions

Single Gaussian

The single Gaussian distribution is used when fitting a distribution which is believed to come from one population where a mean value and some standard deviation from that mean are expected. Examples of types of data that can be fit using a single Gaussian include:

- Single Motor Velocities
- Fluorescence intensity of a single, unchanging type of spot (not exhibiting fluctuation fluorescence values such as in dynamic FRET experiment)
- Step size distributions when only one step size is observed (i.e. no backstepping or sub-steps)

Double Gaussian

The double Gaussian distribution is used when fitting a distribution which is believed to come from two individual populations where each has mean value and some standard deviation from that. Examples of types of data that can be fit using a single Gaussian include:

- Velocities of motors where a fraction of the population of motors is regulated in some way that affects velocity (i.e. half of the motors are phosphorylated, causing increased velocity for that population)
- Fluorescence intensity of a multiple types of puncta or a changing spot (e.g. a dynamic FRET experiment)
- Step size distributions when two step sizes are observed (i.e. with back stepping or sub-steps)

Higher Order Gaussians

Higher order Gaussian functions (3 components, 4 components, etc.) can be generated by the user and used with MEMLET. [For a generalized multicomponent Gaussian PDF compatible with MEMLET's command line function, contact MEMLETInfo@gmail.com]. Such models can be difficult to justify statistically, since there are so many free variables, but are still relevant and applicable for some types of data.

Exponential Distributions

Single Exponential ("Single Exp")

A single exponential distribution most often can be used when studying a single-step kinetic process. This may also include data that observes the result of a competition between two pathways (e.g. processivity, which is a competition between taking another productive step and dissociating). This is the simplest type of distribution for fitting kinetics. Example types of data:

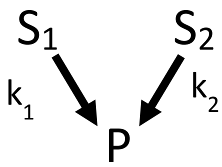
- Molecular attachment durations (from a single step detachment process)
- Dwell times
- Processivity/Run lengths
- Bleaching times

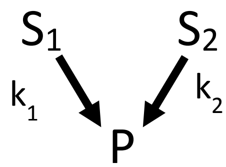
Double Exponential

When a kinetic process consists of two steps a double exponential function may be appropriate. There are several different forms depending on the order and characteristics of the process:

Independent Steps ("Double Exp")

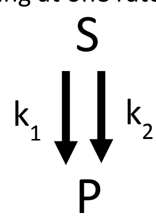
When the variable being observed is the time until some state, P, appears, and P can be formed via two independent paths from S_1 and S_2, as shown below, then the default double exponential function included in MEMLET can be used. An example might be if precursors are being transformed into the same observable product at different rates, such as when two enzymes (at arbitrary concentrations) with different kinetics produce ATP (the observed product). This function gives both rates, as well as A, the relative proportion of the events which transitioned to P at a rate of k1 (1-A gives the proportion with rate k2). In the example, this would give the relative portion of ATP that was formed by each enzyme (which may vary based on each enzyme's concentration, substrate affinity, etc).





Parallel Steps

When the kinetic process involves two parallel path with different rates, where the flux through each pathway is solely determined by the relative proportion of the rates, the parallel path PDF may be useful. This is a special case of the Independent path PDF, where the amplitude, A, is not a fitted variable, because the amplitude only depends on the two rates. This occurs when the choice of pathway a particular transition follows is stochastically determined and does not depend on any other influences. This is usually the case when there are two possible spontaneous or unassisted transitions. An example of this type of data may the dwell time of a molecular motor which can detach from its track (while in a single biochemical state) via two pathways (e.g. ATP binding causes unbinding at one rate, and there is a basal rate of ATP-independent unbinding).



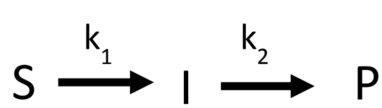
As mentioned, the parallel steps function is a special case of the independent steps pathway with A= k1/(k1+k2)

Sequential Hidden Steps (Hidden Steps)

When two kinetics processes occur in sequence, but the first step is not observed, a double exponential function is also used, but the shape of the distribution is often biphasic. Depending on if the two steps occur at the same or different rates, different forms should be used.

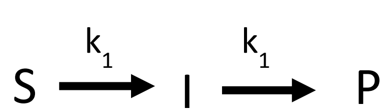
Two phase Hypo-exponential (Hidden Step with unequal rates)

When a processes consists of two steps which occur sequentially at the different rates (that are relatively similar), the "double exponential hidden step" model should be used (also called a hypo-exponential model). This can be used when a processes has to proceed through some unobservable intermediate, such as when acto-myosin has to release ADP (hidden step) and then bind ATP before quickly dissociating (observed step). If the two steps are very different in rate, then the data may appear to fit a single exponential.



Gamma (equal rates)

When a processes two steps which occur sequentially at the same rate, the gamma function is used, which is a special case of the double exponential hidden step. An example of when this can be used is when observing the dwell time of a two-headed motor stepping with alternating heads when only one of the heads is labeled. In such cases, since only one head is labeled, the stepping of the unlabeled head is a hidden step that occurs at the same rate of the labeled head.



Triple Exponentials and higher

If even more kinetic steps are involved, higher order exponential models can be used, and the log-likelihood ratio test can be used to determine if they are statistically justified. The various types of exponentials (sequential, parallel, etc) can be combined in various ways to perform higher order exponentials. Interested users are encouraged to explore the literature or can email MEMLETInfo@gmail.com with questions.

Bell Equation

The Bell Equation is often used when the rate being observed is dependent on a force being applied to the molecule. If the data consists of individual measurements of durations or dwell times and the respective forces, the Bell Equation can be used. An example of this may be attachment durations of a motor in an optical trap experiment where the force applied varies stochastically based on the position of initial binding of a motor. In the Bell Equation model, the rate depends exponentially on force.

Exponential Functions for Camera-based data

When a camera is used to acquire duration data (bleaching experiments, or single FRET experiments using a camera to measure fluorescence), then the durations are necessarily binned based on the time between camera frames. This can be accounted for with specialized discrete PDFs, labeled Camera in the MEMLET program. There are three varieties, corresponding to the models described above.

SingleExp_Camera

DoubleExp_Camera (parallel paths)

DoubleExpHiddenStep_Camera (hidden step)

To access these PDFs, use the Show PDFs menu and check, Show Camera PDFs. For More information about the need for and theory of the camera distributions, see Lewis et. al, Biophysical Journal, 2017.

Other PDFs

The following PDFs are also included in the MEMLET Extra PDFs folder. Users are encouraged to explore more to see if and when they are necessary for their data. Also, arbitrary PDFs maybe entered into MEMLET

Poison

Useful for very rare events such as photon counting or single molecule titrations

Chi-Squared

A variation on the normal distribution where measurements represent the sum of squares of normally distributed variables.

Beta

Useful for variables that are constrained to intervals of finite length. Often used as Priors in Bayesian inference since Beta distributions are often constrained between 0 and 1 and thus can be used for describing probabilities of an occurrence.