Fitting Force-dependent Kinetic Models to Single-Molecule Dwell Time Records

Patrick M. Kain, *Optical Sciences & Engineering,* and Steven Haim, *Optical Sciences & Engineering*

I. Introduction

DEVELOPMENTS in the micro-manipulation field, such as optical tweezers, has made it possible to deliver picoNewton forces and nanometer displacements to microscopic molecules such as DNA. These experiments have enable scientists to study biomolecular machines such as cargo-transporting motor proteins kinesin and myosin-V. Essentially, these experiments specifically study how dwell time varies when assisting or hindering loads are applied. Figure 1 shows the experimental setup.

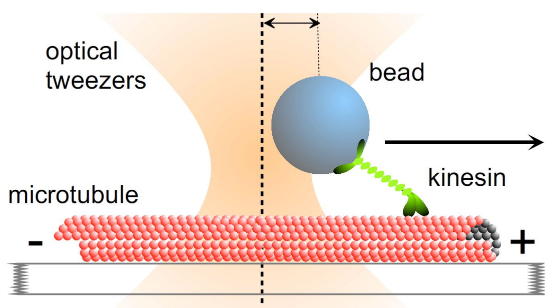

Fig. 1. The kinesin is seen attaching to a bead trapped in the optical tweezers. The dwell time can be observed as the kinesin moves along the microtubule.

The dwell time can be modeled using transition state theory which gives the following equation for dwell time:

$$t_i = \frac{1}{\tau_0} e^{F_i \delta / kT} \tag{1}$$

The nature of optical tweezers does not allow a user to produce a discrete predetermined force; instead, a randomized force in a range of forces can be produced. The current method simply takes the mean of the many measurements in a given interval. Because the dwell time varies nonlinearly with force, simply taking the mean of a given data set will give inaccurate results. The inaccuracy is due to the fact that - for a given range of forces - the dwell time for larger forces will be much longer than the mean force dwell time. Conversely, dwell time will be close to the mean force dwell time for smaller forces. This causes the mean data to be biased towards a longer dwell time.

II. Maximum Likelihood Method

Instead of taking the mean of a given range of data, a method which fits all the data to the known theoretical model through the use of maximum likelihood estimators will provide a more accurate result. The log-likelihood function for this experiment is shown in equation 2 below (with detailed calculations shown in appendix B). Note that each measurement has a interdependently measured dwell time (t) and force (F).

$$\log L(\tau_0, \delta | t, F) = -n \log \tau_0 + \frac{\delta}{kT} \sum_{i=1}^{n} F_i - \sum_{i=1}^{n} t_i e^{F_i \delta / kT} \tag{2}$$

Taking the derivative with respect to the δ and τ_0 parameters and setting this result equal to zero gives the following equations:

$$\frac{\partial}{\partial \delta} \log L(\tau_0, \delta | t, F) = \frac{1}{\tau_0} - \frac{1}{N \tau_0^2} \sum_{i=1}^{n} t_i e^{F_i \delta / kT} = 0 \tag{3}$$

$$\frac{\partial}{\partial \tau_0} \log L(\tau_0, \delta | t, F) = \bar{F} - \frac{1}{N \tau_0} \sum_{i=1}^{n} t_i F_i e^{F_i \delta / kT} = 0 \tag{4}$$

Generally solving for δ in equation 3 and τ_0 in equation 4 would yield the maximum likelihood parameters; however, these equations do not solve for δ very easily. Instead, solving for τ_0 in each equation (shown in equations 5 and 6) while then determining for what value δ the two equations intersect will yield the desired parameters. Section III below will use this method to determine the two distributions of τ_0 as well as the final estimator for δ.

$$\hat{\tau_0} = \frac{1}{N} \sum_{i=1}^{n} t_i e^{F_i \hat{\delta} / kT} \tag{5}$$

$$\hat{\tau_0} = \frac{1}{N \bar{F}} \sum_{i=1}^{n} F_i t_i e^{F_i \hat{\delta} / kT} \tag{6}$$

III. Using R to Determine Dwell Time

This section was focused on using R to simulate the distributions for the two values of $\hat{\tau_0}$ defined in Equations 3 and 4. The final goal is to identify the intersection of the two functions (when plotted as a function of $\hat{\delta}$) to find the estimated value of $\hat{\delta}$.

It is important to note that the magnitudes of the various physical parameters (such as the force, or Boltzmann's constant) were disregarded for this paper just to present the feasibility that the data could be estimated.

First, a time vector was created as a random exponential vector (in accordance with the relationships). Equation 3 could then be estimated in R by a FOR loop iterating the mean as a function of the random exponential time:

```
> for(j in 1:N){t[j]<-rexp(1,exp(F[j]))}
> for (j in 1:N){tau1[j]<-mean(t*exp(delta[j]*F))}
```

Equation 4 could then be estimated by creating a similar loop taking into account the inverse relationship to the mean of the applied force.

```
> for (j in 1:N){tau2[j]<-mean(t*F*exp(delta[j]+
*F))/mean(F)}
```

We now have established an estimator for both values of τ_0 as a function of $\hat{\delta}$. These two relationships were plotted in Figure 2 in order to highlight the two relationships. The intersection is labeled with a vertical dashed line. This location was determined to be $\hat{\delta} = 1.028829$.

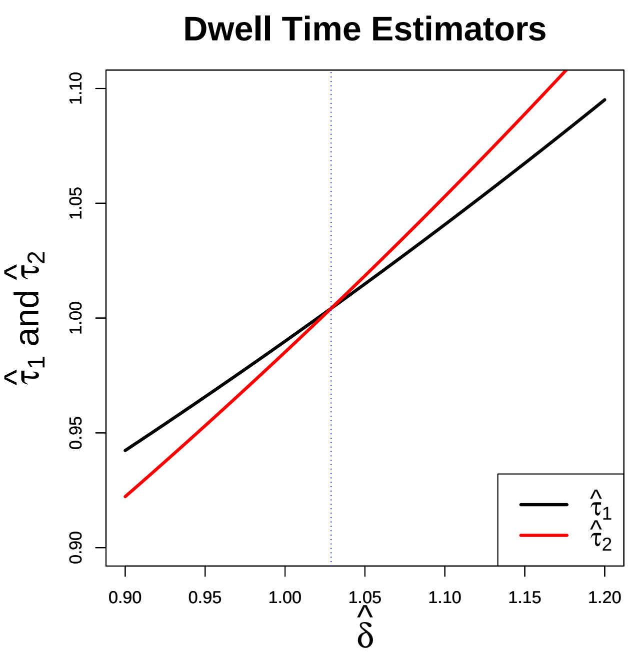

Fig. 2. The two estimators for $\hat{\tau_0}$ plotted as a function of $\hat{\delta}$. The intersection is shown by the vertical dashed blue line.

The location of the intersection was found by taking the iterative (paired) difference between the two vectors for $\hat{\tau_0}$ and finding the value of minimum. By finding the *index* location of this minimum difference, this corresponds to the same index location in the $\hat{\delta}$ vector which will be the x location of the intersection:

```
> for (j in 1:N){diff[j]=abs(tau1[j]-tau2[j])}
> inter = min(diff)
> index = which(diff==inter)
```

It is important to note that is this just one iteration of one value for the estimator. In reality, it would be ideal to simulate and find a number of values for the estimator and take the mean of this result. This was performed in MATLAB once the data was produced in R since working in R is not conducive to correct and useful *nested* FOR loops. One-hundred thirty values for the estimator were found with a mean of 1.0282 and a standard deviation of 0.0707. A histogram of this data compared next to the ideal normal distribution is shown in Figure 3.

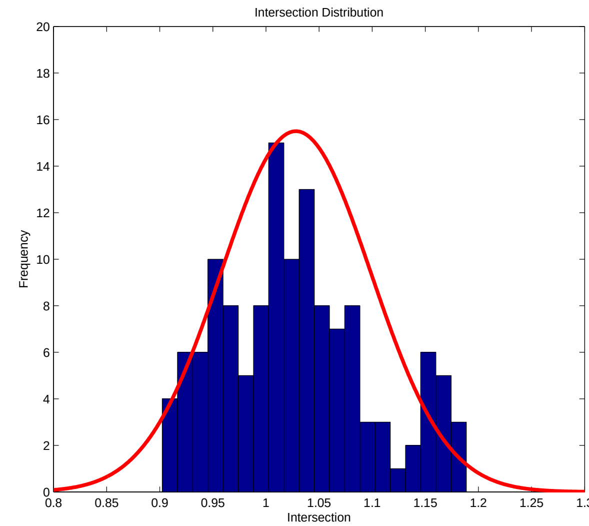

Fig. 3. The histogram of 130 values of $\hat{\delta}$. The red curve highlights the ideal Gaussian distribution.

IV. Conclusion

Using the two relationships for the estimators for $\hat{\tau_0}$, we could successfully find a value for the estimator $\hat{\delta}$ due to the intersection. Once these two estimators are found equation 1 can be used to determine dwell time. By iterating this process a number of times, we could find a result for the mean of the values of the estimators such that $\bar{\hat{\delta}} = 1.0282$ and $\sigma_{\hat{\delta}} = 0.0707$.

The next step in this process would be to provide more iterations (samples) and hopefully complete the project in one software package as well as provide real values for other physical characteristics so that the *magnitude* of the displacement estimator is useful.

References including equation derivations, software code, and variable definitions follow this section.

Appendix A
Variable Definitions

The section below defines the variables used in this paper.

1) τ_0 is the characteristic disassociation time with no force
2) δ is the displacement from equilibrium.
3) F_i is the force associated with the i^{th} observation.
4) t_i is the dwell time associated with the i^{th} observation.
5) k is Boltzmann's constant.
6) T is the absolute temperature

Appendix B
Likelihood Function

The density function is derived from the transition state model where the dwell time varies exponentially as a function of force.

$$f(t|\tau_0, \delta, F) = \tau_0^{-1} e^{F\delta/kT} \exp\left(-\frac{t}{\tau_0} e^{F\delta/kT}\right)$$

To find the likelihood function, the following operator is applied:

$$L(\tau_0, \delta | t, F) = \prod_{i=1}^{n} \tau_0^{-1} e^{F_i \delta / kT} \exp\left(-\frac{t}{\tau_0} e^{F \delta / kT}\right)$$

Taking the log of both sides yields Equation 2

$$\log L(\tau_0, \delta | t, F) = -n \log \tau_0 + \frac{\delta}{kT} \sum_{i=1}^{n} F_i - \sum_{i=1}^{n} t_i e^{F_i \delta / kT}$$

Appendix C
R Code

```
# Clear all
rm(list=ls(all=TRUE))

# Constants
N = 1000
F<- runif(N)
t<- rep(0,N)
tau1 = rep(0,N)
tau2 = rep(0,N)
diff = rep(0,N)
for(j in 1:N){t[j]<-rexp(1,exp(F[j]))}

# Define tau1 and tau2
delta=seq.int(0.9,1.2,length.out=1000)
for (j in 1:N){tau1[j]<-mean(t*exp(delta[j] +
*F))}
for (j in 1:N){tau2[j]<-mean(t*F*exp( +
delta[j]*F))/mean(F)}
for (j in 1:N){diff[j]=abs(tau1[j]-tau2[j])}
inter = min(diff)
index = which(diff==inter)

# Plot the results
par(mar=c(5, 5, 4, 3) + 0.1)
plot(delta,tau1,xlim=c(0.9,1.2),ylim=+
 c(0.9,1.1),type="l",xlab="", +
ylab="",lty=1,lwd=3)
par(new=TRUE)
xlab.name = expression(hat(delta))
ylab.name = expression(paste(hat(tau)[1] +
," and ",hat(tau)[2]))
plot(delta,tau2,xlim=c(0.9,1.2),ylim= +
 c(0.9,1.1),type="l",xlab=xlab.name, +
ylab=ylab.name,col="RED",lty=1,lwd=3, +
cex.lab=2)
title(main="Dwell Time Estimators", +
cex.main=2)
legend("bottomright",c(expression(hat(tau)[1] +
),expression(hat(tau)[2])),col= +
 c("BLACK","RED"),lty=1,lwd=3,cex=1.5)

# Add vertical line
par(new=TRUE)
abline(v=delta[index],lty=3,col="BLUE")
```

Appendix D
MATLAB Code

```
hist(x,20)
hold on
axis([0.8 1.3 0 20])
Mean = mean(x)
Standard_Deviation = std(x)
title('Intersection Distribution')
xlabel('Intersection')
ylabel('Frequency')
x1 = linspace(0.80,1.3,1000);
gauss = 15.5*exp(-(x1-Mean).^2/.01);
p = plot(x1,gauss)
set(p,'Color','red','LineWidth',3)
```

Acknowledgment

The authors would like to thank Dr. J. Watkins for his guidance and teaching during the Spring, 2011 semester at the University of Arizona.

References

[1] *Statistics Case Studies: Single Molecule Dwell Times*. MATH 363, 2009.