Statistical Analysis of SST-Pyr Post-synaptic Distribution

Taisuke Yasuda

July 26, 2017

Contents

1 Framing the Data with a Mathematical Model

As in a typical statistical study, we view the observed data $\left(x^{(i)}\right)_{i=1}^{n}$, where the ith data point (the observed amplitude in mV) is denoted by $x^{(i)}$, as a sequence of observations of a random variable X. The goal of our study is to use this observed data to say something about the behavior of the N contact points, whose individual response amplitudes sum to produce the measured data X. Thus, we may introduce new random variables X_j to model these quantities which are summed and write this knowledge (assumption?) as

$$X = \sum_{j=1}^{N} X_j.$$

Note that in most cases, N is an unknown parameter except for the rare cases in which we have a reconstruction of the cell to provide us with a direct estimate of the true value of N.

1.1 Physical Interpretation of X_j

When we write X as a sum of N random variables $X_1, \ldots, X_N$, we must be careful in the interpretation of the quantities that each of the X_j represent. Although we have implied that we attribute the value of X_j to the jth contact point, we are in fact modeling much more than just the stochastic value produced when the presynaptic signal reaches the jth contact point – if this were the case, then we would have to worry about modeling many other transformations that are applied to that value between the time that it is produced and the time that it gets measured at the cell body. This includes any measurement noise, the decay of the electrical signal as it travels along the dendrite, etc. Instead, we view the quantity X_j as simply the value that is measured at the cell body that can be attributed to the jth contact point. This means that at this point, we are not making any assumptions what happens at all at the jth contact point. The only information that this decomposition of X encodes is that it gets summed at the cell body with the rest of the values attributed to the other contact points.

1.2 Introducing the Release Success Random Variable

At this point, the modeling we have down so far does not take advantage of all the biological knowledge we have – we know that each contact point sometimes fails to fire completely, and that it is known to be reasonable to model this failure as happening with some constant probability. Thus, we introduce another random variable Y_j at each of the N contact points that represents whether there was a release success or not. This corresponds to a Bernoulli variable that takes on the value $Y_j = 1$ (release success) with probability p_j, a parameter of the model, and the value $Y_j = 0$ (release failure) with probability $1 - p_j$. Using this, we model each contact point as producing a value of $X_j = 0$ whenever there is no release, i.e. whenever $Y_j = 0$.

1.3 Specifying the Distribution of $X_j \mid Y_j = 1$

The above states that $(X_j \mid Y_j = 0) \sim 0$, i.e. that given that there is no release success, X_j takes on the value 0 with probability 1. Thus, in order to completely specify the distribution for X_j and thus X, it remains to specify the distribution of $X_j \mid Y_j = 1$, the behavior of X_j when there is a release success. Although this distribution may in general be extremely complicated,

and may encode diverse information as multivesicular release, multimodal behavior, etc. However, we opt to keep the model simple for now and choose a convenient distribution to represent this quantity. We wish to model a continuous and positive quantity, so some simple choices include the exponential distribution, lognormal distribution, and uniform distributions. We work with the lognormal distribution for now.

2   Case Study: 17march2016g

We work with some concrete data to see this model does.

2.1 $N = 1$

We in fact know from cell reconstructions that this cell has 5 contact points. Nonetheless, we first study the data under the assumption that $N = 1$ for a simple example.

2.1.1   A Specialized Approach for $N = 1$

In the special case of $N = 1$, we know that all the nonzero data points can be attributed to the distribution $(X_1 \mid Y_1 = 1)$, so that all the nonzero data points are drawn from some lognormal distribution with parameters μ_1 and σ_1^2. Thus, we may just take the log of the data and estimate these parameters by taking the sample mean and variance of the log data.

2.1.2   A Generalized Approach

In the general case of arbitrary N, we use a variant of method of moments to estimate the parameters. Easily, the mean is given by Npq where q is the mean response at each contact, and the variance is given by Nps^2 where s^2 is the variance of the response at each contact. Then, since we assume a lognormal distribution with parameters μ and σ^2, we have that

$$q = \frac{\overline{X}}{Np} = \exp\left(\mu + \sigma^2/2\right), \qquad s^2 = \frac{S^2}{Np} = \left(\exp(\sigma^2) - 1\right)\exp\left(2\mu + \sigma^2\right).$$

This actually leads to a system of linear equations for μ and σ^2, which turns out to be

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\log\left(\frac{S^2}{Np} + \left(\frac{\overline{X}}{Np}\right)^2\right) \\ 2\log\left(\frac{\overline{X}}{Np}\right) \end{pmatrix}.$$