Research Notebook

Taisuke Yasuda

May 30, 2017

Contents

1 May 20, 2017

1.1 Progress

- set up github for project
- plotted histograms of first trials
- plotted scatter plots of first trials (for stationarity)

1.2 Statistical Model

Recall that our model of the process is

$$X = \sum_{j=1}^{N} Z_j, \qquad Y_j = \text{Bernoulli}(p_j), \qquad (Z_j \mid Y_j = 1) \sim N(\mu_j, \sigma_j^2), \qquad (Z_j \mid Y_j = 0) = 0$$

where $N, \mu_j, \sigma_j^2, p_j$ are all unknown parameters of the model. The Z_j random variable models the response amplitude of a single contact of which there are N, and the Y_j random variable models the release success of a single contact. We assume that all of the Z_j and the Y_j are independent.

1.2.1 Justification

The additivity of the potential is justified by Petterson and Einevoll [6]. The use of a Gaussian distribution for individual contacts is justified by Magee and Cook [4].

2 May 22, 2017

2.1 Progress

- derived point estimates for release probability, mean response amplitude, and response amplitude variance under the constant parameter models

2.2 Point Estimation of Model Parameters

2.2.1 Estimations Under a Simplified Model

We have that the expectation of the model is

$$\mathbb{E}[X] = \sum_{j=1}^{N} \mathbb{E}[Z_j] = \sum_{j=1}^{N} \mu_j \cdot p_j$$

and that the variance is

$$\text{Var}[X] = \sum_{j=1}^{N} \text{Var}[Z_j] = \sum_{j=1}^{N} \sigma_j^2 \cdot p_j.$$

Also note that the failure rate of the model, i.e. the probability that all N contacts fail to release, can be approximated by

$$\mathbb{P}[X=0] \approx \prod_{j=1}^{N}(1-p_j)$$

by assuming that a contact produces a positive response everytime it succeeds in releasing a vesicle. Thus, with simplifying assumptions that all the μ_j, σ_j^2, and p_j are the same across the N points of contact, and by fixing a value of N, we may find a plugin estimator for $\hat{p}$ and method of moments estimators for $\hat{\mu}$ and $\hat{\sigma}^2$ that depend on $\hat{p}$. If we let $\overline{X}$ be the sample mean, S^2 be the sample variance, and p_f be the sample failure rate, these estimates are given by

$$\hat{p} = 1 - \sqrt[N]{p_f}, \qquad \hat{\mu} = \frac{\overline{X}}{N\hat{p}}, \qquad \hat{\sigma}^2 = \frac{S^2}{N\hat{p}}.$$

2.2.2 EM Algorithm

Let us return to the general case. Suppose that we treat the response amplitudes of the individual contacts, the Z_j from $1 \leq j \leq N$, as latent variables. Then, the joint distribution of the latent and observed variables will be an exponential family, so EM algorithm should work very well.

3 May 24, 2017

3.1 Goals

- how was the data collected? can we really justify our model?
- workout details of EM and implement

3.2 Progress

- derive E step of the EM algorithm

3.3 EM Algorithm

We refer to lecture notes by Andrew Ng [5] for the EM reference. Suppose that we fix N and let $(x_i)_{i=1}^n$ be our observations of our model, let $y = (y_j)_{j=1}^N$ be the hidden observed values of the release success of each individual contact, and let $(\mu, \sigma^2, p) = (\mu_j, \sigma_j^2, p_j)_{j=1}^N$. Then, the log-likelihood of the parameters is given by

$$\ell\left(\mu, \sigma^2, p\right) = \sum_{i=1}^{n} \log p(x_i; \mu, \sigma^2, p) = \sum_{i=1}^{n} \log \sum_{y \in \{0,1\}^N} p(x_i \mid y; \mu, \sigma^2, p) \cdot p(y; \mu, \sigma^2, p).$$

In the above, the sum ranges over all possible assignments of the release success of the N individual contacts. Note that $p(z_i \mid y; \mu, \sigma^2, p)$ is simply the pdf of the sum of Gaussian variables which also takes a Gaussian distribution, and $p(y; \mu, \sigma^2, p)$ is simply given by the product of p_j or $1 - p_j$ for each $1 \leq j \leq N$, as appropriate. Also note that this problem is very similar to a Gaussian mixture model, where we sample from 2^N different Gaussian distributions where each Gaussian is a sum of the original N Gaussians. The difference is that the Gaussians that we find may not be arbitrary, but in fact must be generated by N Gaussians in a specific way.

3.3.1 E step

In the E step, we find a probability distribution over the release success of the individual contacts, given the observed amplitude x_i. Denote $\varphi(x; \mu, \sigma^2)$ as the pdf of a Gaussian with mean μ and variance σ^2 evaluated at the point x. Then,

$$Q_i(y) = p(y \mid x_i; \mu, \sigma^2, p) = \frac{p(x_i \mid y; \mu, \sigma^2, p) \cdot p(y; \mu, \sigma^2, p)}{\sum_{z \in \{0,1\}^N} p(x_i \mid z; \mu, \sigma^2, p) \cdot p(z; \mu, \sigma^2, p)}$$

where for all assignments of release successes $z \in \{0, 1\}^N$,

$$p(x_i \mid z; \mu, \sigma^2, p) = \varphi\left(x_i; \sum_{j:z_j=1} \mu_j, \sum_{j:z_j=1} \sigma_j^2\right), \qquad p(z; \mu, \sigma^2, p) = \prod_{j:z_j=1} p_j \prod_{j:z_j=0} (1 - p_j).$$

4 May 27, 2017

4.1 Progress

- derive M step of the EM algorithm for p_l

4.2 EM Algorithm (cont.)

4.2.1 M step

In the M step, we maximize the function

$$\ell(\theta) = \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \log \frac{p(x_i, z; \theta)}{Q_i^{(t)}(z)}$$

and set the value of θ that maximizes the above to the new estimate for θ, where θ is our μ, σ^2, p. For notational convenience, let $\mu_z := \sum_{j:z_j=1} \mu_j$, $\sigma_z^2 := \sum_{j:z_j=1} \sigma_j^2$, and $p_z := \prod_{j:z_j=1} p_j \prod_{j:z_j=0}(1 - p_j)$.

- **Maximization with respect to** μ_l

We have that

$$
\begin{aligned}
\frac{\partial}{\partial \mu_l}\ell(\theta) &= \frac{\partial}{\partial \mu_l}\sum_{i=1}^{n}\sum_{z\in\{0,1\}^N} Q_i^{(t)}(z)\log\frac{p(x_i,z;\theta)}{Q_i^{(t)}(z)} = \frac{\partial}{\partial \mu_l}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(\log p(x_i,z;\theta) - \log Q_i^{(t)}(z)\right)\\
&= \frac{\partial}{\partial \mu_l}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(\log\left(\frac{1}{\sqrt{2\pi\sigma_z^2}}e^{-\frac{(x_i-\mu_z)^2}{2\sigma_z^2}}\cdot p_z\right) - \log Q_i^{(t)}(z)\right)\\
&= \frac{\partial}{\partial \mu_l}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(\log\frac{1}{\sqrt{2\pi\sigma_z^2}} - \frac{(x_i-\mu_z)^2}{2\sigma_z^2} + \log p_z - \log Q_i^{(t)}(z)\right)\\
&= -\frac{\partial}{\partial \mu_l}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\frac{(x_i-\mu_z)^2}{2\sigma_z^2} = -\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\frac{-2(x_i-\mu_z)}{2\sigma_z^2}\\
&= \sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\frac{x_i-\mu_z}{\sigma_z^2}.
\end{aligned}
$$

We cannot isolate μ_l, so we need to solve with the other μ_j as well as the σ_j^2.

- **Maximization with respect to σ_l^2**

  We have that

$$
\begin{aligned}
\frac{\partial}{\partial \sigma_l^2}\ell(\theta) &= \frac{\partial}{\partial \sigma_l^2}\sum_{i=1}^{n}\sum_{z\in\{0,1\}^N} Q_i^{(t)}(z)\log\frac{p(x_i,z;\theta)}{Q_i^{(t)}(z)}\\
&= \frac{\partial}{\partial \sigma_l^2}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(\log\frac{1}{\sqrt{2\pi\sigma_z^2}} - \frac{(x_i-\mu_z)^2}{2\sigma_z^2} + \log p_z - \log Q_i^{(t)}(z)\right)\\
&= \frac{\partial}{\partial \sigma_l^2}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(\log\frac{1}{\sqrt{2\pi\sigma_z^2}} - \frac{(x_i-\mu_z)^2}{2\sigma_z^2}\right)\\
&= \frac{\partial}{\partial \sigma_l^2}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(-\frac{\log(2\pi\sigma_z^2)}{2} - \frac{(x_i-\mu_z)^2}{2\sigma_z^2}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\left(\left(\frac{x_i-\mu_z}{\sigma_z^2}\right)^2 - \frac{1}{\sigma_z^2}\right).
\end{aligned}
$$

  Now we need to solve for all the μ_j and σ_j when the above is set to 0. It is worth noting that the Z-score with respect to $N(\mu_z, \sigma_z^2)$ appears in both of the derivatives for μ_l and σ_l^2.

- **Maximization with respect to p_l**

As before, we have that

$$
\begin{aligned}
\frac{\partial}{\partial p_l}\ell(\theta) &= \frac{\partial}{\partial p_l}\sum_{i=1}^{n}\sum_{z\in\{0,1\}^N} Q_i^{(t)}(z)\log\frac{p(x_i,z;\theta)}{Q_i^{(t)}(z)} \\
&= \frac{\partial}{\partial p_l}\sum_{i=1}^{n}\sum_{z\in\{0,1\}^N} Q_i^{(t)}(z)\left(\log\frac{1}{\sqrt{2\pi\sigma_z^2}} - \frac{(x_i-\mu_z)^2}{2\sigma_z^2} + \log p_z - \log Q_i^{(t)}(z)\right) \\
&= \frac{\partial}{\partial p_l}\sum_{i=1}^{n}\sum_{z\in\{0,1\}^N} Q_i^{(t)}(z)\log p_z \\
&= \frac{\partial}{\partial p_l}\sum_{i=1}^{n}\sum_{z\in\{0,1\}^N} Q_i^{(t)}(z)\left(\sum_{j:z_j=1}\log p_j + \sum_{j:z_j=0}\log(1-p_j)\right) \\
&= \frac{\partial}{\partial p_l}\sum_{i=1}^{n}\left(\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)\log p_l + \sum_{\substack{z\in\{0,1\}^N\\ z_l=0}} Q_i^{(t)}(z)\log(1-p_l)\right) \\
&= \sum_{i=1}^{n}\left(\frac{\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)}{p_l} - \frac{\sum_{\substack{z\in\{0,1\}^N\\ z_l=0}} Q_i^{(t)}(z)}{1-p_l}\right).
\end{aligned}
$$

Setting this to 0 gives us that

$$
0 = \sum_{i=1}^{n}\left(\frac{\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)}{p_l} - \frac{\sum_{\substack{z\in\{0,1\}^N\\ z_l=0}} Q_i^{(t)}(z)}{1-p_l}\right),
$$

which solves to

$$
p_l = \frac{\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)}{\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z) + \sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=0}} Q_i^{(t)}(z)} = \boxed{\frac{\sum_{i=1}^{n}\sum_{\substack{z\in\{0,1\}^N\\ z_l=1}} Q_i^{(t)}(z)}{n}}.
$$

5 May 28, 2017

5.1 Progress

- gave up on a closed form for the M step, decided on gradient descent for the EM algorithm

5.2 EM Algorithm (cont.)

5.2.1 M step (cont.)

Since we cannot solve for the optimal μ_l and σ_l^2 in closed form, we opt for numerically determining the μ_l and σ_l^2 via gradient descent [1]. Recall that we have already computed the gradient:

$$\begin{cases} \frac{\partial}{\partial \mu_l}\ell(\theta) = \sum_{i=1}^n \sum_{\substack{z\in\{0,1\}^N \\ z_l=1}} Q_i^{(t)}(z)\frac{x_i-\mu_z}{\sigma_z^2} \\ \frac{\partial}{\partial \sigma_l^2}\ell(\theta) = \frac{1}{2}\sum_{i=1}^n \sum_{\substack{z\in\{0,1\}^N \\ z_l=1}} Q_i^{(t)}(z)\left(\left(\frac{x_i-\mu_z}{\sigma_z^2}\right)^2 - \frac{1}{\sigma_z^2}\right) \end{cases}.$$

Now, it is straightforward to implement the algorithm.

5.3 Parametric Bootstrap

Once we have an estimator for the parameters of the model, we may use parametric bootstrap to estimate the variance of the estimator, as explained in the classical text by Casella and Berger [2]. Using the parameters $\hat{\theta}$ estimated by the EM algorithm, we may simulate n values $X_1^*, \ldots, X_n^* \sim f(x;\hat{\theta})$ and approximate the MLE using the EM algorithm B times, and use the sample variance of those B estimates of the MLE as the variance of the estimator.

6 May 29, 2017

6.1 Progress

- implement EM algorithm upto expectation step

6.2 Goals

- validate p_j estimates with the sample failure rate
- validate μ_j estimates with the sample mean
- validate σ_j^2 estimates with the sample variance
- after writing slower version of EM algorithm, write a faster version

7 May 30, 2017

7.1 Progress

- derive EM algorithm steps for “binomial distribution” model
- implement draft of EM algorithm for “binomial distribution” model
- test EM algorithm, realize that we need to handle the $z = (0, 0, \ldots, 0)$ case separately

7.2 Goals

- test EM algorithm with data simulated from the generative model
- justify linearity of postsynaptic integration of signals
- investigate emergent properties
- separate out nonzero responses from the zero responses
- show that the classical binomial model does not sufficiently explain the data (figures, graphics)
- investigate results of Turner and West and see if they apply
- also try model that keeps σ^2 constant, and derive the EM algorithm for it
- study Kolmogorov-Smirnov test for testing the equality of the distributions

7.3 Previous Model of EPSP Amplitudes - Binomial Distribution

The previous models of postsynaptic potential amplitudes are usually binomial models [3]. We would like to first show that this model does not fit our data in a satisfactory way. This model assumes that the probabilities of all of the N contacts are the same and that the mean response is also all the same. Let us further suppose that the variances of the Gaussians at each contact are the same and see if we may reject this hypothesis. The insufficiency of this model is also suggested by Turner and West [7], and we wish to support this view.

7.3.1 Statistical Model

With N contacts, the model is given by

$$X = \sum_{j=1}^{N} Z_j, \qquad Y_j = \text{Bernoulli}(p), \qquad (Z_j \mid Y_j = 1) \sim N(\mu, \sigma^2), \qquad (Z_j \mid Y_j = 0) = 0$$

with parameters μ, σ^2, p.

7.3.2 Method of Moments Estimator

Recall that

$$\mathbb{E}[X] = \sum_{j=1}^{N} \mathbb{E}[Z_j] = Np\mu, \qquad \text{Var}[X] = \sum_{j=1}^{N} \text{Var}[Z_j] = Np\sigma^2.$$

Then if we determine p via the failure rate p_f, we find estimates

$$\hat{p} = 1 - \sqrt[N]{p_f}, \qquad \hat{\mu} = \frac{\overline{X}}{N\hat{p}}, \qquad \hat{\sigma}^2 = \frac{S^2}{N\hat{p}},$$

as before.

7.3.3 Maximum Likelihood Estimator

The log likelihood of the above model as a function of the parameters is given by

$$\ell(\mu, \sigma^2, p) = \sum_{i=1}^{n} p(x_i; \mu, \sigma^2, p) = \sum_{i=1}^{n} \log \sum_{z \in \{0,1\}^N} p(x_i, z; \mu, \sigma^2, p).$$

As above, we must employ the EM algorithm to estimate the MLE. Under these assumptions, the E step simplifies to

$$Q_i(y) = p(y \mid x_i; \mu, \sigma^2, p) = \frac{p(x_i \mid y; \mu, \sigma^2, p) \cdot p(y; \mu, \sigma^2, p)}{\sum_{z \in \{0,1\}^N} p(x_i \mid z; \mu, \sigma^2, p) \cdot p(z; \mu, \sigma^2, p)}$$

with

$$p(x_i \mid z; \mu, \sigma^2, p) = \varphi\left(x_i; N_z\mu, N_z\sigma^2\right), \qquad p(z; \mu, \sigma^2, p) = p^{N_z}(1-p)^{N-N_z} = p^{N_z} - p^N.$$

where N_z denotes the number of 1s in $z \in \{0,1\}^N$. In the M step, we have now have closed form solutions to all of the parameters. We have that

$$\frac{\partial}{\partial \mu}\ell(\theta) = \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \frac{-(x_i - N_z\mu)(-2N_z)}{2N_z\sigma^2} = \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \frac{x_i - N_z\mu}{\sigma^2} = 0$$

which gives a solution of

$$\mu = \frac{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i(z) x_i}{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i(z) N_z} = \boxed{\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) N_z}}$$

for μ,

$$\begin{aligned}
\frac{\partial}{\partial \sigma^2}\ell(\theta) &= \frac{\partial}{\partial \sigma^2} \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \left(-\frac{\log(2\pi N_z \sigma^2)}{2} - \frac{(x_i - N_z\mu)^2}{2N_z\sigma^2}\right) \\
&= \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \left(-\frac{1}{2\sigma^2} + \frac{1}{2}\left(\frac{x_i - N_z\mu}{N_z\sigma^2}\right)^2\right) = 0
\end{aligned}$$

which gives a solution of

$$\sigma^2 = \frac{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \left(\frac{x_i - N_z\mu}{N_z}\right)^2}{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z)} = \boxed{\frac{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \left(\frac{x_i - N_z\mu}{N_z}\right)^2}{n}}$$

for σ^2, and

$$\begin{aligned}
\frac{\partial}{\partial p}\ell(\theta) &= \frac{\partial}{\partial p} \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \left(N_z \log p + (N - N_z)\log(1-p)\right) \\
&= \sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) \left(\frac{N_z}{p} - \frac{N - N_z}{1-p}\right) = 0
\end{aligned}$$

which gives a solution of

$$p = \frac{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) N_z}{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) N} = \boxed{\frac{\sum_{i=1}^{n} \sum_{z \in \{0,1\}^N} Q_i^{(t)}(z) N_z}{Nn}}$$

for p.

References

[1] Christopher M Bishop. Pattern recognition. *Machine Learning*, 128:1–58, 2006.

[2] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[3] U Kuhnt, G Hess, and LL Voronin. Statistical analysis of long-term potentiation of large excitatory postsynaptic potentials recorded in guinea pig hippocampal slices: binomial model. *Experimental brain research*, 89(2):265–274, 1992.

[4] Jeffrey C Magee and Erik P Cook. Somatic epsp amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nature neuroscience*, 3(9):895–903, 2000.

[5] Andrew Ng. The em algorithm. University Lecture, 2016.

[6] Klas H Pettersen and Gaute T Einevoll. Amplitude variability and extracellular low-pass filtering of neuronal spikes. *Biophysical journal*, 94(3):784–802, 2008.

[7] Dennis A Turner and Mike West. Bayesian analysis of mixtures applied to post-synaptic potential fluctuations. *Journal of neuroscience methods*, 47(1):1–21, 1993.