Sometimes the spike-triggered average (STA) is just as good as the Poisson GLM for spike-train analysis

M. Rule

October 31, 2012

0.0.1 The Poisson GLM for spiking data

Generalized Linear Models (GLMs) are similar to linear regression, but account for nonlinearities and non-uniform noise in the observations. In neuroscience, it is common to predict a sequence of spikes $Y = \{y_1, .., y_T\}$, $y_i \in \{0, 1\}$, from a series of observations $X = \{\mathbf{x}_1, .., \mathbf{x}_T\}$, using a Poisson GLM:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i \cdot \Delta t) \\ \lambda_i &= \exp\left(\mathbf{a}^\top \mathbf{x}_i + b\right) \end{aligned}$$

These models are fit by minimizing the negative log-likelihood of the observations, given the vector of regression weights $\mathbf{a}$ and mean offset parameter b:

$$\underset{a,b}{\operatorname{argmin}} \langle \lambda - y \ln \lambda \rangle = \underset{\mathbf{a},b}{\operatorname{argmin}} \left\langle e^{\mathbf{a}^\top \mathbf{x}_i + b} - y[\mathbf{a}^\top \mathbf{x}_i + b] \right\rangle$$

Such models are usually optimized via gradient descent, or iterated re-weighted least squares. The gradient

$$\begin{aligned} \partial_{\mathbf{b}}(...) &= \langle \lambda - y \rangle \\ \partial_{\mathbf{a}}(...) &= \left\langle (\lambda - y)\mathbf{x}^\top \right\rangle \end{aligned}$$

This is convex, and the optimum occurs at an inflection point where the gradient is zero, i.e.:

$$\begin{aligned} \langle \lambda \rangle &= \langle y \rangle \\ \left\langle \lambda \mathbf{x}^\top \right\rangle &= \left\langle y \mathbf{x}^\top \right\rangle \end{aligned}$$

For simplicity, assume that $\mathbf{x}$ are a collection of uncorrelated zero-mean Gaussian variables. (If $\mathbf{x}$ has a multivariate Gaussian distribution, we an always transform into coordinates so that this is true by subtracting the mean and whitening the data.)

0.0.2 Linear models are often fine

What happens if we expand our likelihood at first order, and solve for a zero of the gradient? We start by expanding around the point $b_0 = \ln \langle y \rangle$:

$$\begin{aligned} \lambda = e^{\mathbf{a}^\top \mathbf{x} + b} &= e^{\mathbf{a}^\top \mathbf{x} + b - b_0} e^{b_0} \\ &\approx [1 + \mathbf{a}^\top \mathbf{x} + b - b_0] e^{b_0} \\ &= [1 + \mathbf{a}^\top \mathbf{x} + b - b_0] \langle y \rangle \end{aligned}$$

Substituting this approximation in to the identity for the optimal coefficients:

$$\begin{aligned} \left\langle [1 + \mathbf{a}^\top \mathbf{x} + b - b_0] \right\rangle &= 1 \\ \left\langle [1 + \mathbf{a}^\top \mathbf{x} + b - b_0] \mathbf{x}^\top \right\rangle &= \left\langle y \mathbf{x}^\top \right\rangle / \langle y \rangle \end{aligned}$$

If we have chosen $\Delta t \ll 1$, so that y is indeed binary, then the right hand side of the second equation is simply the spike-triggered average. Denote this as $\mathbf{x}_1^\top = \langle y \mathbf{x}^\top \rangle / \langle y \rangle$. Solving for $(\mathbf{a}, b)$ in the above yields:

$$\begin{aligned} b &= b_0 \\ \mathbf{a} &= \mathbf{x}_1 \end{aligned}$$

Which is to say: to a first approximation, the coefficients $\mathbf{a}$ of the Poisson GLM are simply the spike triggered average (of the mean-centered and whitened $\mathbf{x}$).

Intuitively, this means if you're just trying to demonstrate some basic correlation between a variable and neuronal spiking, linear methods like the spike-triggered-average are fine. Moving to a GLM might give you more accuracy, since it has a better model of the observation process, but it often doesn't tell you much more, qualitatively.

0.0.3 For Gaussian covariates, GLMs capture how noise and nonlinearity interact to influence firing rate

Let's see if we can get further intuition if $\mathbf{x}$ is Gaussian, beyond first order. Work in normalized firing rate by defining $\bar{\lambda} = \lambda / \langle y \rangle$. The optimum can then be written as:

$$\begin{aligned} \langle \bar{\lambda} \rangle &= 1 \\ \langle \bar{\lambda} \mathbf{x} \rangle &= \mathbf{x}_1 \end{aligned}$$

If $\mathbf{x}$ is Gaussian, the expectation $\langle \bar{\lambda} \mathbf{x}^\top \rangle$ has a closed form. Defining $\bar{b} = b - b_0 = b - \ln \langle y \rangle$:

$$\langle \bar{\lambda} \mathbf{x} \rangle = \int_{d\mathbf{x}} \mathbf{x} e^{\mathbf{a}^\top \mathbf{x} + \bar{b}} (2\pi)^{-D/2} |\Sigma_{\mathbf{x}}|^{-1/2} \exp\left(-\tfrac{1}{2} (\mathbf{x} - \mu_{\mathbf{x}})^\top \Sigma_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \right),$$

where D is the dimensionality of $\mathbf{x}$. Working in coordinates where $\Sigma_{\mathbf{x}} = I$ and $\mu_{\mathbf{x}} = 0$, this simplifies to:

$$\left\langle \bar{\lambda}\mathbf{x} \right\rangle = (2\pi)^{-D/2} \int_{d\mathbf{x}} \mathbf{x} e^{\mathbf{a}^\top \mathbf{x} + \bar{b} - \frac{1}{2}\mathbf{x}^\top \mathbf{x}}.$$

This integral can be solved by completing the square:

$$\begin{aligned}
\mathbf{a}^\top \mathbf{x} + \bar{b} - \tfrac{1}{2}\mathbf{x}^\top \mathbf{x} &= -\tfrac{1}{2}\left[-2\mathbf{a}^\top \mathbf{x} - 2\bar{b} + \mathbf{x}^\top \mathbf{x}\right] \\
&= -\tfrac{1}{2}\left[\mathbf{x}^\top \mathbf{x} - 2\mathbf{a}^\top \mathbf{x} + \mathbf{a}^\top \mathbf{a} - \mathbf{a}^\top \mathbf{a} - 2\bar{b}\right] \\
&= -\tfrac{1}{2}\left[(\mathbf{x}-\mathbf{a})^\top (\mathbf{x}-\mathbf{a}) - \mathbf{a}^\top \mathbf{a} - 2\bar{b}\right] \\
&= -\tfrac{1}{2}\left[(\mathbf{x}-\mathbf{a})^\top (\mathbf{x}-\mathbf{a})\right] + \left[\tfrac{1}{2}\mathbf{a}^\top \mathbf{a} + \bar{b}\right]
\end{aligned}$$

$$\left\langle \bar{\lambda}\mathbf{x}^\top \right\rangle = (2\pi)^{-D/2} \int_{d\mathbf{x}} \mathbf{x} e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{a})^\top (\mathbf{x}-\mathbf{a})\right] + \left[\frac{1}{2}\mathbf{a}^\top \mathbf{a} + \bar{b}\right]}.$$

This is a Gaussian integral. It evaluates to:

$$\left\langle \bar{\lambda}\mathbf{x} \right\rangle = \mathbf{a} e^{\frac{1}{2}\mathbf{a}^\top \mathbf{a} + \bar{b}}.$$

At the optimum, we therefore have the condition that the optimal decoding weights $\mathbf{a}$ should be proportional to the spike-triggered average $\mathbf{x}_1$:

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{a} e^{\frac{1}{2}\mathbf{a}^\top \mathbf{a} + \bar{b}} \\
&= \mathbf{a} e^{\frac{1}{2}\|\mathbf{a}\|^2 + \bar{b}} \\
\Rightarrow \mathbf{a} &\propto \mathbf{x}_1
\end{aligned}$$

Solving similarly for $\langle \bar{\lambda} \rangle$ gives $\langle \bar{\lambda} \rangle = \exp(\frac{1}{2}\mathbf{a}^\top \mathbf{a} + \bar{b})$, which seems to imply that the Poisson GLM has a closed-form solution in the limit where is y sparse and $\mathbf{x}$ is Gaussian:

$$\begin{aligned}
\mathbf{a} &= \mathbf{x}_1 \\
b &= \ln \langle y \rangle - \tfrac{1}{2}\|\mathbf{x}_1\|^2
\end{aligned}$$

This further supports the idea that the spike-triggered average is an fast, adequate approach to statistics on spike trains. In some cases, it may be unnecessary to train a GLM.

Of course, Poisson GLMs can be fit almost as quickly as linear regressions or spike-triggered averages, using iteratively re-weighted least squares. They can also gracefully and automatically handle cases where the covariates $\mathbf{x}$ are not jointly Gaussian. So perhaps the question should be why *not* use a GLM.