Lecture 11: GLMs
(encoding models, part 2)
neural coding problem

Q: what is the probabilistic relationship between stimuli and spike trains?
Example 1: linear Poisson neuron

spike count  \( y \sim \text{Poisson}(\lambda) \)

spike rate  \( \lambda = \theta x \)

encoding model:  
\[
P(y|x, \theta) = \frac{1}{y!} \lambda^y e^{-\lambda} = \frac{1}{y!} (\theta x)^y e^{-(\theta x)}
\]
\[
\log P(Y \mid X, \theta) = \sum_i \log P(y_i \mid x_i, \theta)
\]
\[
= \sum y_i \log \theta - \theta x_i + c
\]
\[
= \log \theta (\sum y_i) - \theta (\sum x_i)
\]

- Closed-form solution when model in “exponential family”

\[
\frac{d}{d\theta} \log P(Y \mid X, \theta) = \frac{1}{\theta} \sum y_i - \sum x_i = 0
\]
\[
\implies \hat{\theta}_{ML} = \frac{\sum y_i}{\sum x_i}
\]
Example 2: linear Gaussian neuron

spike count \[ y \sim \mathcal{N}(\mu, \sigma^2) \]

spike rate \[ \mu = \theta x \]

parameter stimulus

encoding model:

\[ P(y|x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\theta x)^2}{2\sigma^2}} \]
\[
\text{mean}(y) = \theta x \\
\text{var}(y) = \sigma^2
\]

All slices have same width

encoding distribution

\[p(y|x = 20)\]
\( P(y|x, \theta) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{\frac{-(y-\theta x)^2}{2\sigma^2}} \)

Log-Likelihood

\[
\log P(Y|X, \theta) = -\sum \frac{(y_i - \theta x_i)^2}{2\sigma^2} + c
\]

Do it: differentiate, set to zero, and solve.
$$P(y|x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \theta x)^2}{2\sigma^2}}$$

Log-Likelihood

$$\log P(Y|X, \theta) = - \sum \frac{(y_i - \theta x_i)^2}{2\sigma^2} + c$$

$$\frac{d}{d\theta} \log P(Y|X, \theta) = - \sum \frac{(y_i - \theta x_i)x_i}{\sigma^2} = 0$$

$$\sum y_i x_i - \sum \theta x_i^2 = 0$$

$$\theta \sum x_i^2 = \sum y_i x_i$$

Maximum-Likelihood Estimator:

$$\hat{\theta}_{ML} = \frac{\sum y_i x_i}{\sum x_i^2}$$
Log-Likelihood

$P(y|x, \theta) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(y-\theta x)^2}{2\sigma^2}}$

Log-Likelihood

$$\log P(Y|X, \theta) = -\sum \frac{(y_i - \theta x_i)^2}{2\sigma^2} + c$$

$$\frac{d}{d\theta} \log P(Y|X, \theta) = -\sum \frac{(y_i - \theta x_i)x_i}{\sigma^2} = 0$$

$$\sum y_i x_i - \sum \theta x_i^2 = 0$$

$$\theta \sum x_i^2 = \sum y_i x_i$$

Maximum-Likelihood Estimator:

(“Least squares regression” solution)

$$\hat{\theta} = (X^\top X)^{-1} X^\top Y$$

(Recall that for Poisson, $\hat{\theta}_{ML} = \frac{\sum y_i}{\sum x_i}$)
Example 3: unknown neuron

Be the computational neuroscientist: what model would you use?
Example 3: unknown neuron

More general setup: \[ y \sim \text{Poiss}(\lambda) \]
\[ \lambda = f(\theta x) \] for some nonlinear function \( f \)
Quick Quiz:

The distribution $P(y|x, \theta)$ can be considered as a function of $y$, $x$, or $\theta$.

What is $P(y|x, \theta)$:

1. as a function of $y$?
   Answer: **encoding distribution** - probability distribution over spike counts

2. as a function of $\theta$ ?
   Answer: **likelihood function** - the probability of the data given model params

3. as a function of $x$?
   Answer: **stimulus likelihood function** - useful for ML stimulus decoding!
Stimulus likelihood function
(for decoding)

$P(y = 20| x, \theta)$
Next: Generalized Linear Models (GLMs)

- spike trains (instead of just “counts”) \( P(y(t) \mid \bar{x}(t)) \)
- multi-neuron spike trains \( P(y_1(t), y_2(t), \ldots y_n(t) \mid \bar{x}(t)) \)
Note on GLMs

• Be careful about terminology:

GLM ≠ GLM

General Linear Model ≠ Generalized Linear Model

(Nelder 1972)
Stephen Senn: I must confess to having some confusion when I was a young statistician between general linear models and generalized linear models. Do you regret the terminology?

John Nelder: I think probably I do. I suspect we should have found some more fancy name for it that would have stuck and not been confused with the general linear model, although general and generalized are not quite the same. I can see why it might have been better to have thought of something else.

Senn, (2003). Statistical Science
Moral:
Be careful when naming your model!
1. General Linear Model

\[ \vec{y} = \vec{\theta} \cdot \vec{x} + \epsilon \]

Examples:
1. Gaussian
   \[
   y = \vec{\theta} \cdot \vec{x} + \epsilon
   \]
2. Poisson
   \[
   y \sim \text{Poiss}(\vec{\theta} \cdot \vec{x})
   \]
2. Generalized Linear Model

Examples:

1. Gaussian
   \[ y = f(\theta \cdot \vec{x}) + \epsilon \]

2. Poisson
   \[ y \sim \text{Pois}(f(\theta \cdot \vec{x})) \]
2. Generalized Linear Model

Terminology:

\[ x \rightarrow \text{Linear} \rightarrow \text{Nonlinear} \rightarrow \text{Noise} \rightarrow y \]

- "distribution function"
- "parameter"
- "link function"
From spike counts to spike trains:

\[ y_t = \vec{k} \cdot \vec{x}_t + \text{noise} \]

first idea: linear-Gaussian model!
The response at time $t$ is given by the linear filter $k$ applied to the vector stimulus $\vec{x}_t$ at time $t$:

$$y_t = \vec{k} \cdot \vec{x}_t + \text{noise}$$

We walk through the data one time bin at a time. The stimulus is presented, and the response is recorded at each time step. The distribution of the noise is $N(0, \sigma^2)$. 

Walk through the data one time bin at a time.
$y_t = \vec{k} \cdot \vec{x}_t + \text{noise}$

response at time $t$

linear filter

vector stimulus at time $t$

walk through the data one time bin at a time

$t = 2$

stimulus

response

time

$N(0, \sigma^2)$
The response at time $t$ is given by the linear filter:

$$y_t = \vec{k} \cdot \vec{x}_t + \text{noise}$$

where $\vec{k}$ is the linear filter vector, $\vec{x}_t$ is the vector stimulus at time $t$, and the noise is normally distributed with mean 0 and variance $\sigma^2$.

Walk through the data one time bin at a time for $t = 3$.

The stimulus is shown in the diagram above, and the response is shown below the time axis.

The time axis is labeled with $y_t$. The diagram illustrates how the vector stimulus is filtered to produce the response at each time point.
The response at time $t$ is given by:

$$ y_t = \vec{k} \cdot \vec{x}_t + \text{noise} $$

where $\vec{k}$ is the linear filter, $\vec{x}_t$ is the vector stimulus at time $t$, and $\text{noise}$ is a random variable drawn from a normal distribution $N(0, \sigma^2)$. 

Walk through the data one time bin at a time with $t = 4$. The stimulus and response are shown as follows:
The response at time $t$ can be modeled as:

$$y_t = \vec{k} \cdot \vec{x}_t + \text{noise}$$

where $\vec{k}$ is a linear filter, $\vec{x}_t$ is the vector stimulus at time $t$, and the noise is distributed according to $N(0, \sigma^2)$.

Walk through the data, one time bin at a time, to determine the stimulus $\vec{x}_t$ and the response $y_t$. For example, at $t = 5$:

- The stimulus $\vec{x}_t$ is shown in the diagram.
- The response $y_t$ is also depicted in the diagram.
The response at time $t$ can be modeled as a linear combination of the vector stimulus at time $t$ plus noise:

$$y_t = \vec{k} \cdot \vec{x}_t + \text{noise}$$

where $\vec{k}$ is a linear filter, $\vec{x}_t$ is the vector stimulus at time $t$, and $N(0, \sigma^2)$ represents a normal distribution with mean 0 and variance $\sigma^2$.

To walk through the data one time bin at a time, consider the following example:

- **Stimulus** at time $t = 6$.
- **Response** at time $t = 6$.

The graph illustrates the stimulus and response vectors at time $t = 6$. The stimulus vector $\vec{x}_t$ and the response vector $y_t$ are shown, with the corresponding time bins indicated.
Build up to following matrix version:

\[ Y = X \vec{k} + \text{noise} \]

\[
\begin{bmatrix}
0 \\
0 \\
1 \\
\vdots
\end{bmatrix} = \begin{bmatrix}
\vdots \\
\vdots \\
\vdots
\end{bmatrix} \begin{bmatrix}
\vec{k}
\end{bmatrix}
\]

**design matrix**
Build up to following matrix version:

\[
Y = X\vec{k} + \text{noise}
\]

“Linear-Gaussian” GLM:

(aka “least-squares regression”)

\[
\hat{k} = (X^TX)^{-1}X^TY
\]

stimulus covariance
spike-triggered avg (STA)