Neural encoding models & maximum likelihood

Jonathan Pillow
warming-up quiz (not for credit)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x 1</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>y 2</td>
<td>0.08</td>
<td>0.4</td>
<td>0.32</td>
</tr>
<tr>
<td>y 3</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
</tr>
</tbody>
</table>

1. What is the marginal $P(x)$?
2. What is the marginal $P(y)$?
3. What is the conditional $P(y \mid x = 1)$?
4. Are $x$ and $y$ independent? Why or why not?
1. What is the marginal $P(x)$? 
   $[0.1 \ 0.5 \ 0.4]$
2. What is the marginal $P(y)$? 
   $[0.1 \ 0.8 \ 0.1]$
3. What is the conditional $P(y \mid x = 1)$? 
   $[0.1 \ 0.8 \ 0.1]$
4. Are $x$ and $y$ independent? Why or why not?
   Yes. joint = product of marginals
neural coding problem

Q: what is the probabilistic relationship between stimuli and spike trains?
neural coding problem

Q: What is the probabilistic relationship between stimuli and spike trains?

$$P(y_1, y_2, \ldots, y_n | x, \theta)$$

“encoding model”
today: single-neuron encoding

Question: what criteria for picking a model?
model desiderata

fitability / tractability
(can be fit to data)

sweet spot

GLM

richness / flexibility
(capture realistic neural properties)

linear, Gaussian

multi-compartment Hodgkin-Huxley
Example 1: linear Poisson neuron

- Spike count: $y \sim \text{Poisson}(\lambda)$
- Spike rate: $\lambda = \theta x$

Encoding model: $P(y|x, \theta) = \frac{1}{y!} \lambda^y e^{-\lambda}$

$\lambda = \text{mean} = \text{variance}$
mean\( (y) = \theta x \)

\[ \text{var}(y) = \theta x \]

conditional distribution

\[ p(y|x) \]

conditional distribution

\[ p(y|x = 5) \]
$$\text{mean}(y) = \theta x$$
$$\text{var}(y) = \theta x$$

$${p(y|x)}$$

Conditional distribution

$$p(y|x = 20)$$
mean(\(y\)) = \(\theta x\)

\text{var}(y) = \theta x

\text{conditional distribution} \quad p(y|x)

p(y|x = 35)
Maximum Likelihood Estimation:

• given observed data \((Y, X)\), find \(\theta\) that maximizes \(P(Y|X, \theta)\)

\[
P(Y|X, \theta) = \prod_{i=1}^{N} P(y_i|x_i, \theta)
\]

Q: what assumption are we making about the responses?
A: conditional independence across trials!
Maximum Likelihood Estimation:

- Given observed data \((Y, X)\), find \(\theta\) that maximizes \(P(Y|X, \theta)\).

\[
P(Y|X, \theta) = \prod_{i=1}^{N} P(y_i|x_i, \theta)
\]

Q: What assumption are we making about the responses?
A: Conditional independence across trials!

Q: When do we call \(P(Y|X, \theta)\) a likelihood?
A: When considering it as a function of \(\theta\)!
Maximum Likelihood Estimation:

- given observed data \((Y, X)\), find \(\theta\) that maximizes \(P(Y|X, \theta)\)

- could in theory do this by turning a knob
Maximum Likelihood Estimation:

- given observed data \((Y, X)\), find \(\theta\) that maximizes \(P(Y|X, \theta)\)

\(y \sim \text{Poiss}(\theta x)\)
\(\theta = 1\)

- could in theory do this by turning a knob
Maximum Likelihood Estimation:

• given observed data \((Y, X)\), find \(\theta\) that maximizes \(P(Y|X, \theta)\)

\[ p(y|x) \]

\[ y \sim \text{Poiss}(\theta x) \]
\[ \theta = 0.5 \]

• could in theory do this by turning a knob
Likelihood function: $P(Y|X, \theta)$ as a function of $\theta$

Because data are independent:

$$P(Y|X, \theta) = \prod_i P(y_i|x_i, \theta)$$

$$= \prod \frac{1}{y_i!} (\theta x_i)^{y_i} e^{-\theta x_i}$$
Likelihood function: $P(Y|X, \theta)$ as a function of $\theta$

Because data are independent:

$$P(Y|X, \theta) = \prod_i P(y_i|x_i, \theta)$$

$$= \prod \frac{1}{y_i!} (\theta x_i)^{y_i} e^{-(\theta x_i)}$$

log-likelihood

$$\log P(Y|X, \theta) = \sum_i \log P(y_i|x_i, \theta)$$

$$= \sum y_i \log \theta - \theta x_i + c$$
\[ \log P(Y|X, \theta) = \sum_i \log P(y_i|x_i, \theta) \]
\[ = \sum y_i \log \theta - \theta x_i + c \]
\[ = \log \theta \left( \sum y_i \right) - \theta \left( \sum x_i \right) \]

Do it: solve for \( \theta \)
\[
\log P(Y|X, \theta) = \sum_i \log P(y_i|x_i, \theta) \\
= \sum y_i \log \theta - \theta x_i + c \\
= \log \theta (\sum y_i) - \theta (\sum x_i)
\]

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

\[
\frac{d}{d\theta} \log P(Y|X, \theta) = \frac{1}{\theta} \sum y_i - \sum x_i = 0
\]

\[
\Rightarrow \hat{\theta}_{ML} = \frac{\sum y_i}{\sum x_i}
\]
Properties of the MLE (maximum likelihood estimator)

• consistent
  (converges to true $\theta$ in limit of infinite data)

• efficient
  (converges as quickly as possible, i.e., achieves minimum possible asymptotic error)