Neural encoding models & maximum likelihood

Jonathan Pillow
probability leftovers: sampling vs inference

true mean: [0 0.8]
true cov: [1.0 -0.25
-0.25 0.3]
sample mean: [-0.05 0.83]
sample cov: [0.95 -0.23
-0.23 0.29]
An estimator is a function $f : x \rightarrow \hat{\theta}$

- often we will write $\hat{\theta}(x)$ or just $\hat{\theta}$
Properties of an estimator

bias: \[ b(\theta) = \mathbb{E}[\hat{\theta}] - \theta \]
• “unbiased” if bias=0

variance: \[ \text{var}(\theta) = \mathbb{E} \left[ (\hat{\theta} - \mathbb{E}[\hat{\theta}])^2 \right] \]
• “consistent” if bias and variance both go to zero asymptotically

Q: what is the bias of the estimator \( \hat{\theta}(x) = 7 \) (i.e., estimate is 7 for all datasets \( x \))

Q: what is the variance of that estimator?
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?

\[ P(y_1, y_2, \ldots, y_n | x, \theta) \]
today: single-neuron encoding

X
stimuli

$P(y_i | \vec{x}, \theta)$
“encoding model”

y
spike trains

**Question:** what criteria for picking a model?
model desiderata

fittability / tractability
(can be fit to data)

richness / flexibility
(capture realistic neural properties)

GLM
sweet spot

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

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

Encoding model:

\[
P(y|x, \theta) = \frac{1}{y!} \lambda^y e^{-\lambda} = \frac{1}{y!} (\theta x)^y e^{-(\theta x)}
\]
\[ \text{mean}(y) = \theta x \]
\[ \text{var}(y) = \theta x \]

The conditional distribution \( p(y|x) \) is shown in the graph, with a spike count on the y-axis and contrast on the x-axis. The conditional distribution \( p(y|x = 5) \) is also depicted.
$$\text{mean}(y) = \theta x$$
$$\text{var}(y) = \theta x$$

$p(y|x)$

conditional distribution

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

p(y|x)

conditional distribution
\[ p(y|x = 35) \]
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 \(P(Y|X, \theta)\)

\[y \sim \text{Poiss}(\theta x)\]
\[\theta = 1.5\]

• 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: $P(Y \mid X, \theta)$ as a function of $\theta$

Because data are independent:

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

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

Because data are independent:

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

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

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

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”

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

\[\implies \hat{\theta}_{ML} = \frac{\sum y_i}{\sum x_i}\]
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)
Example 2: linear Gaussian neuron

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

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

parameter stimulus

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

encoding distribution

\[ p(y|x = 20) \]

All slices have same width.
$P(y|x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \theta x)^2}{2\sigma^2}}$

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

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

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

Maximum-Likelihood Estimator:

\[ \hat{\theta}_{ML} = \frac{\sum y_i x_i}{\sum x_i^2} \]

(“Least squares regression” solution)

(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: \[ y \sim Poiss(\lambda) \]

\[ \lambda = f(\theta x) \]

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!