Probabilistic neural models, maximum likelihood and neural coding

Jonathan Pillow
Psychology, Neurobiology, & Division of Statistics & Scientific Computation
The University of Texas at Austin

Nov 2013
neural coding problem

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

potentially

• insight into sensory processing
• sensory prosthetics

\( x \)

novel stimulus
(Dr. Fiete, Cosyne 2012)

Codebook: \( P(y|x) \)

\( y \)
neural coding problem

Bayes’ Rule: \[ P(x|y) \propto P(y|x)P(x) \]

uses: • constraints on neural readout of pop codes
• motor prosthetics
encoding-model approach

Goal: find model that closely approximates true encoding distribution

Question: what criteria for picking a model?

model: \( P_\theta(y|x) \approx P(y|x) \)
Goals for today:

\[ P(y|x, \theta) \]

1. Basic concepts from probability theory
2. Simple neural encoding models (spike counts)
3. Maximum Likelihood methods for fitting
   • estimating parameters \( \theta \) from data
4. Generalized linear models (GLMs)
Probability basics

- Bernoulli distribution
- Gaussian distribution
- Poisson distribution
- random variable
- parameter
- conditional distribution
- likelihood
- independence
Example 1: linear Poisson neuron

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

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

encoding model:

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

conditional distribution

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

conditional distribution

\[ p(y|x) \]

\[ p(y|x = 20) \]
\[
\text{mean}(y) = \theta x \\
\text{var}(y) = \theta x
\]
Maximum Likelihood Estimation:

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

\[ y \sim \text{Poiss}(\theta x) \]
\[ \theta = 1.5 \]
Maximum Likelihood Estimation:

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

\[
p(y|x) \\
y \sim \text{Poiss}(\theta x) \\
\theta = 1
\]
Maximum Likelihood Estimation:

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

\[
p(y|x)
\]

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

\[
\theta = 0.5
\]
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:

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

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

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

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

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

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

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

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

spike rate\quad \mu = \theta x

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

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

Differentiate and set to zero:

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

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

What model would you use to fit this neuron?
More general setup: \( y \sim \text{Poisson}(\lambda) \)

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

for some nonlinear function \( f \)
break?

(no: deep breath)
Aside on GLMs:

1. Be careful about terminology:

GLM ≠ GLM

General Linear Model ≠ Generalized Linear Model

Linear ≠ Linear

(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.
Moral:
Be careful when naming your model!
2. General Linear Model

Examples:

1. Gaussian

\[ y = \theta \cdot x + \sigma^2 \epsilon \]

2. Poisson

\[ y \sim \text{Poisson}(\theta \cdot \bar{x}) \]
3. Generalized Linear Model

Examples:

1. Gaussian \( y = f(\theta \cdot \vec{x}) + \sigma^2 \epsilon \)
2. Poisson \( y \sim \text{Poisss}(f(\theta \cdot \vec{x})) \)
Modeling spike trains with GLMs
Linear-Nonlinear-Poisson model

stimulus $x(t)$ \rightarrow \text{stimulus filter} \rightarrow \text{point nonlinearity} \rightarrow \text{Poisson spiking} \rightarrow y(t)

conditional intensity (spike rate) $\lambda(t) = f(k \cdot x(t))$

$P(y \text{ spikes in } [t, t + \Delta]) = \text{Poiss}(\Delta \lambda(t))$
GLM with history-dependence

output: no longer a Poisson process

interpretation: “soft-threshold” integrate-and-fire model
multi-neuron GLM

stimulus filter

exponential nonlinearity

probabilistic spiking

post-spike filter

coupling filters

stimulus

neuron 1

neuron 2
GLM equivalent diagram:

$$x(t) \quad \text{stimulus} \quad k_1$$

$$y_1(t) \quad \text{cell 1 spikes} \quad h_{11}$$

$$y_2(t) \quad \text{cell 2 spikes} \quad h_{21}$$

conditional intensity (spike rate) \quad \lambda_i(t) = \exp(k_i \cdot x(t) + \sum_j h_{ij} \cdot y(t))