Introduction to Bayesian inference and generative models

Dr. Richard E. Turner (ret26@cam.ac.uk)

Computational and Biological Learning Lab, Department of Engineering
University of Cambridge

Laboratory for Computational Vision, Center for Neural Science
New York University
• Collected inter-spike interval measurements, $x$, from a neuron

• $x$ follows an exponential distribution with a characteristic time-scale $\lambda$, shifted to take account of the absolute refractory period of the neuron, 5ms long.

• ISIs over 50ms were not recorded (short trials used for data-collection)

• $N$ ISIs are observed at \{$x_1, \ldots, x_N$\}. What is $\lambda$?
• Collected inter-spike interval measurements, \( x \), from a neuron

• \( x \) follows an exponential distribution with a characteristic time-scale \( \lambda \), shifted to take account of the absolute refractory period of the neuron, 5\, ms long.

• ISIs over 50\, ms were not recorded (short trials used for data-collection)

• \( N \) ISIs are observed at \{\( x_1, \ldots x_N \}\). What is \( \lambda \)?
Ideas

Idea 1

- bin up into a histogram
  - where do we place the bins
- fit to density
  - what error measure do we minimise?

Idea 2

- construct an estimator e.g. the sample mean \( \mu = \frac{1}{N} \sum_{n=1}^{N} x_n \)
  - which estimator should we choose? mean, variance, higher moments?
- relate to parameters via expectation of estimator e.g. \( \mu \approx \langle x \rangle = f(\lambda) \)
  - small sample effects can be problematic e.g. if \( \mu > \frac{1}{2}(50 + 5) \)ms
A less ad hoc method...probabilities as degrees of belief

Cox’s axioms

• degrees of belief can be represented by real numbers

• take into account all evidence

• consistency: if things can be reasoned in more than one way, each must lead to the same answer

• equivalent states of knowledge $\rightarrow$ equivalent plausibility assignments

Product rule:

$$p(\lambda, x) = p(\lambda | x)p(x) = p(x | \lambda)p(\lambda)$$

Bayes' Rule

Sum rule:

$$p(\lambda) = \sum_x p(\lambda, x)$$

(marginalisation)
A less ad hoc method...probabilities as degrees of belief

Cox’s axioms

• degrees of belief can be represented by real numbers
• take into account all evidence
• consistency: if things can be reasoned in more than one way, each must lead to the same answer
• equivalent states of knowledge → equivalent plausibility assignments

Conclusion: degrees of belief must follow the rules of probability.
A less ad hoc method...probabilities as degrees of belief

**Cox’s axioms**

- degrees of belief can be represented by real numbers
- take into account all evidence
- consistency: if things can be reasoned in more than one way, each must lead to the same answer
- equivalent states of knowledge → equivalent plausibility assignments

**Conclusion:** degrees of belief must follow the rules of probability.

**Product rule:** \( p(\lambda, x) = p(\lambda|x)p(x) = p(x|\lambda)p(\lambda) \)

**Sum rule:** \( p(\lambda) = \sum_x p(\lambda, x) \)
A less ad hoc method...probabilities as degrees of belief

Cox’s axioms

• degrees of belief can be represented by real numbers

• take into account all evidence

• consistency: if things can be reasoned in more than one way, each must lead to the same answer

• equivalent states of knowledge $\rightarrow$ equivalent plausibility assignments

Conclusion: degrees of belief must follow the rules of probability.

Product rule: $p(\lambda, x) = p(\lambda|x)p(x) = p(x|\lambda)p(\lambda)$ $\leftarrow$ Bayes’ Rule

Sum rule: $p(\lambda) = \sum_x p(\lambda, x)$ $\leftarrow$ marginalisation
Mathematical solution

\[ p(x|\lambda) = \begin{cases} \frac{1}{Z(\lambda)} \exp \left( -\frac{x}{\lambda} \right) & \text{if } 5\text{ms} < x < 50\text{ms} \\
0 & \text{otherwise} \end{cases} \]

\[ Z(\lambda) = \int_5^{50} \exp \left( -\frac{x}{\lambda} \right) \, dx = \lambda \left( \exp \left( -\frac{5}{\lambda} \right) - \exp \left( -\frac{50}{\lambda} \right) \right) \]

what the data tell us (likelihood of parameters)
what we knew before hand (prior)

\[ p(\lambda|\{x_1, \ldots, x_N\}) = \frac{p(\{x\}|\lambda)p(\lambda)}{p(\{x\})} \]

\[ \propto \frac{1}{Z(\lambda)^N} \exp \left( -\frac{1}{\lambda} \sum_{n=1}^{N} x_n \right) p(\lambda) \]
Density

\[ p(x|\lambda) \]

- \( p(x|\lambda = 10) \)
- \( p(x|\lambda = 20) \)
- \( p(x|\lambda = 50) \)
Likelihood of the parameters

$\lambda / \text{ms}$

$p(x|\lambda)$

$p(x=15|\lambda)$
Likelihood of the parameters

\[ p(x|\lambda) \]

\[ p(x=15|\lambda) \]

\[ p(x=20|\lambda) \]
Likelihood of the parameters

\[ p(x|\lambda) \]

- \[ p(x=15|\lambda) \]
- \[ p(x=20|\lambda) \]
- \[ p(x=30|\lambda) \]
Posterior distribution: $p(\lambda|x_1)$
Posterior distribution: $p(\lambda | x_1, x_2)$
Posterior distribution: \( p(\lambda | x_1, x_2, x_3) \)
Posterior distribution: $p(\lambda|x_1, x_2, x_3, x_4)$
Posterior distribution: $p(\lambda|x_1, x_2, x_3, x_4, x_5)$
Summarising the posterior distribution

$p(x|\lambda)$

maximum a posteriori (MAP)

error-bars
Summarising the posterior distribution

\[ p(x|\lambda) \]

**Gaussian approximation**
Summarising the posterior distribution

$$\lambda$$ /ms

$p(x|\lambda)$
samples from posterior
Question

- Record inter-spike interval measurements, \( x \)
- As before: absolute refractory of 5 ms & ISIs above 50 ms not recorded
- We know if the neuron is...
  - quiescent: \( x \) follows an exponential distribution with time-scale \( \lambda_0 = 25 \text{ms} \)
  - bursting: \( x \) follows an exponential distribution with time-scale \( \lambda_1 = 5 \text{ms} \)
- You observe a single ISI, \( x = 15 \text{ms} \). Is the neuron in a bursting state?
Question

• Record inter-spike interval measurements, $x$

• As before: absolute refractory of 5ms & ISIs above 50ms not recorded

• We know if the neuron is...
  – quiescent: $x$ follows an exponential distribution with time-scale $\lambda_0 = 25$ms
  – bursting: $x$ follows an exponential distribution with time-scale $\lambda_1 = 5$ms

• You observe a single ISI, $x = 15$ms. Is the neuron in a bursting state?
  
  Intuition: should be close to 50:50
Mathematical solution

Introduce latent variable: \[ b = 0 \] not bursting
\[ b = 1 \] bursting

Generative model

\[ p(b = 0) = \frac{1}{2} \]

\[ p(x|b, \lambda_0, \lambda_1) = \frac{1}{Z(\lambda_b)} \exp\left(-\frac{x}{\lambda_b}\right) \]

Recognition model: inference

\[ p(b = 1|x, \lambda_0, \lambda_1) = \frac{p(x|b=1, \lambda_1)p(b=1)}{p(x|\lambda_0, \lambda_1)} \]

\[ = \frac{1}{1 + \frac{Z(\lambda_1)}{Z(\lambda_0)} \exp\left(-x\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)\right)} \]
Graphical solution
Graphical solution

\[ p(x|b, \lambda_b) \]

\[ p(b=0|x) \]
Graphical solution

\[ p(x|b, \lambda_b) \]

For \( b = 0 \) and \( b = 1 \):

- \[ p(b=0|x) = 0.54 \]
Generative models in neuroscience

- data analysis (spike sorting, fMRI, etc.)
- ideal observer models in psychophysics
- neural encoding models
- neural decoding models
- **Bayesian Brain** - the brain is making inferences about the world using probabilistic calculus