Estimation & Maximum Likelihood

Jonathan Pillow

Mathematical Tools for Neuroscience (NEU 314)
Fall, 2021

lecture 19

warmup problems

P(x,y)

y \ x	0	1	2
2	0	0	0
1	0.25	0.25	0
0	0.25	0	0.25

1) Compute the marginals P(x) and P(y)
2) Compute the means, E(x) and E(y)
3) Compute the covariance, Cov(x,y)
   (are x and y positively correlated, negatively correlated, or uncorrelated?)
4) Compute the P(x)P(y), the independent approximation to P(x,y)
   (Is this joint distribution independent or dependent?)
5) Compute the entropy of P(x)? What is the entropy of P(x l y = 1)
6) Compute the mutual information between x and y using one of the four formulas we discussed in class. (Can you remember them all?)

Estimation

parameter
("stimulus")

θ

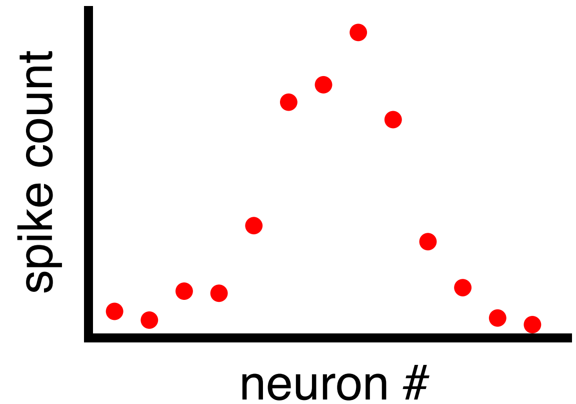
model

$p(m|\theta)$

→

measured dataset
("population response")

$m = \{r_1, r_2, \ldots, r_n\}$



An *estimator* is a function $f : m \longrightarrow \hat{\theta}$

- often we will write $\hat{\theta}(m)$ or just $\hat{\theta}$

Estimation

parameter
("stimulus")

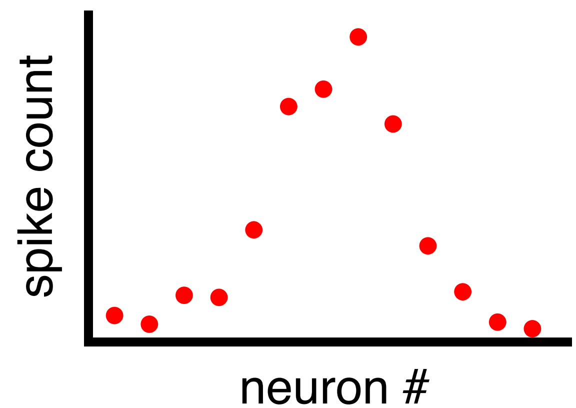
θ

model

$p(m|\theta)$

measured dataset
("population response")

$m = \{r_1, r_2, \ldots, r_n\}$



Maximum likelihood estimator

$\hat{\theta}_{ML}$ = value of θ at which the *likelihood* $p(m|\theta)$ is maximal

Maximum a posteriori (MAP) estimator

$\frac{p(m|\theta)p(\theta)}{p(m)}$

$\hat{\theta}_{MAP}$ = value of θ at which the *posterior* $p(\theta|m)$ is maximal

Properties of an estimator

"expected" value
(average over draws of m)

bias: $$b(\theta) = \mathbb{E}[\hat{\theta}] - \theta$$

- "unbiased" if bias=0

variance: $$\mathrm{var}(\theta) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right]$$

- "consistent" if bias and variance both go to zero asymptotically

Q1: what is the variance of the estimator $\hat{\theta}(m) = 7$
(i.e., estimate is 7 for all datasets *m*)

Q2: what is the bias of this estimator?

Simple Example: Gaussian noise & prior

1. Likelihood $p(m|\theta)$

additive Gaussian noise

$$m = \theta + noise$$

$$p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

2. Prior $p(\theta)$

zero-mean Gaussian

$$p(\theta) = \frac{1}{\sqrt{2\pi}\gamma} e^{\frac{-\theta^2}{2\gamma^2}}$$

posterior distribution

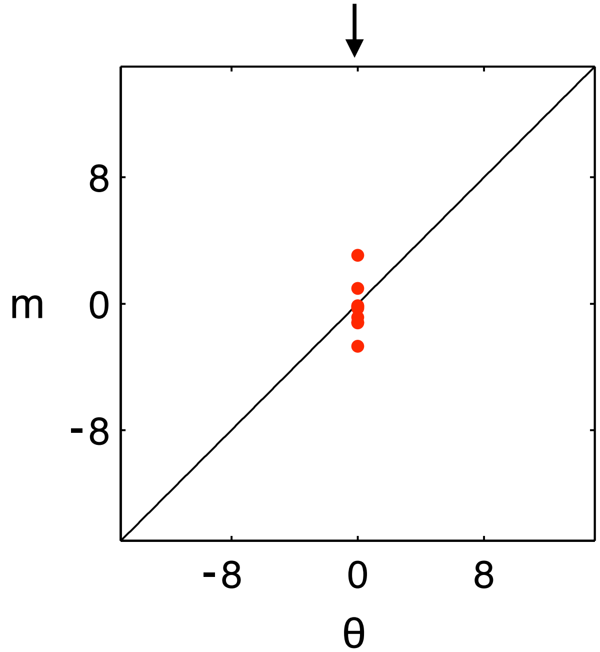
$$p(\theta|m) = \mathcal{N}\left(\frac{\gamma^2}{\sigma^2+\gamma^2} m, \frac{\sigma^2\gamma^2}{\sigma^2+\gamma^2}\right)$$

mean variance

Observation model

$$p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

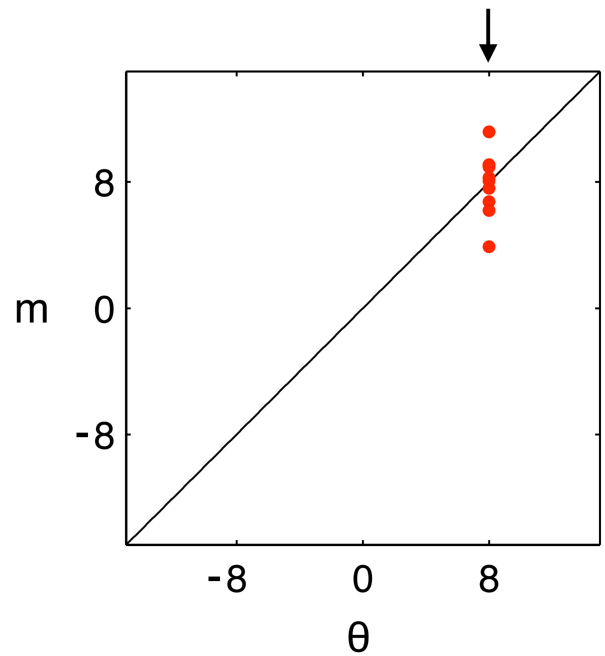
$$m = \theta + noise$$



Observation model

$$p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

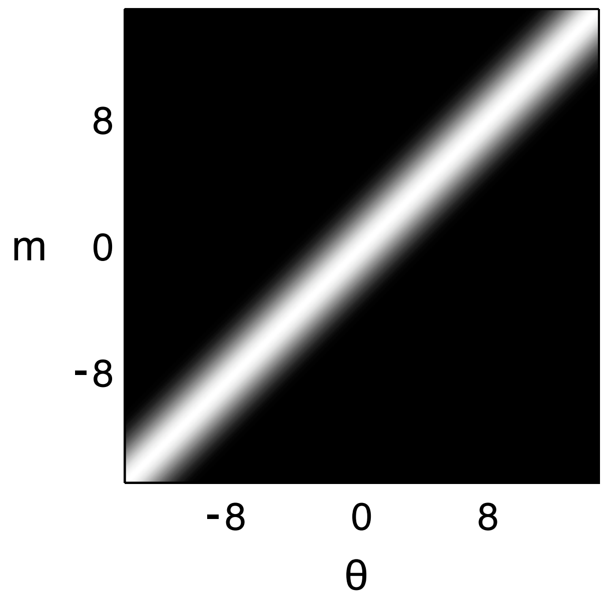
$$m = \theta + noise$$



Observation model

$$p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

$$m = \theta + noise$$



Likelihood: considering as a function of θ

$$p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

$$m = \theta + noise$$