Bayesian Estimation & Information Theory

Jonathan Pillow

Mathematical Tools for Neuroscience (NEU 314)
Spring, 2016

lecture 18

Bayesian Estimation

three basic ingredients:

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

} jointly determine the posterior $p(\theta|m)$

3. Loss function $L(\hat{\theta}, \theta)$ } "cost" of making an estimate $\hat{\theta}$ if the true value is θ

- fully specifies how to generate an estimate from the data

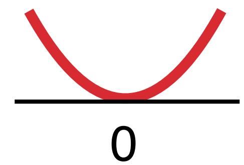
Bayesian estimator is defined as:

$$\hat{\theta}(m) = \arg\min_{\hat{\theta}} \int L(\hat{\theta}, \theta) p(\theta|m) d\theta$$

"Bayes' risk"

Typical Loss functions and Bayesian estimators

1. $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ squared error loss



need to find $\hat{\theta}$ minimizing the expected loss: $\int (\hat{\theta} - \theta)^2 p(\theta|m) d\theta$

Differentiate with respect to $\hat{\theta}$ and set to zero:

$$\int 2(\hat{\theta} - \theta) p(\theta|m) d\theta = 0$$

$$\int \hat{\theta}\, p(\theta|m) d\theta = \int \theta\, p(\theta|m) d\theta$$

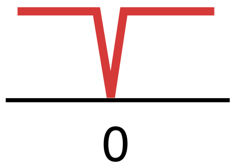
$$\hat{\theta} = \int \theta\, p(\theta|m) d\theta$$ "posterior mean"

also known as Bayes' Least Squares (BLS) estimator

Typical Loss functions and Bayesian estimators

2. $$L(\hat{\theta}, \theta) = 1 - \delta(\hat{\theta} - \theta)$$

"zero-one" loss
(1 unless $\hat{\theta} = \theta$)



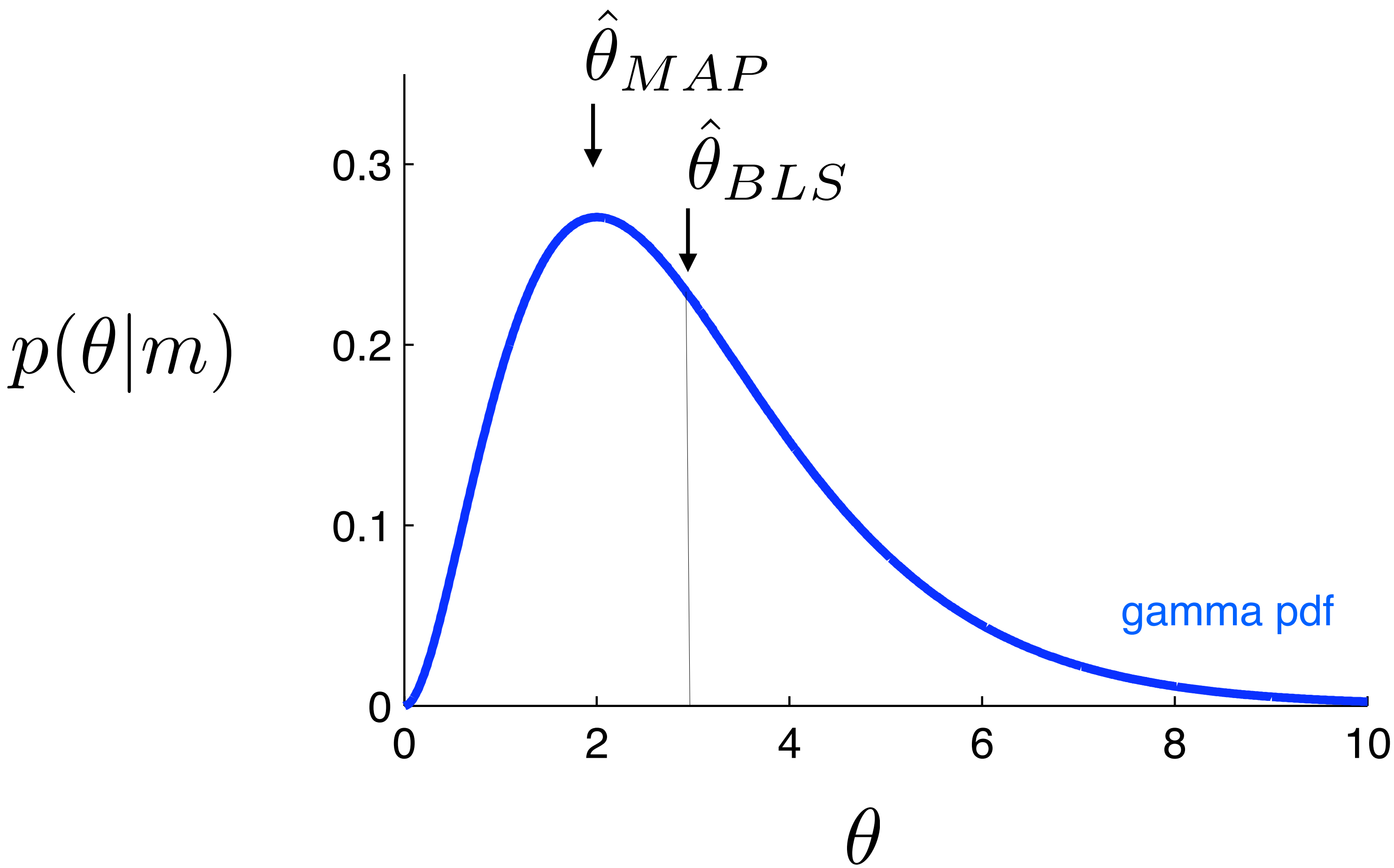
expected loss: $$\int (1 - \delta(\hat{\theta} - \theta)) p(\theta | m) d\theta$$

$$= 1 - p(\hat{\theta} | m)$$

which is minimized by: $$\hat{\theta} = \arg\max_{\theta} p(\theta | m)$$

- posterior maximum (or "mode").
- known as maximum a posteriori (MAP) estimate.

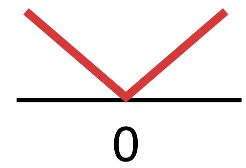
MAP vs. Posterior Mean estimate:



Note: posterior maximum and mean not always the same!

Typical Loss functions and Bayesian estimators

3. $L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$ "L1" loss



expected loss: $$\int |\hat{\theta} - \theta| \; p(\theta|m) d\theta$$

HW problem: What is the Bayesian estimator for this loss function?

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

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

3. Loss function: doesn't matter (all agree here)

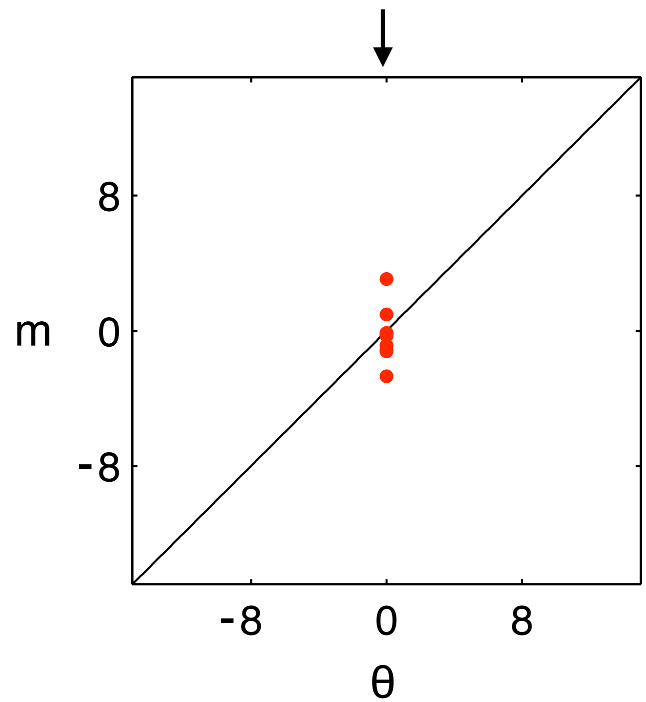
posterior distribution

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

MAP estimate variance

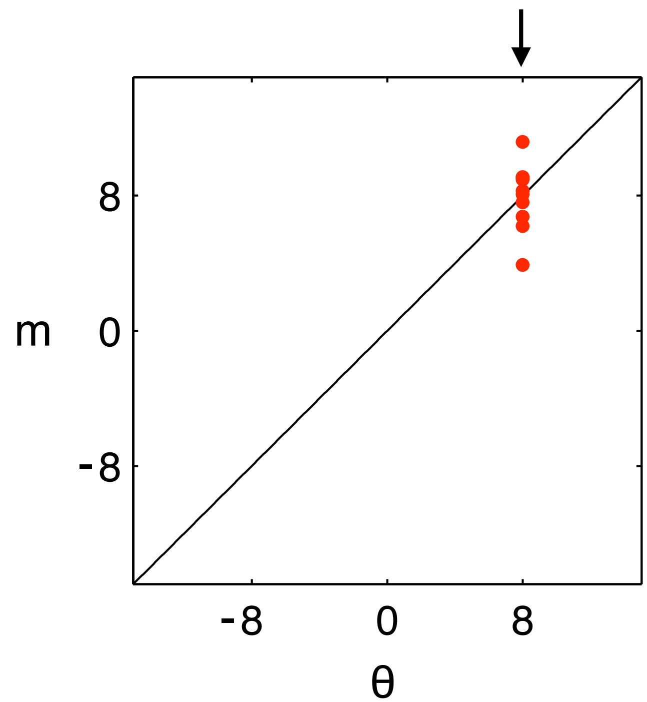
Likelihood

$$m = \theta + noise$$



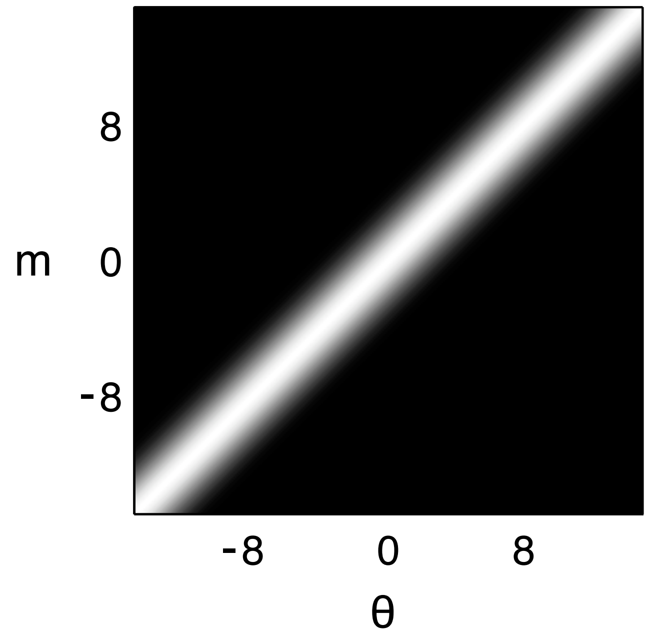
Likelihood

$$m = \theta + noise$$



Likelihood

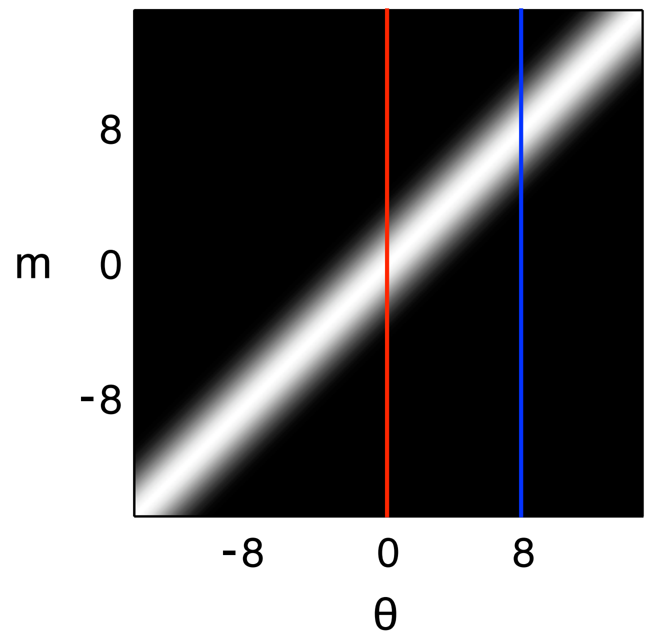
$$m = \theta + noise \qquad p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

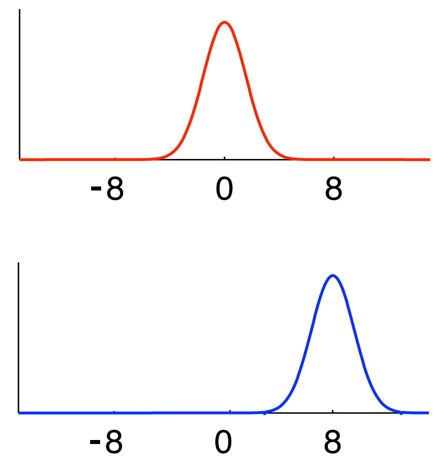


Likelihood

$$m = \theta + noise$$

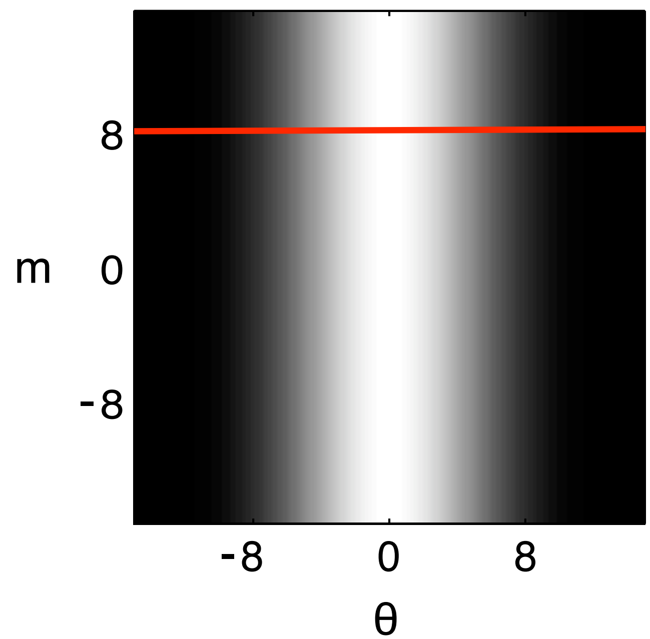
$$p(m|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(m-\theta)^2}{2\sigma^2}}$$

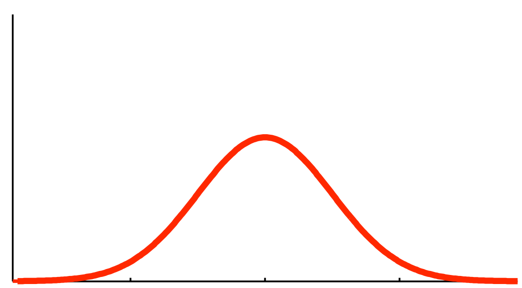




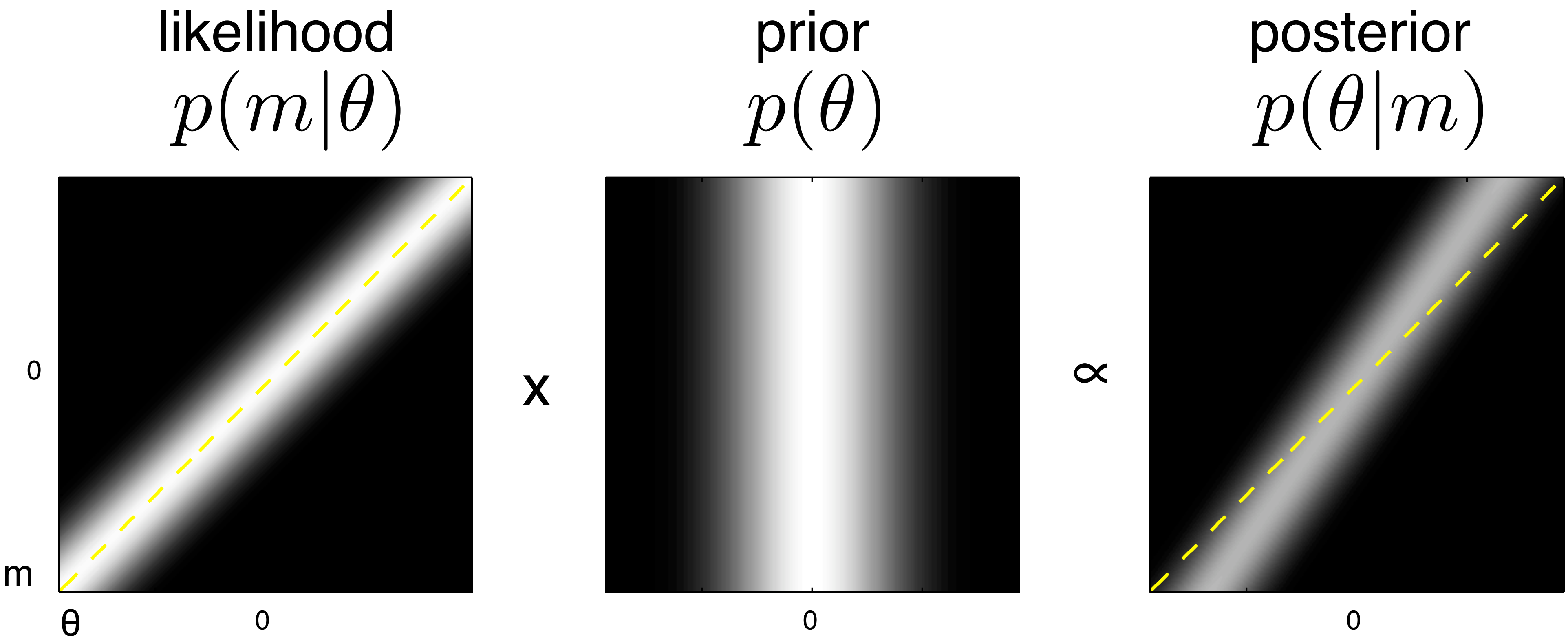
Prior

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

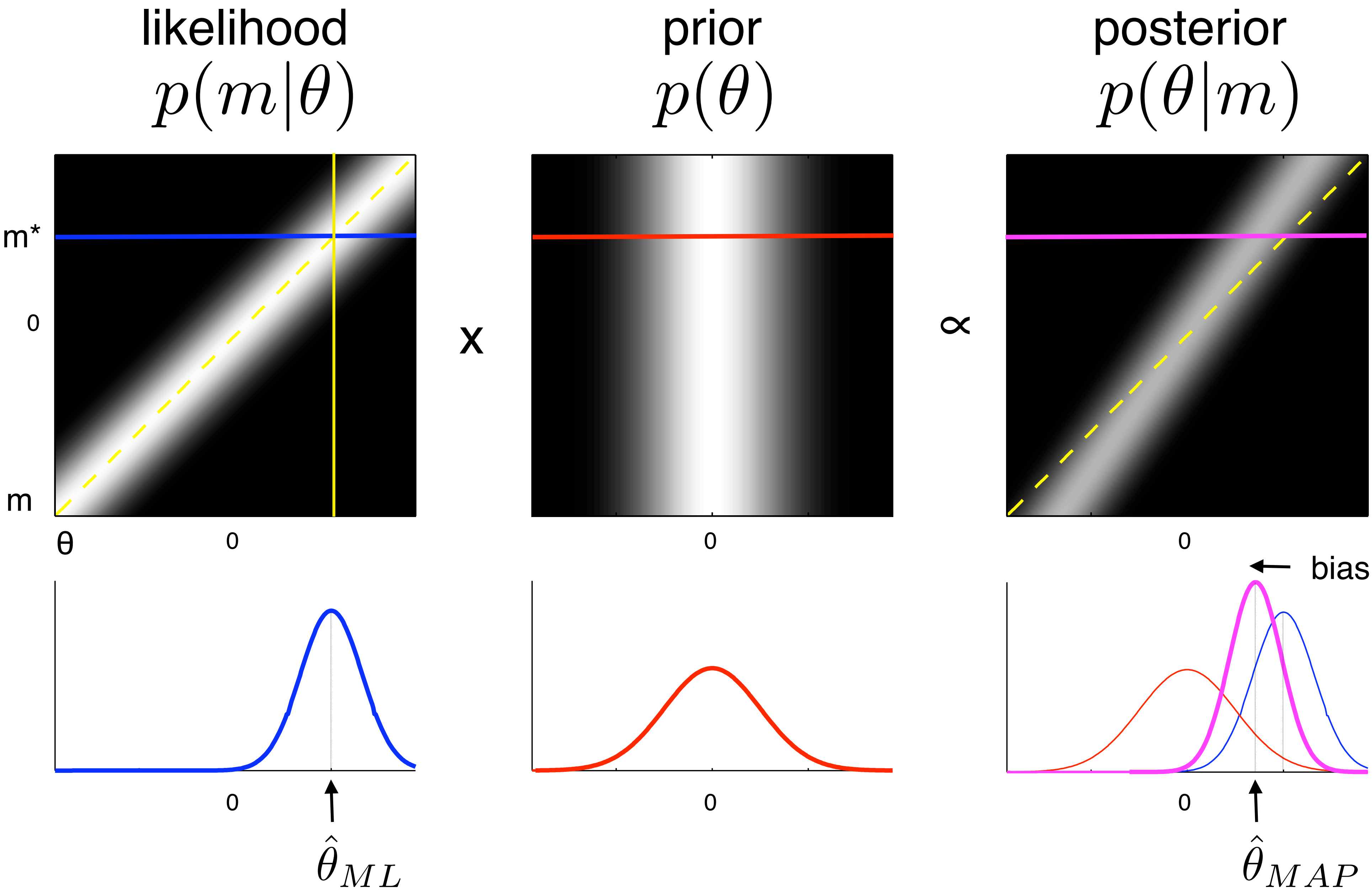




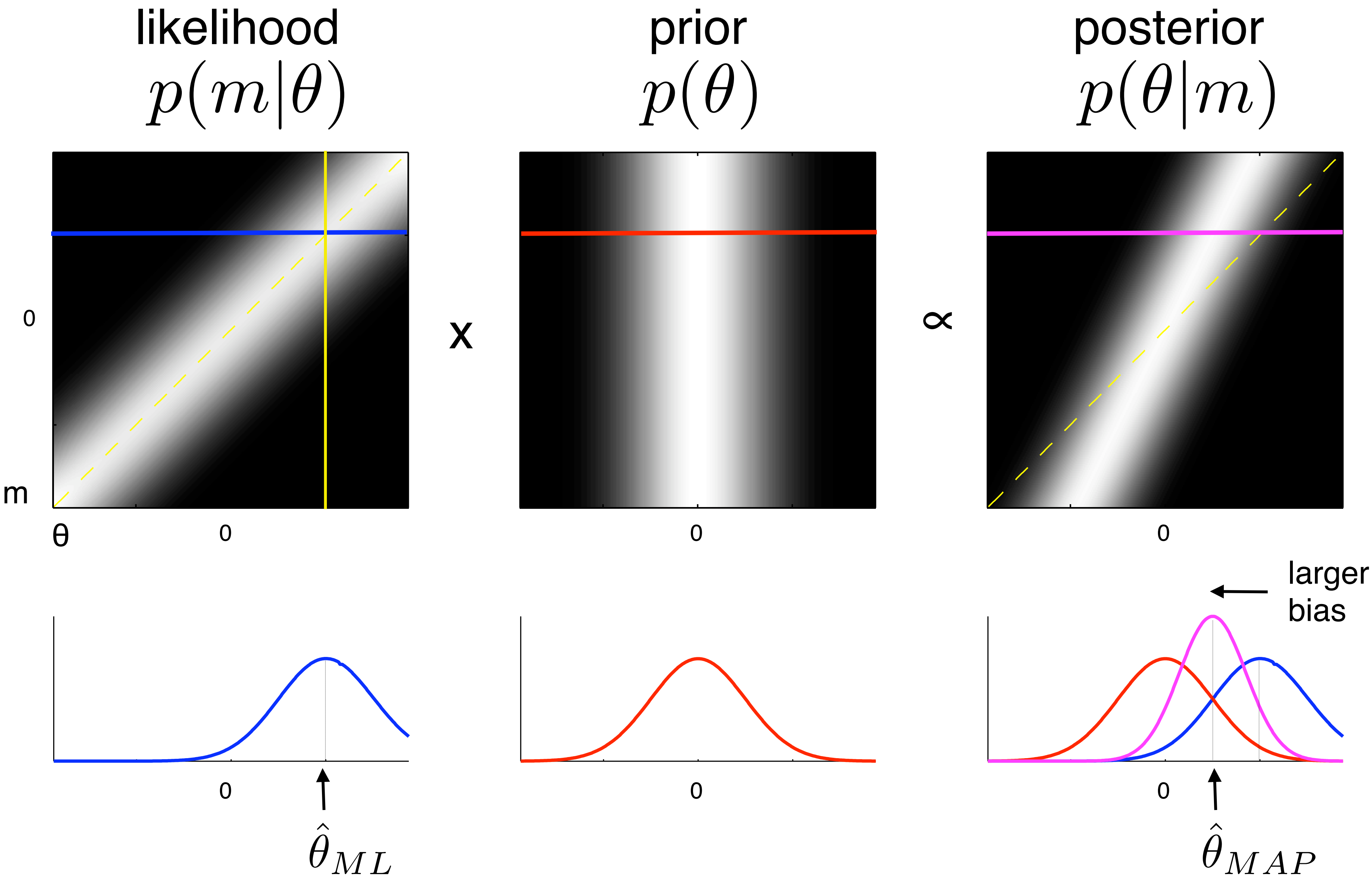
Computing the posterior



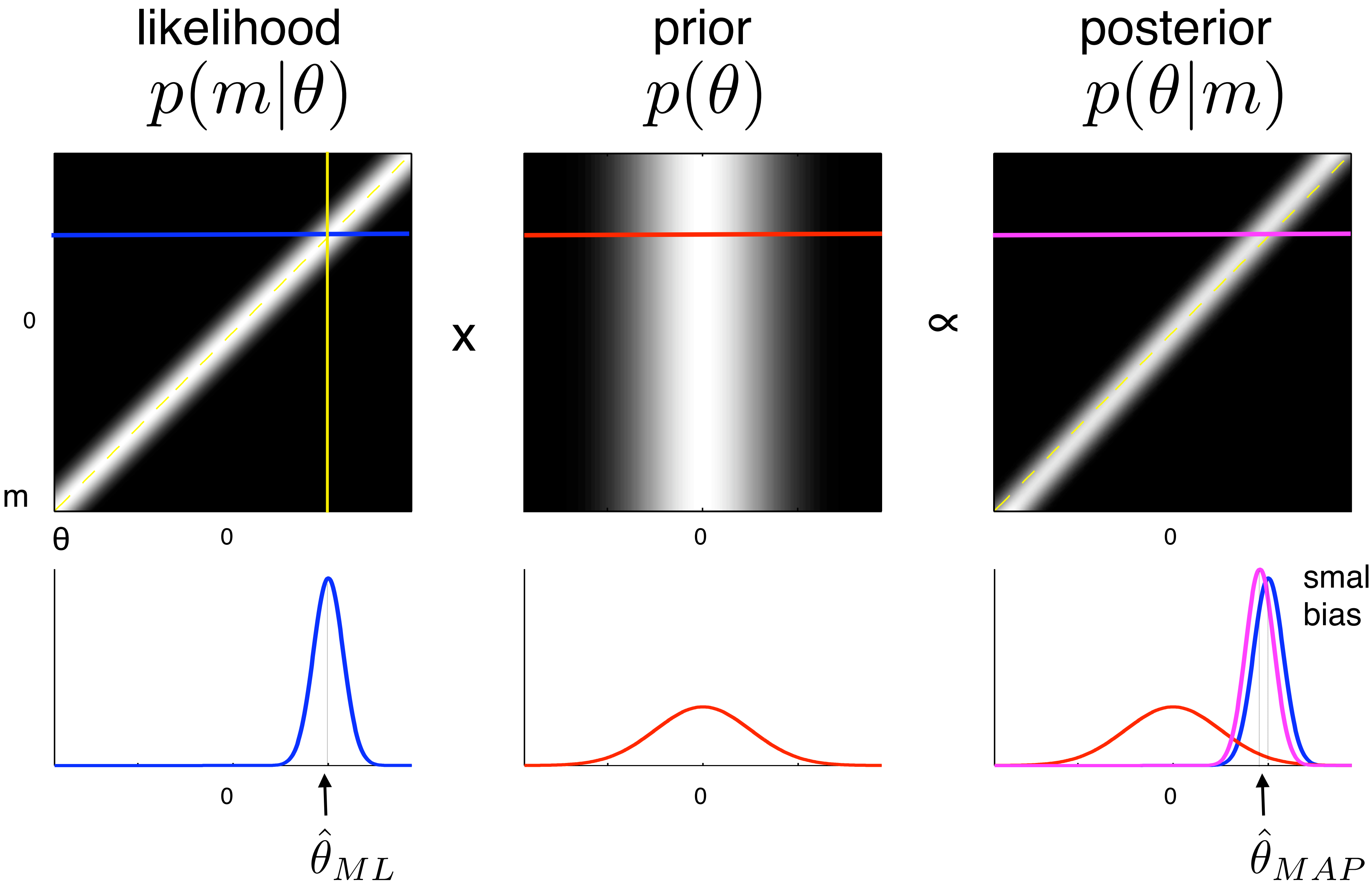
Making an Bayesian Estimate:



High Measurement Noise: large bias



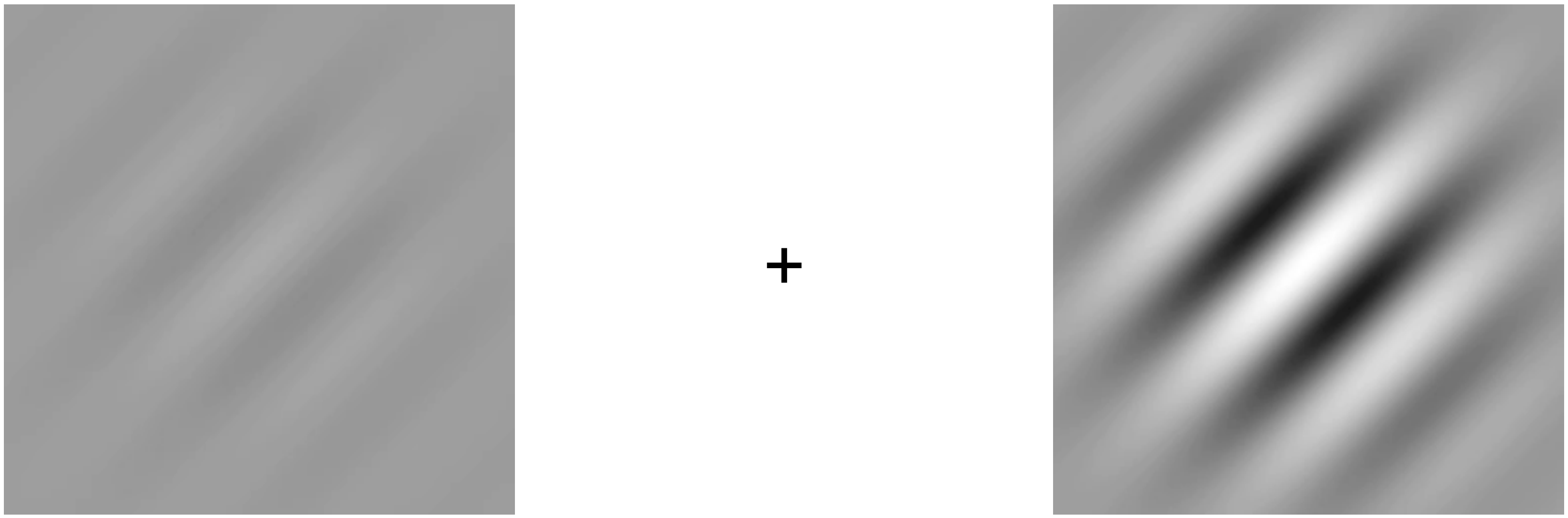
Low Measurement Noise: small bias



Bayesian Estimation:

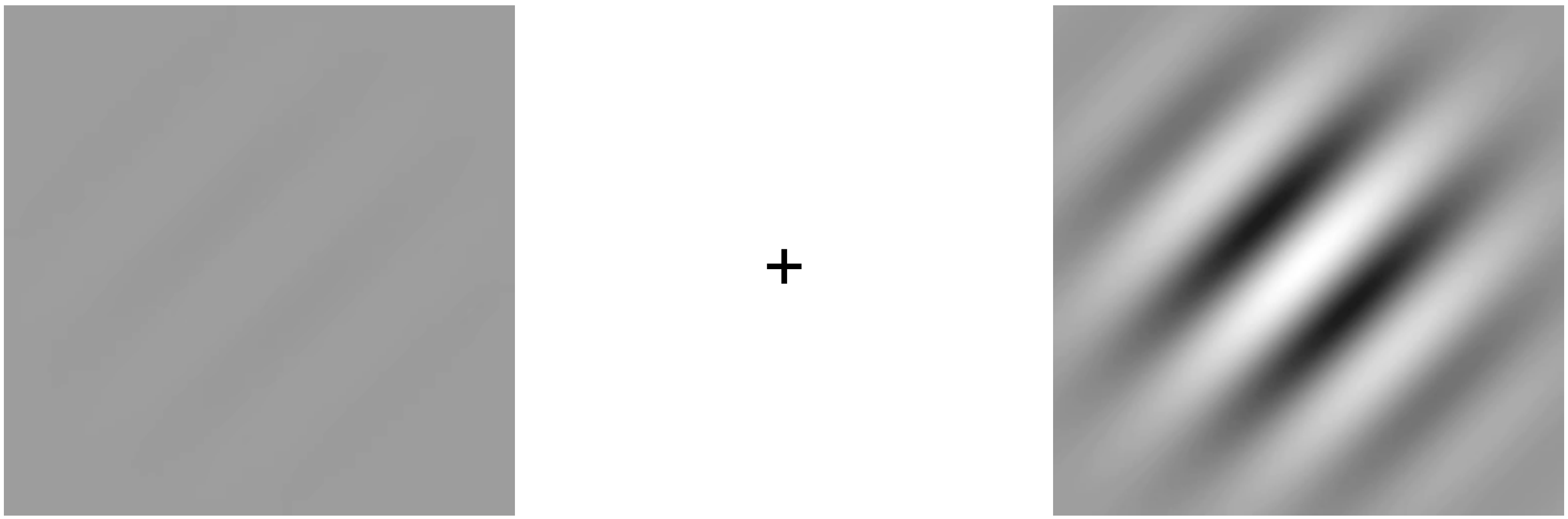
- Likelihood and prior combine to form posterior
- Bayesian estimate is always biased *towards* the prior (from the ML estimate)

Application #1: Biases in Motion Perception



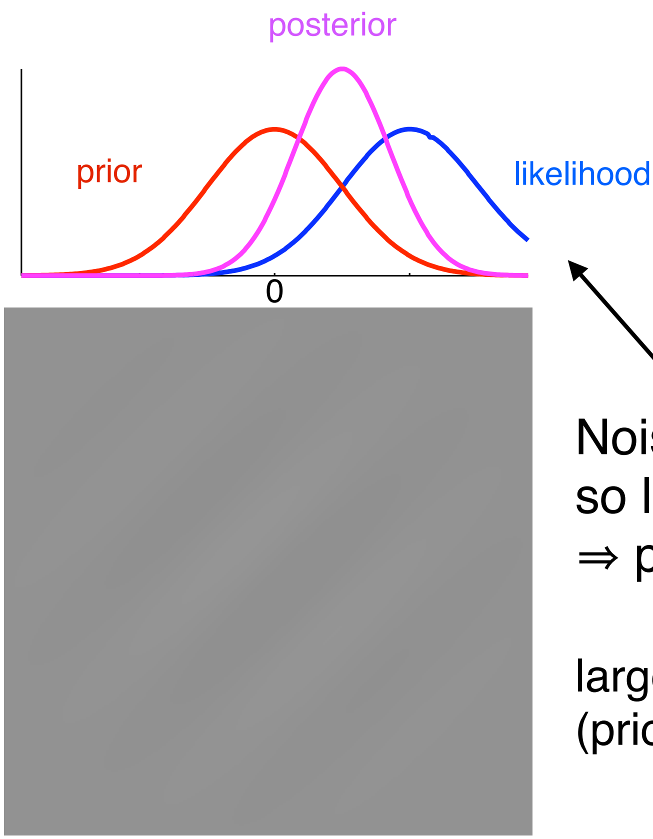
Which grating moves faster?

Application #1: Biases in Motion Perception



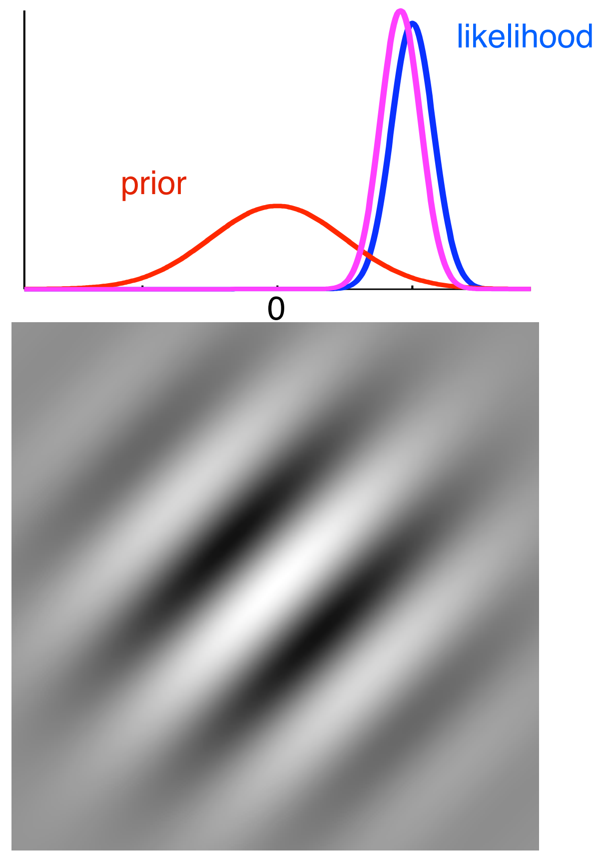
Which grating moves faster?

Explanation from Weiss, Simoncelli & Adelson (2002):



Noisier measurements, so likelihood is broader ⇒ posterior has

larger shift toward 0 (prior = no motion)



- In the limit of a zero-contrast grating, likelihood becomes infinitely broad ⇒ percept goes to zero-motion.

- Claim: explains why people actually speed up when driving in fog!

summary

- 3 ingredients for Bayesian estimation (prior, likelihood, loss)
- Bayes' least squares (BLS) estimator (posterior mean)
- maximum a posteriori (MAP) estimator (posterior mode)
- accounts for stimulus-quality dependent bias in motion perception (Weiss, Simoncelli & Adelson 2002)