# Estimation & Maximum Likelihood

Jonathan Pillow

Mathematical Tools for Neuroscience (NEU 314)
Spring, 2016

lecture 15

# leftovers

- Gaussian facts
- covariance matrices

# the amazing Gaussian

## What else about Gaussians is awesome?

Gaussian family closed under many operations:

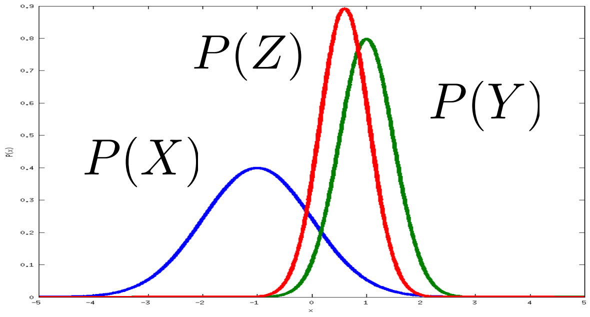
1. scaling: $X \sim Gaussian \implies aX$ is Gaussian
2. sums: $X, Y \sim Gaussian \implies X + Y$ is Gaussian

(thus, any linear function Gaussian RVs is Gaussian)

3. products of Gaussian distributions

Gaussian density

$$X, Y \sim Gaussian \implies P(X)P(Y) \propto P(Z)$$



# the amazing Gaussian

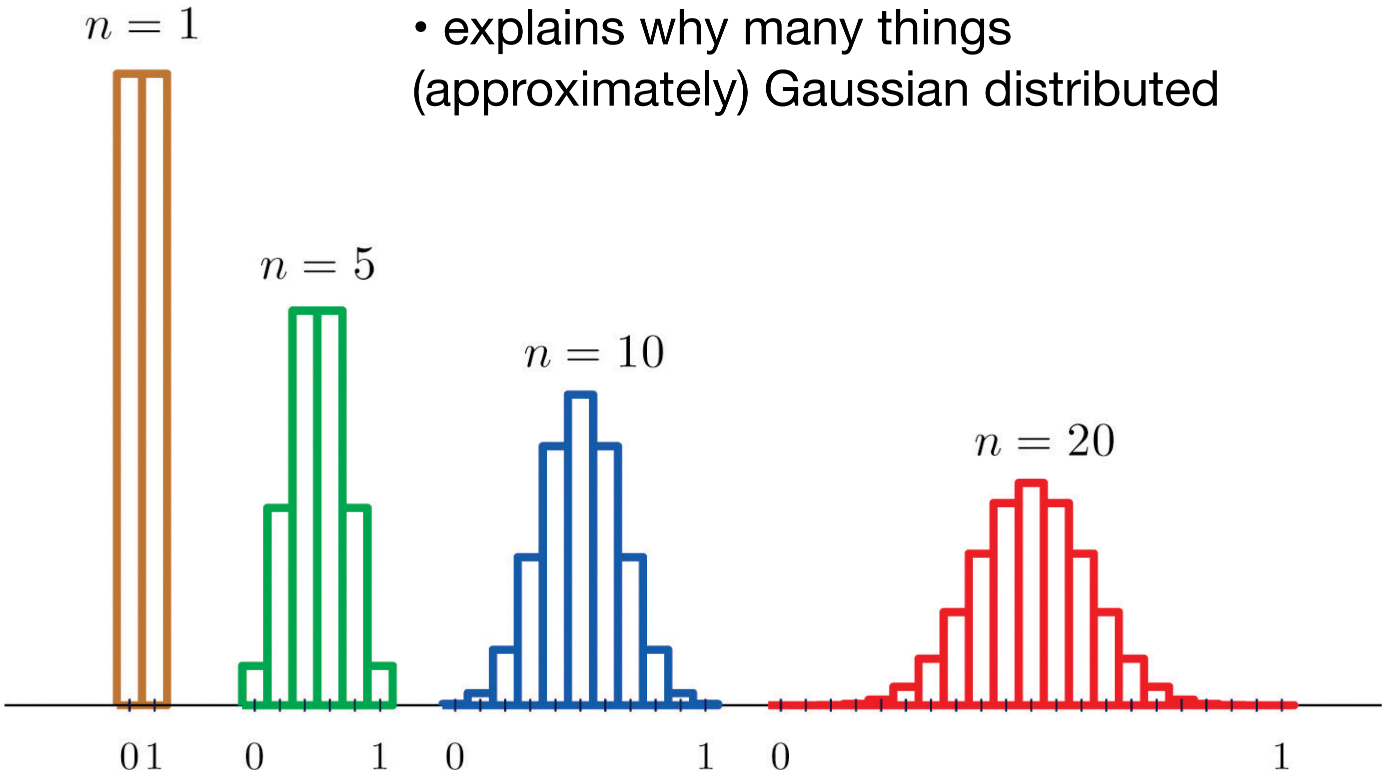
4. Average of many (non-Gaussian) RVs is Gaussian!

Central Limit Theorem: $$S_n = \frac{X_1 + \cdots X_n}{n}$$

$$\sqrt{n}(S_n - \mu) \longrightarrow$$ standard Gaussian

- explains why many things (approximately) Gaussian distributed

coin flipping:



http://statwiki.ucdavis.edu/Textbook_Maps/General_Statistics/Shafer_and_Zhang's_Introductory_Statistics/06%3A_Sampling_Distributions/6.2_The_Sampling_Distribution_of_the_Sample_Mean

# the amazing Gaussian

Multivariate Gaussians: $\vec{X} \sim \mathcal{N}(\overset{\text{mean}}{\vec{\mu}}, \overset{\text{cov}}{C})$

(The random variable X is distributed according to a Gaussian distribution)

$$P(\vec{X} = \vec{x}) = \frac{1}{\sqrt{|2\pi C|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^\top C^{-1}(\vec{x}-\vec{\mu})}$$

5. Marginals and conditionals (“slices”) are Gaussian

6. Linear projections: $\vec{Y} = A\vec{X} \quad \Longrightarrow \quad \vec{Y} \sim \mathcal{N}(A\vec{\mu}, ACA^\top)$

# multivariate Gaussian

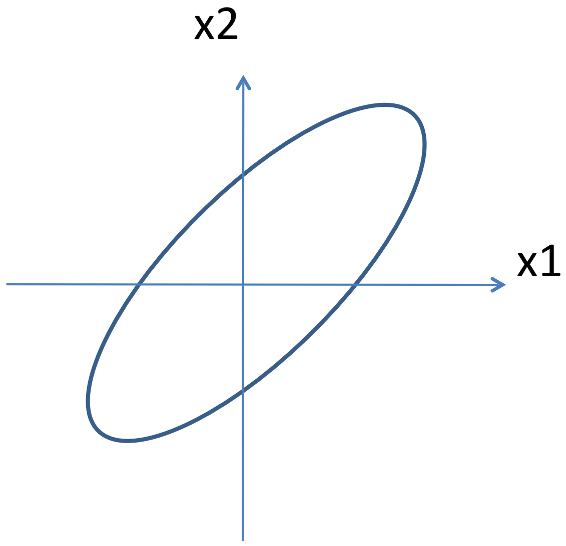
$$P(\vec{X} = \vec{x}) = \frac{1}{\sqrt{|2\pi C|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^\top C^{-1}(\vec{x}-\vec{\mu})}$$

# covariance

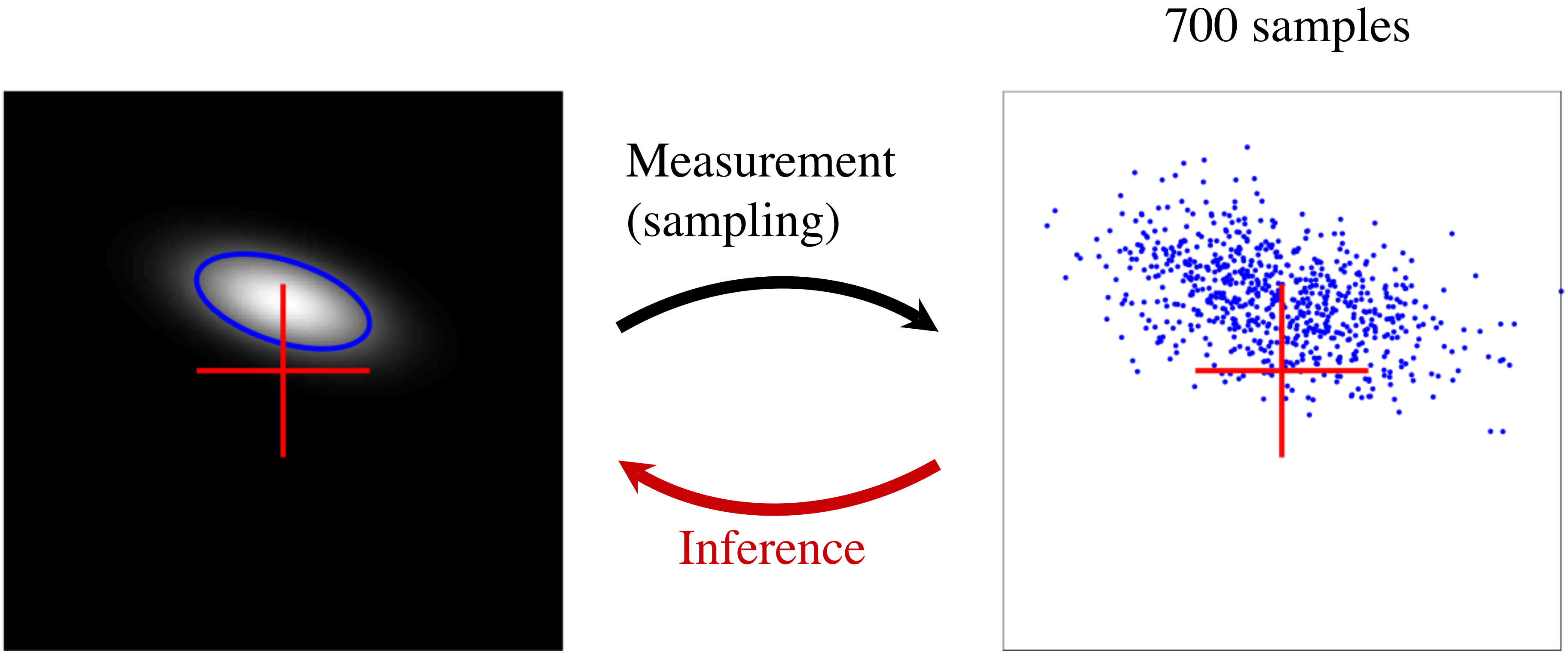
$$E[(\vec{x} - E[\vec{x}])(\vec{x} - E[\vec{x}])^T]$$

after mean correction:

$$\begin{bmatrix} E(x_1^2) & E(x_1 x_2) \\ E(x_1 x_2) & E(x_2^2) \end{bmatrix}$$



# bivariate Gaussian



700 samples

Measurement (sampling)

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]

# Estimation

parameter

("stimulus")

$$\theta$$

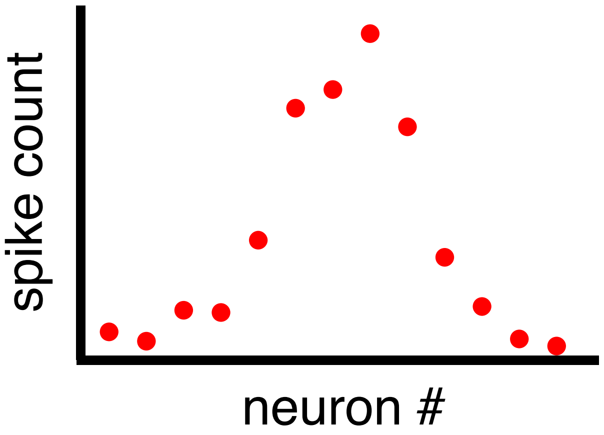
model

$$p(m|\theta)$$

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 $\hat{\theta}$

# 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

**Q**: what is the variance of the estimator $\hat{\theta}(m) = 7$

(i.e., estimate is 7 for all datasets *m*)

# Properties of an estimator

“expected” value
(average over draws of m)

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

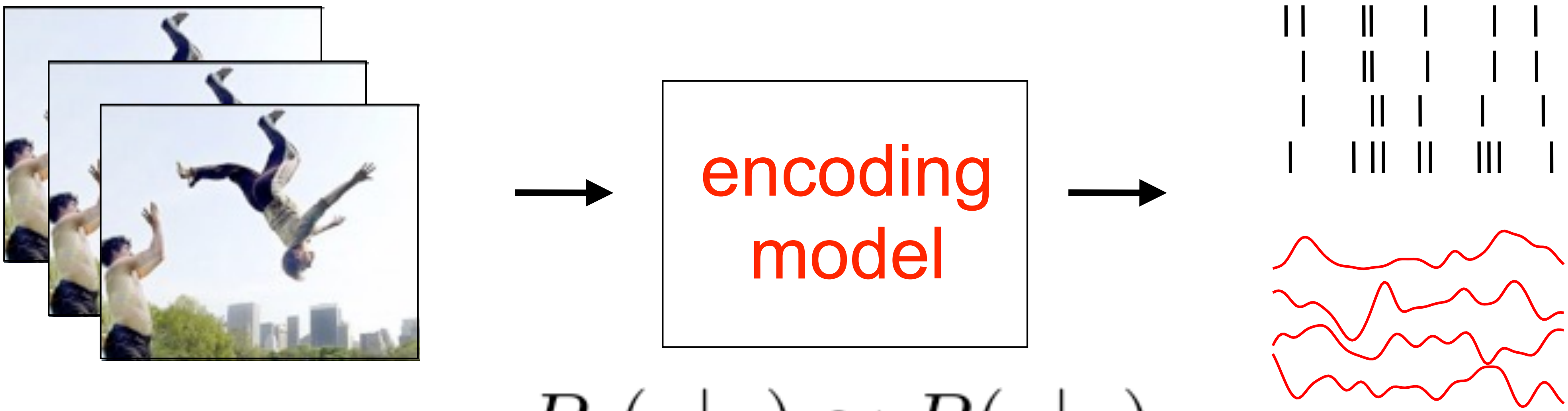
- “unbiased” if bias=0

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

mean squared error (MSE) $\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathrm{bias}^2 + \mathrm{variance}$

# model-based approach



$x$

stimuli

$$P_\theta(y|x) \approx P(y|x)$$

$y$

neural responses

**Goal**: find model that approximates the conditional distribution

(we care about uncertainty as well as the average y given x)

# Example 1: linear Poisson neuron

spike count $y \sim Poiss(\lambda)$

spike rate $\lambda = \theta x$

parameter stimulus

encoding model: $$\begin{aligned} P(y|x,\theta) &= \frac{1}{y!}\lambda^y e^{-\lambda} \\ &= \frac{1}{y!}(\theta x)^y e^{-(\theta x)} \end{aligned}$$

# Summary

- covariance
- Gaussians
- Poisson distribution (mean = variance)
- estimation
- bias
- variance
- maximum likelihood estimator