MEDICAL IMAGING INFORMATICS: Lecture # 1

Basics of Medical Imaging Informatics: Estimation Theory

Norbert Schuff
Professor of Radiology
VA Medical Center and UCSF
Norbert.schuff@ucsf.edu

UCSF VA
Department of
Radiology & Biomedical Imaging

What Is Medical Imaging Informatics?

- Signal Processing
  - Digital Image Acquisition
  - Image Processing and Enhancement
- Data Mining
  - Computational anatomy
  - Statistics
  - Databases
  - Data-mining
  - Workflow and Process Modeling and Simulation
- Data Management
  - Picture Archiving and Communication System (PACS)
  - Imaging Informatics for the Enterprise
  - Image-Enabled Electronic Medical Records
  - Radiology Information Systems (RIS) and Hospital Information Systems (HIS)
  - Quality Assurance
  - Archive Integrity and Security
- Data Visualization
  - Image Data Compression
  - 3D, Visualization and Multi-media
  - DICOM, HL7 and other Standards
- Teleradiology
  - Imaging Vocabularies and Ontologies
  - Transforming the Radiological Interpretation Process (TRIP)[2]
  - Computer-Aided Detection and Diagnosis (CAD).
  - Radiology Informatics Education
- Etc.

UCSF Department of Radiology & Biomedical Imaging
VA

What Is The Focus Of This Course?

Learn using computational tools to maximize information and knowledge gain


Challenge: Maximize Information Gain

1. Q: How can we estimate quantities of interest from a given set of uncertain (noise) measurements?

   A: Apply estimation theory (1st lecture by Norbert)

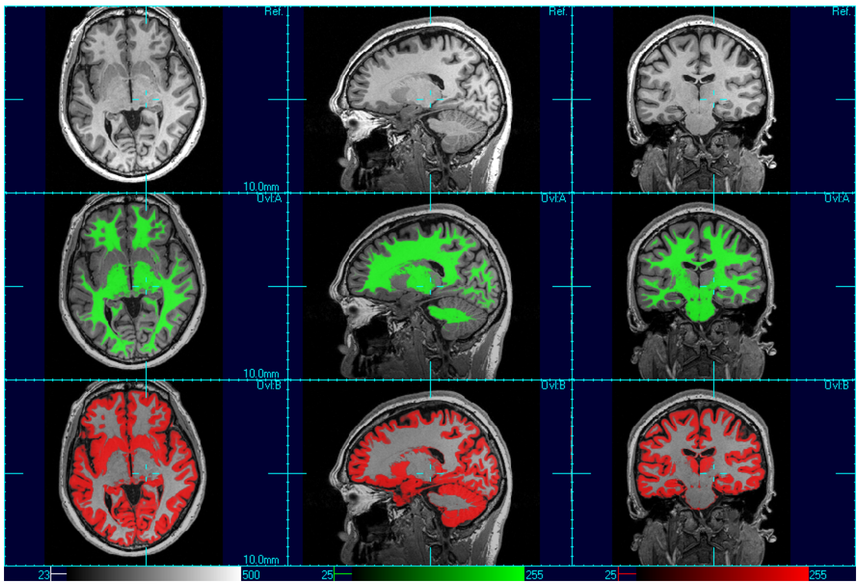
2. Q: How can we measure (quantify) information?

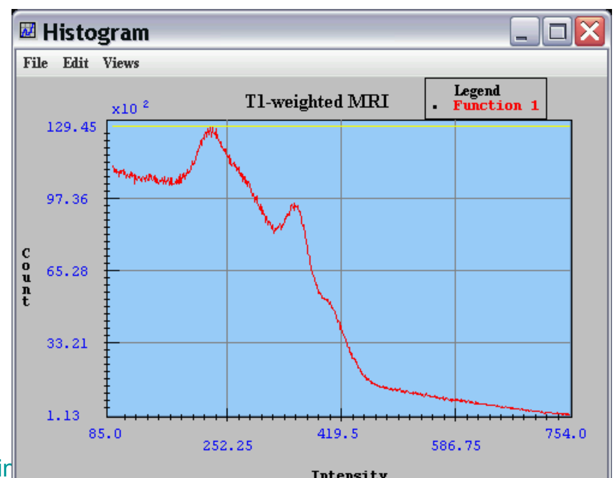
   A: *Apply information theory* (2nd lecture by Wang)

UCSF VA
Department of
Radiology & Biomedical Imaging

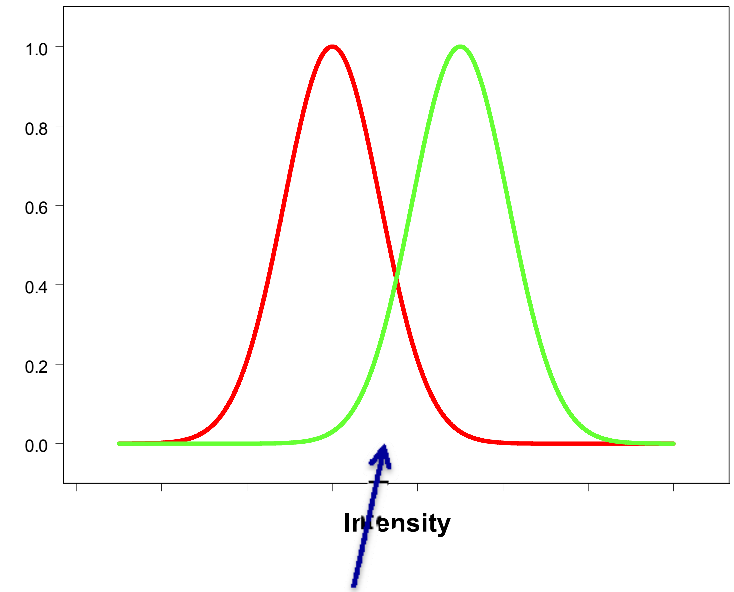
Estimation Theory: Motivation Example I

Gray/White Matter Segmentation





Hypothetical Histogram



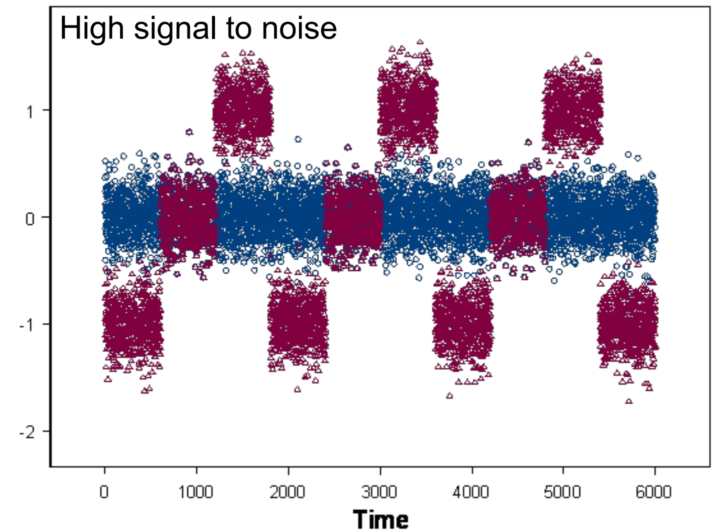
GM/WM overlap 50:50;
Can we do better than flipping a coin?

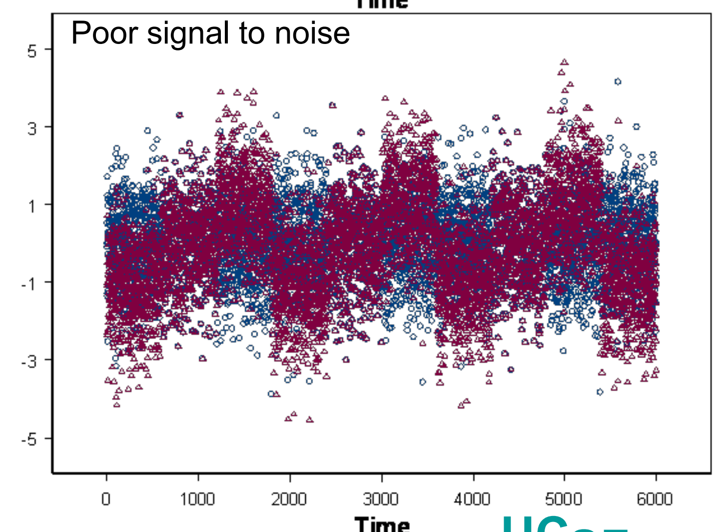
UCSF VA
Department of
Radiology & Biomedical Imaging

Estimation Theory: Motivation Example II

Goal: Capture dynamic signal on a static background

D. Feinberg Advanced MRI Technologies, Sebastopol, CA

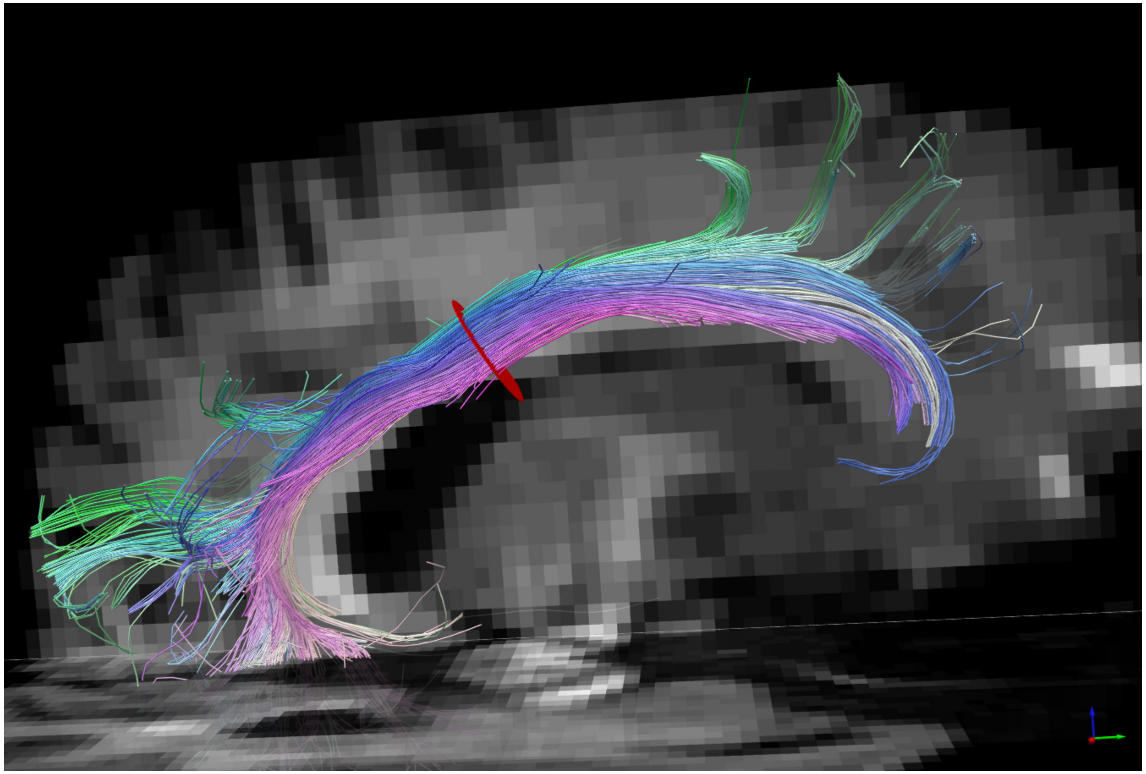




UCSF
Department of
Radiology & Biomedical Imaging
VA

Estimation Theory: Motivation Example III

Diffusion Spectrum Imaging – Human Cingulum Bundle

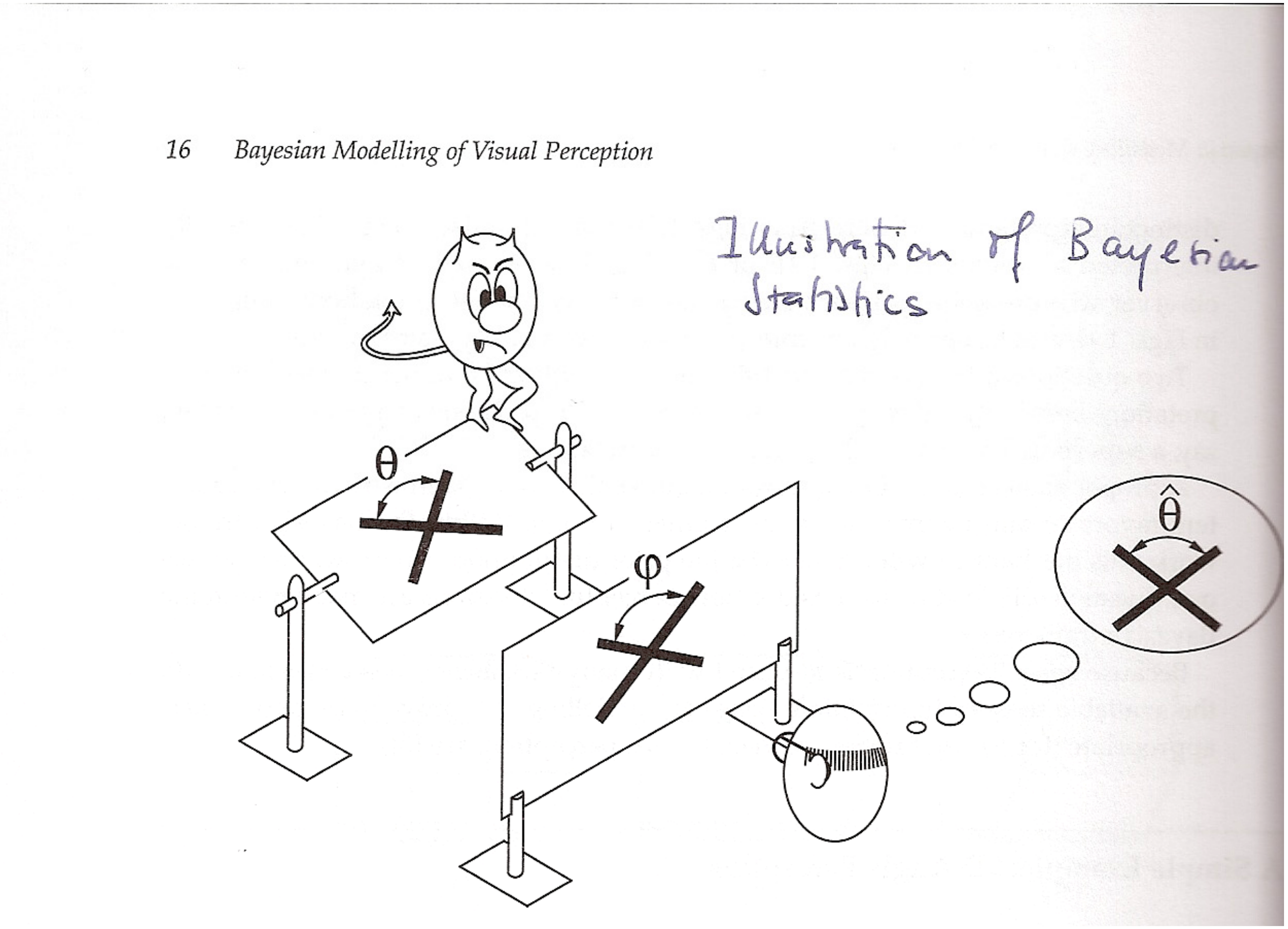


Dr. Van Wedeen, MGH

Goal:
Capture directions of fiber bundles

UCSF
Department of
Radiology & Biomedical Imaging
VA

Basic Concepts of Modeling



Θ: target of interest and unknown

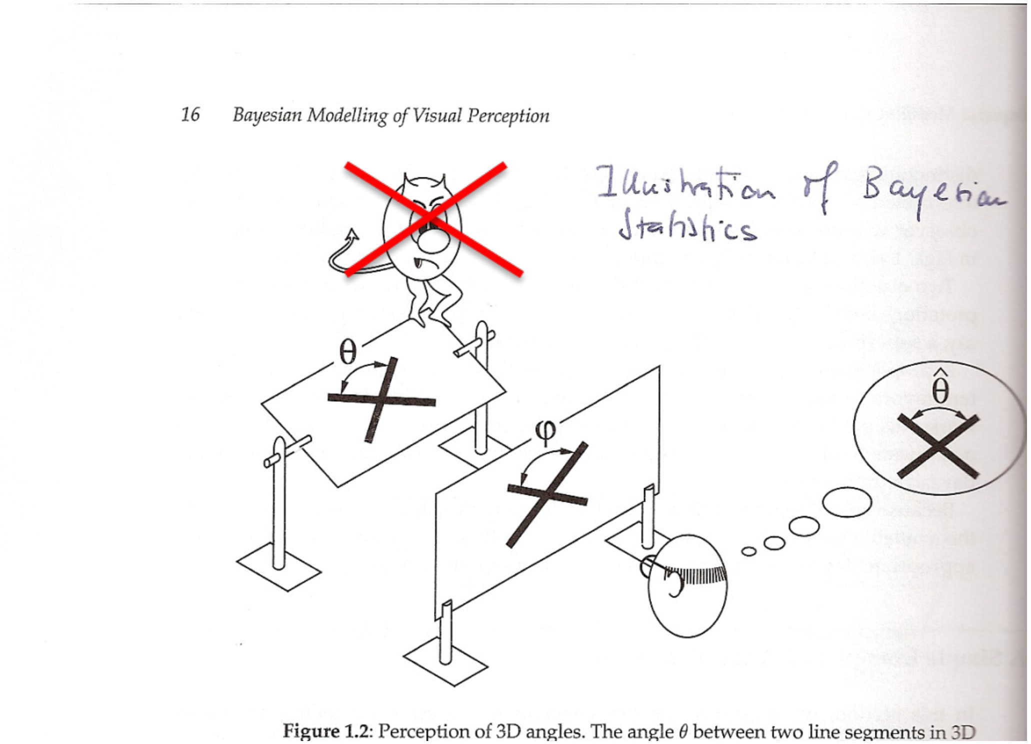
ρ: measurement

$\widehat{\Theta}$: Estimator - a good guess of Θ based on measurements

Cartoon adapted from: Rajesh P. N. Rao, Bruno A. Olshausen Probabilistic Models of the Brain. MIT Press 2002.

UCSF Department of Radiology & Biomedical Imaging
VA

Deterministic Model



$$\varphi_{\mathrm{N}} = \mathbf{H}\boldsymbol{\theta}_{M} + noise_{\mathrm{N}}$$

N = number of measurements
M = number of states, M=1 is possible
Usually N > M and $|noise||^2 > 0$

The model is deterministic, because discrete values of Θ are solutions.

Note:
1) we make no assumption about Θ
2) Each value is as likely as any another value

What is the best estimator under these circumstances?

UCSF
Department of Radiology & Biomedical Imaging
VA

Least-Squares Estimator (LSE)

The best what we can do is minimizing noise:

$$E_{LSE} = \min \frac{1}{2} \left\| noise_N \right\|^2$$

Minimizing E_{LSE} with regard to θ leads to

$$\varphi_N - \mathbf{H}\hat{\boldsymbol{\theta}}_{LSE} = 0$$

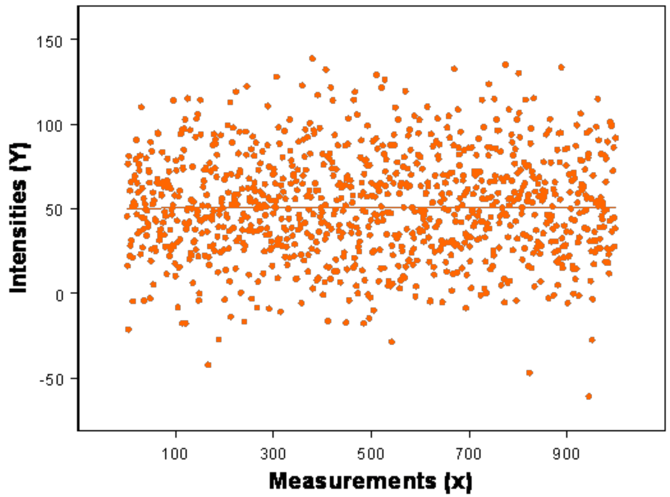
$$\mathbf{H}^T\varphi_N - \left(\mathbf{H}^T\mathbf{H}\right)\hat{\boldsymbol{\theta}}_{LSE} = 0$$

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\varphi_{\mathbf{n}}$$

•LSE is popular choice for model fitting
•Useful for obtaining a descriptive measure
But
•LSE makes no assumptions about distributions of data or parameters
•Has no basis for statistics ➔ “deterministic model”

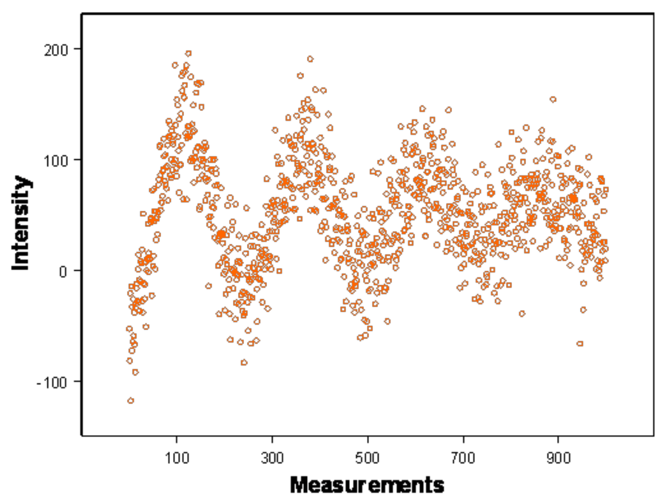
UCSF
Department of
Radiology & Biomedical Imaging
VA

Prominent Examples of LSE



Mean Value: $$\hat{\theta}_{mean} = \frac{1}{N}\sum_{j=1}^{N}\varphi(j)$$

Variance $$\hat{\theta}_{variance} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\varphi(j)-\hat{\theta}_{mean}\right)^2$$



Amplitude $\hat{\theta}_1$

Frequency $\hat{\theta}_2$

Phase $\hat{\theta}_3$

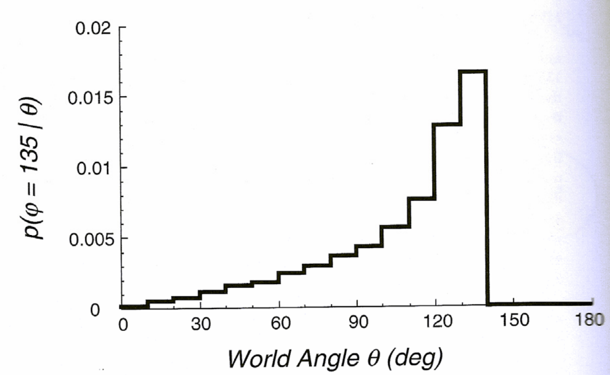
Decay $\hat{\theta}_4$

UCSF VA
Department of
Radiology & Biomedical Imaging

Likelihood Model

Likelihood $L_{\varphi}(\Phi) = p(\varphi \mid \Phi)$

18 *Bayesian Modelling of Visual Perception*



16 *Bayesian Mo*

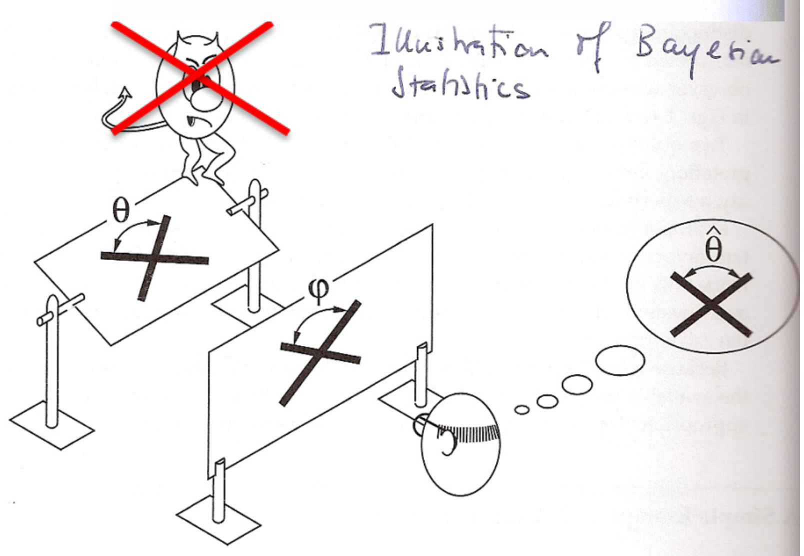


Figure 1.2: Perception of 3D angles. The angle θ between two line segments in 3D

Pretend we know something about Θ

We perform measurements for all possible values of Θ

We obtain the likelihood function of Θ given our measurements ρ

<u>Note:</u>
Θ is random
φ is a fixed parameter
Likelihood is a function of both the unknown Θ and known φ

UCSF Department of Radiology & Biomedical Imaging
VA

Likelihood Model (cont'd)

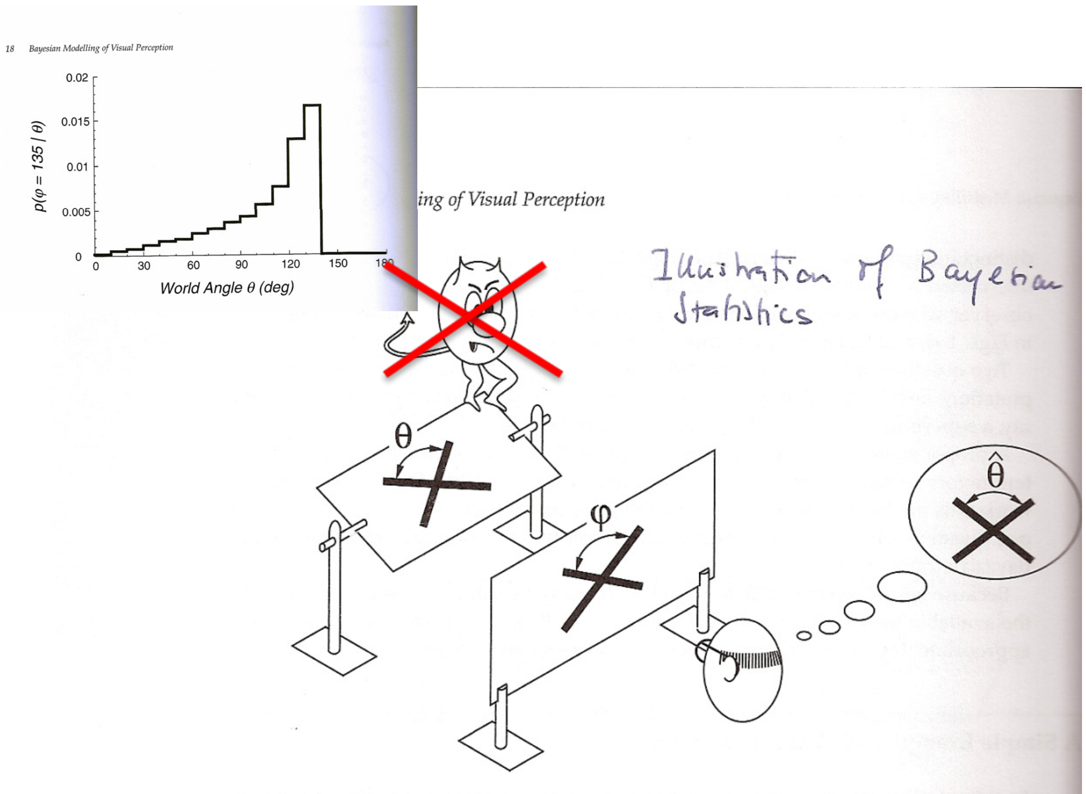


Figure 1.2: Perception of 3D angles. The angle θ between two line segments in 3D

$$L_\varphi(\Phi) = p(\varphi_{\mathbf{N}} \mid \Phi)$$

<u>New Goal:</u>
Find an estimator
which gives the most likely
probability distribution
underlying $L_\varphi(\Phi)$

UCSF
Department of
Radiology & Biomedical Imaging
VA

Maximum Likelihood Estimator (MLE)

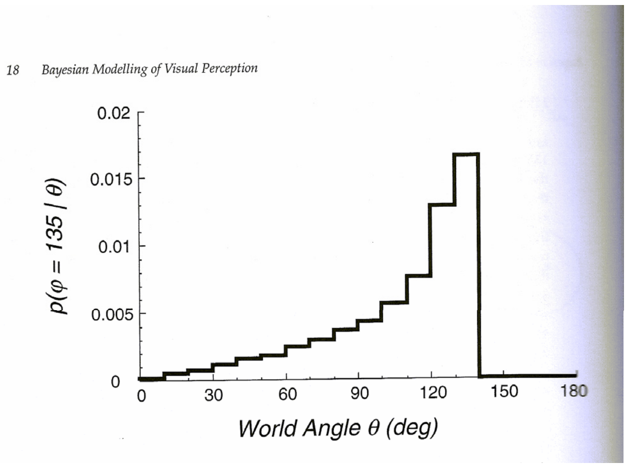
Goal: Find estimator which gives the most likely probability distribution underlying $\mathbf{x_N}$.

$$\hat{\Phi}_{MLE} = \max p\left(\varphi_{\mathbf{N}} \mid \Phi\right)$$

Max likelihood function

θ_{MLE} can be found by

$$\left.\frac{d}{d\Phi} \ln p\left(\varphi_{\mathbf{N}} \mid \Phi\right)\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MLE}} = 0$$



UCSF
Department of
Radiology & Biomedical Imaging
VA

Example I: MLE Of Normal Distribution

Normal distribution

$$p\left(\varphi_{\mathrm{N}} \mid \bar{\Phi}, \sigma^2\right) \approx \exp\left[\frac{1}{2\sigma^2}\sum_{j=1}^{N}\left(\varphi(j) - \bar{\Phi}\right)^2\right]$$

log of the normal distribution (normD)

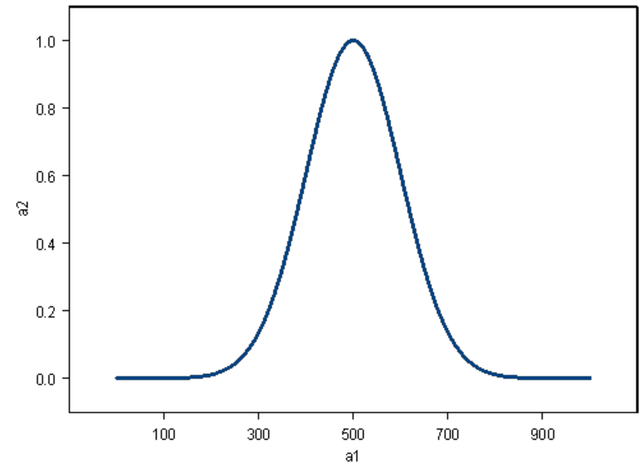
$$\ln p\left(\varphi_{\mathrm{N}} \mid \bar{\Phi}, \sigma^2\right) \approx -\frac{1}{2\sigma^2}\sum_{j=1}^{N}\left(\varphi(j) - \bar{\Phi}\right)^2$$

MLE of the mean (1st derivative):

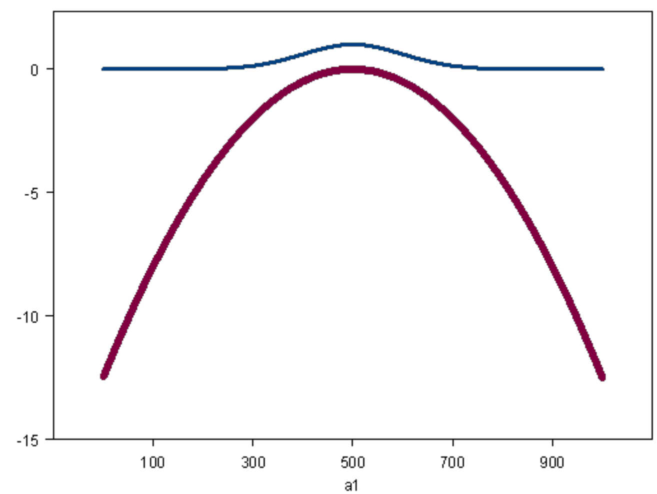
$$\frac{d}{d\widehat{\Phi}_{MLE}} \ln p(\square) = -\frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{N}\left(\varphi(j) - \widehat{\Phi}_{MLE}\right) = 0$$

$$\widehat{\Phi}_{MLE} = \frac{1}{N}\sum_{j=1}^{N}\varphi(j)$$

Normal Distribution



Log Normal Distribution



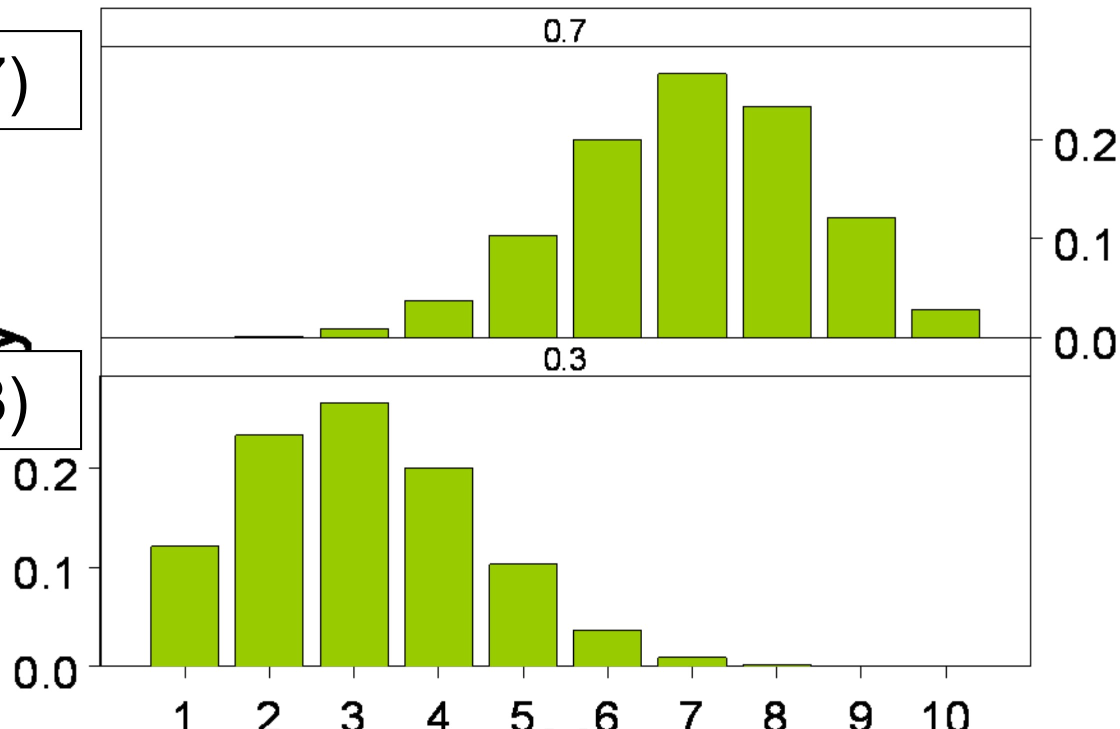
UCSF Department of Radiology & Biomedical Imaging
VA

Example II: MLE Of Binominal Distribution (Coin Toss)

Distribution function f(y|n,w):
n= number of tosses
w= probability of success

f(y|n=10,w=7)

f(y|n=10,w=3)



y

UCSF
Department of
Radiology & Biomedical Imaging
VA

MLE Of Coin Toss (cont'd)

Goal:

Given the observed data *f (y|w=0.7, n=10)*, find the parameter Φ_{MLE} that most likely produced the data.

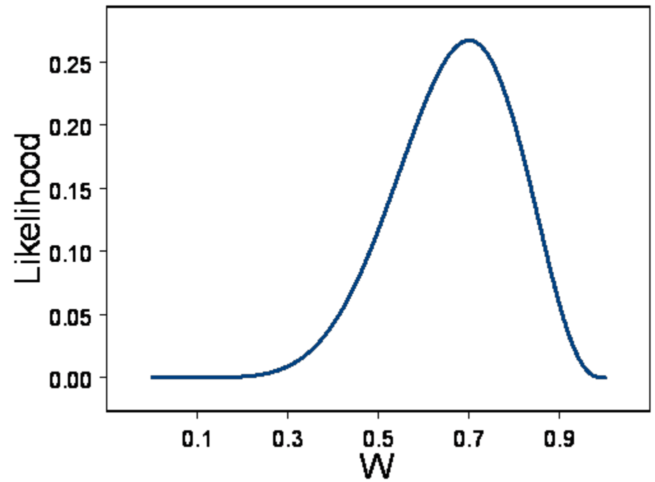
$$L(\Phi_{MLE} \mid y = 7, n = 10)$$

UCSF VA
Department of
Radiology & Biomedical Imaging

MLE Of Coin Toss (cont'd)

Likelihood function of coin tosses

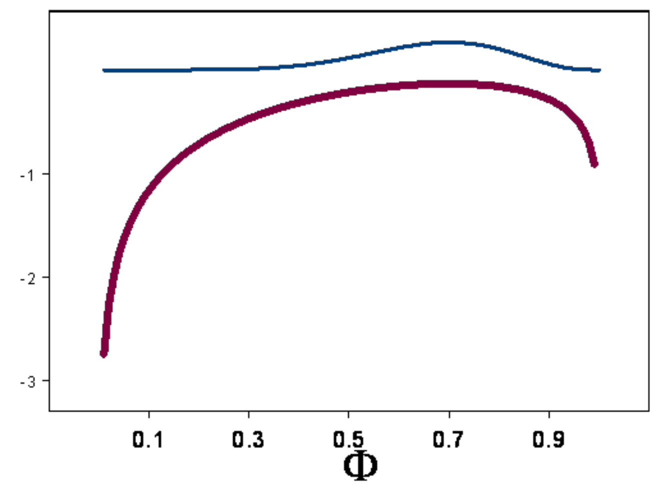
$$L(\Phi \mid y) = \frac{n!}{(y!)(n-y)!} \cdot \Phi^{y} (1-\Phi)^{n-y}$$



log likelihood function

$$\ln L(w \mid y) =$$

$$\ln \frac{n!}{(y!)(n-y)!} + y \ln w + (n-y) \ln (1-w)$$



UCSF
Department of
Radiology & Biomedical Imaging
VA

MLE Of Coin Toss

Evaluate MLE equation (1st derivative)

$$\frac{d\ln L(\square)}{d\Phi_{MLE}} = \frac{y}{\Phi_{MLE}} - \frac{(n-y)}{1-\Phi_{MLE}} = 0$$

$$= \frac{y - n\Phi_{MLE}}{\Phi_{MLE}(1-\Phi_{MLE})} = 0 \Rightarrow \Phi_{MLE} = \frac{y}{n}$$

According to the MLE principle, the distribution f(y/n) for a given n is the most likely distribution to have generated the observed data of y.

UCSF
Department of
Radiology & Biomedical Imaging
VA

Relationship between MLE and LSE

Assume: Θ is independent of noise_N
MLE and noise_N have the same distribution

$$p_{\boldsymbol{\theta}}\left(\varphi_{\mathbf{N}} \mid \Phi\right) = p_{noise}\left(\varphi_{\mathbf{N}} - \mathbf{H}\Phi \mid \Phi\right)$$

noise_N is zero mean and gaussian

$p(\rho|\Theta)$ is maximized when LSE is minimized

UCSF VA
Department of
Radiology & Biomedical Imaging

Bayesian Model

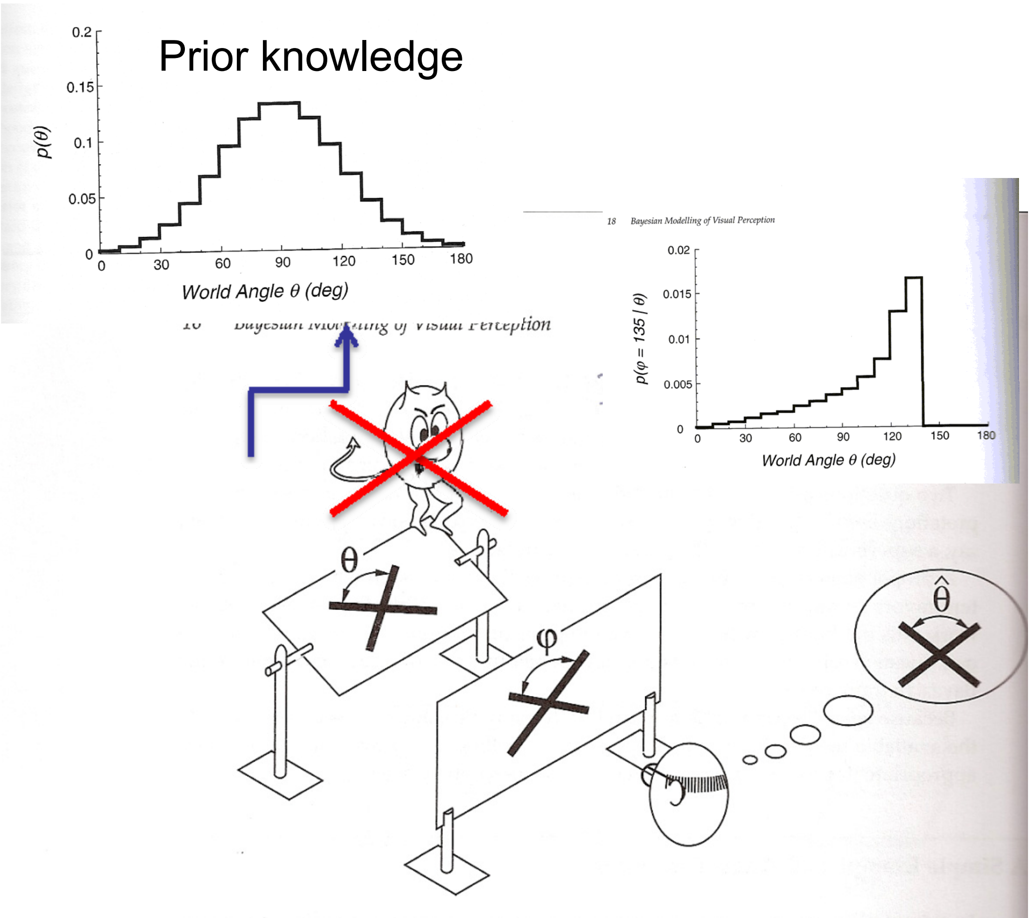


Figure 1.2: Perception of 3D angles. The angle θ between two line segments in 3D

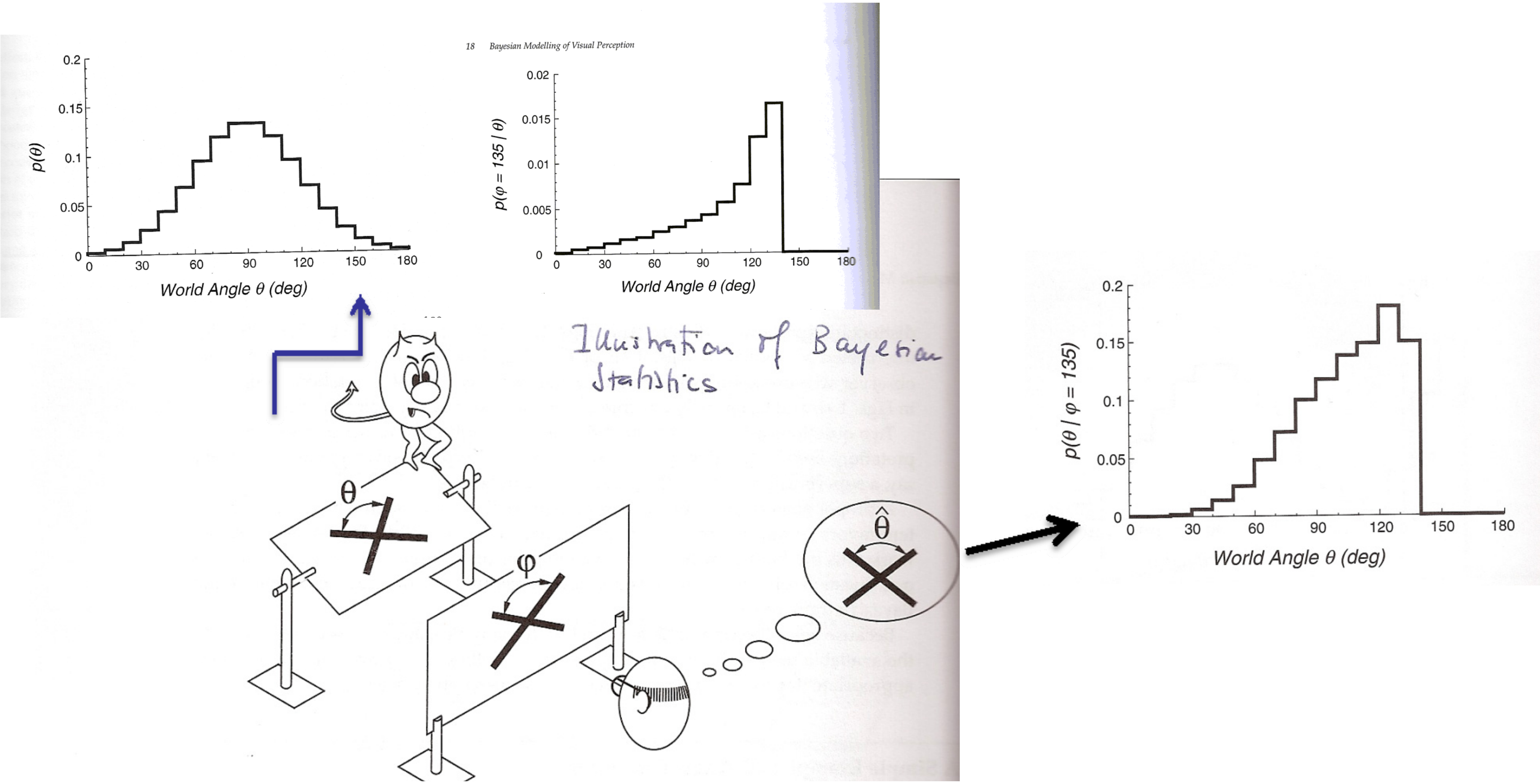
Now, the daemon comes into play, but we know The daemon's preferences for Θ (prior knowledge).

$$prior(\Phi) = p(\Phi)$$

<u>New Goal:</u>
Find the estimator which gives the most likely probability distribution of Θ given everything we know.

UCSF Department of Radiology & Biomedical Imaging
VA

Bayesian Model



$$posterior_{\varphi}(\boldsymbol{\theta}) = C \cdot L_{\varphi}(\varphi_N \mid \Phi) \cdot p(\Phi)$$

UCSF Department of Radiology & Biomedical Imaging
VA

Maximum A-Posteriori (MAP) Estimator

Goal:
Find the most likely Θ_{MAP} (max. posterior density of) given φ.

Maximize joint density

$$\hat{\theta}_{MAP} = \max L\left(\varphi_{\mathbf{N}} \mid \boldsymbol{\theta}\right) p\left(\varphi_N\right)$$

θ_{MAO} can be found by

$$\left.\frac{d}{d\boldsymbol{\theta}} \ln L\left(\varphi_{\mathbf{N}} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)\right| = \frac{\partial}{\partial\boldsymbol{\theta}} \ln p\left(\varphi_{\mathbf{N}} \mid \boldsymbol{\theta}\right) + \frac{\partial}{\partial\boldsymbol{\theta}} \ln p\left(\boldsymbol{\theta}\right) = 0$$

UCSF VA
Department of
Radiology & Biomedical Imaging

Example III: MAP Of Normal Distribution

We have random sample:

$$\frac{\partial}{\partial \hat{u}} \ln L\left(\varphi_{\mathrm{N}} \mid \hat{\mu}, \hat{\sigma}^2\right) p\left(\hat{\mu}, \hat{\sigma}^2\right) = -\frac{1}{\hat{\sigma}_{\varphi}^2} \sum_{j=1}^{N} \left(\varphi(j) - \hat{\mu}_{\varphi}\right) - \frac{1}{\hat{\sigma}_{\boldsymbol{\mu}}^2} \sum_{j=1}^{N} p\left(\hat{\mu}_{MLE}\right) = 0$$

The sample mean of MAP is:

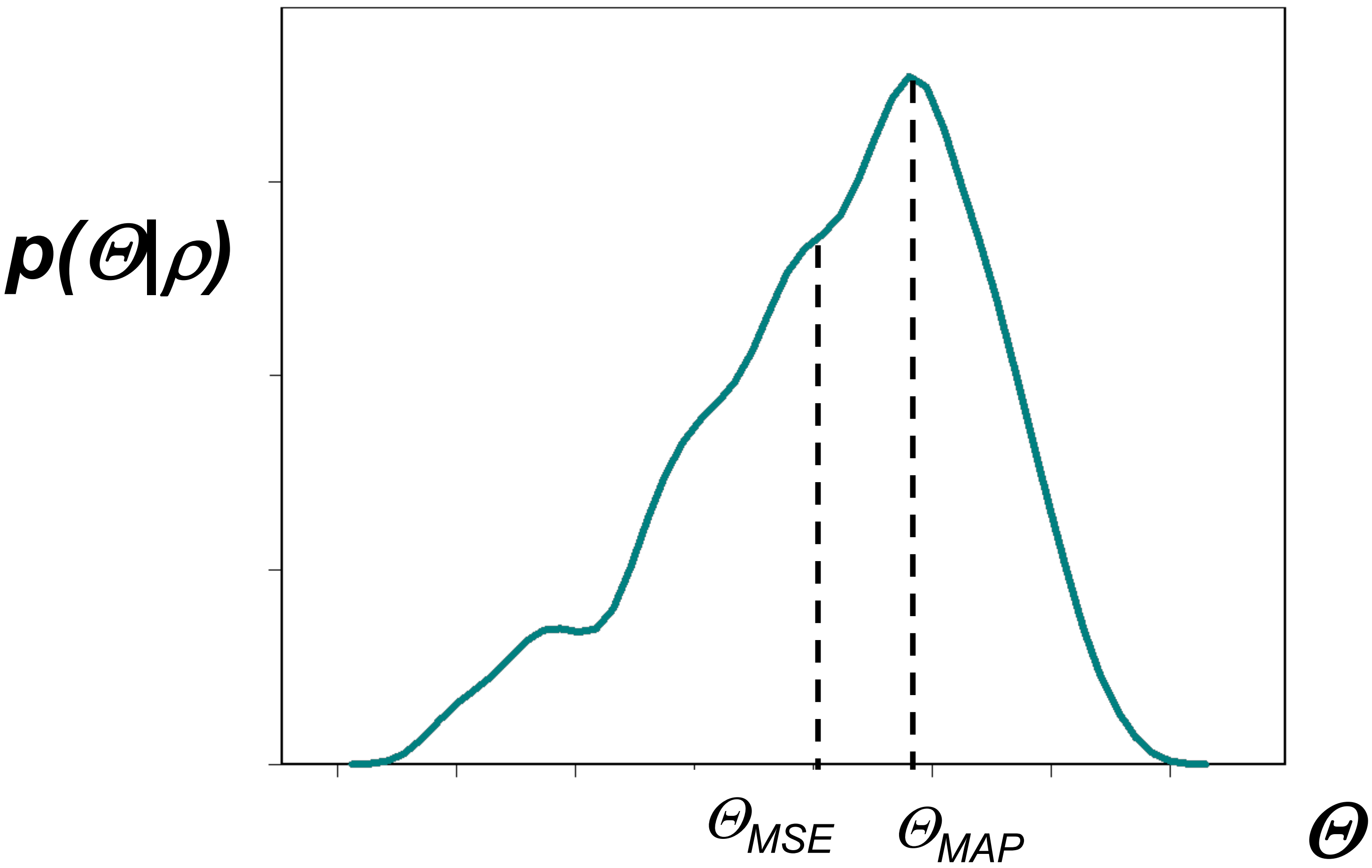
$$\widehat{\Phi}_{\mathbf{MAP}} = \frac{\sigma_{\mu}^2}{\sigma_{\varphi}^2 + T\sigma_{\mu}^2} \sum_{j=1}^{N} \varphi(j)$$

If we do not have prior information on μ, $\sigma_{\mu} \rightarrow$ inf or $T \rightarrow$ inf

$$\hat{\boldsymbol{\mu}}_{\mathrm{MAP}} \Rightarrow \hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \hat{\boldsymbol{\mu}}_{LSE}$$

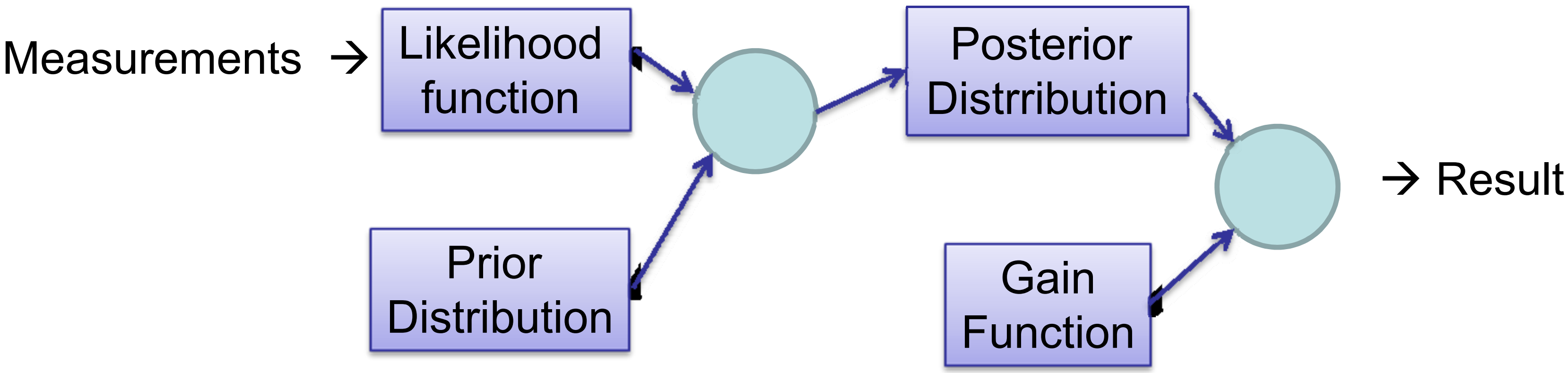
UCSF
Department of
Radiology & Biomedical Imaging
VA

Posterior Distribution and Decision Rules



UCSF
Department of
Radiology & Biomedical Imaging
VA

Decision Rules



θ

UCSF
Department of
Radiology & Biomedical Imaging
VA

Some Desirable Properties of Estimators I:

Unbiased: Mean value of the error should be zero

$$E\left\langle \hat{\Phi} - \Phi \right\rangle = 0 \;\Rightarrow E\left\langle \hat{\Phi} \right\rangle = E\left\langle \theta \right\rangle$$

Consistent: Error estimator should decrease asymptotically as number of measurements increase. (Mean Square Error (MSE))

$$MSE = E\left\langle \left\| \hat{\Phi} - \Phi \right\|^2 \right\rangle \rightarrow 0 \text{ for large N}$$

What happens to MSE when estimator is biased?

$$MSE = E\left\langle \left\| \hat{\Phi} - \Phi - \mathrm{b} \right\|^2 \right\rangle + E\left\langle \left\| \mathrm{b} \right\|^2 \right\rangle$$

variance

bias

UCSF
Department of
Radiology & Biomedical Imaging
VA

Some Desirable Properties of Estimators II:

Efficient: Co-variance matrix of error should decrease asymptotically to its minimal value for large N

$$\mathbf{C}_{\tilde{\boldsymbol{\theta}}} = E\left\langle \left(\widehat{\Phi}_i \text{-} \Phi_i\right)\left(\widehat{\Phi}_k \text{-} \Phi_k\right)^T \right\rangle \geq J_{ik}^{-1}$$

UCSF
Department of
Radiology & Biomedical Imaging
VA

Example:
Properties Of Estimators ***Mean*** and ***Variance***

Mean: $$E\langle \hat{\mu} \rangle = \frac{1}{N}\sum_{j=1}^{N} E\langle \varphi(j) \rangle = \frac{1}{N}\cdot N\mu = \mu$$

The sample mean is an unbiased estimator of the true mean

Variance: $$E\left\langle (\hat{\mu}-\mu)^2 \right\rangle = \frac{1}{N^2}\sum_{j=1}^{N} E\left\langle (\varphi(j)-\mu)^2 \right\rangle = \frac{1}{N^2}\cdot N\sigma^2 = \frac{\sigma^2}{N}$$

The variance is a consistent estimator because
It approaches zero for large number of measurements

UCSF Department of Radiology & Biomedical Imaging
VA

Properties Of MLE

- is consistent: the MLE recovers asymptotically the true parameter values that generated the data for N → inf;
- Is efficient: The MLE achieves asymptotically the minimum error (= max. information)

UCSF VA
Department of
Radiology & Biomedical Imaging

Summary

- *LSE* is a descriptive method to accurately fit data to a model.
- *MLE* is a method to seek the probability distribution that makes the observed data most likely.
- *MAP* is a method to seek the most probably parameter value given prior information about the parameters and the observed data.

- If the influence of prior information decreases, i.e. many measurements, MAP approaches MLE

UCSF VA
Department of
Radiology & Biomedical Imaging

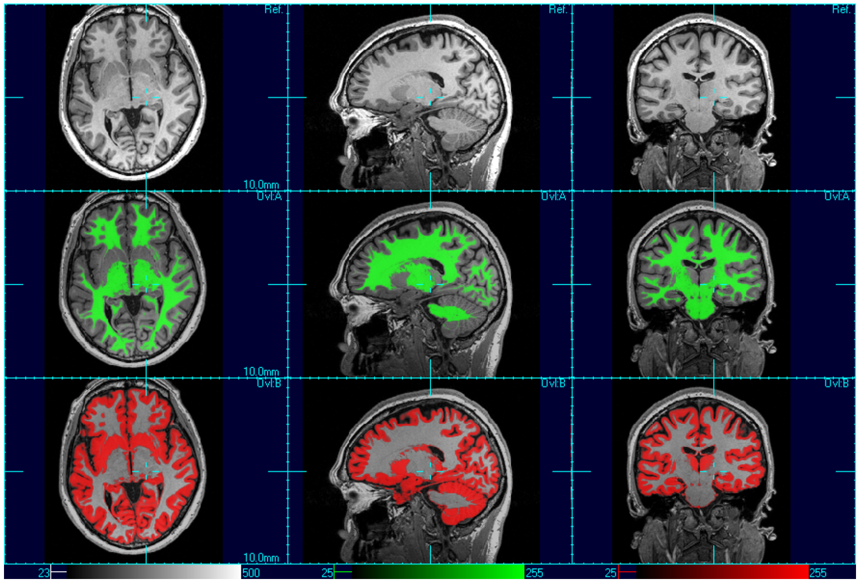
Some Priors in Imaging

- Smoothness of the brain
- Anatomical boundaries
- Intensity distributions
- Anatomical shapes
- Physical models
  - Point spread function
  - Bandwidth limits
- Etc.

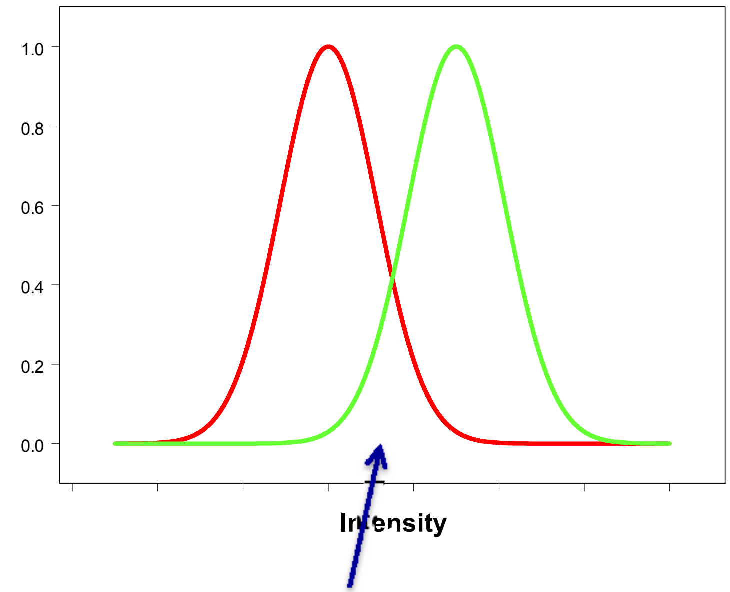
UCSF VA
Department of
Radiology & Biomedical Imaging

Estimation Theory: Motivation Example I

Gray/White Matter Segmentation



Hypothetical Histogram



What works better than flipping a coin?

Design likelihood functions based on
- anatomy
- co-occurance of signal intensities
- others

Determine prior distribution
- population based atlas of regional intensities
- model based distributions of intensities
- others

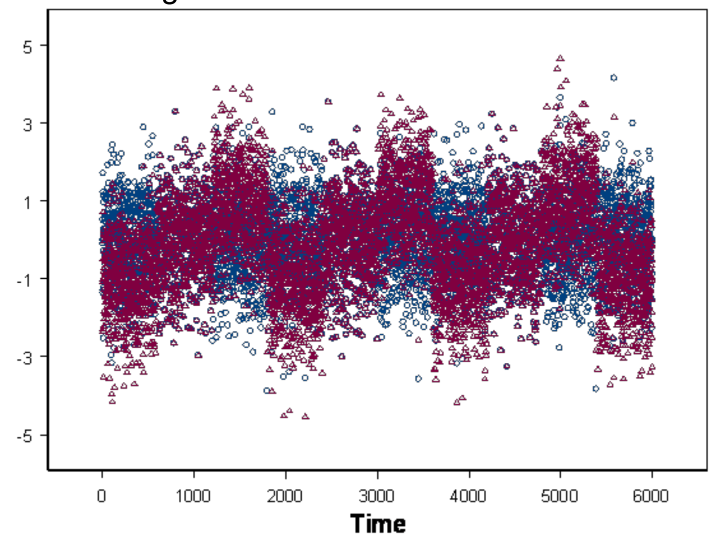
UCSF
Department of
Radiology & Biomedical Imaging
VA

Estimation Theory: Motivation Example II

Goal: Capture dynamic signal on a static background

D. Feinberg Advanced MRI Technologies, Sebastopol, CA

Poor signal to noise



Improvements to identify the dynamic signal:

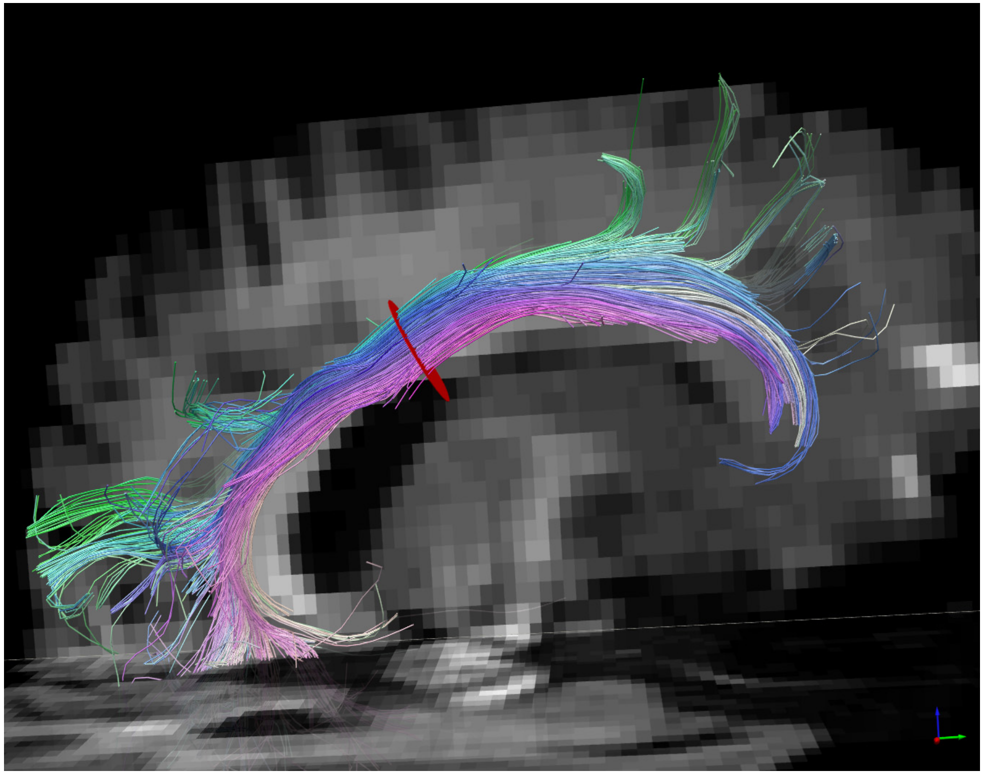
Design likelihood functions based on
- auto-correlations
- anatomical information

Determine prior distributions from
- serial measurements
- multiple subjects
- anatomy

UCSF
VA
Department of
Radiology & Biomedical Imaging

Estimation Theory: Motivation Example III

Diffusion Spectrum Imaging – Human Cingulum Bundle



Dr. Van Wedeen, MGH

Goal:
Capture directions of fiber bundles

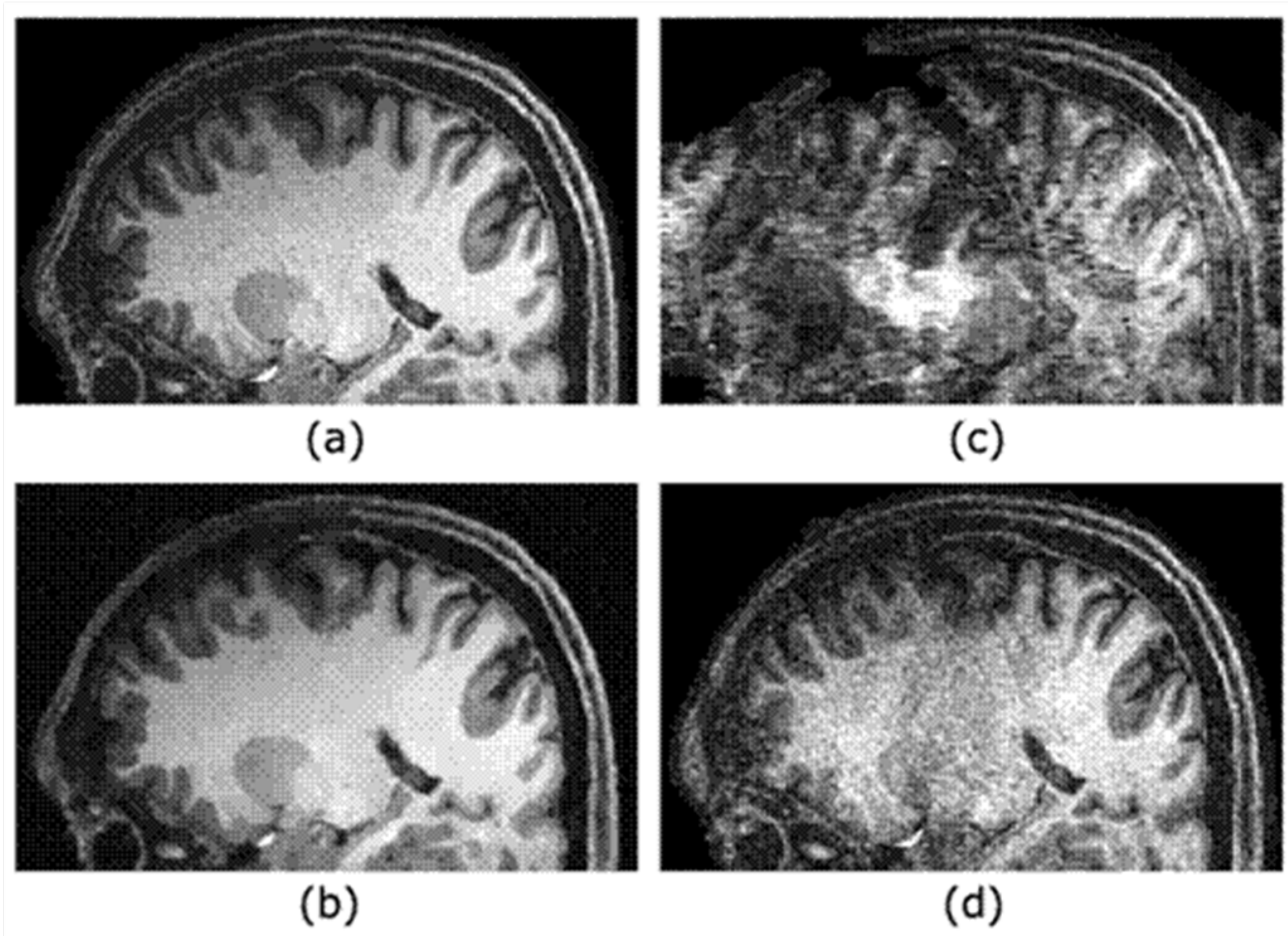
Improvements to identify tracts:

Design likelihood functions based on
- similarity measures of adjacent voxels
- fiber anatomy

Determine prior distributions from
- anatomy
- fiber skeletons from a population
- others

UCSF
Department of
Radiology & Biomedical Imaging
VA

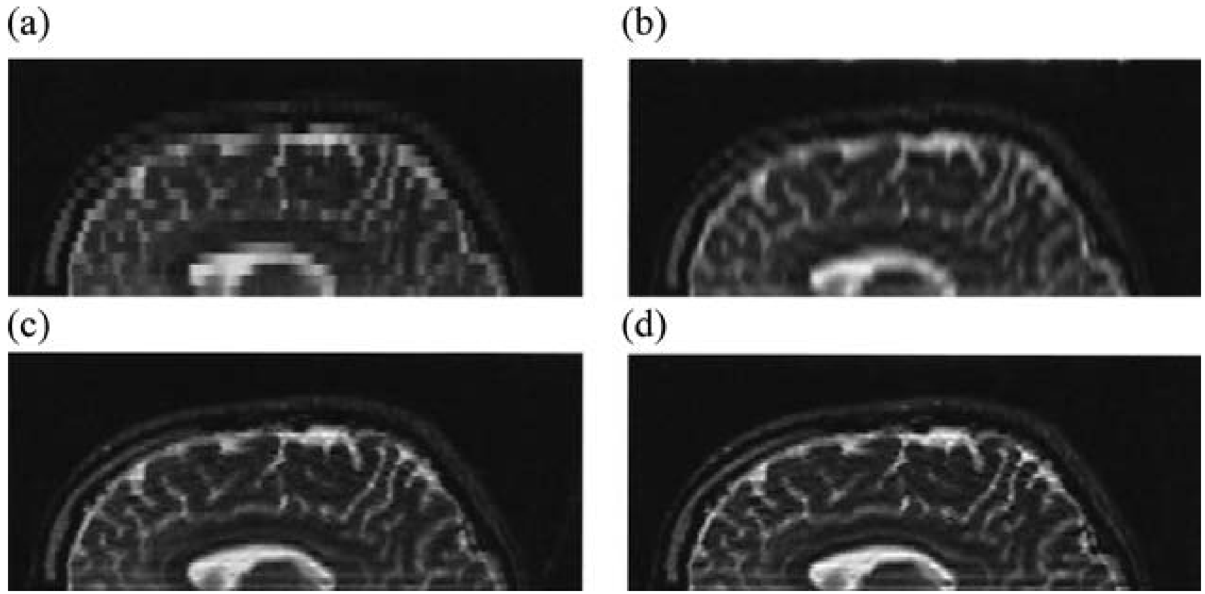
MAP Estimation in Image Reconstructions with Edge-Preserving Priors



Dr. Ashish Raj, Cornell U

UCSF
VA
Department of
Radiology & Biomedical Imaging

MAP Estimation In Image Reconstruction



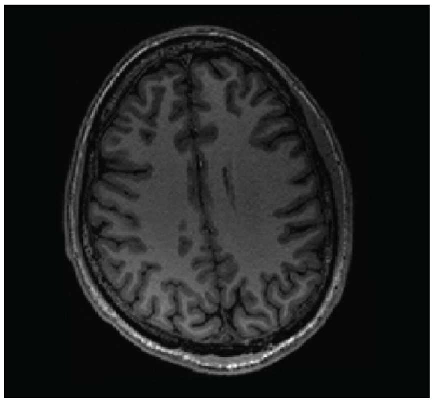
Human brain MRI. (a) The original LR data. (b) Zero-padding interpolation. (c) SR with box-PSF. (d) SR with Gaussian-PSF

From: A. Greenspan in
The Computer Journal Advance Access published February 19, 2008

UCSF Department of Radiology & Biomedical Imaging
VA

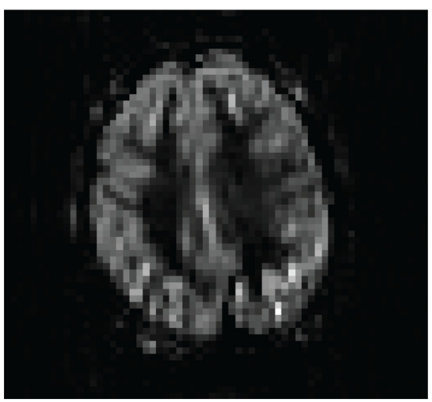
Improved ASL Perfusion Results

MPRAGE

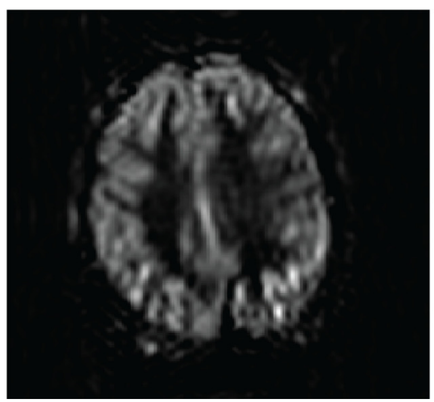


Segmentation

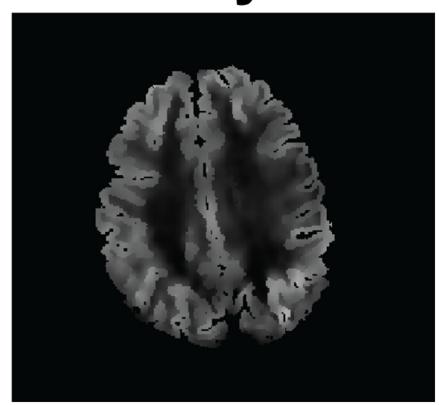
DFT



zDFT



K-Bayes

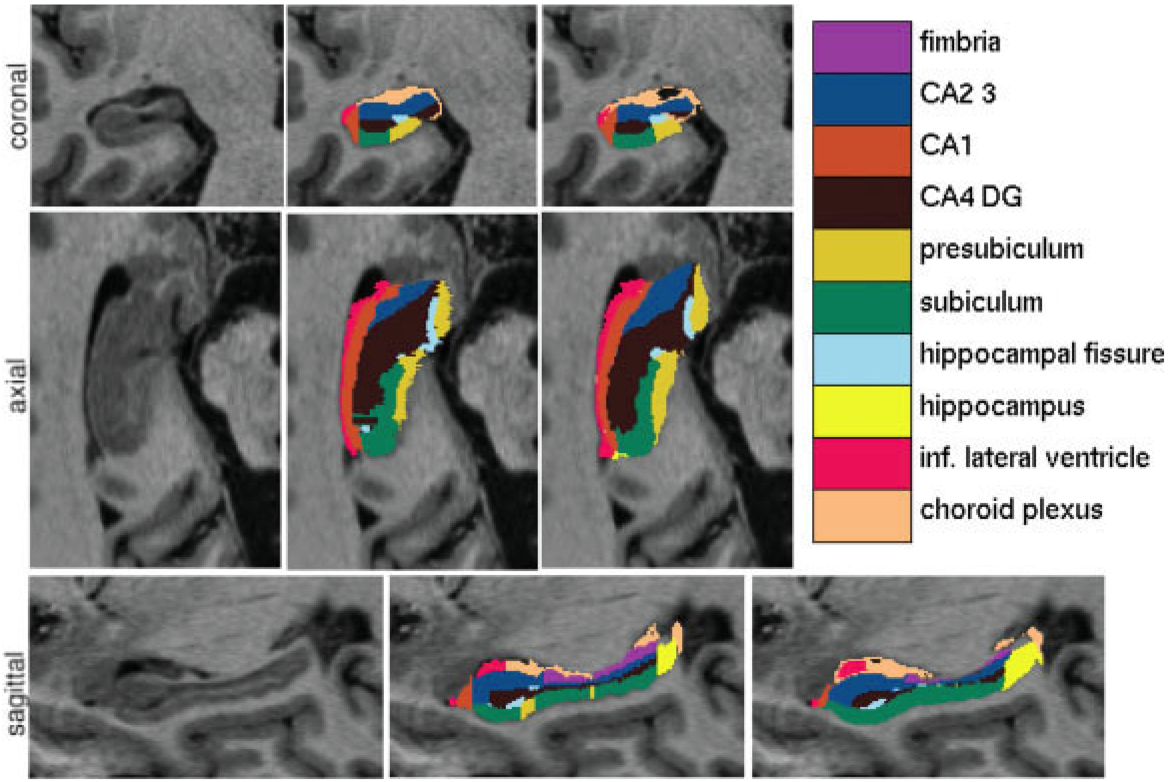


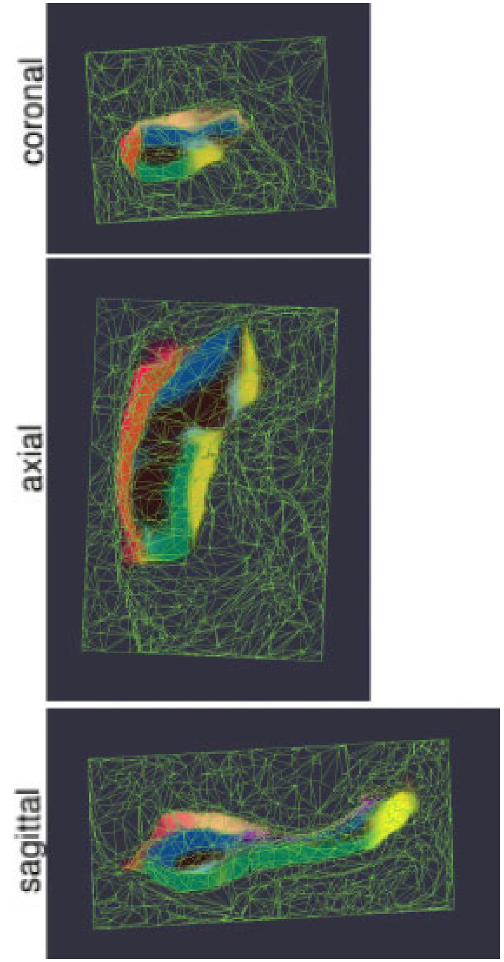
zDFT = zero-filled DFT

By Dr. John Kornak, UCSF

UCSF VA

Bayesian Automated Image Segmentation

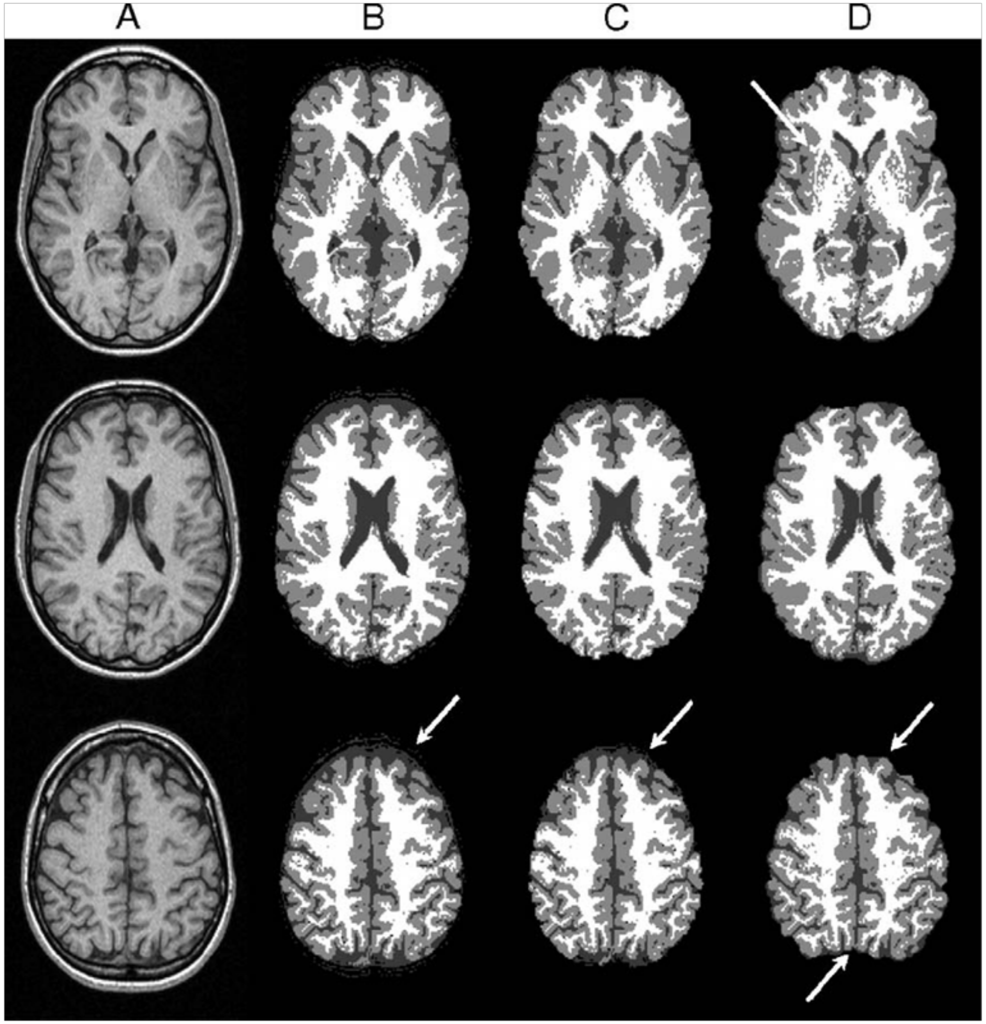




Bruce Fischl, MGH

UCSF
Department of
Radiology & Biomedical Imaging
VA

Segmentation Using MLE

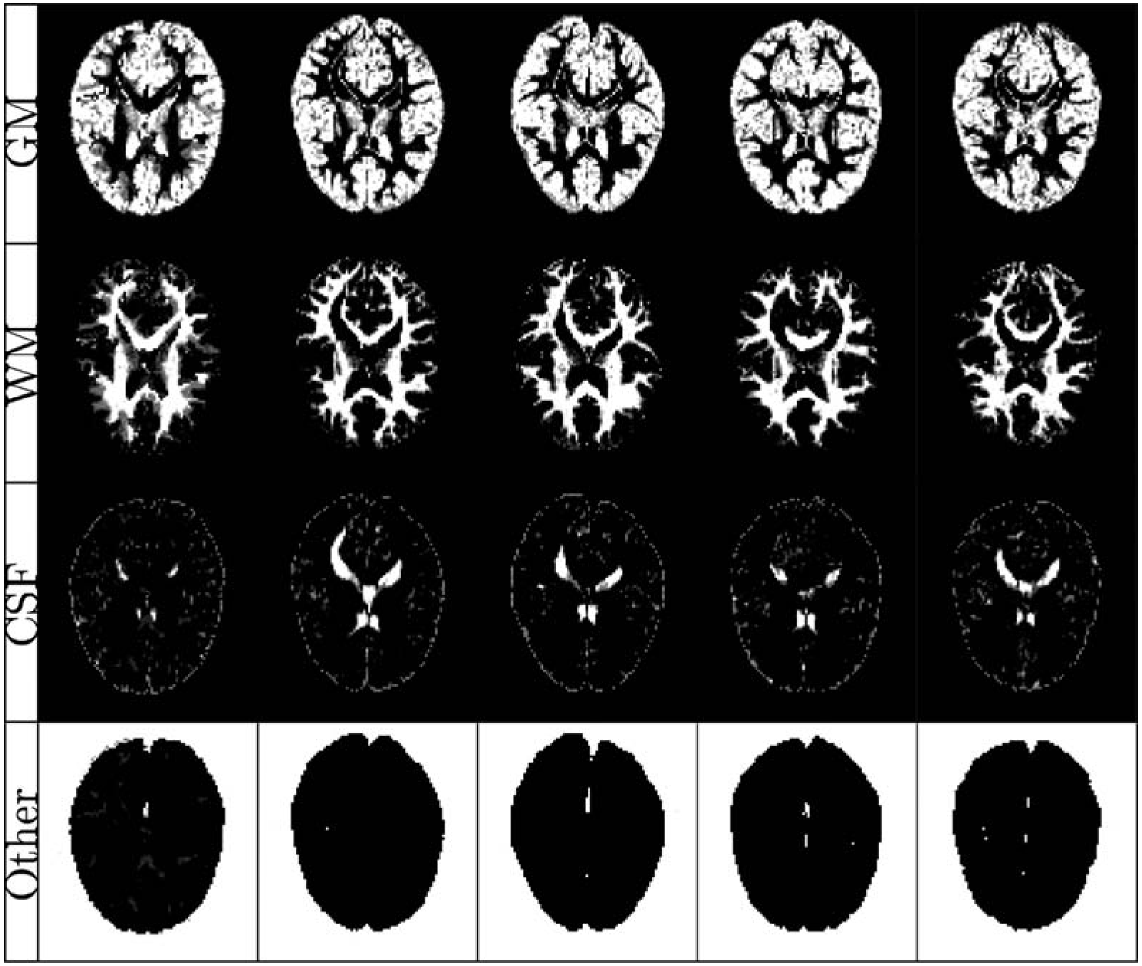


A: Raw MRI
B: SPM2
C: EMS
D: HBSA

from
Habib Zaidi, et al,
NeuroImage 32
(2006) 1591 – 1607

UCSF
Department of
Radiology & Biomedical Imaging
VA

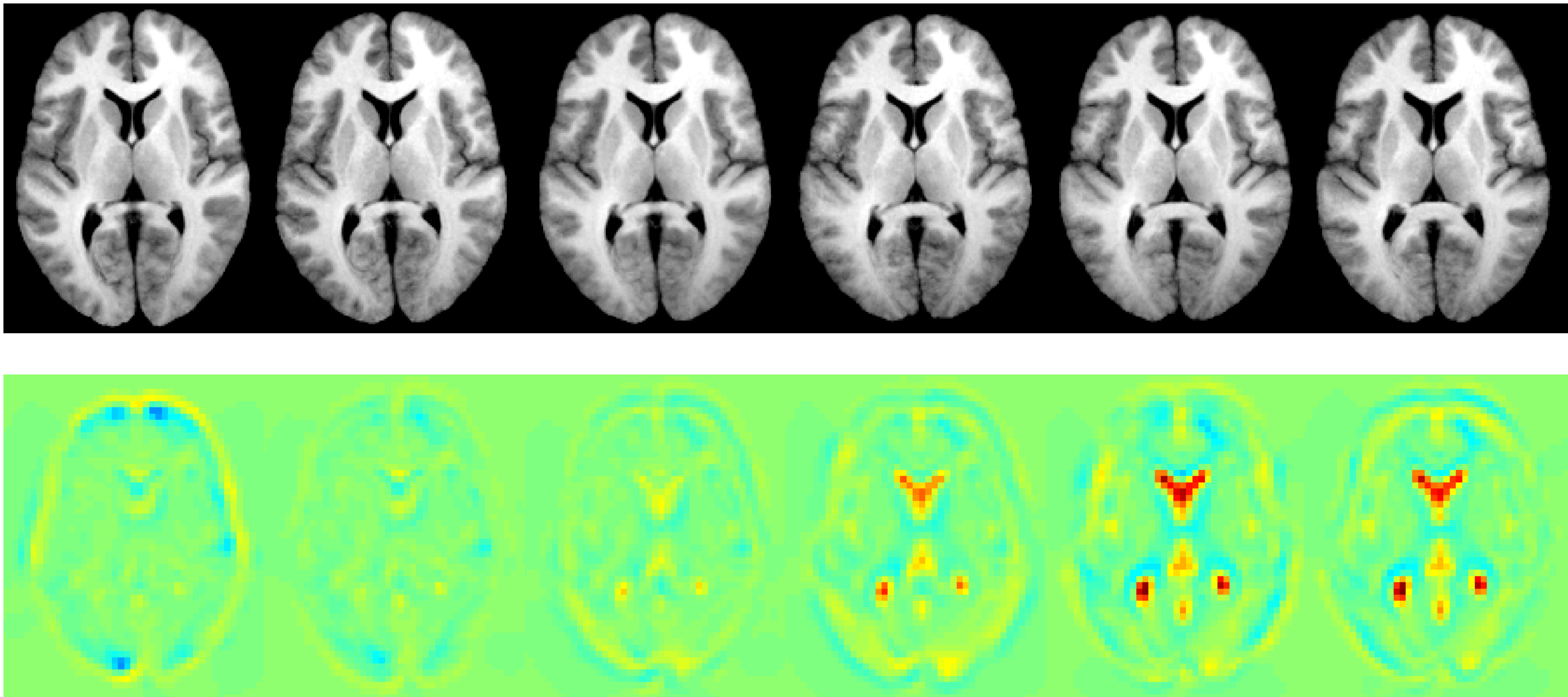
Population Atlases As Priors



Dr. Sarang Joshi, U Utah, Salt Lake City

UCSF
Department of
Radiology & Biomedical Imaging
VA

Population Shape Regressions Based Age-Selective Priors



Age = 29 33 37 41 45 49

Dr. Sarang Joshi, U Utah, Salt Lake City

UCSF Department of Radiology & Biomedical Imaging
VA

Imaging Software Using MLE And MAP

Packages	**Applications**	**Languages**
VoxBo	fMRI	C/C++/IDL
MEDx	sMRI, fMRI	C/C++/Tcl/Tk
SPM	fMRI, sMRI	matlab/C
iBrain		IDL
FSL	fMRI, sMRI, DTI	C/C++
fmristat	fMRI	matlab
BrainVoyager	sMRI	C/C++
BrainTools		C/C++
AFNI	fMRI, DTI	C/C++
Freesurfer	sMRI	C/C++
NiPy		Python

UCSF
VA
Department of
Radiology & Biomedical Imaging

Literature

Mathematical

- H. Sorenson. *Parameter Estimation – Principles and Problems.* Marcel Dekker (pub)1980.

Signal Processing

- S. Kay. Fundamentals of Signal Processing – Estimation Theory. Prentice Hall 1993.
- L. Scharf. Statistical Signal Processing: Detection, Estimation, and Time Series Analysis. Addison-Wesley 1991.

Statistics:

- A. Hyvarinen. Independent Component Analysis. John Wileys & Sons. 2001.
- New Directions in Statistical Signal Processing. From Systems to Brain. Ed. S. Haykin. MIT Press 2007.

UCSF Department of Radiology & Biomedical Imaging
VA