ECE 7251: Signal Detection and Estimation

Spring 2002

Prof. Aaron Lanterman

Georgia Institute of Technology

Lecture 3, 1/9/02:

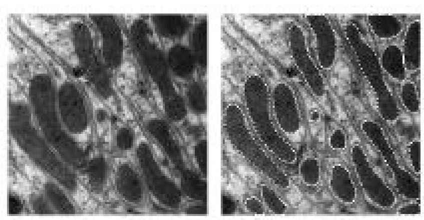
Introduction to Bayesian Estimation

The Players

- Want to estimate a realization q of a random variable Θ from collected data y
- $p(q)$: prior density
- $p(y|q)$: likelihood density
- $p(q|y) = p(y|q)p(q)/p(y)$: posterior density
- $p(y) = \int p(q,y)dq = \int p(y|q)p(q)dq$
  - Marginal density for the data
  - Often called normalizer or partition function
  - Sometimes need explicitly; often don't

Ex: Mitochondria Segmentation

- Data y is an electron micrograph



- q contains:
  - Number of mitochondria
  - Fourier parameterization of mitochondria shapes

Ex: Mitochondria Segmentation

- Data loglikelihood $p(y|q)$ consists of a Gauss-markov random field texture model
  - Mitochondria and cytoplasm have different textures
  - MRF models learned from hand-segmented *training data*
- Prior $p(q)$ learned from training data
  - Derived from over 400 hundred hand-segmented electron micrographs!
- See U. Grenander and M.I. Miller, "Representations of Knowledge in Complex Systems," *J. of the Royal Statistical Society*, B

Ex: Mitochondria Segmentation

- Posterior $p(q|y)$ extremely complicated
- Uses Markov chain Monte Carlo (random sampling) algorithm to draw samples from $p(q|y)$ posterior

Note: Movie removed from powerpoint file on web, since the movie sometimes bombs on my laptop (?)

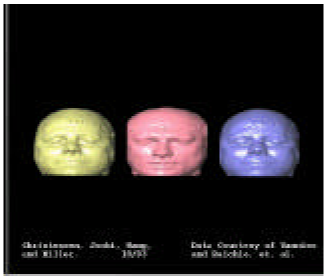
(Images and movie from www.cis.jhu.edu)

Ex: Brain Mapping

- Data y are MRI scans
- Parameters q consists of a vector field mapping one brain to another
- Likelihood $p(y|q)$ has Gaussian form from squared-error metric used to compare 3-D images
- Prior $p(q)$ derived from mechanics
  - Theory of elastics or viscous fluid
  - Prevents brain warping from getting too wacky
- See G.E. Christensen, S.C. Joshi, and M.I. Miller, Volumetric Transformation of Brain Anatomy, *IEEE Transactions on Medical Imaging*. 16(6), Dec. 1997, pp. 864-877.

Ex: Brain Mapping

- Multiscale gradient ascent used to maximize the posterior $p(\theta \mid y)$



(Movie from www.cis.jhu.edu)

Cost Functions

- Measure error between a parameter instance and its estimate
- Some common cost functions for a real or complex parameter:

$$c(\hat{\theta},\theta) = |\hat{\theta} - \theta|^2; \text{ squared error}$$

$$c(\hat{\theta},\theta) = |\hat{\theta} - \theta|; \text{ absolute error}$$

$$c(\hat{\theta},\theta) = I(|\hat{\theta} - \theta| > \epsilon); \text{ "hit or miss" or uniform error}$$

The Bayes Risk

- We seek the estimator which minimizes the <u>Bayes risk</u> or <u>average cost</u>:

$$R \equiv E[C] \equiv \underbrace{E[c(\hat{\theta}(Y),\Theta)]}_{\text{Joint E[] over } Y \text{ and } \boldsymbol{\Theta}}$$

$$= \int\int c(\hat{\theta}(y),\theta)\, p(\theta, y)\, d\theta\, dy$$

$$= \int p(y)\left[\int c(\hat{\theta}(y),\theta)\, p(\theta \mid y)\, d\theta\right] dy$$

$$= \underbrace{E[E[c(\hat{\theta}(Y),\Theta) \mid Y]]}_{\text{"iterating the expectation"}}$$

Navigating the Maze of Many E's

$$\underbrace{E}_{\text{E[] over } \mathbf{\Theta} \text{ cond. on } Y}[c(\hat{\theta}(Y),\Theta) \mid Y] = f(Y) \quad \text{A func. of the r.v. } Y\text{, hence a r.v.}$$

$$\underbrace{E}_{\text{E[] over } Y}[E[c(\hat{\theta}(Y),\Theta) \mid Y]] \quad \text{Ordinary nonrandom number}$$

$$E[c(\hat{\theta}(Y),\Theta) \mid Y = y] = f(y) \quad \text{A func. of the number } y$$

General Trick for Minimizing Risk

- Can do it for each data value individually:

$$R = \int \underbrace{p(y)}_{\text{Since} \geq 0,}\underbrace{\left[\int c(\hat{\theta}(y),\theta)\, p(\theta \mid y)\, d\theta\right]}_{\text{it suffices to minimize bracket for each } y} dy$$

- *Sometimes* can use ordinary calculus:

$$\frac{d}{d\hat{\theta}(y)} \int c(\hat{\theta}(y),\theta)\, p(\theta \mid y)\, d\theta = 0$$

Minimizing the Squared Error Risk

- Let's do it for real parameters:

$$\frac{d}{d\hat{\theta}(y)} \int (\hat{\theta}(y) - \theta)^2\, p(\theta \mid y)\, d\theta = 0$$

$$\int 2(\hat{\theta}(y) - \theta)\, p(\theta \mid y)\, d\theta = 0$$

$$\int \hat{\theta}(y)\, p(\theta \mid y)\, d\theta = \int \theta\, p(\theta \mid y)\, d\theta$$

$$\hat{\theta}(y) \underbrace{\int p(\theta \mid y)\, d\theta}_{=1} = \int \theta\, p(\theta \mid y)\, d\theta$$

Conditional Mean Estimator

- Minimizing squared error risk yields conditional mean estimator:

$$\hat{q}(y) = \int q\, p(q \mid y) dq$$
$$= E[q \mid Y = y] \equiv E[q \mid y]$$

- Hero denotes using $\hat{q}_{CME}(y)$
- See Hero, Sec. 4.2.1, pp. 34-35 for an alternate proof that also applies to complex parameters

Minimizing the Absolute Error Risk

- Clearest derivation seems to be on p. 343-344 of Kay, Vol. I. We want to solve

$$\frac{d}{d\hat{q}(y)} \int |\hat{q}(y) - q \mid p(q \mid y) dq = 0$$

- Split into two terms:

$$\frac{d}{d\hat{q}(y)} \int_{-\infty}^{\hat{q}(y)} (\hat{q}(y) - q) p(q \mid y) dq +$$
$$\frac{d}{d\hat{q}(y)} \int_{\hat{q}(y)}^{\infty} (q - \hat{q}(y)) p(q \mid y) dq = 0$$

Leibnitz's Rule

- Recall fundamental theorem of calculus:

$$\frac{d}{da} \int_{-\infty}^{a} h(b) db = h(a)$$

- Leibnitz's rule extends fundamental theorem of calculus:

$$\frac{d}{da} \int_{f(a)}^{g(a)} h(a,b) db = \int_{f(a)}^{g(a)} \frac{\partial h(a,b)}{\partial a} db$$
$$+ h(a, g(a)) \frac{dg(a)}{da} - h(a, f(a)) \frac{df(a)}{da}$$

Use Leibnitz's Rule on Each Term

$$\frac{d}{d\hat{q}(y)} \int_{-\infty}^{\hat{q}(y)} (\hat{q}(y) - q) p(q \mid y) dq =$$
$$\int_{-\infty}^{\hat{q}(y)} p(q \mid y) dq + \cancel{(\hat{q}(y) - \hat{q}(y)) p(\hat{q}(y) \mid y)} - 0$$

$$\frac{d}{d\hat{q}(y)} \int_{\hat{q}(y)}^{\infty} (q - \hat{q}(y)) p(q \mid y) dq =$$
$$- \int_{\hat{q}(y)}^{\infty} p(q \mid y) dq + 0 - \cancel{(\hat{q}(y) - \hat{q}(y)) p(\hat{q}(y) \mid y)}$$

Conditional Median Estimator

- Minimizing absolute error yields the conditional median estimator:

$$\int_{-\infty}^{\hat{q}(y)} p(q \mid y) dq = \int_{\hat{q}(y)}^{\infty} p(q \mid y) dq$$
$$\Pr\{q \le \hat{q}(y)\} = 1/2$$

- Hero denotes using $\hat{q}_{CmE}(y)$
- See Hero, Sec. 4.2.2, pp. 35-36 for a different kind of proof

Minimizing the Uniform Error Risk

- We want to minimize

$$\int I(|\hat{q}(y) - q| > e) p(q \mid y) dq$$
$$= \int [1 - I(|\hat{q}(y) - q| < e)] p(q \mid y) dq$$
$$= 1 - \underbrace{\int_{\hat{q}(y)-e}^{\hat{q}(y)+e} p(q \mid y) dq}$$

Equivalently, maximize this

Maximum a Posteriori Estimator

- Minimizing uniform or "hit and miss" error, as $\varepsilon \to 0$, yield the <u>maximum a posteriori</u> estimator:

$$\hat{q}(y) = \max_{q} p(q \mid y)$$

- Hero denotes using $\hat{q}_{MAP}(y)$
- Could call it the "conditional mode estimator," but too many things with the acronym CME or CmE already!