LECTURE 22

- **Readings:** Sect. 8.3–8.4; reread Sect. 4.2 and pp. 225–226

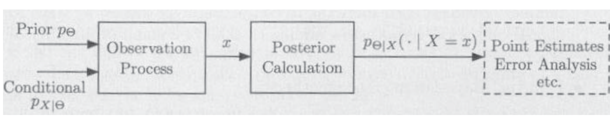
Lecture outline

- End of semester
- Review
- Performance criteria for estimators
- (Bayesian) Least mean squares estimation
- (Bayesian) Linear least mean squares estimation

Semester end game

- Chapter 8 all covered
- Chapter 9
  - Sect. 9.1: covered through middle of p. 470
  - Sect. 9.2: covered through middle of p. 482
  - Sect. 9.3: all covered
  - Sect. 9.4: **not** covered
- Problem set 11 is for practice, not to be turned in
- Final exam: Wednesday, May 19, 9am–noon
- Many office hours between last lecture and final exam
- Course VI Underground Guide Evaluations
  https://sixweb.mit.edu/student/evaluate/6.041-s2010
  until 11:59pm on May 16

Review: Bayesian inference



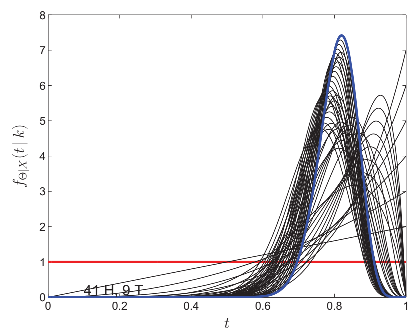
- Posterior computation is use of Bayes' rule, for example

$$p_{\Theta|X}(\theta \mid x) \;=\; \frac{p_\Theta(\theta)\, p_{X|\Theta}(x \mid \theta)}{\sum_k p_\Theta(k)\, p_{X|\Theta}(x \mid k)}$$

- Estimate $\hat{\theta}$ is number computed from posterior
- Maximum a posteriori probability (MAP) rule

$$\hat{\theta}_{\text{MAP}} \;=\; \arg\max_\theta p_{\Theta|X}(\theta \mid x) \quad \text{or} \quad \hat{\theta}_{\text{MAP}} \;=\; \arg\max_\theta f_{\Theta|X}(\theta \mid x)$$

Example: Coin with unknown parameter



Prior: $f_\Theta \sim$ beta(1,1)
(= uniform on $[0,1]$)

Likelihood: Hs in n tosses
$p_{X|\Theta} \sim$ binomial(n,Θ)

Posterior: After k Hs,
$f_{\Theta|X} \sim$ beta($k+1$,$n-k+1$)

$\hat{\theta}_{\text{MAP}}$ is the peak of the posterior, at k/n

Hypothesis testing

- Estimation with discrete Θ called **hypothesis testing**
- Common formulation:
  - θ and $\hat{\theta}$ in $\{1, 2, \ldots, m\}$
  - nonnegative cost c_{ij} for choosing $\hat{\theta} = j$ when $\theta = i$
  - $c_{ii} \le c_{ij}$ for each $j \ne i$; may as well have $c_{ii} = 0$ for each i
  - minimize expected cost:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}\mathbf{P}(\Theta = i, \widehat{\Theta} = j) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}\mathbf{P}(\widehat{\Theta} = j \mid \Theta = i)\mathbf{P}(\Theta = i)$$

  - good to make $\mathbf{P}(\widehat{\Theta} \ne \Theta)$ small, but errors not equally important (costly)
  - equal costs makes MAP rule optimal

Binary hypothesis testing example

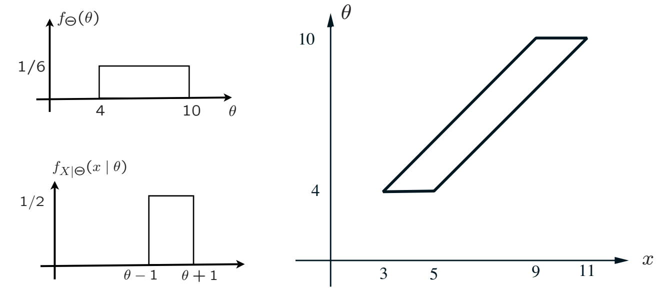
- Prior given: $\mathbf{P}(\Theta = 1) = p, \qquad \mathbf{P}(\Theta = 2) = 1 - p$
- Likelihoods given: $f_{X|\Theta}(x \mid 1), \qquad f_{X|\Theta}(x \mid 2)$
- Costs given: c_{12} (mistake 1 for 2), c_{21} (mistake 2 for 1)
- Minimize expected cost

Bayesian least mean squares (LMS) estimation

- Any estimator is function of observations: $\widehat{\Theta} = g(X)$
- LMS estimator $\widehat{\Theta}_{\mathrm{LMS}}$ minimizes $\mathbf{E}[(\Theta - \widehat{\Theta})^2]$
- LMS estimator is $g_{\mathrm{LMS}}(X) = \mathbf{E}[\Theta \mid X]$

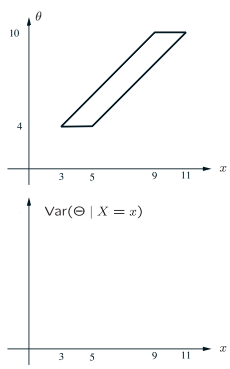
- Recall from L12: For random variable Y and number c
  $\mathbf{E}[(Y-c)^2] \;=\; \mathrm{var}(Y-c) + (\mathbf{E}[Y-c])^2 \;=\; \mathrm{var}(Y) + (\mathbf{E}[Y-c])^2$

LMS estimation example

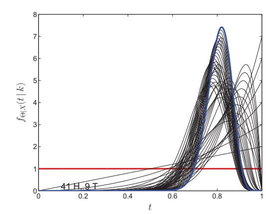


Conditional mean squared error

- $\mathbf{E}[(\Theta - \mathbf{E}[\Theta \mid X])^2 \mid X = x]$
  same as $\mathrm{var}(\Theta \mid X = x)$:
  variance of the conditional distribution of Θ



Example: Coin with unknown parameter



Prior: $f_\Theta \sim$ beta(1,1)

Likelihood: Hs in n tosses
$p_{X|\Theta} \sim$ binomial(n,Θ)

Posterior: After k Hs,
$f_{\Theta|X} \sim$ beta$(k+1, n-k+1)$

$$\mathbf{E}[\mathrm{beta}(\alpha,\beta)] = \frac{\alpha}{\alpha+\beta} \qquad \mathrm{var}(\mathrm{beta}(\alpha,\beta)) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$\hat{\theta}_{\mathrm{LMS}} \;=$

$\mathbf{E}[(\widehat{\Theta}_{\mathrm{LMS}} - \Theta)^2 \mid X = k] \;=$

Some properties of LMS estimation

- Estimator: $\widehat{\Theta} = \mathbf{E}[\Theta \mid X]$
- Estimation error: $\widetilde{\Theta} = \widehat{\Theta} - \Theta$

- $\mathbf{E}[\widetilde{\Theta}] = 0 \qquad\qquad \mathbf{E}[\widetilde{\Theta} \mid X = x] = 0$
- $\mathbf{E}[\widetilde{\Theta}\, h(X)] = 0$, for any function h
- $\mathrm{cov}(\widetilde{\Theta}, \widehat{\Theta}) = 0$
- Since $\Theta = \widehat{\Theta} + \widetilde{\Theta}$:
  $\mathrm{var}(\Theta) = \mathrm{var}(\widehat{\Theta}) + \mathrm{var}(\widetilde{\Theta})$

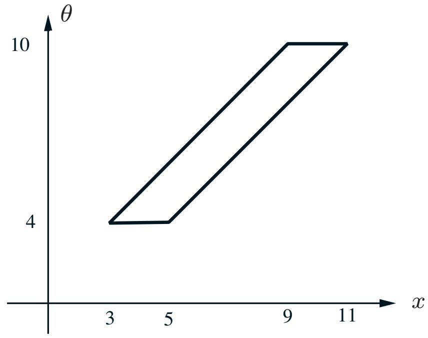
Linear LMS

- Consider estimators of the form $\widehat{\Theta}_{\mathrm{LLMS}} = aX + b$
- Minimize $\mathbf{E}\left[(\Theta - aX - b)^2\right]$
- Best choice of a,b; best linear estimator:

$$\widehat{\Theta}_{\mathrm{LLMS}} = \mathbf{E}[\Theta] + \frac{\mathrm{cov}(X,\Theta)}{\mathrm{var}(X)}(X - \mathbf{E}[X])$$

$$\mathbf{E}[(\widehat{\Theta}_{\mathrm{LLMS}} - \Theta)^2] = (1-\rho^2)\sigma_\Theta^2$$

Linear LMS: Example



Linear LMS with more data

- Consider estimators of the form:

$$\widehat{\Theta} = a_1 X_1 + \cdots + a_n X_n + b$$

- Find best choices of $a_1, \ldots, a_n, b$
- Minimize:

$$\mathbf{E}[(a_1 X_1 + \cdots + a_n X_n + b - \Theta)^2]$$

- Set derivatives to zero
  linear system in b and the a_i
- Only means, variances, covariances matter