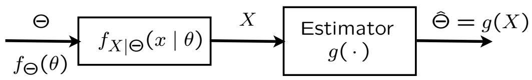
LECTURE 22

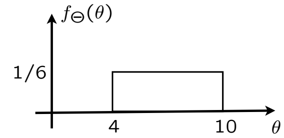
- **Readings:** pp. 225-226; Sections 8.3-8.4

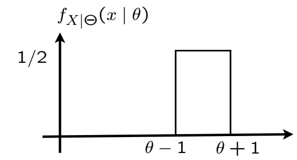
Topics

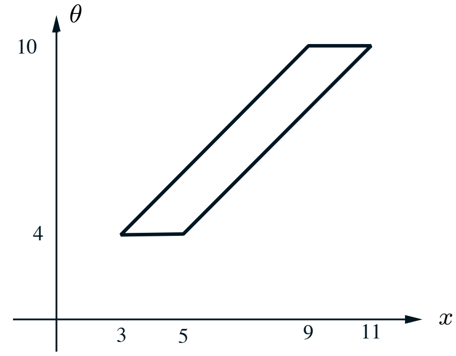
- (Bayesian) Least means squares (LMS) estimation
- (Bayesian) Linear LMS estimation



- MAP estimate: $\hat\theta_{\text{MAP}}$ maximizes $f_{\Theta|X}(\theta \mid x)$
- LMS estimation:
  - $\hat\Theta = \mathbf{E}[\Theta \mid X]$ minimizes $\mathbf{E}\left[(\Theta - g(X))^2\right]$ over all estimators $g(\cdot)$
  - for any x, $\hat\theta = \mathbf{E}[\Theta \mid X = x]$ minimizes $\mathbf{E}\left[(\Theta - \hat\theta)^2 \mid X = x\right]$ over all estimates $\hat\theta$

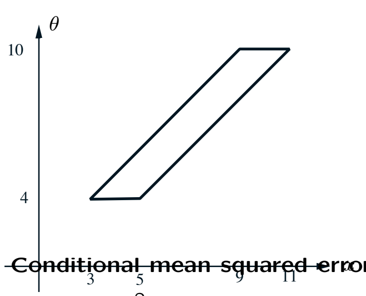






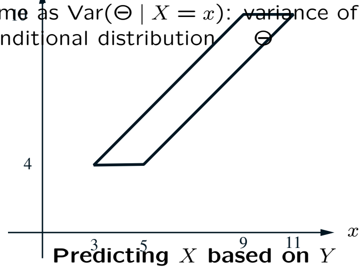
Conditional mean squared error

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



Conditional mean squared error

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



Predicting X based on Y

- Two r.v.'s X, Y
- we observe that $Y = y$
  - new universe: condition on $Y = y$
- $\mathbf{E}\left[(X - c)^2 \mid Y = y\right]$ is minimized by $c =$

Some properties of LMS estimation

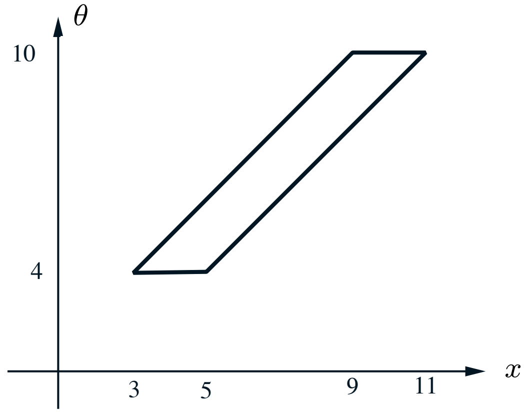
- Estimator: $\hat\Theta = \mathbf{E}[\Theta \mid X]$
- Estimation error: $\tilde\Theta = \hat\Theta - \Theta$

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

Linear LMS

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

$$\hat{\Theta}_L = \mathbf{E}[\Theta] + \frac{\text{Cov}(X,\Theta)}{\text{var}(X)}(X - \mathbf{E}[X])$$



Linear LMS properties

$$\hat{\Theta}_L = \mathbf{E}[\Theta] + \frac{\text{Cov}(X,\Theta)}{\text{var}(X)}(X - \mathbf{E}[X])$$

$$\mathbf{E}[(\hat{\Theta}_L - \Theta)^2] = (1 - \rho^2)\sigma_\Theta^2$$

Linear LMS with multiple data

- Consider estimators of the form:

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

- 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

The cleanest linear LMS example

$X_i = \Theta + W_i, \qquad \Theta, W_1, \ldots, W_n$ independent
$\Theta \ \sim \ \mu,\ \sigma_0^2 \qquad W_i \ \sim \ 0,\ \sigma_i^2$

$$\hat{\Theta}_L = \frac{\mu/\sigma_0^2 + \sum_{i=1}^{n} X_i/\sigma_i^2}{\sum_{i=0}^{n} 1/\sigma_i^2}$$

(weighted average of $\mu, X_1, \ldots, X_n$)

- If all normal, $\hat{\Theta}_L = \mathbf{E}[\Theta \mid X_1, \ldots, X_n]$

Choosing X_i in linear LMS

- $\mathbf{E}[\Theta \mid X]$ is the same as $\mathbf{E}[\Theta \mid X^3]$
- Linear LMS is different:
  - $\hat{\Theta} = aX + b$ versus $\hat{\Theta} = aX^3 + b$
  - Also consider $\hat{\Theta} = a_1 X + a_2 X^2 + a_3 X^3 + b$

Big picture

- **Standard examples:**
  - X_i uniform on $[0, \theta]$;
    uniform prior on θ
  - X_i Bernoulli(p);
    uniform (or Beta) prior on p
  - X_i normal with mean θ, known variance σ^2;
    normal prior on θ;
    $X_i = \Theta + W_i$
- **Estimation methods:**
  - MAP
  - MSE
  - Linear MSE