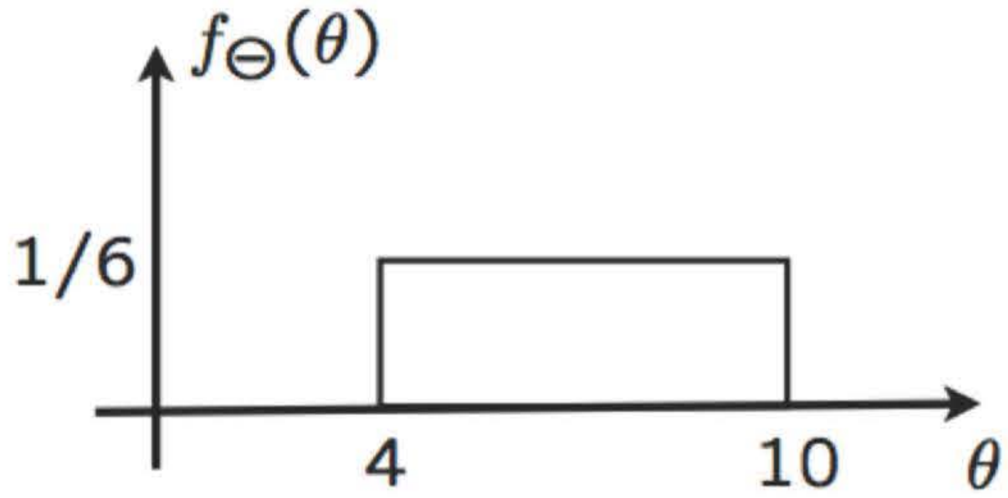
LECTURE 16: Least mean squares (LMS) estimation

- minimize (conditional) mean squared error $\mathbf{E}\left[(\Theta - \widehat{\theta})^2 \,|\, X = x\right]$
  - solution: $\widehat{\theta} = \mathbf{E}[\Theta \,|\, X = x]$
  - general estimation method
- Mathematical properties
- Example

LMS estimation in the absence of observations

- unknown Θ; prior $p_\Theta(\theta)$
  - interested in a point estimate $\hat{\theta}$
  - no observations available
  - MAP rule:
  - (Conditional) expectation:



- Criterion: Mean Squared Error (MSE): $\mathbf{E}\left[(\Theta - \hat{\theta})^2\right]$

  minimize mean squared error

LMS estimation in the absence of observations

- Least mean squares formulation:

minimize mean squared error (MSE), $\mathbf{E}\left[(\Theta - \hat{\theta})^2\right]$: $\hat{\theta} = \mathbf{E}[\Theta]$

- Optimal mean squared error: $\mathbf{E}\left[(\Theta - \mathbf{E}[\Theta])^2\right] = \text{var}(\Theta)$

LMS estimation of Θ based on X

- unknown Θ; prior $p_\Theta(\theta)$
  - interested in a point estimate $\hat{\theta}$
- observation X; model $p_{X|\Theta}(x \mid \theta)$
  - observe that $X = x$

minimize mean squared error (MSE), $\mathbf{E}\left[(\Theta - \hat{\theta})^2\right]$: $\hat{\theta} = \mathbf{E}[\Theta]$

minimize conditional mean squared error, $\mathbf{E}\left[(\Theta - \hat{\theta})^2 \mid X = x\right]$: $\hat{\theta} = \mathbf{E}[\Theta \mid X = x]$

- LMS estimate: $\hat{\theta} = \mathbf{E}[\Theta \mid X = x]$

  estimator: $\widehat{\Theta} = \mathbf{E}[\Theta \mid X]$

LMS estimation of Θ based on X

- $\mathbf{E}[\Theta]$ minimizes $\mathbf{E}\big[(\Theta - \widehat{\theta})^2\big]$

- $\mathbf{E}[\Theta \mid X = x]$ minimizes $\mathbf{E}\big[(\Theta - \widehat{\theta})^2 \mid X = x\big]$

$\widehat{\Theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X]$ minimizes $\mathbf{E}\big[(\Theta - g(X))^2\big]$, over all estimators $\widehat{\Theta} = g(X)$

LMS performance evaluation

- LMS estimate: $\widehat{\theta} = \mathbf{E}[\Theta \mid X = x]$

  estimator: $\widehat{\Theta} = \mathbf{E}[\Theta \mid X]$

– Expected performance, once we have a measurement:

$$\text{MSE} = \mathbf{E}\Big[\big(\Theta - \mathbf{E}[\Theta \mid X = x]\big)^2 \mid X = x\Big] = \text{var}(\Theta \mid X = x)$$

– Expected performance of the design:

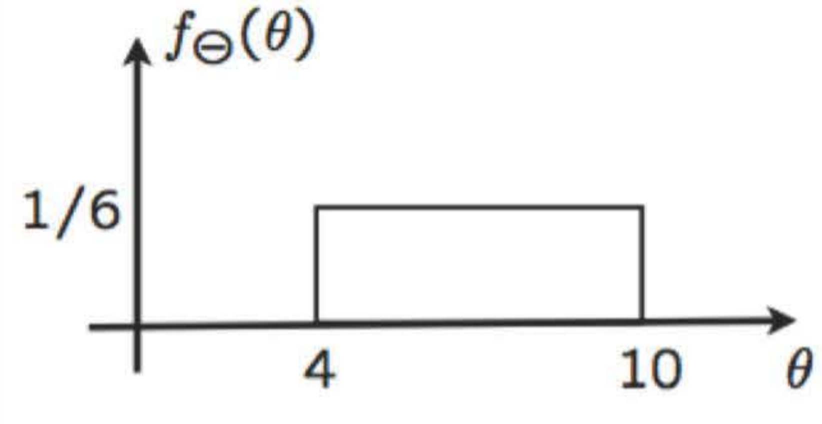
$$\text{MSE} = \mathbf{E}\Big[\big(\Theta - \mathbf{E}[\Theta \mid X]\big)^2\Big] = \mathbf{E}\big[\text{var}(\Theta \mid X)\big]$$

LMS estimation of Θ based on X

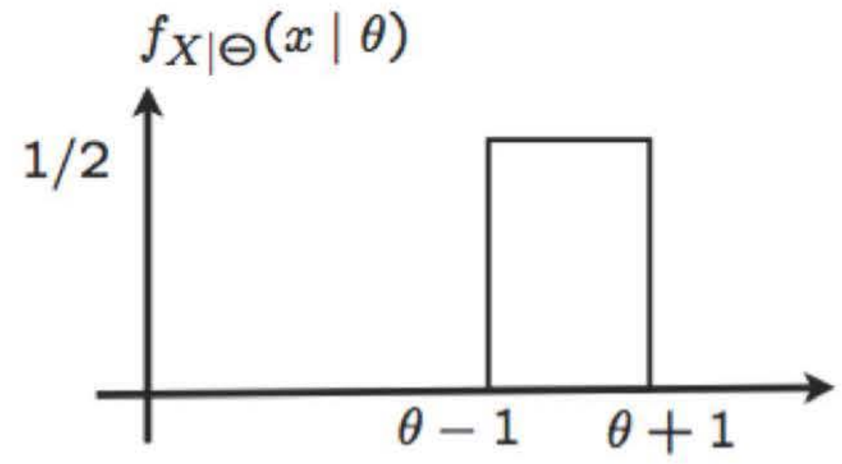
- LMS relevant to estimation (not hypothesis testing)

- Same as MAP if the posterior is unimodal and symmetric around the mean
  - e.g., when posterior is normal (the case in "linear–normal" models)

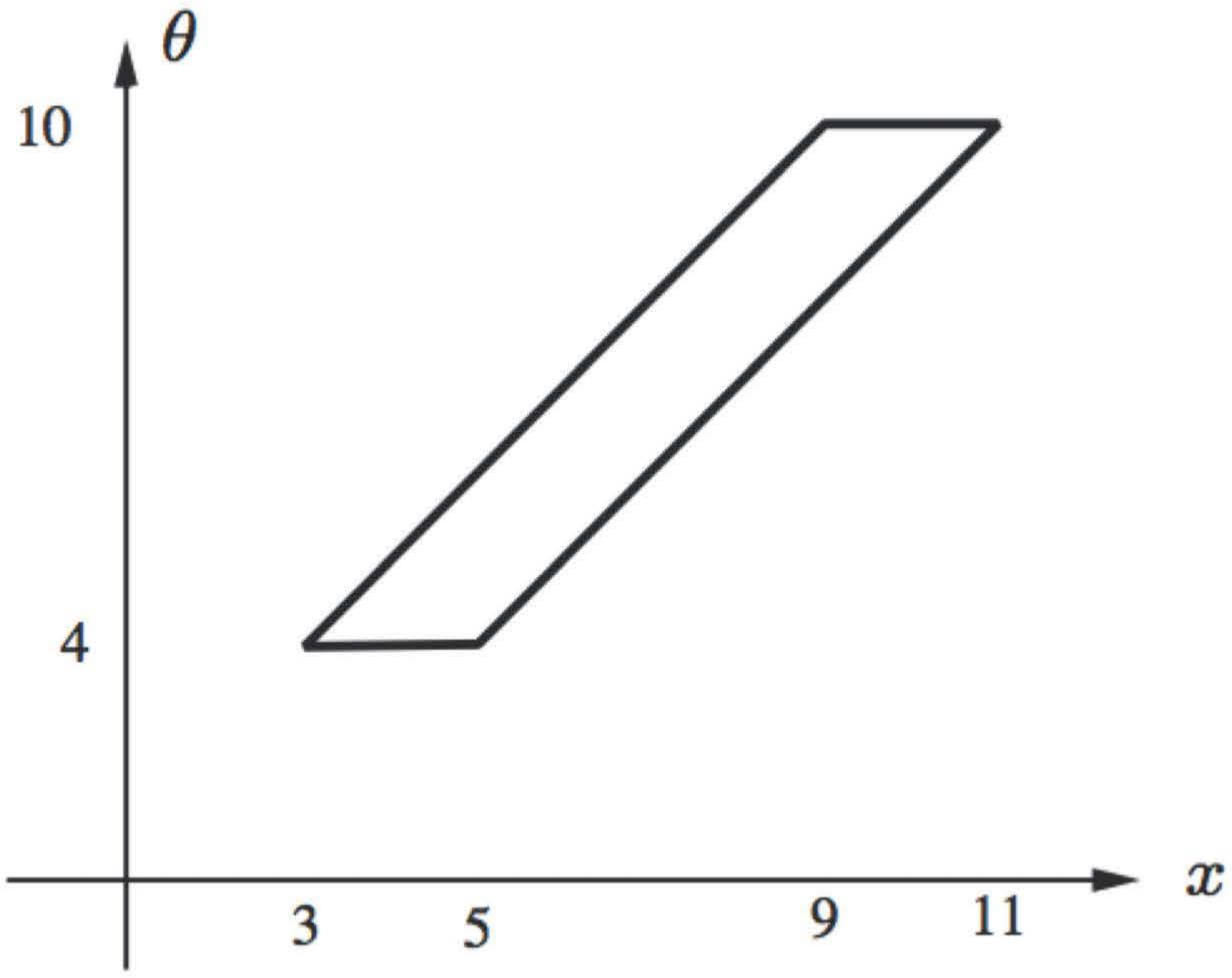
Example



$f_\Theta(\theta)$
1/6
4
10
θ

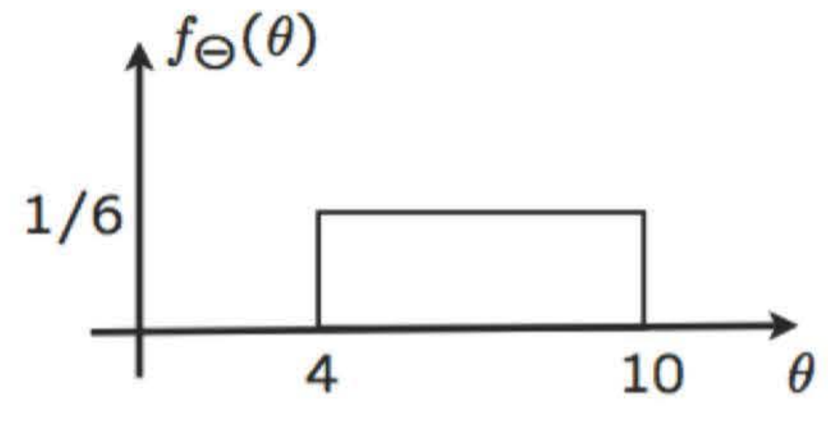


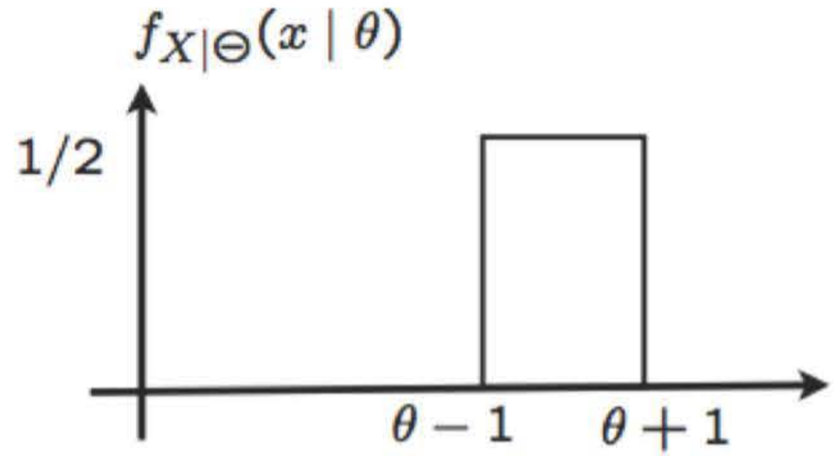
$f_{X|\Theta}(x \mid \theta)$
1/2
$\theta - 1$
$\theta + 1$

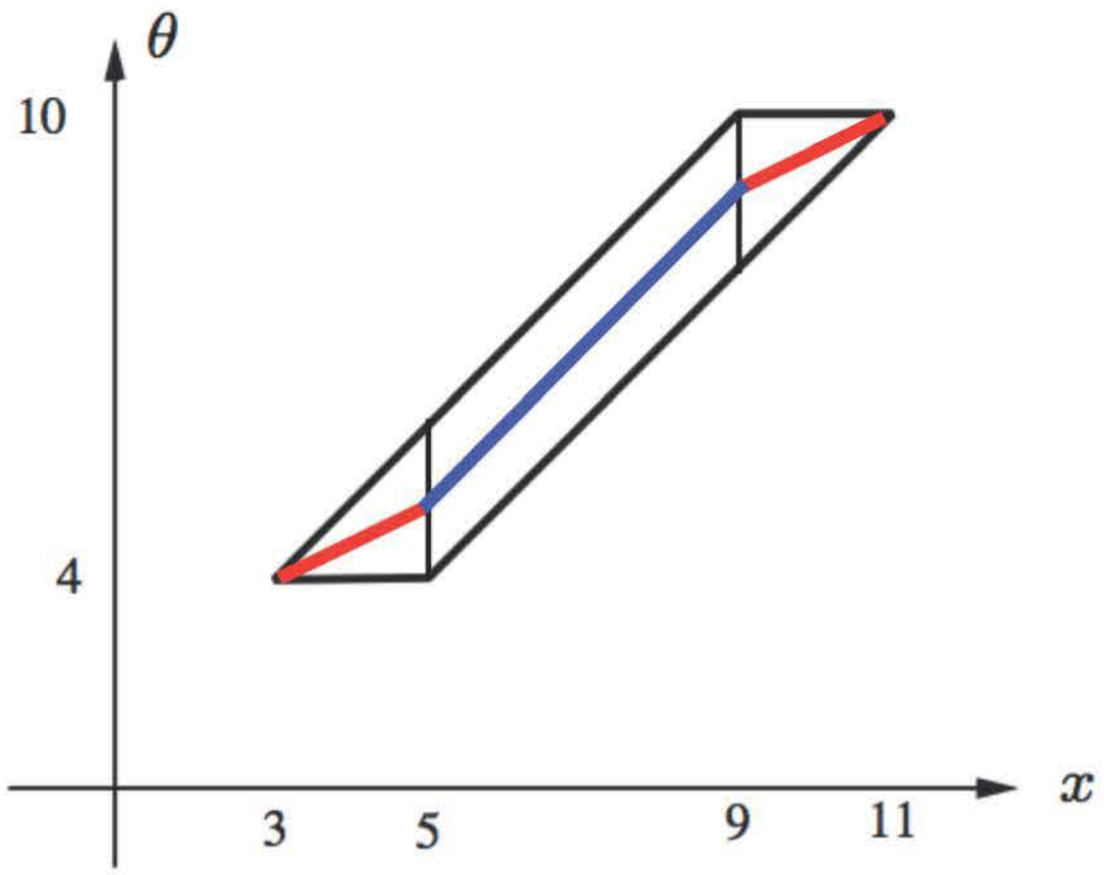


θ
10
4
x
3
5
9
11

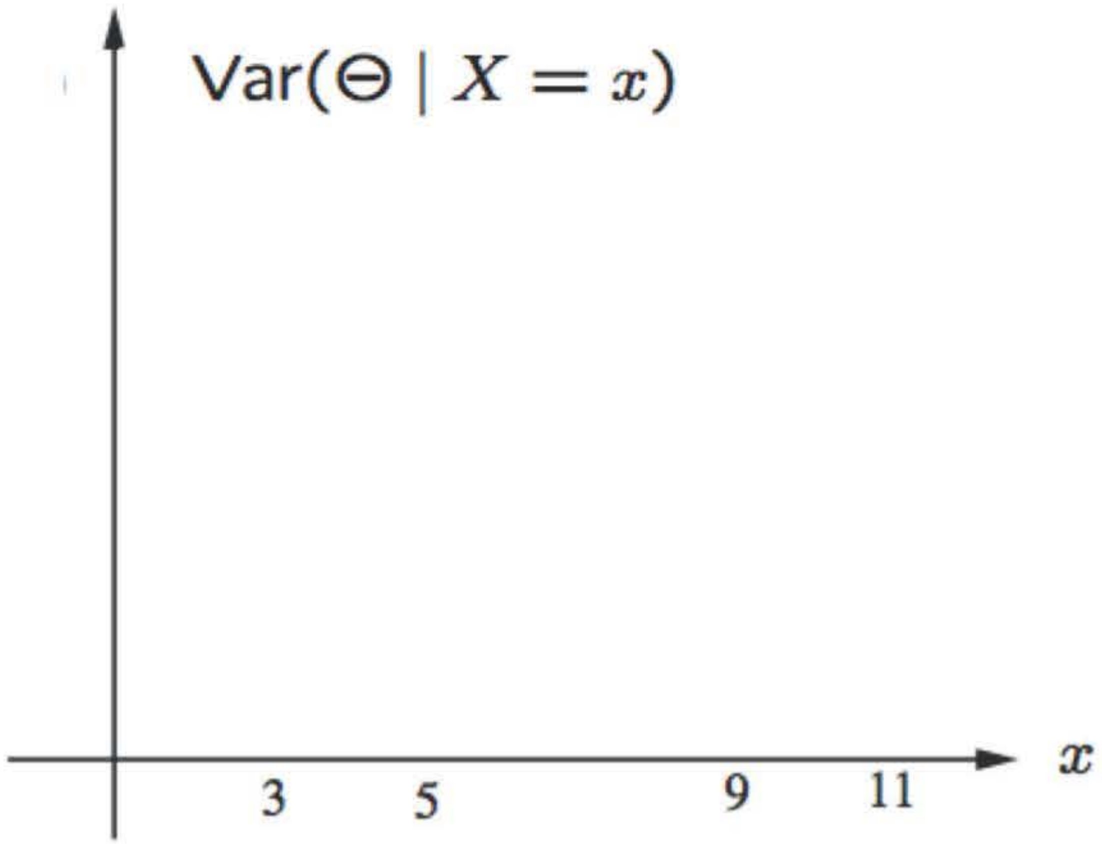
Conditional mean squared error







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



LMS estimation with multiple observations or unknowns

- unknown Θ; prior $p_{\Theta}(\theta)$
  - interested in a point estimate $\widehat{\theta}$

- observations $X = (X_1, X_2, \ldots, X_n)$; model $p_{X|\Theta}(x \mid \theta)$
  - observe that $X = x$
  - new universe: condition on $X = x$

- LMS estimate: $\mathbf{E}[\Theta \mid X_1 = x_1, \ldots, X_n = x_n]$

- If Θ is a vector, apply to each component separately

Some challenges in LMS estimation

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$f_X(x) = \int f_\Theta(\theta') f_{X|\Theta}(x \mid \theta')\, d\theta'$$

- Full correct model, $f_{X|\Theta}(x \mid \theta)$, may not be available
- Can be hard to compute/implement/analyze

Properties of the estimation error in LMS estimation

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

$$\mathbf{E}[\widetilde{\Theta} \mid X = x] = 0$$

$$\text{cov}(\widetilde{\Theta}, \widehat{\Theta}) = 0$$

$$\text{var}(\Theta) = \text{var}(\widehat{\Theta}) + \text{var}(\widetilde{\Theta})$$