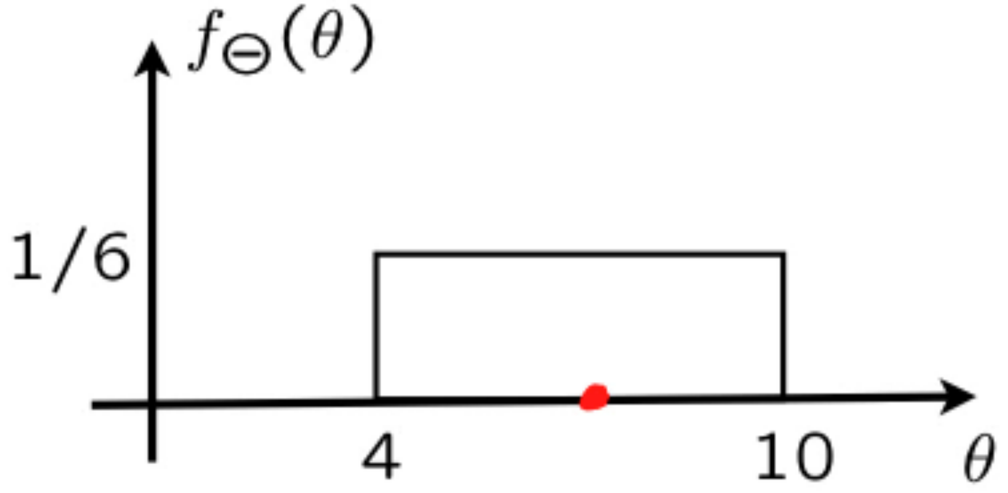
LECTURE 16: Least mean squares (LMS) estimation

- 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: any $\hat{\theta} \in [4, 10]$
  - (Conditional) expectation: $\hat{\theta} = 7$



- 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), $\mathbf{E}\left[(\Theta-\widehat{\theta})^2\right]$: $\widehat{\theta}=\mathbf{E}[\Theta]$

$$\mathbf{E}[\Theta^2] - 2\mathbf{E}[\Theta]\hat{\theta} + \hat{\theta}^2 \qquad \frac{d}{d\hat{\theta}} = 0: \; -2\mathbf{E}[\Theta] + 2\hat{\theta} = 0$$

$$\hat{\theta} = \mathbf{E}[\Theta]$$

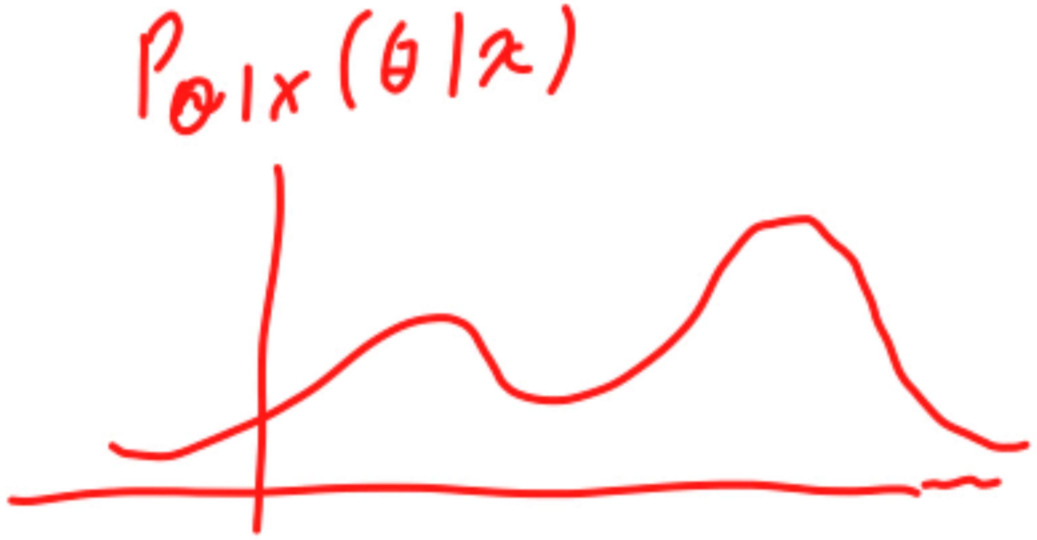
$$\underset{\displaystyle \mathrm{Var}(\Theta)}{\mathrm{Var}(\Theta - \hat{\theta})} + \underbrace{(\mathbf{E}[\Theta - \hat{\theta}])^2}_{}$$

minimized when $\hat{\theta} = \mathbf{E}[\Theta]$

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

LMS estimation of Θ based on X

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



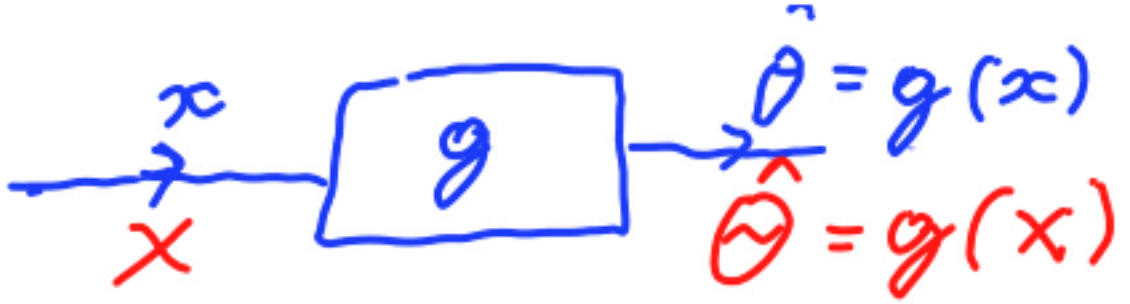
minimize mean squared error (MSE), $\mathbf{E}\left[(\Theta - \hat{\theta})^2\right]$: $\hat{\theta} = \mathbf{E}[\Theta]$

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

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

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

LMS estimation of Θ based on X



- $\mathbf{E}[\Theta]$ minimizes $\mathbf{E}\left[(\Theta - \hat{\theta})^2\right]$

$$E\left[(\Theta - E[\Theta])^2\right] \le E\left[(\Theta - c)^2\right], \text{ for all } c$$

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

$$E\left[(\Theta - E[\Theta \mid X = x])^2 \mid X = x\right] \le E\left[(\Theta - g(x))^2 \mid X = x\right] \text{ for all } x$$

$$E\left[(\Theta - E[\Theta \mid X])^2 \mid X\right] \le E\left[(\Theta - g(X))^2 \mid X\right]$$

$$E\left[(\Theta - \underline{E[\Theta \mid X]})^2\right] \le E\left[(\Theta - g(X))^2\right]$$

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

LMS performance evaluation

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

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

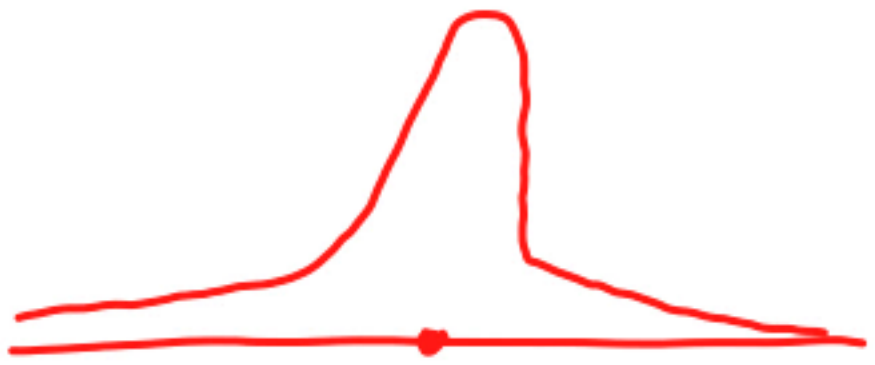
– Expected performance, once we have a measurement:

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

– Expected performance of the design:

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

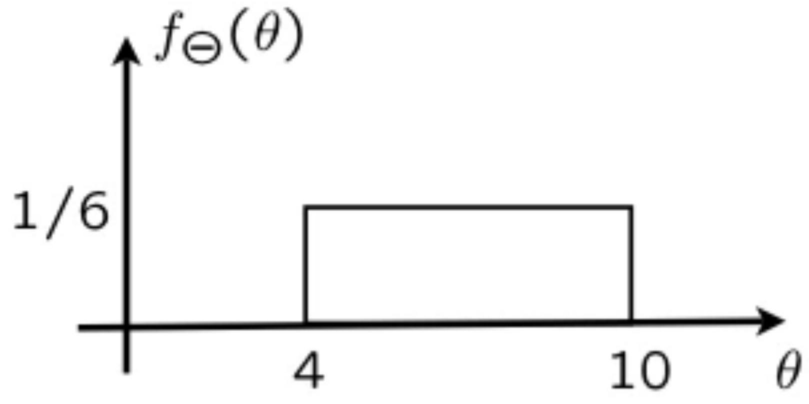
LMS estimation of Θ based on X

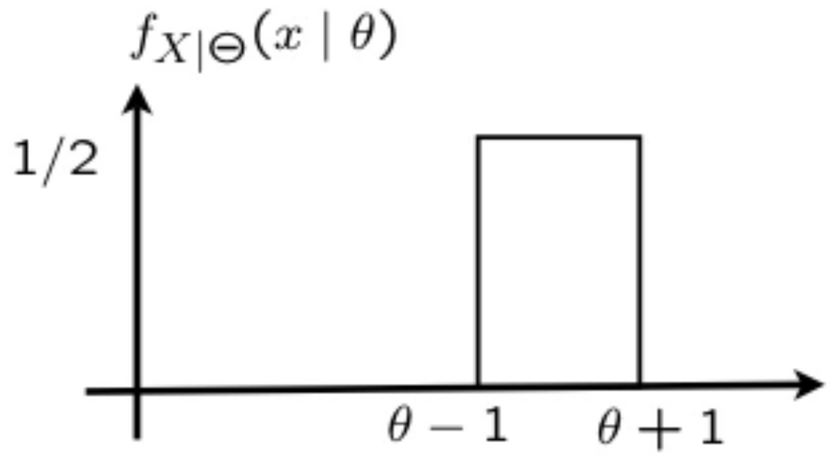


- LMS relevant to estimation (not hypothesis testing)

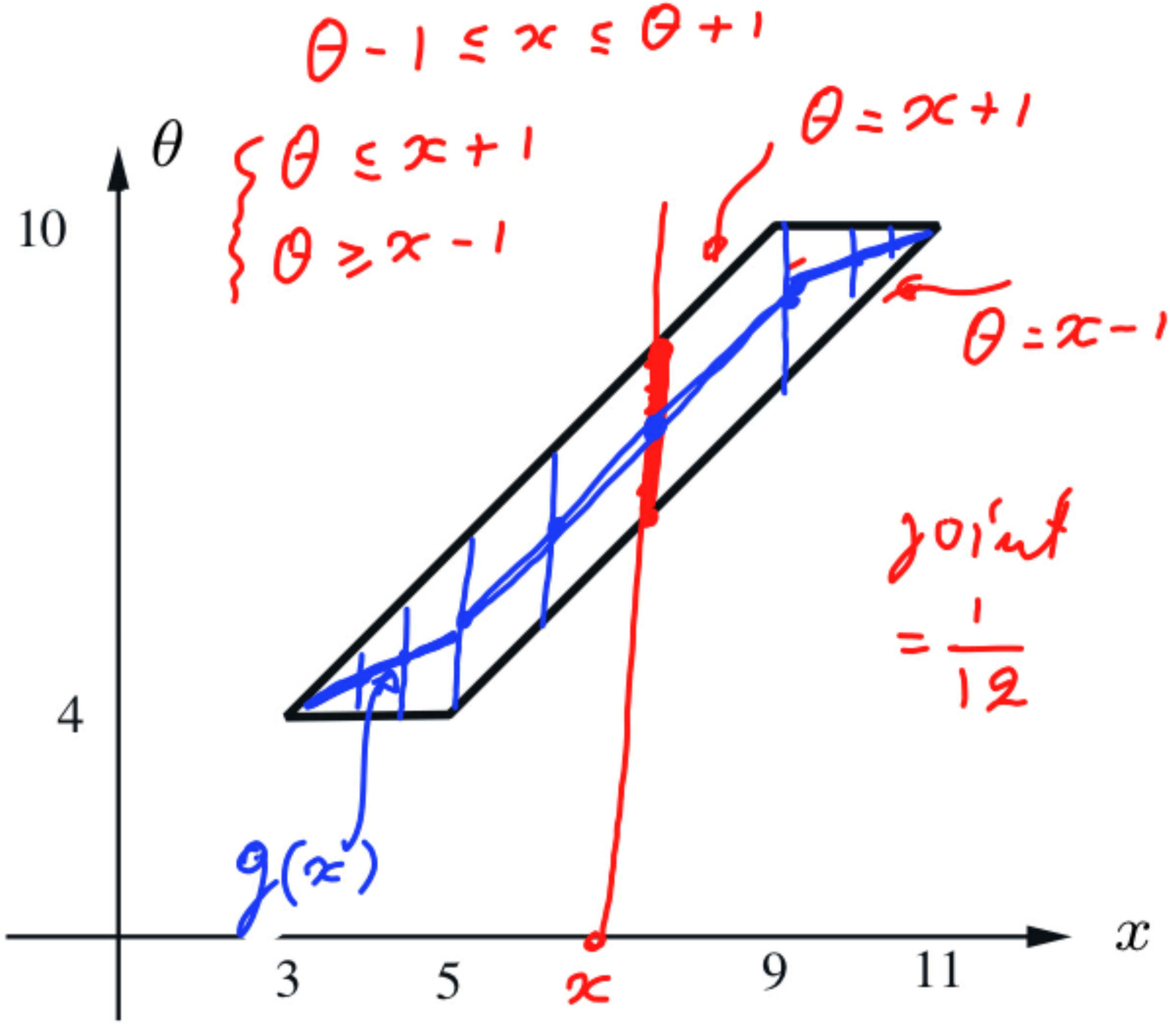
- Same as MAP if the posterior is unimodal and symmetric around the mean
  - e.g., when posterior is normal (the case in "linear–normal" models)

Example

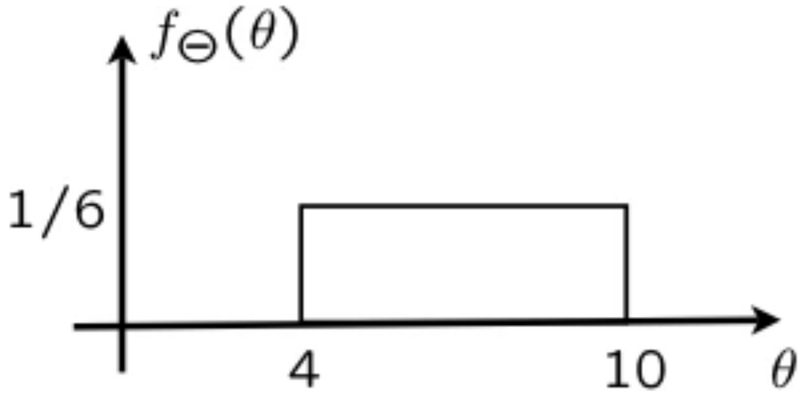


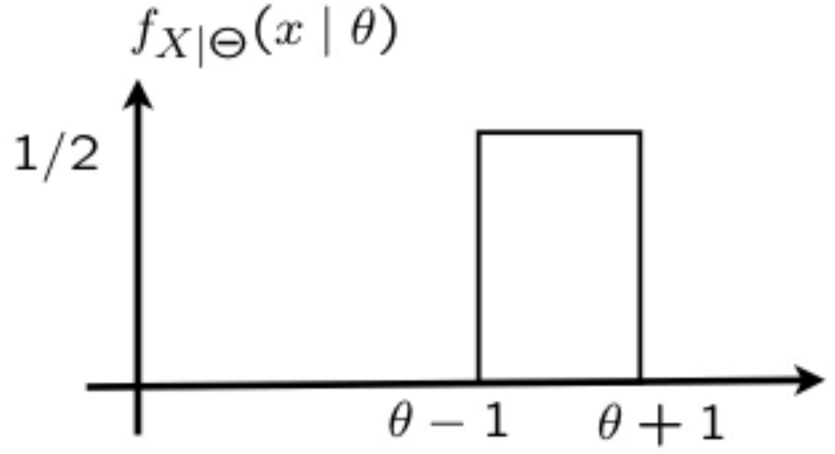


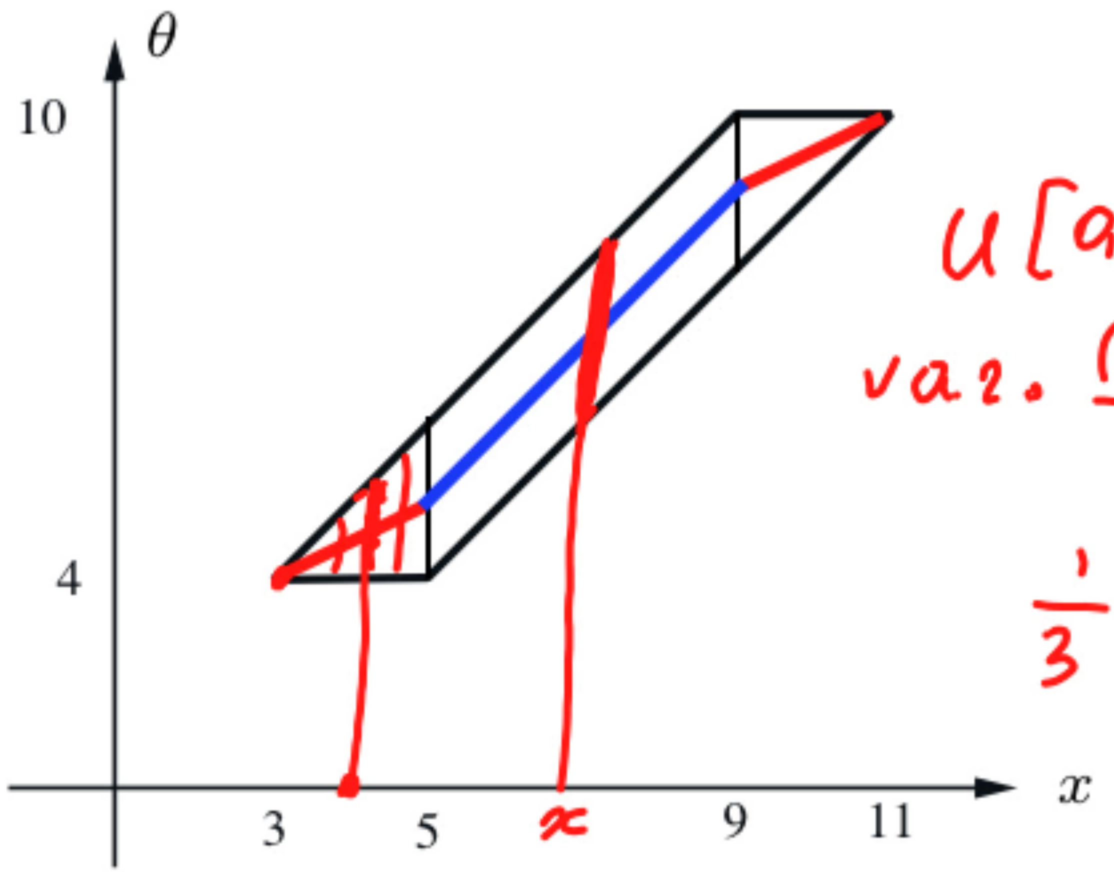
$x = \Theta + U \qquad U \sim \text{unif}(-1, 1)$



Conditional mean squared error







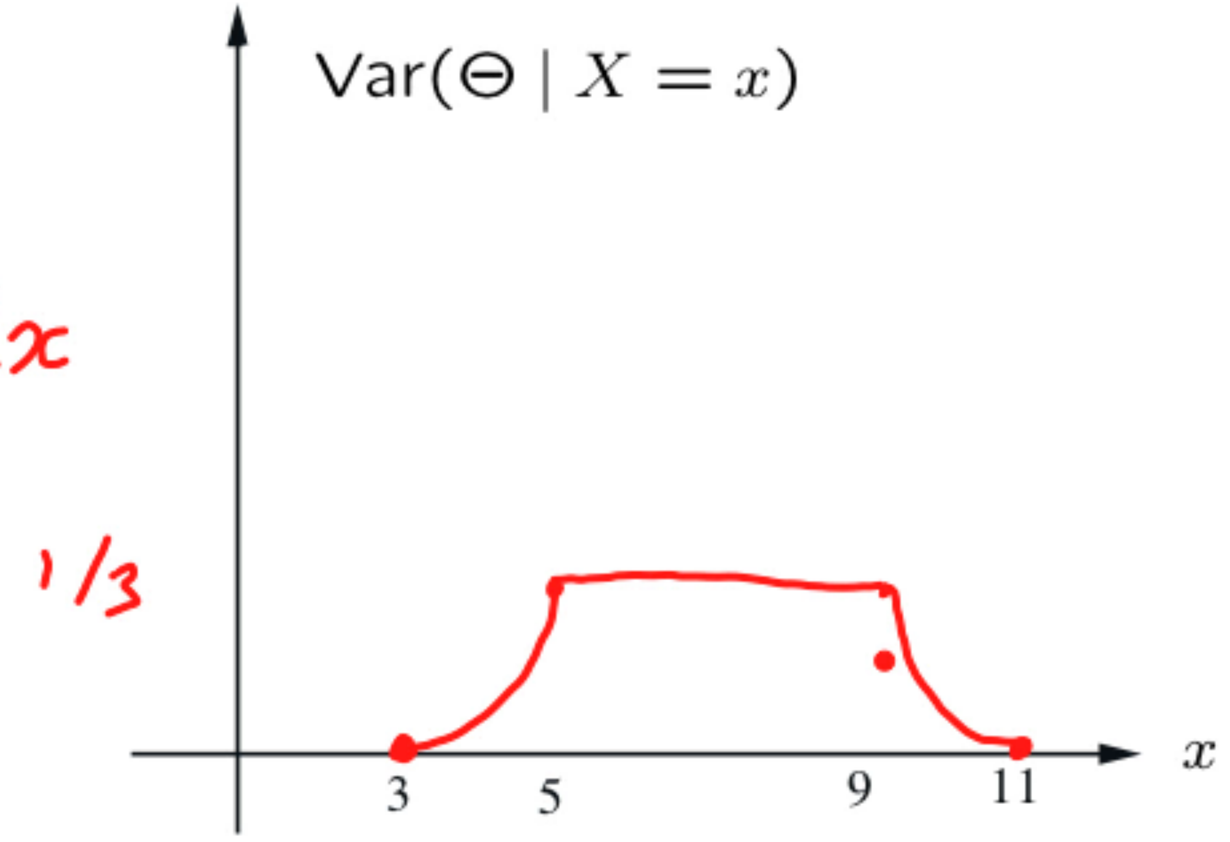
$U[a,b]$

$\text{var} = \dfrac{(b-a)^2}{12}$

$\dfrac{1}{3}$

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

$E[\mathrm{Var}(\Theta|X)] = \int f_X(x)\,\mathrm{Var}(\Theta|X=x)\,dx$



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

$$\Theta = (\Theta_1, \ldots, \Theta_m) \qquad \hat{\Theta}_j = E[\Theta_j \mid X_1 = x_1, \ldots, X_n = x_n]$$

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

$$E[\Theta_j \mid X = x] = \iiint \theta_j \, f_{\Theta|X}(\theta \mid x)\, d\theta_1 \cdots d\theta_m$$

Properties of the estimation error in LMS estimation

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

$E[\hat{\Theta}] = E[\Theta]$

$E[\tilde{\Theta}] = 0$

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

$E[\hat{\Theta} - \Theta \mid X = x] = \hat{\Theta} - E[\Theta \mid X = x] = 0$

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

$E[\tilde{\Theta}\hat{\Theta}] - \cancel{E[\tilde{\Theta}]} E[\hat{\Theta}]$ 0

$E[\tilde{\Theta}\hat{\Theta} \mid X = x] = \hat{\Theta} E[\tilde{\Theta} \mid X = x] = 0$

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

$\Theta = \hat{\Theta} - \tilde{\Theta}$