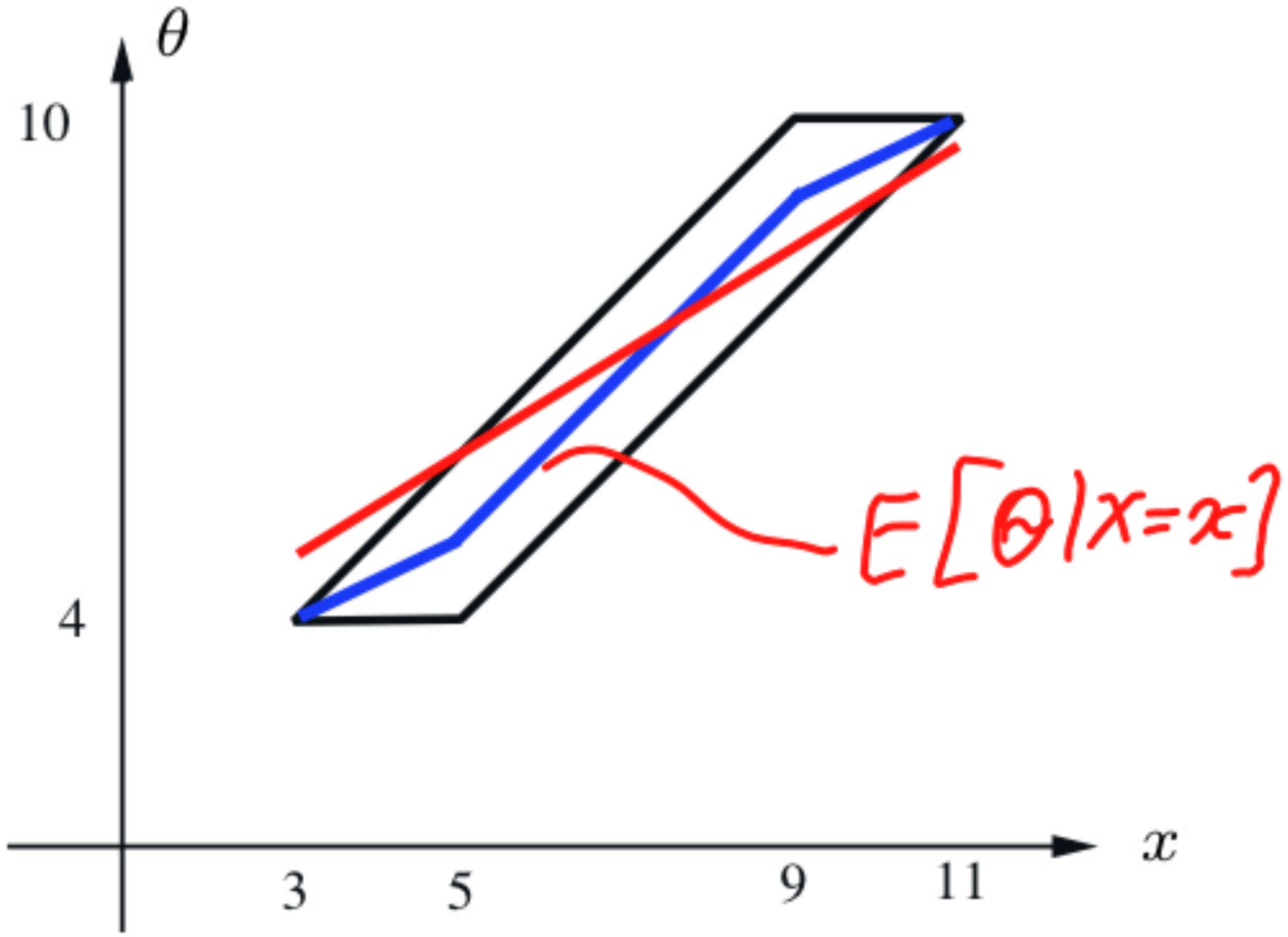
LECTURE 17: Linear least mean squares (LLMS) estimation

- Conditional expectation $\mathbf{E}[\Theta \mid X]$ may be hard to compute/implement
- Restrict to estimators $\widehat{\Theta} = aX + b$
  - minimize mean squared error
- Simple solution
- Mathematical properties
- Example

LLMS formulation

- Unknown Θ; observation X



- Minimize $\mathbf{E}\Big[(\widehat{\Theta}-\Theta)^2\Big]$
- Estimators $\widehat{\Theta}=g(X) \longrightarrow \widehat{\Theta}_{\mathrm{LMS}}=\mathbf{E}[\Theta \mid X]$
- Consider estimators of Θ, of the form $\widehat{\Theta}=aX+b$
- Minimize $\mathbf{E}\Big[(\Theta-aX-b)^2\Big]$, w.r.t. a, b
- If $\mathbf{E}[\Theta \mid X]$ is linear in X, then $\widehat{\Theta}_{\mathrm{LMS}}=\widehat{\Theta}_{\mathrm{LLMS}}$

Solution to the LLMS problem

- Minimize $\mathbf{E}\left[(\underbrace{\Theta - aX} - b)^2\right]$, w.r.t. a, b
  - suppose a has already been found: $b = E[\Theta] - aE[X]$

$$\min \; E\left[(\Theta - aX - E[\Theta - aX])^2\right] = \mathrm{var}(\Theta - aX)$$

$$= \mathrm{var}(\Theta) + a^2 \mathrm{var}(X) - 2a\,\mathrm{cov}(\Theta, X)$$

$$\frac{d}{da} = 0 : \; 2a\,\mathrm{var}(X) - 2\,\mathrm{cov}(\Theta, X) = 0$$

$$a = \mathrm{cov}(\Theta, X)/\mathrm{var}(X)$$

$$\rho = \frac{\mathrm{cov}(\Theta, X)}{\sigma_\Theta \sigma_X}$$

$$a = \frac{\rho \sigma_\Theta \sigma_X}{\sigma_X^2}$$

$$\widehat{\Theta}_L = \mathbf{E}[\Theta] + \frac{\mathrm{Cov}(\Theta, X)}{\mathrm{var}(X)}\big(X - \mathbf{E}[X]\big) = \mathbf{E}[\Theta] + \rho\frac{\sigma_\Theta}{\sigma_X}\big(X - \mathbf{E}[X]\big)$$

Remarks on the solution and on the error variance

$$\boxed{\widehat{\Theta}_L = \mathbf{E}[\Theta] + \frac{\mathsf{Cov}(\Theta, X)}{\mathsf{var}(X)}\big(X - \mathbf{E}[X]\big) = \mathbf{E}[\Theta] + \rho\frac{\sigma_\Theta}{\sigma_X}\big(X - \mathbf{E}[X]\big)}$$

- Only means, variances, covariances matter
- $\rho > 0$: $X > E[X] \Rightarrow \hat{\Theta}_L > E[\Theta]$
- $\rho = 0$: $\hat{\Theta}_L = E[\Theta]$

$|\rho| = 1$

$\hat{\Theta}_L = \Theta$

$$\boxed{\mathbf{E}\Big[(\widehat{\Theta}_L - \Theta)^2\Big] = (1 - \rho^2)\,\mathsf{var}(\Theta)}$$

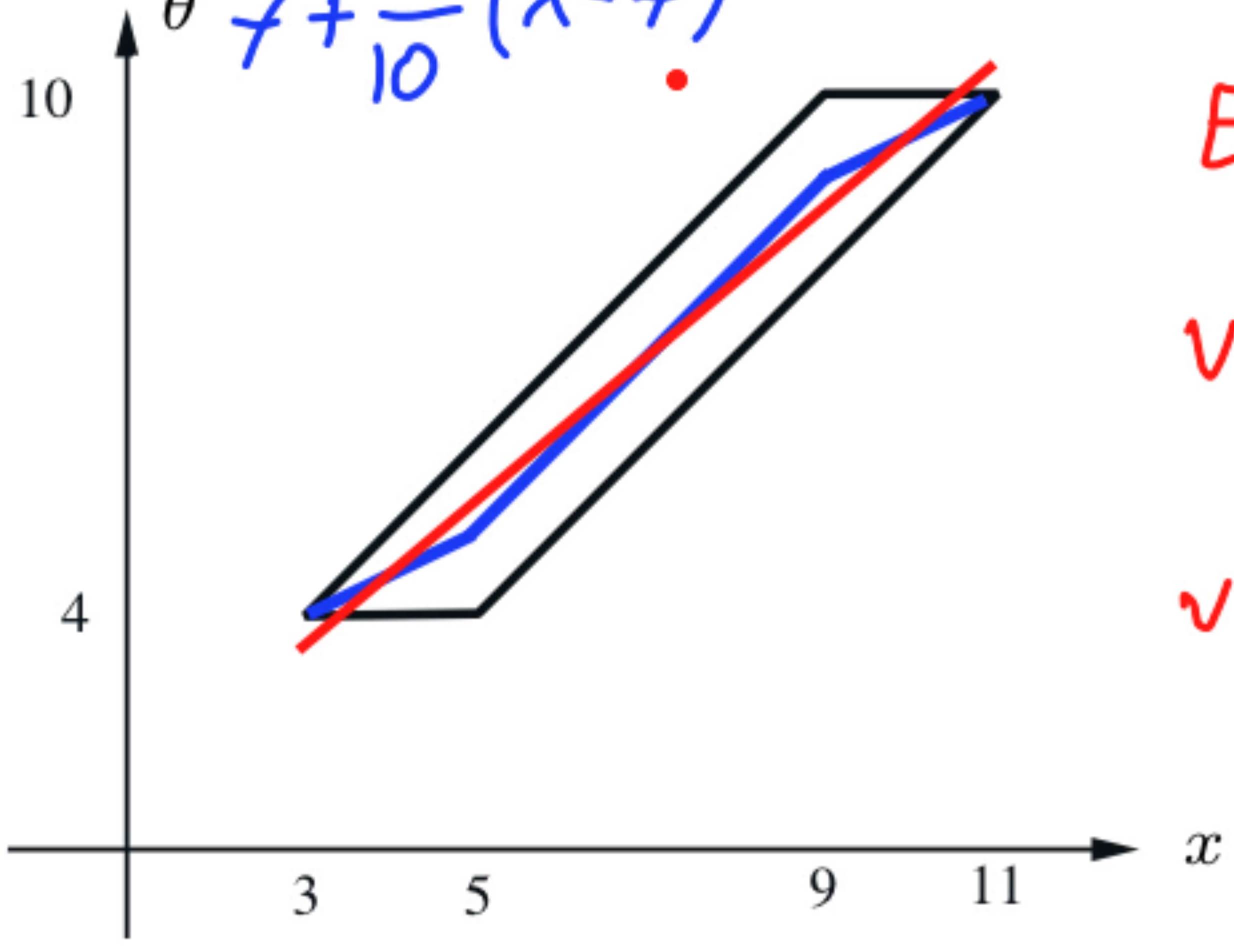
assume $E[\Theta] = E[X] = 0$

$$E\left[\left(\Theta - \rho\frac{\sigma_\Theta}{\sigma_X}X\right)^2\right] = \sigma_\Theta^2 - 2\rho\frac{\sigma_\Theta}{\cancel{\sigma_X}}\rho\sigma_\Theta\cancel{\sigma_X} + \rho^2\frac{\sigma_\Theta^2}{\cancel{\sigma_X^2}}\cancel{\sigma_X^2}$$

Example

$$\widehat{\Theta}_L = \mathbf{E}[\Theta] + \frac{\mathsf{Cov}(\Theta, X)}{\mathsf{var}(X)}\big(X - \mathbf{E}[X]\big) = \mathbf{E}[\Theta] + \rho\frac{\sigma_\Theta}{\sigma_X}\big(X - \mathbf{E}[X]\big)$$

$7 + \frac{9}{10}(x-7)$



$E[\Theta] = 7 \qquad E[U] = 0 \qquad E[X] = 7$

$\mathrm{var}(\Theta) = \frac{6^2}{12} = 3 \qquad \mathrm{var}(U) = \frac{2^2}{12} = \frac{1}{3}$

$\mathrm{var}(X) = 3 + \frac{1}{3} = \frac{10}{3}$

$\mathrm{cov}(\Theta, \Theta + U) =$

$= \mathrm{cov}(\Theta, \Theta) + \cancel{\mathrm{cov}(\Theta, U)} = 3$

Θ : uniform $[4, 10]$

$X = \Theta + U \qquad U$ uniform $[-1, 1]$

Θ, U independent

LLMS for inferring the parameter of a coin

- Standard example:
  - coin with bias Θ; prior $f_\Theta(\cdot)$
  - fix n; $X =$number of heads

- Assume $f_\Theta(\cdot)$ is uniform in $[0, 1]$

$$\widehat{\Theta}_{\text{LMS}} = \frac{X+1}{n+2} = \widehat{\Theta}_{\text{LLMS}}$$

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

LLMS for inferring the parameter of a coin

- Θ: uniform on $[0,1]$ $\quad \mathbf{E}[\Theta] = \frac{1}{2}$ $\quad \text{var}(\Theta) = \frac{1}{12}$ $\quad \mathbf{E}[\Theta^2] = \frac{1}{12} + \frac{1}{2^2} = \frac{1}{3}$
- $p_{X|\Theta}$: $\text{Bin}(n, \Theta)$ $\quad \mathbf{E}[X \mid \Theta] = n\Theta$ $\quad \text{var}(X \mid \Theta) = n\Theta(1-\Theta)$

$$\mathbf{E}[X] = E[n\Theta] = n/2 \qquad \mathbf{E}[X^2 \mid \Theta] = n\Theta(1-\Theta) + n^2\Theta^2$$

- $$\mathbf{E}[X^2] = E\big[E[X^2 \mid \Theta]\big] = E\big[n\Theta + (n^2-n)\Theta^2\big] = \frac{n}{2} + \frac{n^2-n}{3} = \frac{n}{6} + \frac{n^2}{3}$$

$$\text{var}(X) = E[X^2] - (E[X])^2 = \frac{n}{6} + \frac{n^2}{3} - \frac{n^2}{4} = \frac{n}{6} + \frac{n^2}{12} = \frac{n(n+2)}{12}$$

$$\mathbf{E}[\Theta X \mid \Theta] = \Theta E[X \mid \Theta] = n\Theta^2$$

$$\mathbf{E}[\Theta X] = E\big[E[\Theta X \mid \Theta]\big] = E[n\Theta^2] = n/3$$

$$\text{cov}(\Theta, X) = E[\Theta X] - E[\Theta]E[X] = \frac{n}{3} - \frac{n}{4} = \frac{n}{12}$$

LLMS for inferring the parameter of a coin

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

$$\text{cov}(\Theta, X) = \frac{n}{12} \qquad \text{var}(X) = \frac{n(n+2)}{12} \qquad \mathbf{E}[X] = \frac{n}{2}$$

$$\widehat{\Theta}_{\text{LLMS}} = \frac{X+1}{n+2} = \hat{\Theta}_{LMS}$$

LLMS with multiple observations

- Unknown Θ; observations $X = (X_1, \ldots, X_n)$

- 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: $\mathbf{E}\Big[(a_1 X_1 + \cdots + a_n X_n + b - \Theta)^2\Big] = a_1^2 E[X_1^2] + 2a_1 a_2 E[X_1 X_2] + \cdots + a_1 E[X_1 \Theta] + \cdots$

- If $\mathbf{E}[\Theta \mid X]$ is linear in X, then $\widehat{\Theta}_{\mathsf{LMS}} = \widehat{\Theta}_{\mathsf{LLMS}}$

- Solve linear system in b and the a_i

- Only means, variances, covariances matter

- If multiple unknown Θ_j, apply to each one, separately

The simplest LLMS example with multiple observations

$$X_1 = \Theta + W_1 \qquad \Theta \sim x_0,\ \sigma_0^2 \qquad W_i \sim 0,\ \sigma_i^2$$

$$\vdots$$

$$X_n = \Theta + W_n \qquad \Theta, W_1, \ldots, W_n \text{ uncorrelated}$$

- Suppose $\Theta, W_1, \ldots, W_n$ are independent normal

$$\hat{\theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X = x] = \frac{\displaystyle\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} \qquad \widehat{\Theta}_{\text{LMS}} = \mathbf{E}[\Theta \mid X] = \frac{\dfrac{x_0}{\sigma_0^2} + \displaystyle\sum_{i=1}^{n} \frac{X_i}{\sigma_i^2}}{\displaystyle\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} = \widehat{\Theta}_{\text{LLMS}}$$

- Suppose general (not normal) distributions,
  but same means, variances, as in normal example
  - all covariances also the same
  - solution must be the same

The representation of the data matters in LLMS

- Estimation based on X versus X^3
  - LMS: $\mathbf{E}[\Theta \mid X]$ is the same as $\mathbf{E}[\Theta \mid X^3]$
  - LLMS is different: estimator $\widehat{\Theta} = aX + b$ versus $\widehat{\Theta} = aX^3 + b$

    $\mathrm{cov}(\Theta, X^3)$ $\quad \mathrm{var}(X^3)$

  - can also consider $\widehat{\Theta} = a_1 X + a_2 X^2 + a_3 X^3 + b$
  - can also consider $\widehat{\Theta} = a_1 X + a_2 e^X + a_3 \log X + b$