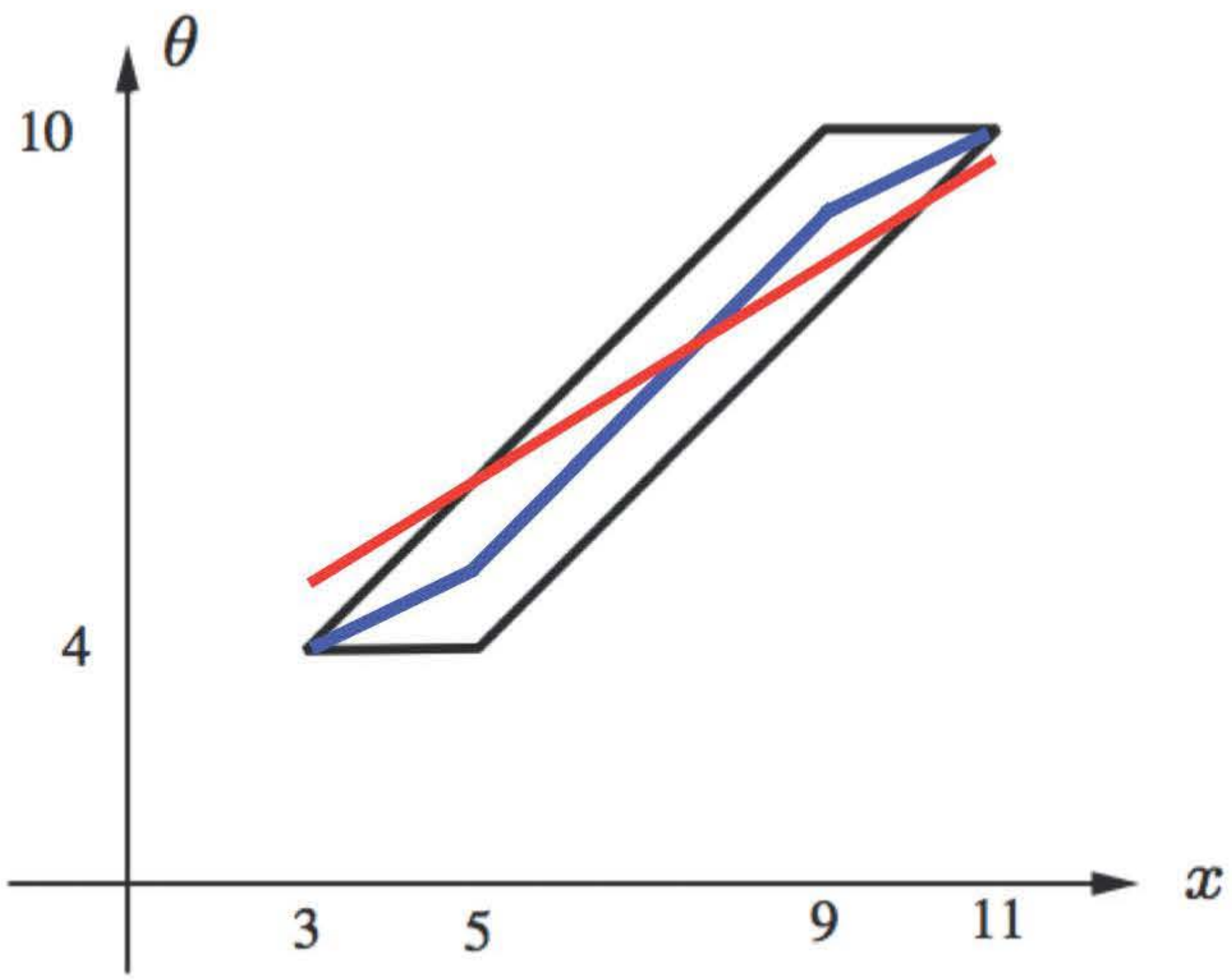
LECTURE 17: Linear least mean squares (LLMS) estimation

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

Solution to the LLMS problem

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

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

Remarks on the solution and on the error variance

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

- Only means, variances, covariances matter

- $\rho > 0$:

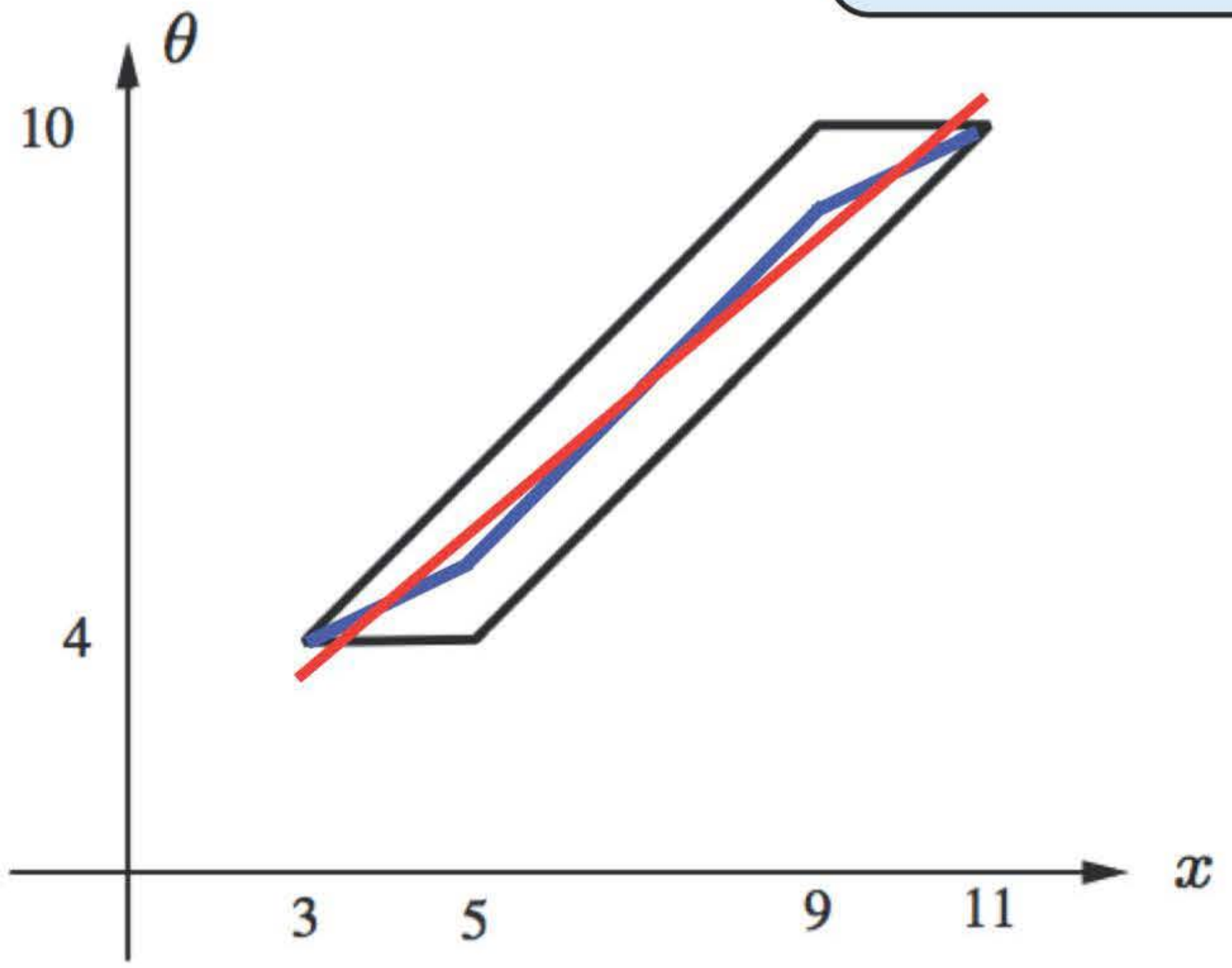
- $\rho = 0$:

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

- $|\rho| = 1$:

Example

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



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

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

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

$\mathrm{var}(X) =$

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

$\mathbf{E}[\Theta X] =$

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

LLMS for inferring the parameter of a coin

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

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

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

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: $\mathbf{E}\Big[(a_1 X_1 + \cdots + a_n X_n + b - \Theta)^2\Big]$
- If $\mathbf{E}[\Theta \mid X]$ is linear in X, then $\widehat{\Theta}_{\text{LMS}} = \widehat{\Theta}_{\text{LLMS}}$
- 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

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

- 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 $\widehat{\Theta} = aX^3 + b$
  - can also consider $\widehat{\Theta} = a_1X + a_2X^2 + a_3X^3 + b$
  - can also consider $\widehat{\Theta} = a_1X + a_2e^X + a_3 \log X + b$