LECTURE 22: LINEAR LEAST MEAN-SQUARED ERROR (LMS) ESTIMATION

November 26, 2008

Reading: Sect. 8.4

Given Y, we want an estimator $\widehat{X}(Y) = g(Y)$ that gives a good estimate for X. Here, ***good*** means it has the LMS (least mean squared) error property.

If we find the best ***linear*** estimator (i.e., one with less LMS error than any other ***linear*** estimator), then

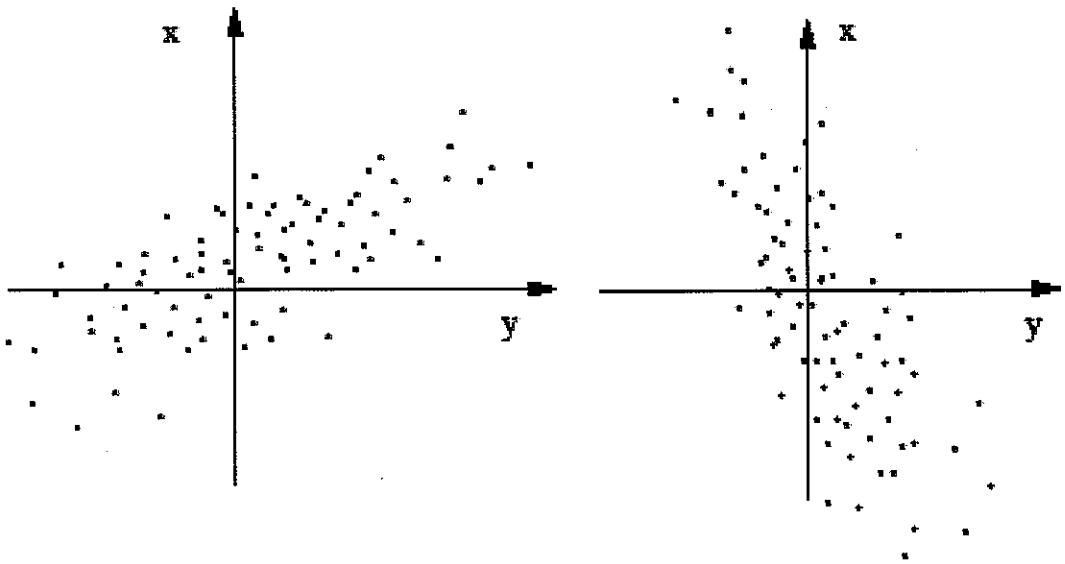
1) (***the bad news***) poorer performance results: the optimum (i.e., minimum) expected squared error ***increases*** over that of the conditional mean (or stays the same if the conditional mean was linear in the first place), and

2) (***the good news***) the estimator is generally simpler to find, requires much less knowledge of the joint distribution of X and Y, and is easier to implement.

Covariance

- $\operatorname{cov}(X,Y) \triangleq \boldsymbol{E}\left[(X-\boldsymbol{E}[X])\bullet(Y-\boldsymbol{E}[Y])\right]$ (1)

- Zero-mean case: $\operatorname{cov}(X,Y)=\boldsymbol{E}\left[XY\right]$



- $\operatorname{cov}(X,Y)=\boldsymbol{E}[XY]-\boldsymbol{E}[X]\boldsymbol{E}[Y]$

- $\operatorname{var}\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n}\operatorname{var}(X_i)+2\sum_{i,j:\, i<j}\operatorname{cov}(X_i,X_j)$

- independent $\Rightarrow \operatorname{cov}(X,Y)=0$
  (converse is not true)

To find the linear LMS (least mean squared) error estimator of X given Y, minimize

$$E[(X-(aY+b))^2]. \qquad (2)$$

Use the zero-mean variables $(X-\mu_X)$ and $(Y-\mu_Y)$ and set the derivative of $E[((X-\mu_X) - a(Y-\mu_Y))^2]$ with respect to a to zero:

$$0 = \frac{d}{da} E[((X-\mu_X) - a(Y-\mu_Y))^2] =$$

$$\frac{d}{da}\left\{E[(X-\mu_X)^2 - 2aE[(X-\mu_X)(Y-\mu_Y)] + a^2 E[(Y-\mu_Y)^2]\right\} =$$

$$-2E[(X-\mu_X)(Y-\mu_Y)] + 2aE[(Y-\mu_Y)^2] = 0$$

$$a = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{E[(Y-\mu_Y)^2]} = \frac{\operatorname{cov}(X,Y)}{\sigma_Y^2}$$

Thus the linear LMS estimator of $(X-\mu_X)$ is:

$$\widehat{X}_{LMMSE}(Y)-\mu_X = a(Y-\mu_Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_Y^2}(Y-\mu_Y).$$

Therefore

$$\boxed{\widehat{X}_{linearLMS}(Y) = \mu_X + \frac{\mathrm{cov}(X,Y)}{\sigma_Y^2}(Y-\mu_Y) \qquad (3)}$$

Is the linear LMS estimator biased?

$$\text{error} \triangleq e \triangleq \widehat{X}_{Linear\ \mathrm{LMS}}(Y) - X = \mu_X + \frac{\mathrm{cov}(X,Y)}{\sigma_Y^2}(Y - \mu_Y) - X$$

$$E(e) = \mu_X + \frac{\mathrm{cov}(X,Y)}{\sigma_Y^2}(\mu_Y - \mu_Y) - \mu_X = 0,$$

i.e.,

$$E[\widehat{X}_{Linear\ \mathrm{LMS}}(Y)] = E[X]$$

so the linear LMS estimator is ***unbiased.***

How Well Does the Linear Estimator Work?

Mean-squared error in the linear LMS estimator:

$$E[e^2] = E[(\mu_X + \frac{\text{cov}(X,Y)}{\sigma_Y^2}(Y-\mu_Y) - X)^2] =$$

$$E[(\frac{\text{cov}(X,Y)}{\sigma_Y^2}(Y-\mu_Y) - (X-\mu_X))^2] =$$

$$[\frac{\text{cov}(X,Y)}{\sigma_Y^2}]^2 \sigma_Y^2 - 2\frac{\text{cov}(X,Y)}{\sigma_Y^2}\text{cov}(X,Y) + \sigma_X^2 =$$

$$\sigma_X^2 - \frac{[\text{cov}(X,Y)]^2}{\sigma_Y^2} = \sigma_X^2[1 - \frac{[\text{cov}(X,Y)]^2}{\sigma_X^2\sigma_Y^2}], \text{ i.e.,}$$

(since the ***correlation coefficient*** $\rho_{X,Y} \triangleq \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$)

$$\boxed{\sigma_e = \sigma_X \sqrt{1-\rho_{X,Y}^2}} \qquad (4)$$

$$\boxed{\sigma_e = \sigma_X \sqrt{1 - \rho_{X,Y}^2}} \qquad (4)$$

$\rho_{X,Y} \triangleq \dfrac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$, $(-1 \le \rho_{X,Y} \le 1)$, is the ***correlation coefficient*** relating X and Y

- $$\sigma_e \le \sigma_X$$

- $$\sigma_e = \sigma_X \Leftrightarrow \mathrm{cov}(X,Y) = 0$$

- Construction of LMMSE estimator

$$\widehat{X}_{LMMSE}(Y) = \mu_X + \frac{\rho_{X,Y}\sigma_X}{\sigma_Y}(Y - \mu_Y)$$

requires knowing only five expectations:

E[X], E[Y], E[X^2], E[Y^2], E[XY]

- $-1 \le \rho_{X,Y} \le 1$, X,Y indep. $\Rightarrow$ $\rho_{X,Y} = 0$.

- $\mathrm{var}(\widehat{X}_{LMMSE}) = \rho^2_{X,Y}\,\mathrm{var}(X) \le \mathrm{var}(X)$

- If X and Y are jointly Gaussian (as in tire pressure example in Lecture 21), then the MMSE estimator turns out to be automatically linear.

In the Gaussian case:

LMS estimator = Linear LMS estimator.

Simple Example

X is a uniform random variable, $-1 \le X \le 1$

Observation Y = X + N, where N is a uniform random variable, $-a \le N \le a$, independent of X.

E[X] = E[Y] = E[N] = 0.

$$\sigma_N^2 = \int_{-a}^{a} \frac{1}{2a} x^2 \, dx = a^2/3, \quad \sigma_X^2 = 1/3,$$

$$\sigma_Y^2 = \sigma_X^2 + \sigma_N^2 = (a^2+1)/3$$

$$\mathrm{cov}(X,Y) = E[XY] = E[X(X+N)] = E[X^2] = 1/3$$

$$\widehat{X}_{linearLMS}(Y) = \mu_X + \frac{\mathrm{cov}(X,Y)}{\sigma_Y^2}(Y-\mu_Y) = \left(\frac{1}{a^2+1}\right)Y$$

$$\rho_{X,Y} \triangleq \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1/3}{\sqrt{1/3}\sqrt{1/3(a^2+1)}} = \frac{1}{\sqrt{a^2+1}}$$

$$E[(\widehat{X}_{linearLMS}(Y)-X)^2] = \sigma_X^2(1-\rho^2{}_{X,Y}) = \frac{1}{3}\frac{a^2}{a^2+1}$$

Much of this is quite intuitive. The expected squared error is always less than the expected squared error for your best estimate made with no data, i.e., it is always less than σ_X^2, and it approaches σ_X^2 from below as a becomes so large the measurement becomes useless.

The largest possible value of Y is 1 + a. For certain values of a, e.g., a = 0.5, the largest value of $\widehat{X}$ is

$$\widehat{X}(Y_{\max}) = \widehat{X}(1+a) = \frac{1+a}{1+a^2} = 6/5,$$

which is larger than the largest possible value of X (which is 1).

Back to the Dating Example

Your date has a maximum amount of time θ (s)he will show up late, with θ uniformly distributed in [0, 1 hour]. On each date (s)he will show up late by an amount X, where X is uniformly distributed in [0, θ]. Given your first date (which includes a single measurement X), find the best ***linear*** estimate of θ, given X.

This is the same problem as the one in which a stick is broken twice. The first break is uniform in [0,1], like θ, and the second is uniform in [0, θ], like X.

$$\hat{\theta}_{linearLMS}(X) = \mu_\theta + \frac{\text{cov}(X,\theta)}{\sigma_X^2}(X - \mu_X)$$

In the stick problem (Ex. 4.17) we showed that

$$\mu_X = 1/4, \quad \sigma_X^2 = 7/144$$
$$\mu_\theta = 1/2, \quad \sigma_\theta^2 = 1/12$$

To find cov(X, θ), we use

$$\mathrm{cov}(X, \theta) = E[\theta X] - E[\theta]E[X]$$

and

$$E[\theta^2] = \sigma_\theta^2 + (\mu_\theta)^2 = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}.$$

$$E[\theta X] = \underset{\theta}{E}[\underset{X}{E}[\theta X \mid \theta] =$$

$$\underset{\theta}{E}[\theta \underset{X}{E}[X \mid \theta] = \underset{\theta}{E}[\theta \frac{\theta}{2}] = E[\frac{\theta^2}{2}] = 1/6$$

Thus

$$\mathrm{cov}(X, \theta) = E[\theta X] - E[\theta]E[X] =$$

$$(1/6) - (1/2)(1/4) = 1/24$$

and finally,

$$\widehat{\theta}linearLMS(X)=\mu_\theta+\frac{\mathrm{cov}(X,\theta)}{\sigma_X^2}(X-\mu_X) =$$

$$\frac{1}{2}+\frac{1/24}{7/144}(X-1/4)=(\frac{6}{7})X+2/7.$$

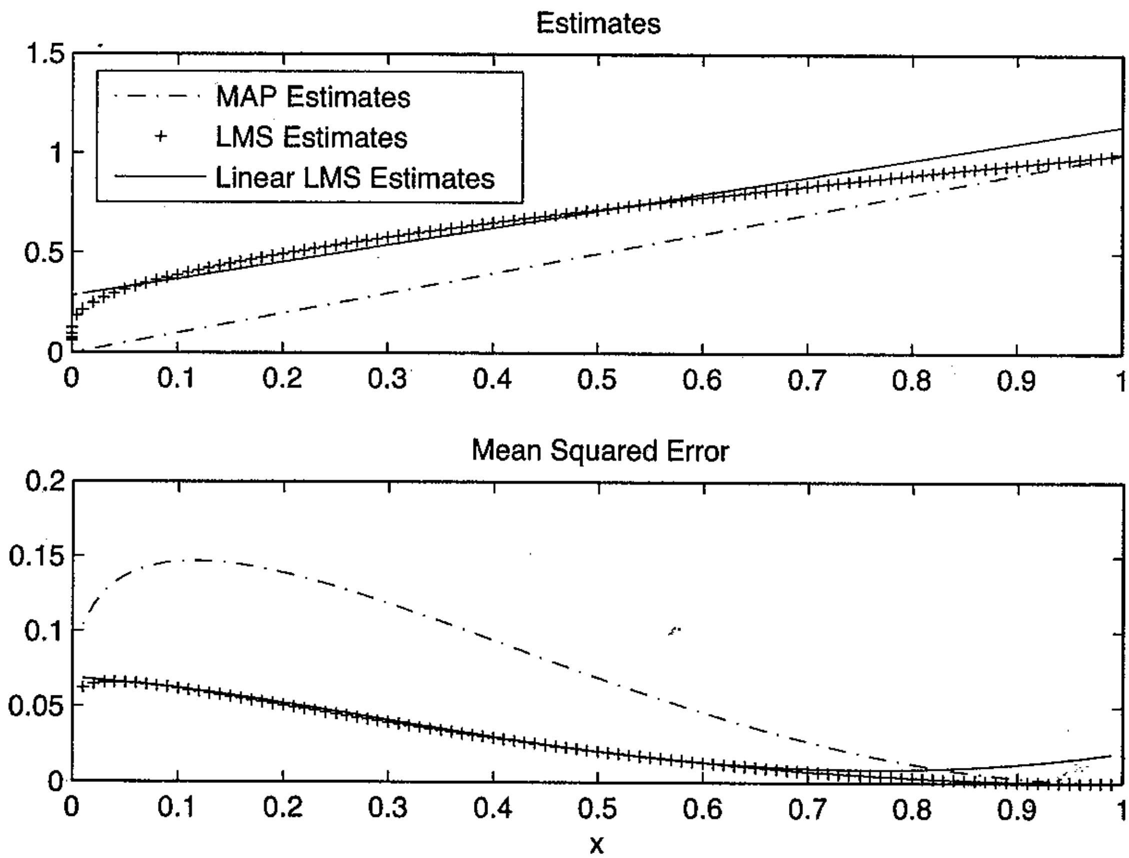


Figure 8.12. Three estimators and their mean squared errors, as functions of the observed value x, for the problem in Example 8.15.