E9 211: Adaptive Signal Processing

Lecture 6: Optimal estimation (scalar-valued data)

INDIAN INSTITUTE OF SCIENCE

Outline

1. Estimation without observations

2. Estimation given dependent observations

3. Gaussian random variables (optimal estimators $=$ affine)

Estimation

- Suppose that all we know about a real-valued random variable x is its mean and variance $\{\bar{x}, \sigma_x^2\}$
- We wish to estimate the value of x in a given experiment. Denote the estimate of x as $\hat{x}$.
- How do we come up with a value $\hat{x}$?
- How do we decide whether this value is optimal or not? If optimal, in what sense?

Mean squared error (lack of observations)

- We shall adopt the mean-squared-error as the design criterion with the error signal

$$\tilde{x} := x - \hat{x}$$

and mean-squared-error

$$E(\tilde{x}^2) := E(x - \hat{x})^2.$$

- We compute $\hat{x}$ by minimizing the mean-squared-error (m.s.e.)

$$\underset{\hat{x}}{\text{minimize}} \;\; E(\tilde{x}^2)$$

$$E(\tilde{x}^2) = E(x - \bar{x} + \bar{x} - \hat{x})^2 = \sigma_x^2 + (\bar{x} - \hat{x})^2$$

- Only the second term $(\bar{x} - \hat{x})^2$ depends on $\hat{x}$ and is annihilated by choosing

$$\hat{x} = \bar{x}$$

Mean squared error (lack of observations)

- Intuitively, the best the guess for x is what we would expect for x on average.
- The criterion is forcing estimation error to assume values close to its mean
$$E(\tilde{x}) = E(x - \tilde{x}) = \bar{x} - \bar{x} = 0$$
Therefore attempting to increase the likelihood of small errors.
- The resulting minimum mean-square error (m.m.s.e) is
$$\text{m.m.s.e.} := E(\tilde{x}^2) = \sigma_x^2$$
The initial uncertainty is not reduced: $\sigma_{\tilde{x}}^2 = \sigma_x^2$.

Given dependent observations

- Suppose we have access to observations

$$y = x + v$$

where v is the noise (or the disturbance), and y is linearly dependent on x.

- How to compute the optimal estimator of x given y:

$$\hat{x} = h(y)$$

for some function $h(\cdot)$ to be determined.

- Different realizations of y lead to different $\hat{x}$.

Least-mean-squared estimator

- We find $\hat{x}$ by minimizing the mean-square-error over all possible functions $h(\cdot)$:

$$\underset{h(\cdot)}{\text{minimize}} \quad E(\tilde{x}^2)$$

- The optimal estimator, i.e., least-mean-squares estimator (l.m.s.e) is given by:

$$\hat{x} = E(x|y) = \int_{S_x} x f_{x|y}(x|y) dx$$

  where S_x is the support of the random variable x and $f_{x|y}(x|y)$ is the conditional density function.

- The estimator is unbiased: $E(\hat{x}) = E(x)$

- The resulting minimum cost is $E(\tilde{x}^2) = \sigma_x^2 - \sigma_{\hat{x}}^2$.

- Often $E(x|y)$ is a nonlinear function of the data or closed-form expression does not exist.

Gaussian random variable case

- We limit to a subclass of estimators that are affine :

$$h(y) = Ky + b$$

for some constants K and b to be determined.

- Although affine estimators are not always optimal, there is an important special case for which the optimal estimator turns out to be affine in y.

- Suppose x and y are jointly Gaussian with the density function

$$f_{x,y}(x,y) = \frac{1}{2\pi}\frac{1}{\sqrt{\det \mathbf{R}}}\exp\left\{-\frac{1}{2}\begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^{\mathrm{T}} \mathbf{R}^{-1} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}\right\}$$

with

$$\mathbf{R} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

where $\{\sigma_x^2, \sigma_y^2, \sigma_{xy}\}$ denote the variances and cross-correlation of x and y, respectively.

Gaussian random variable case

- The l.m.s.e. is given by the affine relation

$$\hat{x} = E(x|y) = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

- m.m.s.e is given by

$$\sigma_{\tilde{x}}^2 = \sigma_x^2 - \frac{\sigma_{xy}}{\sigma_y^2}$$

- Note that the m.m.s.e is smaller than σ_x^2.