ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

PRESENTATION TITLE

Presentation Subtitle

TITLE AND CONTENT SLIDE

- Text

V. Statistical errors and estimation

- Concept of statistical errors
- Estimation of mean and mean square
- Mean square estimation

V. Statistical errors and estimation

- Concept of Statistical Errors
  - Scope of studying
    - In this chapter we deals with accuracy of estimates of parameters for continuous stochastic processes over the time interval with length equals T.
    - Assume that, observations are samples of continuous stationary stochastic process with arbitrary means.
    - We deals with statistical errors in estimating such parameters as: mean, mean square, probability density, covariance function and spectrum density function.

V. Statistical errors and estimation

- Accuracy of some characteristic of stochastic process received from observations of samples functions is described by mean square error.

$$E\{(\hat{\theta}-\theta)^2\}$$

  - Where $\theta^\wedge$ is some estimate of parameter θ.
  - After expansion of mean square error, we have:

$$E\{(\hat{\theta}-\theta)^2\} = E\{(\hat{\theta}-E\{\hat{\theta}\}+E\{\hat{\theta}\}-\theta)^2\} =$$

$$E\{(\hat{\theta}-E\{\hat{\theta}\})^2\} + 2E\{(\hat{\theta}-E\{\hat{\theta}\})(E\{\hat{\theta}\}-\theta)\} + E\{(E\{\hat{\theta}\}-\theta)^2\} =$$

$$E\{\hat{\theta}-E\{\hat{\theta}\}\} = E\{\hat{\theta}\}-E\{\hat{\theta}\} = 0$$

$$E\{(\hat{\theta}-\theta)^2\} = E\{(\hat{\theta}-E\{\hat{\theta}\})^2\} + E\{(E\{\hat{\theta}\}-\theta)^2\}$$

V. Statistical errors and estimation

$$E\{(\hat{\theta}-\theta)^2\} = E\{(\hat{\theta}-E\{\hat{\theta}\})^2\} + E\{(E\{\hat{\theta}\}-\theta)^2\}$$

- This result means: mean square error consists of two parts:
  - the first part– variance of error, which describe level of randomness of error.

$$Var\{\hat{\theta}\} = E\{(\hat{\theta}-E\{\hat{\theta}\})^2\} = E\{\hat{\theta}^2\} - E^2\{\hat{\theta}\}$$

  - The second part – bias square of estimate, which describes systematic bias.
  - So as: $b^2[\hat{\theta}] = E\{b^2[\hat{\theta}]\} = E\{(E\{\hat{\theta}\}-\theta)^2\}$

$$E\{(\hat{\theta}-\theta)^2\} = Var\{\hat{\theta}\} + b^2[\hat{\theta}]$$

  - Standard error deviation – standard error:

$$\sigma(\hat{\theta}) = \sqrt{E\{\hat{\theta}^2\} - E^2\{\hat{\theta}\}}$$

V. Statistical errors and estimation

- The error bias:

$$b[\hat{\theta}] = E\{\hat{\theta}\} - \theta$$

- Normalized standard error is

- $$\varepsilon_r = \frac{\sigma[\hat{\theta}]}{\theta} = \frac{\sqrt{E\{\hat{\theta}^2\} - E^2\{\hat{\theta}\}}}{\theta}$$

$$\varepsilon_b = \frac{E\{\hat{\theta}\}}{\theta} - 1$$

$$\varepsilon_r = \frac{\sqrt{\sigma^2[\hat{\theta}] + b^2[\hat{\theta}]}}{\theta} = \frac{\sqrt{E\{(\hat{\theta} - \theta)^2\}}}{\theta}$$

V. Statistical errors and estimation

- When ε_r is small, we can use:
- So as:

$$\hat{\theta}^2 = \theta^2(1 \pm \varepsilon_r)$$

- Hence:

$$\hat{\theta} = \theta(1 \pm \varepsilon_r)^{1/2} \approx \theta\left(1 \pm \frac{\varepsilon_r}{2}\right)$$

$$\varepsilon_r[\hat{\theta}^2] \approx 2\varepsilon_r[\hat{\theta}]$$

- If the bias of estimate $\theta^\wedge$ is negligible small, i. e b[$\theta^\wedge$] and normalized mean square error $\varepsilon=\varepsilon[\theta^\wedge]=\sigma[\theta^\wedge]/\theta$ is small too, for example $\varepsilon \leq 0.2$, then probability density of this estimate can be approximately gaussian with mean equals $E[\theta^\wedge]=\theta$ and standard deviation $\sigma[\theta^\wedge]=\varepsilon\theta$.

V. Statistical errors and estimation

- Estimation of mean and mean square
  - Assume that, each sample function of process x(t) is defined on finite time interval with length T.
  - Sample mean of the process is:

$$\hat{\mu}_X = \frac{1}{T}\int_0^T x(t)dt$$

  - Actual average value: $\mu_X = E\{X(t)\}$ which does not depend on time t if the process is stationary.

V. Statistical errors and estimation

- Mathematical expectation of estimate $\hat{\mu}_x$ equals:

$$E\{\hat{\mu}_X\} = E\left\{\frac{1}{T}\int_0^T x(t)dt\right\} = \frac{1}{T}\int_0^T E\{x(t)\}dt = \frac{1}{T}\int_0^T \mu_x dt = \mu_x$$

- Hence, $\hat{\mu}_x$ is unbiased estimate of parameter μ_x.
- Due to $\hat{\mu}_x$ is unbiased estimate, so as mean square error of estimate $\hat{\mu}_x$ is equals variance, i. e

$$Var\{\hat{\mu}_x\} = E\{(\hat{\mu}_x - \mu_x)^2\} = E\{\hat{\mu}_x^2\} - \mu_x^2$$

- Therefore

$$E\{\hat{\mu}_x^2\} = \frac{1}{T^2}\int_0^T\int_0^T E\{x(\xi)x(\eta)\}d\eta d\xi$$

V. Statistical errors and estimation

- Autocorrelation function of stationary process X(t)

$$R_{XX}(\tau)=E\{x(t)x(t+\tau)\}$$

- Autocovariance function:

$$C_{XX}(\tau)=R_{XX}(\tau) - \mu^2_X$$

- If autocovariance function is integrable, then process X(t) is ergodic.
- Variance of estimate can be defined using autocovariance function, so that:

$$Var\{\hat{\mu}_x\} = E\{(\hat{\mu}_x - \mu_x)^2\} = \frac{1}{T^2}\int_0^T\int_0^T C_{xx}(\eta-\xi)d\eta d\xi =$$

$$= \frac{1}{T^2}\int_0^T\int_{-\xi}^{T-\xi} C_{xx}(\tau)d\tau d\xi = \frac{1}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)C_{xx}(\tau)d\tau$$

V. Statistical errors and estimation

- When T tends to infinity, we get:

$$\lim_{T\to\infty} T Var\{\hat{\mu}_x\} = \int_{-\infty}^{\infty} C_{xx}(\tau)\,d\tau < \infty$$

- This equation shows, when T is large enough and $|\tau| << T$, variance of estimate can be approximated by

$$Var\{\hat{\mu}_x\} \approx \frac{1}{T}\int_{-\infty}^{\infty} C_{xx}(\tau)\,d\tau$$

- When $T \to \infty$, $Var\{\hat{\theta}\} \to 0$: the estimate is consistent

V. Statistical errors and estimation

- Mean square estimate
  - x(t) is sample function of stationary and ergodic process X(t).
  - Mean square of X(t): $$\hat{\psi}_x^2 = \frac{1}{T}\int_0^T x^2(t)dt$$
  - True mean square: $$\psi_x^2 = E\{x^2(t)\}$$
  - Mean of mean square estimate: $$E\{\hat{\psi}_x^2\} = \frac{1}{T}\int_0^T E\{x^2(t)\}dt = \frac{1}{T}\int_0^T \psi_x^2 dt = \psi_x^2$$
  - $\psi\hat{}^2_x$ is unbiased estimator of ψ^2_x

V. Statistical errors and estimation

- Mean square error of this estimate is equal:

$$Var\{\hat{\psi}_X^2\} = E\{(\hat{\psi}_X^2 - \psi_X^2)^2\} = E\{\hat{\psi}_X^4\} - \psi_X^4 =$$

$$= \frac{1}{T^2}\int_0^T\int_0^T \left(E\{x^2(\xi)x^2(\eta)\} - \psi_X^4\right) d\eta d\xi$$

- If X(t) is Gaussian with nonzero mean $\mu_X \neq 0$, we have:

$$Var\{\hat{\psi}_X^2\} = \frac{2}{T^2}\int_{-T}^{T}\left(1 - \frac{|\tau|}{T}\right)\left(C_{xx}^2(\tau) + 2\mu_x^2 C_{xx}(\tau)\right)d\tau$$

- If T is very large and $|\tau| << T$, we have:

$$Var\{\hat{\psi}_X^2\} = \frac{2}{T^2}\int_{-T}^{T}\left(C_{xx}^2(\tau) + 2\mu_x^2 C_{xx}(\tau)\right)d\tau$$

Mean Square Estimation

Given some information that is related to an unknown quantity of interest, the problem is to obtain a good estimate for the unknown in terms of the observed data.

Suppose $X_1, X_2, \cdots, X_n$ represent a sequence of random variables about whom one set of observations are available, and Y represents an unknown random variable. The problem is to obtain a good estimate for Y in terms of the observations $X_1, X_2, \cdots, X_n$.

Let

$$\hat{Y} = \varphi(X_1, X_2, \cdots, X_n) = \varphi(\underline{X}) \qquad (16\text{-}1)$$

represent such an estimate for Y.

Note that $\varphi(\cdot)$ can be a linear or a nonlinear function of the observation $X_1, X_2, \cdots, X_n$. Clearly

$$\varepsilon(\underline{X}) = Y - \hat{Y} = Y - \varphi(\underline{X}) \qquad (16\text{-}2)$$

represents the error in the above estimate, and $|\varepsilon|^2$ the square of the error. Since ε is a random variable, $E\{|\varepsilon|^2\}$ represents the mean square error. One strategy to obtain a good estimator would be to minimize the mean square error by varying over all possible forms of $\varphi(\cdot)$, and this procedure gives rise to the Minimization of the Mean Square Error (MMSE) criterion for estimation. Thus under MMSE criterion,the estimator $\varphi(\cdot)$ is chosen such that the mean square error $E\{|\varepsilon|^2\}$ is at its minimum.

Next we show that the conditional mean of Y given $\underline{X}$ is the best estimator in the above sense.

Theorem1: Under MMSE criterion, the best estimator for the unknown Y in terms of $X_1, X_2, \cdots, X_n$ is given by the conditional mean of Y gives $\underline{X}$. Thus

$$\hat{Y} = \varphi(\underline{X}) = E\{Y \mid \underline{X}\}. \tag{16-3}$$

Proof : Let $\hat{Y} = \varphi(\underline{X})$ represent an estimate of Y in terms of $\underline{X} = (X_1, X_2, \cdots, X_n)$.Then the error $\varepsilon = Y - \hat{Y}$, and the mean square error is given by

$$\sigma_\varepsilon^2 = E\{|\varepsilon|^2\} = E\{|Y - \hat{Y}|^2\} = E\{|Y - \varphi(\underline{X})|^2\} \tag{16-4}$$

Since

$$E[z] = E_X[E_z\{z \mid \underline{X}\}] \qquad (16\text{-}5)$$

we can rewrite (16-4) as

$$\sigma_\varepsilon^2 = E\{\underbrace{|Y - \varphi(\underline{X})|^2}_{z}\} = E_{\underline{X}}[E_Y\{\underbrace{|Y - \varphi(\underline{X})|^2}_{z} | \underline{X}\}]$$

where the inner expectation is with respect to *Y*, and the outer one is with respect to $\underline{X}$.

Thus

$$\sigma_\varepsilon^2 = E[E\{|Y - \varphi(\underline{X})|^2 \mid \underline{X}\}]$$

$$= \int_{-\infty}^{+\infty} E\{|Y - \varphi(\underline{X})|^2 \mid \underline{X}\} f_X(\underline{X})dx. \qquad (16\text{-}6)$$

To obtain the best estimator φ, we need to minimize σ_ε^2 in (16-6) with respect to φ. In (16-6), since $f_{\underline{X}}(\underline{X}) \geq 0$, $E\{|Y - \varphi(\underline{X})|^2 \mid \underline{X}\} \geq 0$, and the variable φ appears only in the integrand term, minimization of the mean square error σ_ε^2 in (16-6) with respect to φ is equivalent to minimization of $E\{|Y - \varphi(\underline{X})|^2 \mid \underline{X}\}$ with respect to φ.

Since $\underline{X}$ is fixed at some value, $\varphi(\underline{X})$ is no longer random, and hence minimization of $E\{|Y-\varphi(\underline{X})|^2 \,|\underline{X}\}$ is equivalent to

$$\frac{\partial}{\partial \varphi} E\{|Y-\varphi(\underline{X})|^2 \,|\underline{X}\} = 0. \qquad (16\text{-}7)$$

This gives

$$E\{|Y-\varphi(\underline{X})| \,|\, \underline{X}\} = 0$$

or

$$E\{Y \,|\, \underline{X}\} - E\{\varphi(\underline{X}) \,|\, \underline{X}\} = 0. \qquad (16\text{-}8)$$

But

$$E\{\varphi(\underline{X}) \,|\, \underline{X}\} = \varphi(\underline{X}), \qquad (16\text{-}9)$$

since when $\underline{X} = \underline{x}, \varphi(\underline{X})$ is a fixed number $\varphi(\underline{x})$. Using (16-9) in (16-8) we get the desired estimator to be

$$\hat{Y} = \varphi(\underline{X}) = E\{Y \mid \underline{X}\} = E\{Y \mid X_1, X_2, \cdots, X_n\}. \tag{16-10}$$

Thus the conditional mean of Y given $X_1, X_2, \cdots, X_n$ represents the best estimator for Y that minimizes the mean square error.

The minimum value of the mean square error is given by

$$\sigma^2_{\min} = E\{|Y - E(Y \mid X)|^2\} = E[\underbrace{E\{|Y - E(Y \mid X)|^2 \mid \underline{X}\}}_{\operatorname{var}(Y \mid \underline{X})}]$$

$$= E\{\operatorname{var}(Y \mid \underline{X})\} \geq 0. \tag{16-11}$$

As an example, suppose $Y = X^3$ is the unknown. Then the best MMSE estimator is given by

$$\hat{Y} = E\{Y \mid X\} = E\{X^3 \mid X\} = X^3. \tag{16-12}$$

Clearly if $Y = X^3$, then indeed $\hat{Y} = X^3$ is the best estimator for Y

in terms of X. Thus the best estimator can be nonlinear.

Next, we will consider a less trivial example.

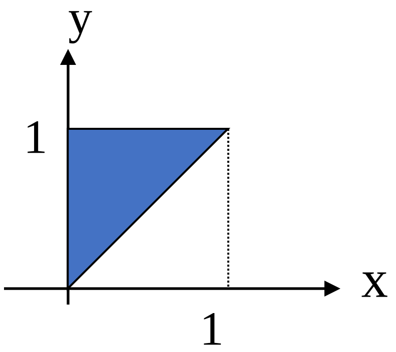
Example : Let

$$f_{X,Y}(x,y) = \begin{cases} kxy, & 0 < x < y < 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $k > 0$ is a suitable normalization constant. To determine the best estimate for Y in terms of X, we need $f_{Y|X}(y \mid x)$.

$$f_X(x) = \int_x^1 f_{X,Y}(x,y)dy = \int_x^1 kxydy$$

$$= \frac{kxy^2}{2}\Bigg|_x^1 = \frac{kx(1-x^2)}{2}, \quad 0 < \text{x} < 1.$$



Thus

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{kxy}{kx(1-x^2)/2} = \frac{2y}{1-x^2}; \quad 0 < x < y < 1.$$

(16-13)

Hence the best MMSE estimator is given by

$$\hat{Y} = \varphi(X) = E\{Y \mid \underline{X}\} = \int_x^1 y\, f_{Y|X}(y \mid x)dy$$

$$= \int_x^1 y \tfrac{2y}{1-x^2} dy = \tfrac{2}{1-x^2} \int_x^1 y^2 dy$$

$$= \frac{2}{3} \frac{y^3}{1-x^2} \Bigg|_x^1 = \frac{2}{3} \frac{1-x^3}{1-x^2} = \frac{2}{3} \frac{(1+x+x^2)}{1-x^2}. \qquad (16\text{-}14)$$

Once again the best estimator is nonlinear. In general the best estimator $E\{Y \mid \underline{X}\}$ is difficult to evaluate, and hence next we will examine the special subclass of best linear estimators.

Best Linear Estimator

In this case the estimator $\hat{Y}$ is a linear function of the observations $X_1, X_2, \cdots, X_n$. Thus

$$\hat{Y}_l = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i. \qquad (16\text{-}15)$$

where $a_1, a_2, \cdots, a_n$ are unknown quantities to be determined. The mean square error is given by $(\varepsilon = Y - \hat{Y}_l)$

$$E\{|\varepsilon|^2\} = E\{|Y - \hat{Y}_l|^2\} = E\{|Y - \sum a_i X_i|^2\} \qquad (16\text{-}16)$$

and under the MMSE criterion $a_1, a_2, \cdots, a_n$ should be chosen so that the mean square error $E\{|\varepsilon|^2\}$ is at its minimum possible value. Let σ_n^2 represent that minimum possible value. Then

$$\sigma_n^2 = \min_{a_1, a_2, \cdots, a_n} E\{|Y - \sum_{i=1}^{n} a_i X_i|^2\}. \qquad (16\text{-}17)$$

To minimize (16-16), we can equate

$$\frac{\partial}{\partial a_k} E\{|\varepsilon|^2\} = 0, \qquad k = 1, 2, \cdots, n. \qquad (16\text{-}18)$$

This gives

$$\frac{\partial}{\partial a_k} E\{|\varepsilon|^2\} = E\left\{\frac{\partial |\varepsilon|^2}{\partial a_k}\right\} = 2E\left[\varepsilon \left\{\frac{\partial \varepsilon}{\partial a_k}\right\}^*\right] = 0. \qquad (16\text{-}19)$$

But

$$\frac{\partial \varepsilon}{\partial a_k} = \frac{\partial (Y - \sum_{i=1}^{n} a_i X_i)}{\partial a_k} = \frac{\partial Y}{\partial a_k} - \frac{\partial (\sum_{i=1}^{n} a_i X_i)}{\partial a_k} = -X_k. \tag{16-20}$$

Substituting (16-19) in to (16-18), we get

$$\frac{\partial E\{|\varepsilon|^2\}}{\partial a_k} = -2E\{\varepsilon\, X_k^*\} = 0,$$

or the best linear estimator must satisfy

$$E\{\varepsilon\, X_k^*\} = 0, \quad k = 1, 2, \cdots, n. \tag{16-21}$$

Notice that in (16-21), ε represents the estimation error $(Y - \sum_{i=1}^{n} a_i X_i)$, and X_k, $k = 1 \to n$ represents the data. Thus from (16-21), the error ε is orthogonal to the data X_k, $k = 1 \to n$ for the best linear estimator. This is the **orthogonality principle**.

In other words, in the linear estimator (16-15), the unknown constants $a_1, a_2, \cdots, a_n$ must be selected such that the error

$\varepsilon = Y - \sum_{i=1}^{n} a_i X_i$ is orthogonal to every data $X_1, X_2, \cdots, X_n$ for the best linear estimator that minimizes the mean square error.

Interestingly a general form of the orthogonality principle holds good in the case of nonlinear estimators also.

Nonlinear Orthogonality Rule: Let $h(\underline{X})$ represent *any* functional form of the data and $E\{Y \mid \underline{X}\}$ the best estimator for Y given $\underline{X}$. With $e = Y - E\{Y \mid \underline{X}\}$ we shall show that

$$E\{eh(\underline{X})\} = 0, \qquad (16\text{-}22)$$

implying that

$$e = Y - E\{Y \mid \underline{X}\} \quad \perp \quad h(\underline{X}).$$

This follows since

$$\begin{aligned} E\{eh(\underline{X})\} &= E\{(Y - E[Y \mid \underline{X}])h(\underline{X})\} \\ &= E\{Yh(\underline{X})\} - E\{E[Y \mid \underline{X}]h(\underline{X})\} \\ &= E\{Yh(\underline{X})\} - E\{E[Yh(\underline{X}) \mid \underline{X}]\} \\ &= E\{Yh(\underline{X})\} - E\{Yh(\underline{X})\} = 0. \end{aligned}$$

Thus in the nonlinear version of the orthogonality rule the error is orthogonal to *any* functional form of the data.

The orthogonality principle in (16-20) can be used to obtain the unknowns $a_1, a_2, \cdots, a_n$ in the linear case.

For example suppose $n = 2$, and we need to estimate Y in terms of X_1 and X_2 linearly. Thus

$$\hat{Y}_l = a_1 X_1 + a_2 X_2$$

From (16-20), the orthogonality rule gives

$$E\{\varepsilon \, \mathrm{X}_1^*\} = E\{(Y - a_1 X_1 - a_2 X_2) X_1^*\} = 0$$
$$E\{\varepsilon \, \mathrm{X}_2^*\} = E\{(Y - a_1 X_1 - a_2 X_2) X_2^*\} = 0$$

Thus

$$E\{| X_1 |^2\} a_1 + E\{X_2 X_1^*\} a_2 = E\{Y X_1^*\}$$
$$E\{X_1 X_2^*\} a_1 + E\{| X_2 |^2\} a_2 = E\{Y X_2^*\}$$

or

$$\begin{pmatrix} E\{|X_1|^2\} & E\{X_2X_1^*\} \\ E\{X_1X_2^*\} & E\{|X_2|^2\} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} E\{YX_1^*\} \\ E\{YX_2^*\} \end{pmatrix} \qquad (16\text{-}23)$$

(16-23) can be solved to obtain a_1 and a_2 in terms of the cross-correlations.

The minimum value of the mean square error σ_n^2 in (16-17) is given by

$$\begin{aligned} \sigma_n^2 &= \min_{a_1,a_2,\cdots,a_n} E\{|\varepsilon|^2\} \\ &= \min_{a_1,a_2,\cdots,a_n} E\{\varepsilon\varepsilon^*\} = \min_{a_1,a_2,\cdots,a_n} E\{\varepsilon(Y-\sum_{i=1}^n a_iX_i)^*\} \\ &= \min_{a_1,a_2,\cdots,a_n} E\{\varepsilon\, Y^*\} - \min_{a_1,a_2,\cdots,a_n} \sum_{i=1}^n a_i E\{\varepsilon\, X_l^*\}. \end{aligned} \qquad (16\text{-}24)$$

But using (16-21), the second term in (16-24) is zero, since the error is orthogonal to the data X_i, where $a_1, a_2, \cdots, a_n$ are chosen to be optimum. Thus the minimum value of the mean square error is given by

$$\sigma_n^2 = E\{\varepsilon Y^*\} = E\{(Y - \sum_{i=1}^{n} a_i X_i)Y^*\}$$

$$= E\{|Y|^2\} - \sum_{i=1}^{n} a_i E\{X_i Y^*\} \qquad (16\text{-}25)$$

where $a_1, a_2, \cdots, a_n$ are the optimum values from (16-21).

Since the linear estimate in (16-15) is only a special case of the general estimator $\varphi(\underline{X})$ in (16-1), the best linear estimator that satisfies (16-20) cannot be superior to the best nonlinear estimator $E\{Y \mid \underline{X}\}$. Often the best linear estimator will be inferior to the best estimator in (16-3).

This raises the following question. Are there situations in which the best estimator in (16-3) also turns out to be linear ? In those situations it is enough to use (16-21) and obtain the best linear estimators, since they also represent the best global estimators. Such is the case if Y and $X_1, X_2, \cdots, X_n$ are distributed as jointly Gaussian

We summarize this in the next theorem and prove that result.

Theorem2: If $X_1, X_2, \cdots, X_n$ and Y are jointly Gaussian zero mean random variables, then the best estimate for Y in terms of $X_1, X_2, \cdots, X_n$ is always linear.

Proof : Let

$$\hat{Y} = \varphi(X_1, X_2, \cdots, X_n) = E\{Y \mid \underline{X}\} \qquad (16\text{-}26)$$

represent the best (possibly nonlinear) estimate of Y, and

$$\hat{Y}_l = \sum_{i=1}^{n} a_i X_i \qquad (16\text{-}27)$$

the best linear estimate of Y. Then from (16-21)

$$\varepsilon \triangleq Y - Y_l = Y - \sum_{i=1}^{n} a_i X_i \qquad (16\text{-}28)$$

is orthogonal to the data $X_k, \quad k = 1 \to n.$ Thus

$$E\{\varepsilon \mathrm{X}_\mathrm{k}^*\} = 0, \quad k = 1 \to n. \qquad (16\text{-}29)$$

Also from (16-28),

$$E\{\varepsilon\} = E\{Y\} - \sum_{i=1}^{n} a_i E\{X_i\} = 0. \qquad (16\text{-}30)$$

Using (16-29)-(16-30), we get

$$E\{\varepsilon X_k^*\} = E\{\varepsilon\}E\{X_k^*\} = 0, \quad k = 1 \to n. \qquad (16\text{-}31)$$

From (16-31), we obtain that ε and X_k are zero mean uncorrelated random variables for $k = 1 \to n.$ But ε itself represents a Gaussian random variable, since from (16-28) it represents a linear combination of a set of jointly Gaussian random variables. Thus ε and $\underline{X}$ are jointly Gaussian and uncorrelated random variables. As a result, ε and $\underline{X}$ are independent random variables. Thus from their independence

$$E\{\varepsilon \mid \underline{X}\} = E\{\varepsilon\}. \qquad (16\text{-}32)$$

But from (16-30), $E\{\varepsilon\} = 0,$ and hence from (16-32)

$$E\{\varepsilon \mid \underline{X}\} = 0. \qquad (16\text{-}33)$$

Substituting (16-28) into (16-33), we get

$$E\{\varepsilon \mid \underline{X}\} = E\{Y - \sum_{i=1}^{n} a_i X_i \mid \underline{X}\} = 0$$

or

$$E\{Y \mid \underline{X}\} = E\{\sum_{i=1}^{n} a_i X_i \mid \underline{X}\} = \sum_{i=1}^{n} a_i X_i = Y_l. \qquad (16\text{-}34)$$

From (16-26), $E\{Y \mid \underline{X}\} = \varphi(\underline{x})$ represents the best possible estimator, and from (16-28), $\sum_{i=1}^{n} a_i X_i$ represents the best linear estimator. Thus the best linear estimator is also the best possible overall estimator in the Gaussian case.

Next we turn our attention to prediction problems using linear estimators.

Linear Prediction

Suppose $X_1, X_2, \cdots, X_n$ are known and X_{n+1} is unknown. Thus $Y = X_{n+1}$, and this represents a one-step prediction problem. If the unknown is X_{n+k}, then it represents a *k*-step ahead prediction problem. Returning back to the one-step predictor, let $\hat{X}_{n+1}$ represent the best linear predictor. Then

$$\hat{X}_{n+1} \triangleq -\sum_{i=1}^{n} a_i X_i, \tag{16-35}$$

where the error

$$\begin{aligned} \varepsilon_n &= X_{n+1} - \hat{X}_{n+1} = X_{n+1} + \sum_{i=1}^{n} a_i X_i \\ &= a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + X_{n+1} \\ &= \sum_{i=1}^{n+1} a_i X_i, \qquad\qquad a_{n+1} = 1, \end{aligned} \tag{16-36}$$

is orthogonal to the data, i.e.,

$$E\{\varepsilon_n X_k^*\} = 0, \qquad k = 1 \to n. \tag{16-37}$$

Using (16-36) in (16-37), we get

$$E\{\varepsilon_n X_k^*\} = \sum_{i=1}^{n+1} a_i E\{X_i X_k^*\} = 0, \quad k = 1 \to n. \tag{16-38}$$

Suppose X_i represents the sample of a wide sense stationary stochastic process $X(t)$ so that

$$E\{X_i X_k^*\} = R(i-k) = r_{i-k} = r_{k-i}^* \tag{16-39}$$

Thus (16-38) becomes

$$E\{\varepsilon_n X_k^*\} = \sum_{i=1}^{n+1} a_i r_{i-k} = 0, \quad a_{n+1} = 1, \quad k = 1 \to n. \tag{16-40}$$

Expanding (16-40) for $k = 1, 2, \cdots, n,$ we get the following set of linear equations.

$$\begin{aligned}
&a_1 r_0 + a_2 r_1 + a_3 r_2 + \cdots + a_n r_{n-1} + r_n = 0 \leftarrow k = 1 \\
&a_1 r_1^* + a_2 r_0 + a_3 r_1 + \cdots + a_n r_{n-2} + r_{n-1} = 0 \leftarrow k = 2 \\
&\vdots \\
&a_1 r_{n-1}^* + a_2 r_{n-2}^* + a_3 r_{n-3}^* + \cdots + a_n r_0 + r_1 = 0 \leftarrow k = n.
\end{aligned} \tag{16-41}$$

Similarly using (16-25), the minimum mean square error is given by

$$\sigma_n^2 = E\{|\varepsilon|^2\} = E\{\varepsilon_n Y^*\} = E\{\varepsilon_n X^*_{n+1}\}$$

$$= E\{(\sum_{i=1}^{n+1} a_i X_i) X^*_{n+1}\} = \sum_{i=1}^{n+1} a_i r^*_{n+1-i}$$

$$= a_1 r_n^* + a_2 r_{n-1}^* + a_3 r_{n-2}^* + \cdots + a_n r_1 + r_0. \qquad (16\text{-}42)$$

The *n* equations in (16-41) together with (16-42) can be represented as

$$\begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ r_2^* & r_1^* & r_0 & \cdots & r_{n-2} \\ & & \vdots & & \\ r_{n-1}^* & r_{n-2}^* & \cdots & r_0 & r_1 \\ r_n^* & r_{n-1}^* & \cdots & r_1^* & r_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_n^2 \end{pmatrix}. \qquad (16\text{-}43)$$

Let

$$T_n = \begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_n^* & r_{n-1}^* & \cdots & r_1^* & r_0 \end{pmatrix}. \tag{16-44}$$

Notice that T_n is Hermitian Toeplitz and positive definite. Using (16-44), the unknowns in (16-43) can be represented as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \\ 1 \end{pmatrix} = T_n^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_{\mathrm{n}}^2 \end{pmatrix} = \sigma_n^2 \begin{pmatrix} \text{Last} \\ \text{column} \\ \text{of} \\ T_n^{-1} \end{pmatrix} \tag{16-45}$$

Let

$$
T_n^{-1} = \begin{pmatrix} T_n^{11} & T_n^{12} & \cdots & T_n^{1,n+1} \\ T_n^{21} & T_n^{22} & \cdots & T_n^{2,n+1} \\ & & \vdots & \\ T_n^{n+1,1} & T_n^{n+1,2} & \cdots & T_n^{n+1,n+1} \end{pmatrix}. \qquad (16\text{-}46)
$$

Then from (16-45),

$$
\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ 1 \end{pmatrix} = \sigma_n^2 \begin{pmatrix} T_n^{1,n+1} \\ T_n^{2,n+1} \\ \vdots \\ T_n^{n+1,n+1} \end{pmatrix}. \qquad (16\text{-}47)
$$

Thus

$$
\sigma_n^2 = \frac{1}{T_n^{n+1,n+1}} > 0, \qquad (16\text{-}48)
$$

and

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \frac{1}{T_n^{n+1,n+1}} \begin{pmatrix} T_n^{1,n+1} \\ T_n^{2,n+1} \\ \vdots \\ T_n^{n+1,n+1} \end{pmatrix}. \qquad (16\text{-}49)$$

Eq. (16-49) represents the best linear predictor coefficients, and they can be evaluated from the last column of T_n in (16-45). Using these, The best one-step ahead predictor in (16-35) taken the form

$$\hat{X}_{n+1} = -\left(\frac{1}{T_n^{n+1,n+1}}\right)\sum_{i=1}^{n} (T_n^{i,n+1}) X_i. \qquad (16\text{-}50)$$

and from (16-48), the minimum mean square error is given by the $(n+1, n+1)$ entry of T_n^{-1}.

From (16-36), since the one-step linear prediction error

$$\varepsilon_n = X_{n+1} + a_n X_n + a_{n-1} X_{n-1} + \cdots + a_1 X_1, \qquad (16\text{-}51)$$

we can represent (16-51) formally as follows

$$X_{n+1} \rightarrow \boxed{1+a_n z^{-1}+a_{n-1}z^{-2}+\cdots+a_1 z^{-n}} \rightarrow \varepsilon_n$$

Thus, let

$$A_n(z) = 1+a_n z^{-1}+a_{n-1}z^{-2}+\cdots+a_1 z^{-n}, \qquad (16\text{-}52)$$

them from the above figure, we also have the representation

$$\varepsilon_n \rightarrow \boxed{\frac{1}{A_n(z)}} \rightarrow X_{n+1}.$$

The filter

$$H(z) = \frac{1}{A_n(z)} = \frac{1}{1+a_n z^{-1}+a_{n-1}z^{-2}+\cdots+a_1 z^{-n}} \qquad (16\text{-}53)$$

represents an $AR(n)$ filter, and this shows that linear prediction leads to an auto regressive (AR) model.

The polynomial $A_n(z)$ in (16-52)-(16-53) can be simplified using (16-43)-(16-44). To see this, we rewrite $A_n(z)$ as

$$A_n(z) = a_1 z^{-n} + a_2 z^{-(n-1)} + \cdots + a_{n-1} z^{-2} + a_n z^{-1} + 1$$

$$= [z^{-n}, z^{-(n-1)}, \cdots, z^{-1}, 1] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ 1 \end{bmatrix} = [z^{-n}, z^{-(n-1)}, \cdots, z^{-1}, 1]\, T_n^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_{\mathrm{n}}^2 \end{bmatrix} \qquad (16\text{-}54)$$

To simplify (16-54), we can make use of the following matrix identity

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -AB \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & D - CA^{-1}B \end{bmatrix}. \qquad (16\text{-}55)$$

Taking determinants, we get

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A|\,|D - CA^{-1}B|. \qquad (16\text{-}56)$$

In particular if $D \equiv 0$, we get

$$|CA^{-1}B| = \frac{(-1)^n}{|A|} \begin{vmatrix} A & B \\ C & 0 \end{vmatrix}. \qquad (16\text{-}57)$$

Using (16-57) in (16-54), with

$$C = [z^{-n}, z^{-(n-1)}, \cdots, z^{-1}, 1], \;\; A = T_n, \;\; B = \begin{bmatrix} 0 \\ \vdots \\ \sigma_n^2 \end{bmatrix}$$

we get

$$A_n(z) = \frac{(-1)^n}{|T_n|}\left|\begin{array}{c:c} T_n & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_n^2 \end{matrix} \\ \hdashline z^{-n},\cdots,z^{-1},1 & 0 \end{array}\right| = \frac{\sigma_n^2}{|T_n|}\begin{vmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_{n-1}^* & r_{n-2}^* & \cdots & r_0 & r_1 \\ z^{-n} & z^{-(n-1)} & \cdots & z^{-1} & 1 \end{vmatrix}. \quad (16\text{-}58)$$

Referring back to (16-43), using Cramer’s rule to solve for $a_{n+1}(=1)$, we get

$$a_{n+1} = \frac{\sigma_n^2 \begin{vmatrix} r_0 & \cdots & r_{n-1} \\ & \vdots & \\ r_{n-1} & \cdots & r_0 \end{vmatrix}}{|T_n|} = \sigma_n^2 \frac{|T_{n-1}|}{|T_n|} = 1$$

or

$$\sigma_n^2 = \frac{|T_n|}{|T_{n-1}|} > 0. \tag{16-59}$$

Thus the polynomial (16-58) reduces to

$$A_n(z) = \frac{1}{|T_{n-1}|} \begin{vmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_{n-1}^* & r_{n-2}^* & \cdots & r_0 & r_1 \\ z^{-n} & z^{-(n-1)} & \cdots & z^{-1} & 1 \end{vmatrix} \tag{16-60}$$

$$= 1 + a_n z^{-1} + a_{n-1} z^{-2} + \cdots + a_1 z^{-n}.$$

The polynomial $A_n(z)$ in (16-53) can be alternatively represented as in (16-60), and $H(z) = \dfrac{1}{A_n(z)} \sim AR(n)$ in fact represents a stable

AR filter of order *n*, whose input error signal ε_n is white noise of constant spectral height equal to $|T_n|/|T_{n-1}|$ and output is X_{n+1}. It can be shown that $A_n(z)$ has all its zeros in $|z|>1$ provided $|T_n|>0$ thus establishing stability.

Linear prediction Error

From (16-59), the mean square error using n samples is given by

$$\sigma_n^2 = \frac{|T_n|}{|T_{n-1}|} > 0. \qquad (16\text{-}61)$$

Suppose one more sample from the past is available to evaluate X_{n+1} (i.e., $X_n, X_{n-1}, \cdots, X_1, X_0$ are available). Proceeding as above the new coefficients and the mean square error σ_{n+1}^2 can be determined. From (16-59)-(16-61),

$$\sigma_{n+1}^2 = \frac{|T_{n+1}|}{|T_n|}. \qquad (16\text{-}62)$$

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for your attentions!