Optimal Estimation

4/8

Simplest Case: Estimation without any observation.

- X is random variable with pdf $f_X(x)$

Let $\hat{X}$ be estimate of X (fixed value)

- Error $\tilde{X} = X - \hat{X}$

Note $\tilde{X}$ is Random Variable.

- Want to minimize mean squared error (MMSE)

Problem: choose $\hat{X}$ such that $E(\tilde{X}^2)$ is minimum.

Solution: Optimal estimate $\hat{X} = E(X)$

Corresponding min. mean squared error

$$E(\tilde{X}^2) = Var(X)$$

Proof: (Approach 1)

Let $\alpha = E(X)$

Now, $E(\tilde{x}^2) = E\{(x-\hat{x})^2\}$

$$= E\{[(x-\alpha) + (\alpha-\hat{x})]^2\}$$

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

$$= E(x-\alpha)^2 + E((\alpha-\hat{x})^2) + 2(\alpha-\hat{x})\cancel{E(x-\alpha)}^{0}$$

$$= E[(x-\alpha)^2] + \underbrace{E(\alpha-\hat{x})^2}_{=(\alpha-\hat{x})^2}$$

$$E(\tilde{x}^2) \geq E((x-\alpha)^2\}$$

equality if and only if $\alpha - \hat{x} = 0$

So optimal estimate $\hat{x} = \alpha = E(x)$

In this case $\hat{x} = E(x)$,

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

$$= Var(x)$$

x ———————————— x

Approach 2

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

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

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

differentiate $E(\tilde{x}^2)$ wrt $\hat{x}$ and equate to zero

$$\frac{d\,E(\tilde{x}^2)}{d\hat{x}} = 2\hat{x} - 2E(x) = 0$$

$$\Rightarrow \hat{x} = E(x).$$

Example: x is binary $+1$ or -1

$$\Pr\{x=1\} = 3/4$$

$$\Pr\{x=-1\} = 1/4$$

MMSE estimate (given no observation)

$$\hat{x} = E(x) = 3/4\,(1) + \frac{1}{4}(-1) = 1/2.$$

(In this example, MMSE estimate is not very meaningful as x will never take value of $1/2$)

Optimal Estimation given observation

X and Y are two random variables

$f_{XY}(x,y)$ is joint pdf of X and Y.

Want an estimator of X given Y.

Answer is related conditional expectation

Given $Y=y$, we have

$$E(X|Y=y) = \int_{-\infty}^{\infty} x\, f_X(x|y)\, dx$$

where $f_X(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$

This is a function of y.

Say, $E(X|Y=y) = g(y)$.

Conditional Expectation of X given Y is defined

as $E(X|Y) = g(Y)$ which is

a transformed Random Variable

This transformed RV takes value $E(X|Y=y)$ when Y takes value y.

Now, consider expected value of this transformed RV $E(X|Y)$

Thm. $$E\left[E(X|Y)\right] = E(X)$$

iterative expectation identity.

Proof: $$E(g(Y)) = \int_{-\infty}^{\infty} g(y)\, f_Y(y)\, dy \; ;$$

for any $g(\cdot)$.

Take $g(Y) = E(X|Y)$

$$E\left[E(X|Y)\right] = \int_{-\infty}^{\infty} \underbrace{E(X|Y=y)}_{\shortparallel}\, f_Y(y)\, dy$$

$$= \int_{y=-\infty}^{\infty} \left[\int_{x=-\infty}^{\infty} x \cdot f_X(x|y)\, dx\right] f_Y(y)\, dy$$

$$= \int_{x=-\infty}^{\infty} x \left[\int_{y=-\infty}^{\infty} \underbrace{f_X(x|y)\, f_Y(y)}_{f_{XY}(x,y)}\, dy\right] dx$$

$$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx = E(X).$$

x ——— y

Example: Suppose N is a RV which takes positive integer values.

$Y_1, Y_2, Y_3, \ldots$ is a sequence of identically distributed RV

Consider $S = \sum_{i=1}^{N} Y_i$

Want to find $E(S)$

find pdf of S $f_S(s)$

$E(S) = \int_{-\infty}^{\infty} s\, f_S(s)\, ds$

$$E(S) = E\left[E(S|N)\right]$$

$$E(S|N=n) = E(Y_1 + Y_2 + \cdots Y_n)$$

$$= E(Y_1) + E(Y_2) + \cdots + E(Y_n)$$

$$= n\, E(Y_1) \quad \text{[Since } Y_i \text{ have same mariginal pdf]}$$

$$E(S|N) = N\, E(Y_1)$$

$$E(E(S|N)] = E[N\, E(Y_1)]$$

$$\boxed{E(S) = E(Y_1)\, E(N)} \rightarrow \text{Wald's identity}$$

Optimal estimator.

X and Y are two random variables

Given Y, find an estimator $\hat{X}$ of X

such that $E\left\{(X-\hat{X})^2\right\}$ is minimum.

Theorem:

Optimal Estimator $\hat{X} = E(X|Y)$

Moreover, for this optimal estimator

(a) $E(\hat{X}) = E(X)$

(b) $E\left\{(X-\hat{X})^2\right\} = Var(X) - Var(\hat{X})$

x ———— x

Proof: Consider $X - E(X|Y)$

Claim 1 : $X - E(X|Y)$ is zero mean RV.

$$E\left[X - E(X|Y)\right]$$

$$= E(X) - E\left(E(X|Y)\right)$$

$$= E(X) - E(X) = 0$$

Claim 2 : $X - E(X|Y)$ is orthogonal to $g(Y)$ for any function $g(\cdot)$.

Proof: Need to show

$$E\left\{ [X - E(X|Y)]\, g(Y) \right\} = 0$$

or equivalently,

$$E\{X\, g(Y)\} = E\left\{ E(X|Y)\, g(Y) \right\}.$$

Consider. $E\{X\, g(Y)\}$ Let $Z = X\, g(Y)$

$$E(Xg(Y)) = E(Z) = E\left(E(Z|Y) \right)$$

$$= E\left(E(X g(Y) | Y) \right)$$

$$= E\left(g(Y)\, E(X|Y) \right)$$

So $E(X g(Y)) = E\left(E(X|Y)\, g(Y) \right)$ for any $g(\cdot)$.

So $E\left([X - E(X|Y)]\, g(Y) \right) = 0 \quad \forall\, g(\cdot)$.

Since $X - E(X|Y)$ is zero mean,

orthogonality $\Rightarrow$ uncorrelatedness

$X - E(X|Y)$ is uncorrelated with $g(Y)$ for any $g(\cdot)$

Proof of our theorem:

for some estimator $\hat{X}$ (some function of Y)

$$E\left((X-\hat{X})^2\right)$$

$$= E\left\{\left[X - E(X|Y) + \underbrace{E(X|Y) - \hat{X}}_{= h(Y) \text{ for some } h(\cdot).}\right]^2\right\}$$

Since $h(Y)$ is uncorrelated with $X - E(X|Y)$

$$= E\left\{\left[X - E(X|Y)\right]^2\right\} + E\left\{h^2(Y)\right\}$$

$$\geq E\left\{\left\{X - E(X|Y)\right\}^2\right\}$$

with equality if and only if

$h(Y) = 0 \Rightarrow \hat{X} = E(X|Y)$

6/8 Need to show, for optimal estimator $\hat{X} = E(X|Y)$

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

Proof.

$$\tilde{X} = X - \underbrace{E(X|Y)}_{\hat{X}}$$

$$E(\tilde{X}) = E(X) - E\big(E(X|Y)\big) = 0$$

From orthogonality result,

notation for orthogonality

$$\tilde{X} \perp g(Y) \quad \text{for any } g(\cdot).$$

As a special case,

$$\tilde{X} \perp \hat{X} \quad (\text{Since } \hat{X} \text{ is a function of } Y)$$

Now, $$X = \tilde{X} + \hat{X}$$

$$Var(X) = Var(\tilde{X}) + Var(\hat{X})$$

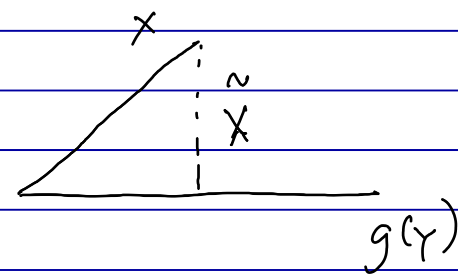
Since $\tilde{X}$ & $\hat{X}$ are uncorrelated.

So $$E\{\tilde{X}^2\} = Var(X) - Var(\hat{X})$$

Geometric Interpretation of Orthogonality Principle.

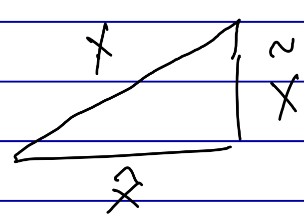
$$\hat{X} = E(X|Y) \qquad \tilde{X} = X - E(X|Y)$$

Orthog. Principle says $\tilde{X}$ is $\perp$ $g(Y)$.

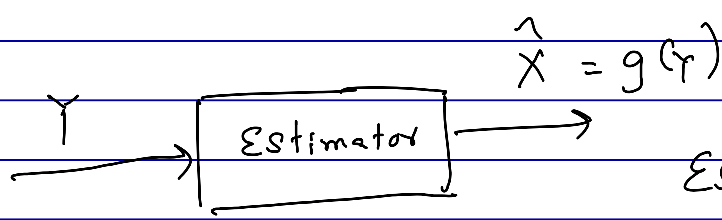


Interpretation: Any processing of data $(g(Y))$ does not help in reducing the (optimal) error $\tilde{X}$ further [since $g(Y) \perp \tilde{X}$]

Special case $g(Y) = \hat{X} = E(X|Y)$



$\rightarrow$ for the optimal MMSE estimator.



Estimator $\hat{X}$ is a function of Y

$\hat{X}$ is a random variable.

for a specific outcome $Y = y$,

$\hat{X} = g(y)$ is an <u>estimate</u> of X

Estimate is a scalar

<u>Definition</u>. $\hat{X}$ is called a unbiased estimator

if $E(\hat{X}) = E(X)$

<u>Remark:</u> Optimal MMSE estimator is <u>unbiased</u>

<u>Converse</u> of <u>orthogonality</u> <u>principle</u>.

Suppose $\hat{X}$ is an <u>unbiased</u>

estimator of X using

Observation Y.

Suppose $X - \hat{X} \perp g(Y)$ for any $g(\cdot)$.

Then $\hat{X} = E(X|Y)$. → optimal MMSE estimator

<u>Proof:</u>

Let us define $Z = \hat{X} - E(X|Y)$

Need to show $Z = 0$ (with probability one)

Note that $E(Z) = E(\hat{X}) - E(E(X|Y))$

$$= E(X) - E(X)$$

$$= 0$$

We have (given) $X - \hat{X} \perp g(Y)$ for any $g(.)$.

We also know

$$X - E(X|Y) \perp g(Y)$$

Subtracting

$$E(X|Y) - \hat{X} \perp g(Y)$$

So $$E\{\underbrace{[\hat{X} - E(X|Y)]}_{Z} g(Y)\} = 0 \quad \forall g(.)$$

Note Z is a function of Y.

Take $g(Y)$ to be Z.

$$E(Z.Z) = 0 \Rightarrow E(Z^2) = 0.$$

$$Var(Z) = 0 \Rightarrow Z \text{ is constant}$$

$$E[Z] = 0 \Rightarrow \underline{\underline{Z = 0}}.$$

so $\hat{X} = E(X|Y)$

Examples of MMSE estimator

① X and Y are jointly Gaussian

Parameters $\quad m_x = E(X) \qquad \sigma_x^2 = Var(X)$

$\qquad\qquad\quad m_y = E(Y) \qquad \sigma_Y^2 = Var(Y)$

Cross correlation $\rightarrow \sigma_{XY} = E\big((X - m_x)(Y - m_y)\big)$

Covariance matrix $C = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$

joint pdf of X & Y

$$f_{XY}(x,y) = \frac{1}{2\pi \underbrace{|C|}_{\text{determinant}}^{1/2}} e^{-\frac{1}{2}\left\{\underbrace{[x-m_x \;\; y-m_y]}_{1\times 2 \text{ row vector}} C^{-1} \underbrace{\begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}}_{\substack{\text{column vector} \\ 2\times 1}}\right\}}$$

To find $E(X|Y=y)$

We need $f_X(x|Y=y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$

$f_Y(y)$ is Gaussian with mean m_y and Variance σ_y^2

$$\det(C) = |C| = \sigma_x^2 \sigma_y^2 - \sigma_{xy}^2$$

$$C^{-1} = \frac{1}{|C|} \begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix}$$

$$[x-m_x \quad y-m_y] \, C^{-1} \begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix} = \sigma_y^2 (x-m_x)^2 + \sigma_x^2 (y-m_y)^2 - 2\sigma_{xy}(y-m_y)(x-m_x)$$

Doing some algebra,

$$f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi}\sqrt{\frac{|C|}{\sigma_y^2}}} e^{-\frac{1}{2}\left\{\frac{\left(x - \left[m_x + \frac{\sigma_{xy}}{\sigma_y^2}[y-m_y]\right]\right)^2}{|C|/\sigma_y^2}\right\}}$$

This conditional pdf is Gaussian

Mean is $m_x + \frac{\sigma_{xY}}{\sigma_Y^2}[y - m_y]$

Variance is $\frac{|C|}{\sigma_y^2} = \sigma_x^2 - \frac{\sigma_{xY}^2}{\sigma_Y^2}$

So MMSE estimate $\hat{X} = E(X|Y=y)$

$$= m_x + \frac{\sigma_{xY}}{\sigma_y^2}[y - m_y]$$

So MMSE estimator $\hat{X} = E(X) + \frac{\sigma_{XY}}{\sigma_Y^2}[Y - E(Y)]$

it is a linear function of Y.

(Important Special case)

$$MMSE = E(X-\hat{X})^2$$

$$= Var(X) - Var(\hat{X})$$

$$= \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^{\cancel{4}2}} \cdot \underbrace{Var(Y)}_{= \cancel{\sigma_Y^2}}$$

$$\text{min. MSE} = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2}$$

x ——— x

Example 2. $Y = X + V$

X is binary $\{\pm 1\}$ $\quad P_r\{X=1\} = P_r\{X=-1\} = \frac{1}{2}$

V is Gaussian with zero mean Variance σ^2

$$V \sim \mathcal{N}(0, \sigma^2)$$

X and V are independent

Find $E(X \mid Y=y)$.

Need $\Pr\{X=1 \mid Y=y\}$, $\Pr\{X=-1 \mid Y=y\}$

Now, $\Pr\{X=+1 \mid Y=y\} \underset{\text{Baye's Rule}}{=} \dfrac{f_Y(y \mid X=1) \cdot \Pr\{X=1\}}{f_Y(y)}$

Bayes Rule

$$\Pr\{A \mid B\} = \frac{\Pr\{B \mid A\} \cdot P\{A\}}{\Pr\{B\}}$$

Given $X=1$

$$Y = 1 + V$$

$$f_Y(y \mid X=1) \sim N(1, \sigma^2)$$

$$f_Y(y) \underset{\text{total prob. theorem}}{=} f_Y(y \mid X=1) \cdot \Pr\{X=1\} + f_Y(y \mid X=-1) \Pr\{X=-1\}$$

$$\Pr\{X=1 \mid Y=y\} = \frac{\frac{1}{2} \cdot f_Y(y \mid X=1)}{\frac{1}{2} \underbrace{f_Y(y \mid X=1)}_{N(1,\sigma^2)} + \frac{1}{2} \underbrace{f_Y(y \mid X=-1)}_{N(-1,\sigma^2)}}$$

$$\Pr\{X=1 \mid Y=y\} = e^{-\frac{1}{2\sigma^2}(y-1)^2} \Big/ \left(e^{-\frac{1}{2\sigma^2}(y-1)^2} + e^{-\frac{1}{2\sigma^2}(y+1)^2}\right)$$

$$\Pr\{X=-1 \mid Y=y\} = \frac{e^{-\frac{1}{2\sigma^2}(y+1)^2}}{e^{-\frac{1}{2\sigma^2}(y-1)^2} + e^{-\frac{1}{2\sigma^2}(y+1)^2}}$$

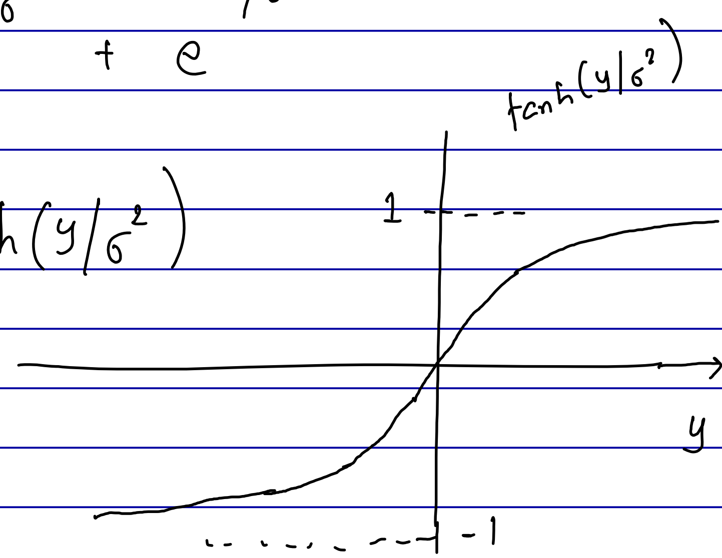
$$E(X \mid Y=y) = (+1)\Pr\{X=+1 \mid Y=y\} + (-1)\Pr\{X=-1 \mid Y=y\}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(y-1)^2} - e^{-\frac{1}{2\sigma^2}(y+1)^2}}{e^{-\frac{1}{2\sigma^2}(y-1)^2} + e^{-\frac{1}{2\sigma^2}(y+1)^2}}$$

with some algebra

$$= \frac{e^{y/\sigma^2} - e^{-y/\sigma^2}}{e^{y/\sigma^2} + e^{-y/\sigma^2}}$$

$$E(X \mid Y=y) = \tanh(y/\sigma^2)$$



$$\hat{X} = E(X \mid Y) = \tanh(Y/\sigma^2)$$

nonlinear function of Y.

8|8

Extension to Vector Case.

Case 1: X is a scalar ranom Variable (unknown)

$Y_1, Y_2, \ldots, Y_M$ are (observed) random Variables

Want to estimate X using $\{Y_1, Y_2, \ldots, Y_M\}$.

Let $\hat{X}$ be estimator of X given $Y_1, Y_2, \ldots, Y_M$

Want to minimize $E(X-\hat{X})^2$

Theorem: Optimal Estimator $\hat{X} = E(X \mid Y_1, Y_2, \ldots, Y_M)$

Note:
$$f_X(x \mid y_1, y_2, \ldots, y_m) = \frac{f_{XY_1,\ldots,Y_M}(x, y_1, y_2, \ldots, y_m)}{f_{Y_1 Y_2 \ldots Y_M}(y_1, y_2, \ldots, y_m)}$$

Proof. Exactly same as previous scalar case.

Following orthogonality condition also holds

$$E\left\{\left[X - E(X \mid Y_1, Y_2, \ldots Y_M)\right] g(Y_1, \ldots, Y_M)\right\} = 0$$

for any $g(\cdot, \ldots, \cdot)$

Example:

$$Y_1 = X + V_1$$

$$Y_2 = X + V_2$$

X is binary $\{\pm 1\}$ equally likely

$$V_1 \sim N(0, \sigma_1^2) \qquad V_2 \sim N(0, \sigma_2^2)$$

X, V_1, V_2 are all independent

MMSE Estimator $\hat{X} = E(X \mid Y_1, Y_2)$

try this yourself. $\longrightarrow$ $= \tanh\left(\dfrac{Y_1}{\sigma_1^2} + \dfrac{Y_2}{\sigma_2^2}\right)$

Case 2 $X_1, X_2, \ldots, X_L$ are (unknown) random Variables

$Y_1, Y_2, \ldots, Y_M$ are (observed) random Variables

Define vectors $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_L \end{bmatrix}$ $L \times 1$ vector

(underscore denotes vector)

$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix}$ $M \times 1$ vector.

Want an estimator of $\underline{X}$ given $\underline{Y}$.

Let $\hat{\underline{X}} = g(\underline{Y})$

$$\underline{\hat{X}} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_L \end{bmatrix}$$

Error Vector

$$\underline{\tilde{X}} = \underline{X} - \underline{\hat{X}}$$

$$= \begin{bmatrix} X_1 - \hat{X}_1 \\ \vdots \\ X_L - \hat{X}_L \end{bmatrix}$$

Recall
$$\underline{\hat{X}} = \begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_L \end{bmatrix} = g(\underline{Y})$$

$$= \begin{bmatrix} g_1(\underline{Y}) \\ g_2(\underline{Y}) \\ \vdots \\ g_L(\underline{Y}) \end{bmatrix}$$

<u>Goal:</u> Minimize total MSE

ie) $E\left\{ \tilde{X}_1^2 + \tilde{X}_2^2 + \cdots + \tilde{X}_L^2 \right\}$

by finding optimal $\{ g_1(\cdot), \ldots, g_L(\cdot) \}$

<u>Note:</u> Minimizing each term $E\{\tilde{X}_i^2\}$ separately will lead to minimal total MSE.

$$\tilde{X}_i = X_i - g_i(\underline{Y})$$

We already know (from case 1) that

$E\{\tilde{X}_i^2\}$ is minimized if

$$g_i(\underline{Y}) = E\{X_i \mid \underline{Y}\}$$

So

Optimal estimator $\hat{\underline{X}} = \begin{bmatrix} E(X_1 \mid \underline{Y}) \\ E(X_2 \mid \underline{Y}) \\ \vdots \\ E(X_L \mid \underline{Y}) \end{bmatrix}$

Optimal estimation of $\underline{X}$ given $\underline{Y}$ is equivalent to optimal estimation of each X_i separately using $\underline{Y}$

x ——————— x