Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute
Shanghai Jiao Tong University

Oct. 27 2020

Ve501 2020-2021 Fall

Based on Lecture Notes by Prof. Chong Han

Outline

- 3. Multiple Random Variables
  - Two Random Variables
  - Marginal PDF
  - Functions of Two Random Variables
  - Conditional PDF
  - Joint Moments
  - Mean Square Error Estimation
  - Probability bound
  - Random Vectors
  - Sample Mean
  - Convergence of Random Sequences
  - Central Limit Theorem

Joint Moments

Joint Moments

Let $Q(X, Y)$ be any function of RV's X and Y with joint PDF $f_{XY}(x, y)$.

$$E\big(Q(X,Y)\big) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} Q(x, y)\, f_{XY}(x, y)\,dx\,dy$$

Consider:

$$Q(X,Y) = X^{j}Y^{k}$$

$$E\left[X^{j}Y^{k}\right] = \mathrm{jk}^{\mathrm{th}}\ \text{moment of } (X,Y)$$

$$E\left[(X-\mu_X)^{j}(Y-\mu_Y)^{k}\right] = \mathrm{jk}^{\mathrm{th}}\ \text{central moment}$$

Special Cases

$\underline{j = k = 1:}$

$E[XY] =$ Correlation of X and Y

$\mathrm{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] =$ Covariance of X and Y

$= E[XY] - \mu_x \mu_y$

Alternative formula:

$$\mathrm{cov}(X,Y) = E(XY) - \mu_X \mu_Y$$

A normalized version:

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Correlation Coefficient

Correlation Coefficient

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Measures affine dependence between X and Y, that is, how well Y is predicted by aX+b, given an observation of X.

$$|\rho| = 1 \quad \Leftrightarrow \quad Y = aX + b$$ totally linear

$$\rho = 1 \quad \Rightarrow \quad a > 0$$ positively

$$\rho = -1 \quad \Rightarrow \quad a < 0$$ negatively

$$\rho = 0 \quad \Rightarrow \quad X \text{ and } Y$$ are uncorrelated

$|\rho_{XY}| \leq 1$

Can you prove $|\boldsymbol{\rho}_{XY}| \leq \mathbf{1}$?

- **Cauchy-Schwarz inequality** says that

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}$$

where equality holds iff $Y = aX$ for some constant a, i.e., X and Y are linearly related.

This result provides an important bound on the correlation between two random variables.

Proof

Let $Z = X - \lambda Y$ *where* λ *is a constant. Then*

$$0 \le \mathrm{E}\left[Z^2\right] = \mathrm{E}\left[(X - \lambda Y)^2\right]$$

$$= \mathrm{E}\left[X^2 + \lambda^2 Y^2 - 2\lambda XY\right] \quad (\textit{equalivancy of expectation})$$

$$= \mathrm{E}\left[X^2\right] + \lambda^2 \mathrm{E}\left[Y^2\right] - 2\lambda \mathrm{E}[XY], \quad (\textit{linearity})$$

Proof

Consider the RHS as a polynomial in λ*, since*

$$\lambda^2 \mathrm{E}\left[Y^2\right] - 2\lambda \mathrm{E}[XY] + \mathrm{E}\left[X^2\right] \geq 0$$

always, the discriminant $b^2 - 4ac$

$$(-2\,\mathrm{E}[XY])^2 - 4\,\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right] \leq 0$$

$(b^2 - 4ac \leq 0,$ *there are no x-intercepts*$)$

$$\Longrightarrow [\mathrm{E}[XY]]^2 \leq \mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]$$

Proof

Apply now the Cauchy-Schwarz inequality to the r.v.'s in the definition of correlation coefficient.

$$Z_1 = \frac{X - E[X]}{\sigma_X}, \quad Z_2 = \frac{Y - E[Y]}{\sigma_Y}$$

Note that

$$\begin{aligned} E[Z_1] &= 0, \quad \mathrm{Var}\{Z_1\} = 1 = E\left[Z_1^2\right] \\ E[Z_2] &= 0, \quad \mathrm{Var}\{Z_2\} = 1 = E\left[Z_2^2\right] \end{aligned}$$

Proof

By the Cauchy-Schwarz inequality

$$[E[Z_1Z_2]]^2 \leq \overset{=1}{E\left[Z_1^2\right]}\,\overset{=1}{E\left[Z_2^2\right]} = 1$$

$$\Longrightarrow |E[Z_1Z_2]| \leq 1$$

$$|\rho_{XY}| = \left|E\left[\left(\frac{X-m_X}{\sigma_X}\right)\left(\frac{Y-m_Y}{\sigma_Y}\right)\right]\right| \leq 1$$

Uncorrelated RV's

X and Y are uncorrelated when

correlation: $$E(XY) = E(X)E(Y)$$

Recalling, $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$,

we see there are two more ways to indicate uncorrelatedness:

$$\text{cov}(X,Y) = 0 \qquad \rho_{XY} = 0$$

Important Relations

Independence --> Uncorrelation

Proof:

$$E[XY]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_{XY}(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)dxdy$$

$$=\int_{-\infty}^{+\infty} xf_X(x)dx\int_{-\infty}^{+\infty} yf_Y(y)dy=E(X)E(Y)$$

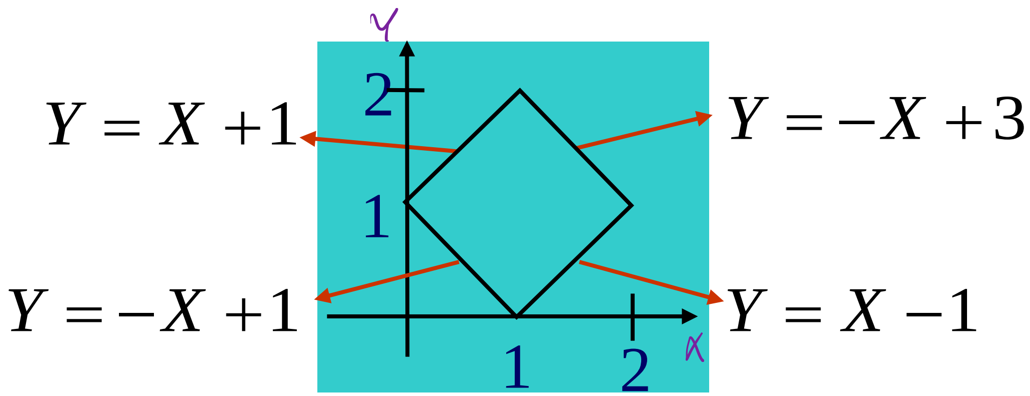
Uncorrelation ~~-->~~Independence

$E[XY] = E[X][Y]$ 不 in general

$E[XY] = E[X]E[Y]$
+ Gaussian => Independence

Example

Let $f_{XY}(x,y)$ be constant (uniform) over the diamond:



By observation, $f_X(x)$ and $f_Y(y)$, are the same, and symmetrical about 1, thus $E(X)=E(Y)=1$.

The height of $f_{XY}(x,y)$ is ½.

Example, Continued

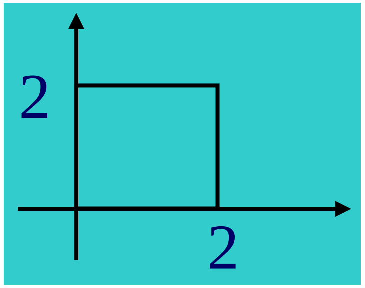
$$E(XY) = \iint\limits_{\substack{\text{over}\\ \text{diamond}}} \frac{1}{2} xy\,dx\,dy = 1$$

(Verified numerically)

$\therefore$ X and Y are uncorrelated since

$$E(XY) = E(X)E(Y)$$

However, X and Y are not independent because the ROS of $f_X(x)$, $f_Y(y)$, covers the square:



Example

$$\begin{tikzcd} X \arrow[r] & \bigotimes \arrow[r] & Y = X + N \\ & N \arrow[u] & \end{tikzcd}$$

Suppose X and N are uncorrelated and N has zero mean. Show that

$$\mathrm{E}\left[Y^2\right] = \mathrm{E}\left[X^2\right] + \mathrm{E}\left[N^2\right].$$

$$E[Y^2] = E[(X+N)^2] \qquad E[X]\,\overset{0}{E[N]}$$

$$= E[X^2] + 2E[XN] + E[N^2]$$

$$= E[X^2] + E[N^2]$$

Orthogonality

- E(XY) qualifies as an inner product or

$$E(XY) = \langle X, Y \rangle$$

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} XY f_{XY}(x,y)\,dx\,dy$$

- X and Y are orthogonal when E(XY)=0

- Will be useful in linear mean square estimation

Jointly Normal RVs

Let the covariance matrix C be defined:

$$C = \begin{bmatrix} E[(X-\eta_X)(X-\eta_X)] & E[(X-\eta_X)(Y-\eta_Y)] \\ E[(Y-\eta_Y)(X-\eta_X)] & E[(Y-\eta_Y)(Y-\eta_Y)] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_X^2 & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & \sigma_Y^2 \end{bmatrix}$$

in general

$C = E[(Z-\eta_Z)\cdot(Z-\eta_Z)^T]$

Let $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ $\quad$ mean $\quad \eta_Z = \begin{bmatrix} \eta_X \\ \eta_Y \end{bmatrix}$

好像这里的 η 就是 μ

Then X and Y are jointly Gaussian iff

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{\det C}} \exp\left\{ -\frac{[Z-\eta_Z]^T C^{-1} [Z-\eta_Z]}{2} \right\}$$

Expanded Form

$$f_{XY}(x,y)$$

$$= \frac{\exp\left\{-\frac{1}{2\left(1-\rho_{XY}^2\right)}\left[\left(\frac{X-\eta_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{X-\eta_X}{\sigma_X}\right)\left(\frac{Y-\eta_Y}{\sigma_Y}\right) + \left(\frac{Y-\eta_Y}{\sigma_Y}\right)^2\right]\right\}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}$$

if uncorrelated $X, Y \Rightarrow \rho_{XY} = 0.$

[illegible]

This expression has the interesting property that X and Y always appear in centered, normalized forms

any matrix

$$Z = A\begin{bmatrix} X \\ Y \end{bmatrix}$$

if $\begin{bmatrix} X \\ Y \end{bmatrix}$ is joint Gaussian.

Z is also joint Gaussian.

Uncorrelated Gaussians

If X and Y are uncorrelated, then $\rho_{XY} = 0$

$$C = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix},$$

and

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{-\frac{(x-\eta_X)^2}{2\sigma_X^2} - \frac{(y-\eta_Y)^2}{2\sigma_Y^2}\right\}$$

$$= f_X(x) f_Y(y)$$

where $X \sim N(\eta_X, \sigma_X^2)$, $Y \sim N(\eta_Y, \sigma_Y^2)$

REMEMBER

Gaussian & Uncorrelated --> Independent

Short Summary

Several joint moments discussed:

- Correlation
- Covariance
- Correlation Coefficient
- Covariance Matrix

Independence implies uncorrelatedness

- But not vice versa

Correlation is a type of inner product

Jointly Gaussian RVs

- Gaussian & Uncorrelated --> Independent

Mean Square Error Estimation

Linear Mean Square Error (MSE) Estimation

Given: μ_X μ_Y, σ_X σ_Y ρ_{XY} and an observation of X.

Goal: Get an estimate of Y in the form:

$$\hat{Y}_{LNH} = aX + b$$

Linear non-homogenous (LNH)

$$\hat{Y}_{LH} = aX$$

Linear homogenous (LH)

Intuition: If X and Y are well correlated, $\hat{Y}_{LNH}$ should be a "good" estimator.

Applications

One step predictor: $x_1, x_2, x_3, \ldots$ is a sequence of correlated random variables (NASDAQ Composite?)

$$\hat{X}_{n+1} = aX_n + b$$

Weight, W, and cholesterol level, C

$$\hat{C} = aW + b$$

Mean Squared Error

MMSE : minimized MSE

‖

Goodness is measured in mean squared error (MSE). Let *e* be the estimation error. Then,

$$MSE = E\left[\varepsilon^2\right] = E\left[\left(Y - \hat{Y}\right)^2\right]$$

$$= \text{“average error power”}$$

Pick coefficients a and b (or just a for homogenous case) to minimize MSE.

Linear Non-Homogenous Estimation

$$MSE = E\left\{\left[Y - \left(aX + b\right)\right]^2\right\}$$

$$= E\left[Y^2\right] - 2aE\left[XY\right] - 2bE\left[Y\right] + a^2E\left[X^2\right] + 2abE\left[X\right] + b^2$$

$$\left.\begin{aligned} \frac{\partial MSE}{\partial a} &= -2E\left[XY\right] + 2aE\left[X^2\right] + 2bE\left[X\right] = 0 \\ \frac{\partial MSE}{\partial b} &= -2E\left[Y\right] + 2aE\left[X\right] + 2b = 0 \end{aligned}\right\} \text{使 MSE minimized}$$

$$\rho_{XY} = \frac{E[XY]}{\sigma_X \sigma_Y}$$

Using $E[XY] = \rho_{XY}\sigma_X\sigma_Y + \mu_X\mu_Y$

$$a = \frac{\sigma_Y}{\sigma_X}\rho_{XY} \qquad b = E\left[Y\right] - aE\left[X\right]$$

$= a(X - \mu_X) + \mu_Y$

Key result: $\hat{Y}_{LNH} = \frac{\sigma_Y}{\sigma_X}\rho_{XY} X + \mu_Y - a\mu_X$

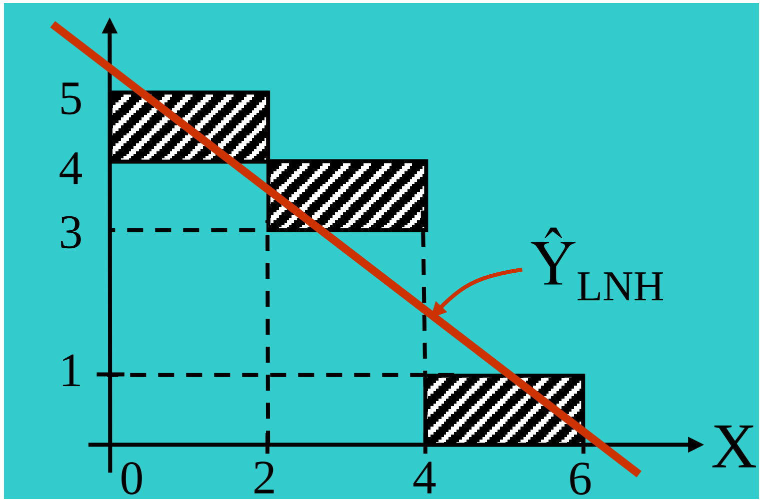
Rearrangement and interpretations:

$$\hat{Y}_{LNH} = \sigma_Y \overbrace{\rho_{XY} \underbrace{\left(\frac{X - \mu_X}{\sigma_X}\right)}_{\text{Zero-mean, unit variance version of X}}}^{\text{Zero-mean, unscaled version of } \hat{Y}} + \mu_Y$$

scaling slope offset

Example of LNH Estimator

Let X and Y be uniformly distributed over the shaded region:



$\hat{Y}_{LNH} = \frac{\sigma_Y}{\sigma_X}\rho_{XY} X + \mu_Y - a\mu_X$

where $a = \frac{\sigma_Y}{\sigma_X}\rho_{XY}$

Needed moments: $\mu_X = 3, \sigma_X = \sqrt{3},$

$\Rightarrow a = -0.889$,

$$\mu_Y = \frac{17}{6}, \sigma_Y = 1.724, \rho_{XY} = -0.893$$

$$\hat{Y}_{LNH} = -0.889X + 5.5$$

$MSE = E[\varepsilon^2]$

$= \ldots$ 见前页 a, b 求导

Orthogonality Condition

Recall the optimal "a" for $\hat{Y}_{\text{LNH}}$ solves: $$\frac{\partial}{\partial a}E\left[\varepsilon^2\right]=0$$

$$\frac{d}{da}E\left[\varepsilon^2\right]=E\left[2\varepsilon\left(\frac{d}{da}\varepsilon\right)\right]$$

μ_X

$$=2E\left\{\varepsilon\left(\frac{d}{da}\left[Y-a\left(X-m_X\right)-m_Y\right]\right)\right\}$$

$$=2E\left\{\varepsilon\left(X-m_X\right)\right\}$$

$$\Rightarrow E\left[\varepsilon\left(X-m_X\right)\right]=0$$

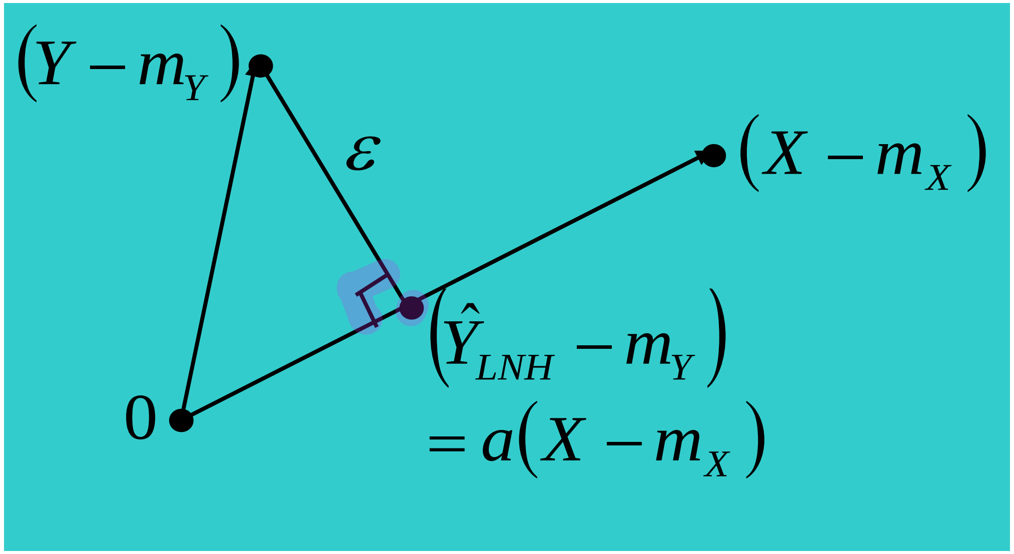
Orthogonal

Also, because $E\left[\varepsilon\right]=0$ then we have

$$E\left[\varepsilon X\right]=0$$ Orthogonality between error and "data"

Geometrical View – Non-homogeneous Case

$$\left(\hat{Y}_{LNH} - m_Y\right) = a\left(X - m_X\right)$$
$$\varepsilon = \left(Y - m_Y\right) - a\left(X - m_X\right)$$



The estimator is the point in the space spanned by $(X\text{-}m_X)$ that is nearest to $(Y\text{-}m_Y)$.

Performance of $\hat{Y}_{LNH}$

$$MSE_{opt} = E\{\varepsilon[(Y - m_Y) - a(X - m_X)]\}$$

orthogonal

$$= E\{\varepsilon(Y - m_Y)\}$$

$$= E\{[(Y - m_Y) - a(X - m_X)](Y - m_Y)\}$$

$$= \sigma_Y^2 - a\,\mathrm{cov}(X, Y)$$

$$= \sigma_Y^2 - \frac{\sigma_Y}{\sigma_X}\rho_{XY}\,\mathrm{cov}(X, Y)$$

$$\boxed{= \sigma_Y^2(1 - \rho_{XY}^2)}$$

Observations About Optimal MSE

$$MSE_{opt} = \sigma_Y^2 (1 - \rho_{XY}^2)$$

Lowest when $|\rho_{XY}| = 1$ (Perfect correlation with $Y=aX+b$)

Highest when $\rho_{XY} = 0$ (Uncorrelated)

"When X and Y are uncorrelated, linear estimation is worthless."

Worst case:

$$\rho_{XY} = 0 \Rightarrow \hat{Y}_{LNH} = m_Y, \quad MSE = \sigma_Y^2$$

Linear Homogenous Estimation

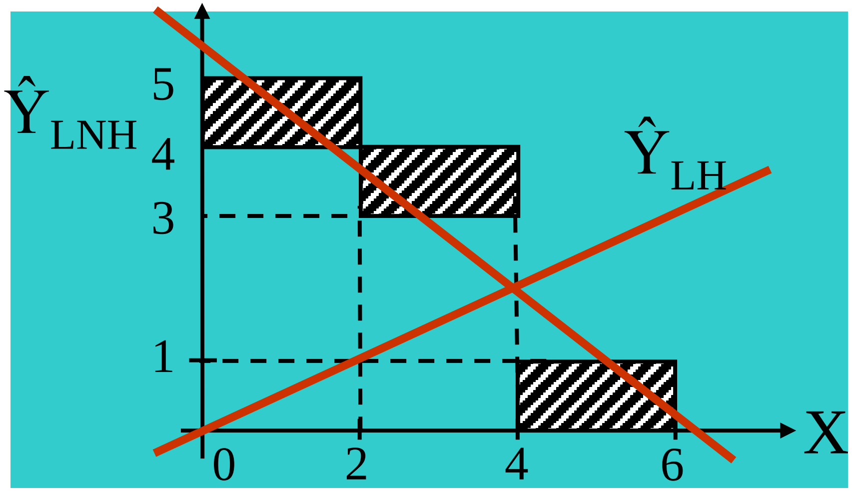
This has the form: $\hat{Y}_{LH} = aX$

"a" minimizes the MSE: $\dfrac{d}{da} E\left[\varepsilon^2\right] = 0 \Rightarrow a = \dfrac{E(XY)}{E(X^2)}$

$$MSE_{opt} = E\left[Y^2\right]\left[1 - \frac{E^2(XY)}{E(X^2)E(Y^2)}\right]$$

Observe that all of this is a special case of $\hat{Y}_{LNH}$ when $m_X = m_Y = 0$

Earlier Example Cont'd



$$MSE_{opt,LNH} = 0.602$$

$$MSE_{opt,LH} = 10.97$$

这个不太行，显然

$$\hat{Y}_{LH} = 0.486X$$

REMEMBER

Linear homogenous estimators are best for zero-mean joint distributions.

Orthogonality Condition for the Homogeneous Case

Recall the optimal "a" for $\hat{Y}_{LH}$ solves: $$\frac{\partial}{\partial a} E\left[\varepsilon^2\right] = 0$$

$$\frac{d}{da} E\left[e^2\right] = E\left[2e\left(\frac{d}{da}e\right)\right]$$

$$= 2E\left\{e\left(\frac{d}{da}\left[Y - aX\right]\right)\right\}$$

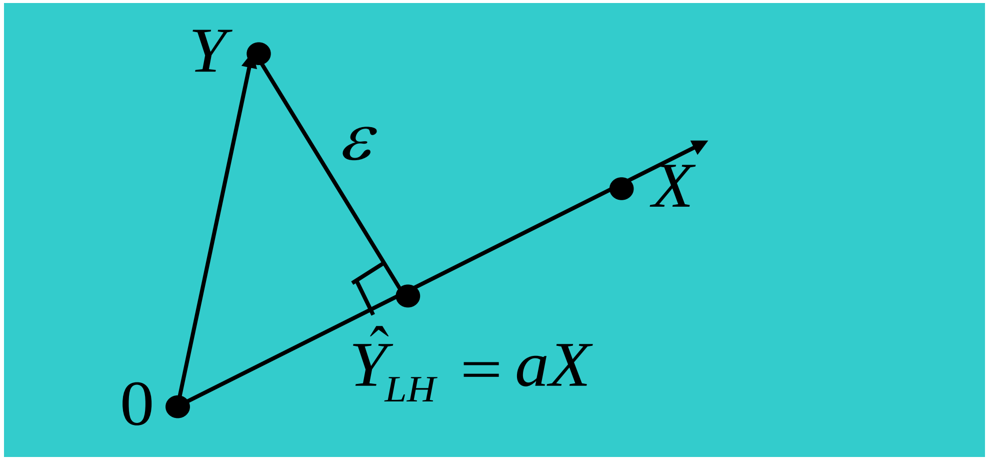
$$= 2E\left\{eX\right\}$$

$$= 0$$

Geometrical View – Homogeneous Case

$$\hat{Y}_{LH} = aX$$

$$\varepsilon = Y - aX$$



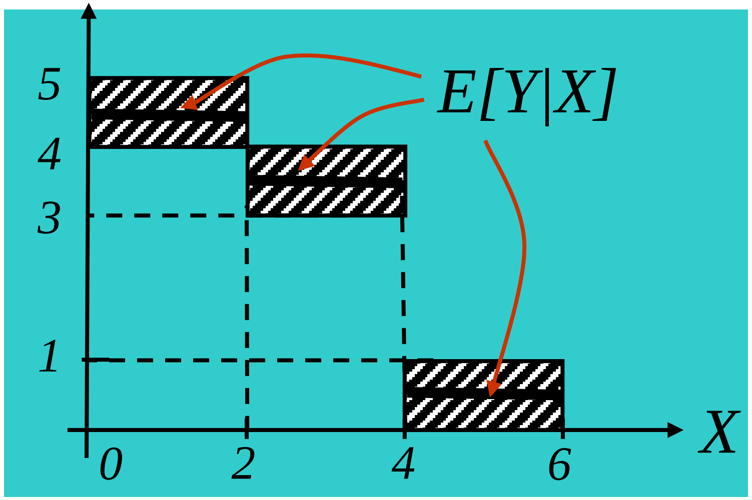
The estimator is the point in the space spanned by X that is nearest to Y.

Nonlinear MSE Estimation

Now we remove the constraint that $\hat{Y}$ must be a linear function of X. We will show that the optimal estimator is

$$\boxed{\hat{Y}_{NL} = E\left(Y \mid X\right)}$$

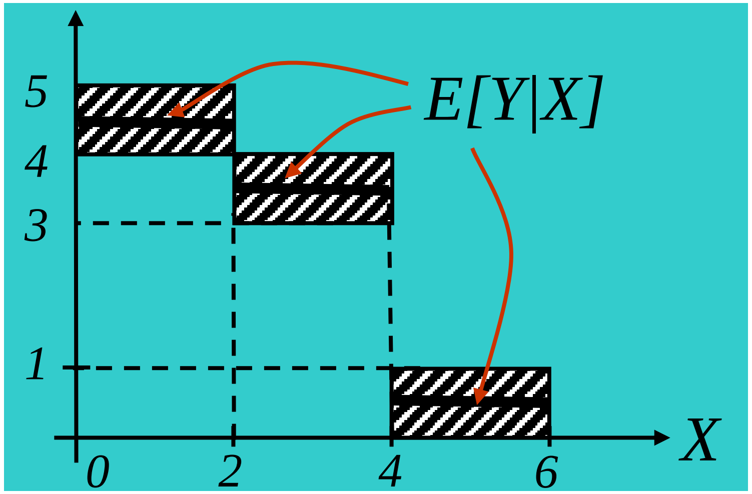
$E\left(Y \mid X\right)$ for the previous example is indicated in bold:



X and Y are uniformly distributed over the shaded region.

Performance

Typically, a double integral is required to calculate the optimal MSE_{NL}.



~uniform.

For this example,

$$MSE_{NL} = \int_0^2 \int_4^5 (y-4.5)^2 \frac{1}{6} dydx$$

$$+ \int_2^4 \int_3^4 (y-3.5)^2 \frac{1}{6} dydx + \int_4^6 \int_0^1 (y-0.5)^2 \frac{1}{6} dydx$$

$$= 0.08\overline{3}$$

Recall MSE_{LH}=10.97 and MSE_{LNH}=0.602.

Proof That $\hat{Y}_{NL}$=E[Y|X]

The proof includes an interesting use of iterated expectation.

Begin with $\hat{Y}_{NL} = H(X)$, some arbitrary function of X.

any function

We want $H(X)$ to minimize

just subtract and add it

$$
\begin{aligned}
MSE_{NL} &= E\left\{\left(Y - H(X)\right)^2\right\} \\
&= E\left\{\left[Y - E(Y \mid X) + E(Y \mid X) - H(X)\right]^2\right\} \\
&= E\left\{\left[Y - E(Y \mid X)\right]^2\right\} + 2E\left\{\left[Y - E(Y \mid X)\right]\left[E(Y \mid X) - H(X)\right]\right\} \\
&\quad + E\left\{\left[E(Y \mid X) - H(X)\right]^2\right\}
\end{aligned}
$$

≥ 0

≥ 0

Will address the second term=0 next

证明这个.

$H(X) = E(Y|X)$

自然是 MSE↓ 的解

Proof, Cont'd

Use iterated expectation on the second term:

$$E\{[Y-E(Y|X)][E(Y|X)-H(X)]\}$$

$E[Y]$

$= E[E[Y|X]]$

$$= E\{E([Y-E(Y|X)]\underbrace{[E(Y|X)-H(X)]}\cancel{\}}|X)\}$$

just a function of X, so it comes out of the conditional expectation.

$$= E\{\overbrace{\underbrace{E[(Y-E[Y|X])|X]}}^{0}[E(Y|X)-H(X)]\}$$

This equals: $E[Y|X]-E[Y|X]=0$

so the second term is zero

Proof, Concluded

The first and third terms remain:

$$MSE_{NL} = \underbrace{E\left\{\left[Y - E(Y \mid X)\right]^2\right\}}_{\text{Ignore this term; it is not affected by } H(X).} + \underbrace{E\left\{\left[E(Y \mid X) - H(X)\right]^2\right\}}_{\text{This is minimized by setting } H(X) = E[Y \mid X]}$$

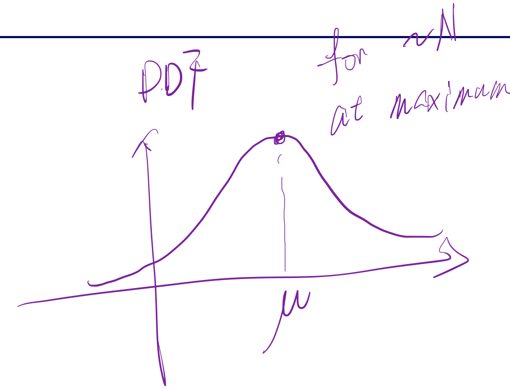
Ignore this term; it is not affected by *H(X)*.

This is minimized by setting

$$H(X) = E[Y \mid X]$$

$$\therefore \hat{Y}_{NL} = E(Y \mid X)$$

Nonlinear MSE Estimator for Gaussians



$E(Y|X)$ is the mean of $f_{Y|X}(y|x)$

$$f_{Y|X}(y\,|\,x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= A(x)\exp\left\{-\frac{1}{2\left(1-\rho_{XY}^2\right)}\left[B(x) - 2\rho_{XY}\left(\frac{X-\eta_X}{\sigma_X}\right)\left(\frac{Y-\eta_Y}{\sigma_Y}\right) + \left(\frac{Y-\eta_Y}{\sigma_Y}\right)^2\right]\right\}$$

Exponent is quadratic in y; leading term is negative → $f_{Y|X}(y|x)$ is a Gaussian PDF for y.

Because it is Gaussian

We can find the mean by maximizing $f_{Y|X}(y|x)$, which is equivalent to minimizing the y-dependent portion of the exponent:

$$\left[-2\rho_{XY}\left(\frac{X-\eta_X}{\sigma_X}\right)\left(\frac{Y-\eta_Y}{\sigma_Y}\right)+\left(\frac{Y-\eta_Y}{\sigma_Y}\right)^2\right]$$

minimize

$e^{-\frac{\cdots}{\cdots}}$

for Joint Gaussian

Non linear

best

LNH

The minimization yields

$$\hat{Y}_{NL}=\frac{\sigma_Y\rho_{XY}}{\sigma_X}\left(X-\eta_X\right)+\eta_Y$$

LINEAR NON-HOMOGENEOUS!

The linear non-homogeneous estimator is the best estimator when X and Y are jointly Gaussian

Mean Square Error Estimation

Linear MSE estimator

Non-homogeneous $\hat{Y}_{LNH} = \frac{s_Y}{s_X} r_{XY} X + m_Y - a m_X$

Homogeneous $\hat{Y}_{LH} = \frac{E(XY)}{E(X^2)} X$

Orthogonality condition

Non-linear MSE estimator $\hat{Y}_{NL} = E(Y \mid X)$

The linear non-homogeneous estimator is the best estimator when X and Y are jointly Gaussian

Thank You!