Lecture 19: Linear Transformation of a Gaussian Random Vector

Let $\vec{X}$ $(n \times 1)$ be Gaussian random vector, (μ_X, C_X). Given a $m \times n$ matrix A, with $\text{rank}(A) = m$,

$\vec{Y} = A\vec{X} + b$ is

an m-dim Gaussian r.vector with

$$\boxed{\begin{aligned} \mu_Y &= A\mu_X + b, \\ C_Y &= A C_X A^T \end{aligned}}$$

Proof

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{|\det(A)|} \cdot f_{\vec{X}}\left(A^{-1}(\vec{y} - b)\right)$$

$$= \frac{1}{(2\pi)^{n/2} |\det(A)| \cdot |\det(C_X)|^{1/2}} \exp\left\{-\frac{1}{2}\left[A^{-1}(\bar{y}-b) - \mu_X\right]^T C_X^{-1} \left[A^{-1}(\bar{y}-b) - \mu_X\right]\right\}$$

Note $\Rightarrow$ $A^{-1}(\vec{y} - b) - \mu_X = A^{-1}\left(\vec{y} - (A\mu_X + b)\right) = A^{-1}(\vec{y} - \mu_Y)$

$$\Rightarrow \quad f_{\vec{Y}}(\vec{y}) = \frac{\exp\left\{-\frac{1}{2}\left(A^{-1}(\vec{y} - \mu_Y)\right)^T C_X^{-1}\left(A^{-1}(\vec{y} - \mu_Y)\right)\right\}}{(2\pi)^{n/2} |\det(A)| \, |\det(C_X)|^{1/2}}$$

Using the identities

$$\left\{\begin{aligned} &\left.|\det(A)|\,|\det(C_X)|^{1/2} = |\det(A C_X A^T)|^{1/2}\right\} \text{ (Check yourself)} \\ &\text{and } \quad (A^{-1})^T = (A^T)^{-1} \end{aligned}\right.$$

$$f_{\vec{Y}}(\vec{y}) = \frac{\exp\left\{-\frac{1}{2}(\vec{y} - \mu_Y)^T (A C_X A^T)^{-1} (\vec{y} - \mu_Y)\right\}}{(2\pi)^{n/2} |\det(A C_X A^T)|^{1/2}}.$$

Standard Normal Random Vector Z
is a n-dim Gaussian random vector, with

$$E[Z] = \vec{0} \quad \text{and} \quad C_Z = I$$

($n \times n$ identity matrix).

Theorem

For a Gaussian (μ_X, C_X) random vector X, let A be a $n \times n$ matrix with the property $C_X = AA^T$, then the random vector

$$Z = A^{-1}(X - \mu_X)$$

is a standard normal random vector.

Theorem

Given the n-dimensional standard normal random vector Z, an invertible matrix A, and a n-dim vector b,

$$X = AZ + b$$

is a n-dim Gaussian random vector with

$$\mu_X = b$$

$$C_X = AA^T$$

<u>Theorem</u> For a Gaussian random vector X, with covariance matrix C_X, there always exists a matrix A such that

$$C_X = A A^T.$$

To prove this, we will establish a sequence of important facts.

<u>Fact 1</u>: For a random vector X with correlation matrix R_X and covariance matrix C_X, R_X and C_X are both positive semidefinite.

A matrix M is positive semi-definite if for any non-zero vector a,

$$a^T M a \geq 0$$

<u>Let us prove Fact 1</u>:

Given a random vector X, we can define $Y = X - \mu_X$, so that,

$$C_X = E\left[(X-\mu_X)(X-\mu_X)^T\right] = E\left[YY^T\right] = R_Y$$

$\boxed{\text{Thus, it follows that the covariance matrix } (C_X) \text{ is positive semi-definite if and only if the correlation matrix } (R_Y) \text{ is PSD.}}$

$\Rightarrow$ It suffices to show that <u>every</u> correlation matrix is PSD (whether it is denoted by R_Y or R_X).

To show R_Y is Positive Semi-definite, we write

$$a^T R_Y a = a^T E[Y Y^T] a$$
$$= E[a^T Y Y^T a]$$
$$= E[(a^T Y)(a^T Y)^T]$$

If we denote $W = a^T Y$, then $a^T Y = W$ is a random variable.

$$\Rightarrow \quad a^T R_Y a = E[W W^T] = E[W^2]$$

however $E[W^2] \geq 0$ for any random variable W

$$\Rightarrow \quad a^T R_Y a = E[W^2] \geq 0$$

$\Rightarrow$ R_Y is positive semi-definite.

$\Rightarrow$ For a PSD matrix, all eigenvalues are ≥ 0 i.e nonnegative.

The definition of the Gaussian vector PDF requires the existence of C_x^{-1}. (or $\det(C_x) > 0$)

$\Rightarrow$ all eigenvalues are non-zero (of C_x)

{Why?? $\Rightarrow$} Since $\det(C_x) > 0$ and $\det(C_x) =$ {Product of its eigenvalues}

$\Rightarrow$ all eigenvalues of C_x are <u>positive</u>.

Since C_x is a real symmetric matrix, it has a singular value decomposition (SVD)

$$C_X = U D U^T, \quad \text{where}$$

$$D = \text{diag}[\lambda_1 \ \ \lambda_2 \ \ldots \ \lambda_n]$$ is the diagonal matrix of eigenvalues of C_x.

Since each d_i is positive, we can define

$$D^{1/2} = \text{diag}[\sqrt{\lambda_1} \ \ \sqrt{\lambda_2} \ \ldots \ \sqrt{\lambda_n}], \quad \text{and then}$$

$$C_X = U D^{1/2} D^{1/2} U^T = (U D^{1/2})(U D^{1/2})^T$$

$$\Rightarrow \boxed{A = U D^{1/2}}.$$

From these Theorems & facts, we can see that any Gaussian (μ_X, C_X) random vector X can be written as a linear transformation of uncorrelated Gaussian $(0,1)$ random variables

Since $C_X = UDU^T = (UD^{1/2})(UD^{1/2})^T$,
$= AA^T$,

$$\boxed{X = AZ + \mu_X}$$

$$\Rightarrow \boxed{X = UD^{1/2}Z + \mu_X}$$

Note that U has orthonormal colums $\vec{u}_1, \dots, \vec{u}_n$.
When $\mu_X = 0$, then

$$X = UD^{1/2}Z$$

$$X = \sum_{i=1}^{n} \sqrt{d_i}\, \vec{u}_i \cdot Z_i$$

the interpretation of the above eq^n is that X is a combination of orthogonal vectors $\sqrt{d_i}\, \vec{u}_i$, each scaled by an independent Gaussian random variable Z_i.

Lecture 24

24

Lecture 24 10-23-2015

Linear MS Estimation

Returning to Mean Squared Estimation

* Recall that we showed that to estimate Y from X such that the MSE was minimized, the optimal estimator is $E[Y|X]$.

* $E[Y|X]$ may not be linear in X.

An easier problem is to find the best linear MS estimator of Y from X. In other words, the estimator is of the form $AX+B$. and we want to find A and B to minimize MSE

$$MSE = E\left[(Y-(AX+B))^2\right]$$
$$= E\left[(Y-AX-B)^2\right]$$

$\Rightarrow$ for a given A,

$$\boxed{\begin{aligned} B &= E(Y-AX) \text{ minimizes the above} \\ &= \mu_Y - A\mu_X \end{aligned}}$$

$$\Rightarrow MSE = E\left[(Y-AX-\mu_Y-A\mu_X)^2\right]$$
$$= E\left[((Y-\mu_Y)-A(X-\mu_X))^2\right]$$
$$= \sigma_Y^2 + A^2\sigma_X^2 - 2A\,Cov(X,Y)$$

$$\frac{\partial MSE}{\partial A} = 2A\sigma_X^2 - 2Cov(X,Y) = 0$$

$$\Rightarrow \boxed{A = \frac{Cov(X,Y)}{\sigma_X^2} = \frac{\rho\sigma_X\sigma_Y}{\sigma_X^2} = \frac{\rho\sigma_Y}{\sigma_X}}$$

Linear-Minimum Mean Sq. Estimate

(2)

$\Rightarrow$ Optimal linear (MS) estimator (or LMMSE)

$$AX + B$$

with $A = \frac{\rho \sigma_Y}{\sigma_X}$; $B = \mu_Y - \frac{\rho \sigma_Y}{\sigma_X} \mu_X$

Resulting Error $= \sigma_Y^2 + \frac{\rho^2 \sigma_Y^2}{\sigma_X^2} \cdot \sigma_X^2 - 2 \frac{\rho \sigma_Y}{\sigma_X} \cdot \rho \sigma_Y \sigma_X$

The error is E[(Y-Y_L

$$= \sigma_Y^2 + \rho^2 \sigma_Y^2 - 2 \rho^2 \sigma_Y^2$$

$$= \sigma_Y^2 - \rho^2 \sigma_Y^2$$

$$= \sigma_Y^2 (1 - \rho^2)$$

Fact: For jointly normal random variables (X, Y), nonlinear and linear MS estimates are identical

[Convince Yourself......]

MMSE=L-MMSE

$$\hat{Y}_{LMMSE} = AX + B$$

$$= \left(\frac{\rho \sigma_Y}{\sigma_X}\right) X + \mu_Y - \left(\frac{\rho \sigma_Y}{\sigma_X}\right) \mu_X$$

$$= \left(\frac{\rho \sigma_Y}{\sigma_X}\right) (X - \mu_X) + \mu_Y$$

$$\hat{Y}_{LMMSE} = \frac{\mathrm{cov}(X, Y)}{\sigma_X^2} (X - \mu_X) + \mu_Y$$

Example $f_{X,Y}(x,y) = \begin{cases} 2e^{-x}e^{-y} & 0 \le y \le x < \infty \\ 0 & \text{otherwise.} \end{cases}$

(1) Find the MMSE of X given Y

$$\hat{X}_{MMSE} = E[X|Y] = Y+1 \quad (\text{check!!})$$

(2) Find the Linear MMSE (or LMMSE) of X given Y

Since $\hat{X}_{MMSE}$ is already linear

$$\Rightarrow \quad \hat{X}_{LMMSE} = \hat{X}_{MMSE} = Y+1.$$

(3) Find the LMMSE of Y given X

To find the LMMSE,
we need: μ_X, μ_Y, σ_X^2, σ_Y^2, $Cov(X,Y)$

$\mu_X = 3/2 \qquad \sigma_X^2 = 5/4$

$\mu_Y = 1/2 \qquad \sigma_Y^2 = 1/4$

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y$$

$$E(XY) = \int_0^\infty \int_0^x 2xy\, e^{-x} e^{-y}\, dy\, dx = 1$$

$\Rightarrow \quad Cov(X,Y) = 1 - \frac{3}{2} \times \frac{1}{2} = 1/4$

$$\boxed{\Rightarrow \hat{Y}_{LMMSE} = \left(\frac{Cov(X,Y)}{\sigma_X^2}\right)(X - \mu_X) + \mu_Y}$$

$$= \left(\frac{1/4}{5/4}\right)(x - 3/2) + 1/2$$

$$= \frac{x}{5} - \frac{1}{5}.$$

optimal
MMSE of Y given X

$$\hat{Y}_{MMSE} = E[Y|X]$$

$$= \int y \, f_{Y|X}(y|x) \, dy$$

$$\vdots$$

$$= \underbrace{1 - \frac{xe^{-x}}{1-e^{-x}}}_{\text{non-linear function.}}$$

Error MMSE <= Error L-MMSE

Orthogonality Principle

For Linear MSE estimator (LMMSE),

$$\boxed{E\left[(\text{error}) \times \underbrace{X}_{\text{data}}\right] = 0}$$

i.e error and X are orthogonal.

$$E\left[(Y - AX + B)\, X\right] = 0$$

with $A = \rho \dfrac{\sigma_Y}{\sigma_X}$; $B = \mu_Y - A\mu_X$

For optimal (non-linear) MS estimator, (MMSE)

$$E\left[(\text{error}) \times \underbrace{f(x)}_{\text{any function of data}}\right] = 0$$

i.e., error and any function of X are orthogonal.

Proof: $E\left[\underbrace{(Y - E(Y|X))\, f(x)}_{Z}\right]$

$= E[Z]$

$= \underbrace{E}_{\text{over } X}\left[\underbrace{E[Z|X]}_{\text{this is a r.v. which is a function of X.}}\right]$ [by Iterated Expectation Theorem].

$$\Rightarrow E\Big[\underbrace{\big(Y - E[Y|X]\big) f(X)}_{Z}\Big] = \underset{\substack{\downarrow \\ \text{over} \\ X}}{E}\Big[E[Z|X]\Big]$$

Let us look at $E[Z|X]$

for $X = x$, $E[Z|X] = E[Z \mid X = x]$

$$= E\Big[\big(Y - E(Y|X=x)\big) f(X=x) \Big| X=x\Big]$$

$$= E\Big[f(x)Y - f(x)E[Y|X=x] \Big| X=x\Big]$$

$$= E\big[f(x)Y \big| X=x\big] - E\big[f(x)E(Y|X=x) \big| X=x\big]$$

$$= f(x)E(Y|X=x) - f(x)E(Y|X=x)$$

$$= 0$$

$\Rightarrow$ for $X = x$, $E[Z|X=x] = 0$

$$\Rightarrow \underset{\substack{\downarrow \\ \text{over} \\ X}}{E}\Big[E[Z|X]\Big] = 0$$

$\Rightarrow$ For the optimal (non-linear) MS estimator, we have shown a more general orthogonality principle. as the error is orthogonal to <u>any</u> function of X.

Vector Space Picture for Correlation Properties of R.V.'s.

If we interpret random variables X and Y as vectors $\vec{X}$ and $\vec{Y}$ in some abstract vector space, we can define the inner product in this vec. space as:

$$\boxed{\langle \vec{X}, \vec{Y} \rangle \triangleq E(XY) = R_{XY}}$$

Such a definition satisfies the standard properties of an inner product:

Symmetry: $\langle \vec{X}, \vec{Y} \rangle = \langle \vec{Y}, \vec{X} \rangle$

Linearity: $\langle \vec{X}, a_1\vec{Y_1} + a_2\vec{Y_2} \rangle = a_1\langle \vec{X}, \vec{Y_1} \rangle + a_2\langle \vec{X}, \vec{Y_2} \rangle$

Positivity: $\langle \vec{X}, \vec{X} \rangle$ is positive for $\vec{X} \neq 0$ and 0 otherwise.

Two r.v.'s X, Y are orthogonal, if $\langle \vec{X}, \vec{Y} \rangle = 0$
i.e. $E(XY) = 0$

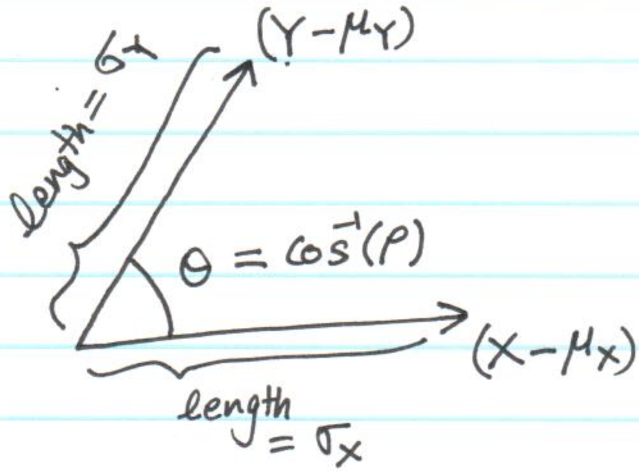
We can represent $X - \mu_X$ and $Y - \mu_Y$ by vectors $\tilde{\vec{X}}$, $\tilde{\vec{Y}}$,

$$\left.\begin{aligned} \langle \tilde{\vec{X}}, \tilde{\vec{X}} \rangle &= E((X-\mu_X)^2) = \sigma_X^2 \\ \langle \tilde{\vec{Y}}, \tilde{\vec{Y}} \rangle &= E((Y-\mu_X)^2) = \sigma_Y^2 \end{aligned}\right\} \Rightarrow$$

Length of $\tilde{\vec{X}} = \sigma_X$,

Length of $\tilde{\vec{Y}} = \sigma_Y$

$$\langle \tilde{\vec{X}}, \tilde{\vec{Y}} \rangle = E\left[(X-\mu_X)(Y-\mu_Y)\right] = \mathrm{cov}(X,Y)$$
$$= \rho\, \sigma_X\, \sigma_Y$$
$$= (\text{length of } \tilde{\vec{X}}) \times (\text{Length of } \tilde{\vec{Y}}) \times \cos(\theta),$$



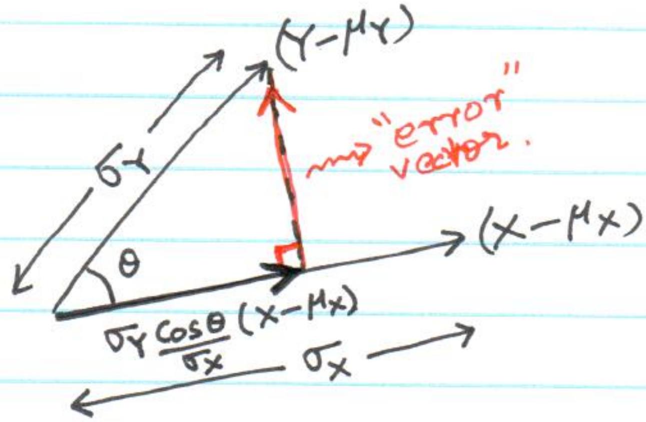
$$\rho = \cos(\theta) \quad \& \quad -1 \le \rho \le 1$$

$\rho = 1 \rightarrow$ "vectors" aligned in same direction

$\rho = -1 \rightarrow$ " " " opposite "

$\rho = 0 \rightarrow$ " are orthogonal.

Interpretation of orthogonality Principle for LMMSE via vector-space



$\Rightarrow$ LMMSE seeks to find a vector $\alpha \vec{\tilde{X}}$ which lies along $\vec{\tilde{X}}$, such that the squared length of the error $(\vec{\tilde{Y}} - \alpha \vec{\tilde{X}})$ vector is minimized.

Such a vector is the orthogonal projection of $\vec{\tilde{Y}}$ on $\vec{\tilde{X}}$ and is given as

$$\alpha \vec{\tilde{X}} = \frac{\langle \vec{\tilde{Y}}, \vec{\tilde{X}} \rangle}{\langle \vec{\tilde{X}}, \vec{\tilde{X}} \rangle} \vec{\tilde{X}}$$

$$\Rightarrow \quad \boxed{\widehat{(Y-\mu_Y)}_{LMMSE} = \frac{\rho \sigma_X \sigma_Y}{\sigma_X^2} (X - \mu_X).}$$

and the error is orthogonal.

LMMSE for a (Scalar) random variable Y from a random vector $\vec{x}$

$$\hat{Y}_{LMMSE} = C_{XY}^T C_X^{-1}(\vec{x} - \vec{\mu}_x) + \mu_Y$$

Proof: we will prove the above for the case when $\mu_Y = 0$, and $\vec{\mu}_x = \vec{0}$. In this special case,

$$\hat{Y} = a^T \vec{x} = a_1 x_1 + a_2 x_2 \cdots + a_n x_n$$

By orthogonality principle

$$E[\underbrace{(Y - a_1 x_1 - a_2 x_2 \cdots - a_n x_n)}_{\text{error}} \overbrace{X_i}^{\text{data}}] = 0$$

for $i = 1, 2, 3, \ldots, n$.

$$\Rightarrow \quad E[Y x_i - x_i(a_1 x_1 + \cdots + a_n x_n)] = 0$$

$$\Rightarrow \quad a_1 E(x_i x_1) + a_2 E(x_i x_2) + \cdots + a_n E(x_i x_n) = E(Y x_i)$$

$i = 1, 2, \ldots, n$

n-equations in n-variables.

Writing in a matrix form

$$\underbrace{\begin{bmatrix} E(x_1^2) & E(x_1 x_2) & \cdots & E(x_1 x_n) \\ \vdots & E(x_2^2) & & \vdots \\ \vdots & & \ddots & \\ \cdot & \cdots & & E(x_n^2) \end{bmatrix}}_{C_X} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}}_{a} = \underbrace{\begin{bmatrix} E(Y x_1) \\ E(Y x_2) \\ \vdots \\ E(Y x_n) \end{bmatrix}}_{C_{YX}}$$

$$\Rightarrow \quad C_X a = C_{XY}$$

$$\Rightarrow \quad a = C_X^{-1} C_{XY}$$

$$\Rightarrow \hat{Y}_{LMMSE} = a^T \vec{X}$$

$$= \left(C_X^{-1} C_{XY}\right)^T \vec{X}$$

$(AB)^T = B^T A^T$

$$= C_{XY}^T \left(C_X^{-1}\right)^T \vec{X}$$

$$= C_{XY}^T C_X^{-1} \vec{X}$$

Since C_X is a symmetric, non-singular matrix & its inverse C_X^{-1} is also symmetric, i.e $(C_X^{-1})^T = C_X^{-1}$.

To prove the general result, (non-zero means) apply the above technique to

$$\left(\begin{aligned} \vec{X}' &= (\vec{X} - \vec{\mu}_X) \\ Y' &= (Y - \mu_Y) \end{aligned}\right).$$

$(\vec{X}', Y')$ are zero-mean...