Probability and Random Process

Aimin Tang

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

Oct. 29 2020

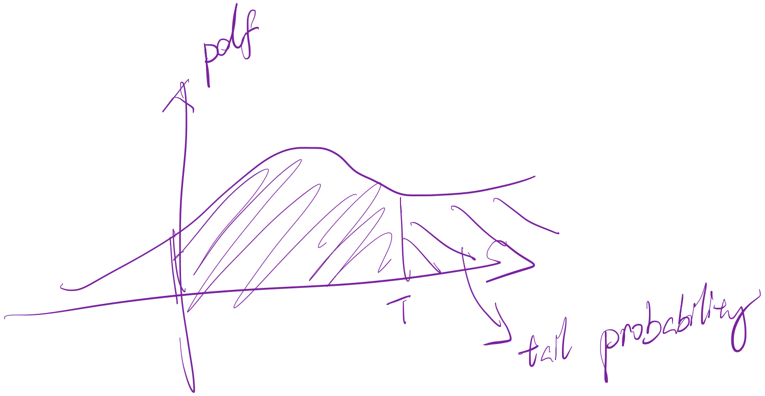
Ve501 2020-2021 Fall

Based on Lecture Notes by Prof. Chong Han

Outline

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

Probability bounds



The Markov and Chebyshev Inequalities

- These inequalities give us loose bounds on certain probabilities and require only mean and variance.

- Markov's Inequality:
  - If X is a random variable that takes only non-negative values, then for any $a>0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

An upper bound on a tail probability

Proof of Markov's Inequality

$$E(X) = \int_0^a x f_X(x)dx + \int_a^\infty x f_X(x)dx$$

$$\geq \int_a^\infty x f_X(x)dx \qquad \text{(Just drop first integral)}$$

$$\geq \int_a^\infty a f_X(x)dx$$

(Because $x \geq a$ over this domain of integration)

$$= aP(X \geq a)$$

So $P(X \geq a) \leq \frac{E(X)}{a}$

Example

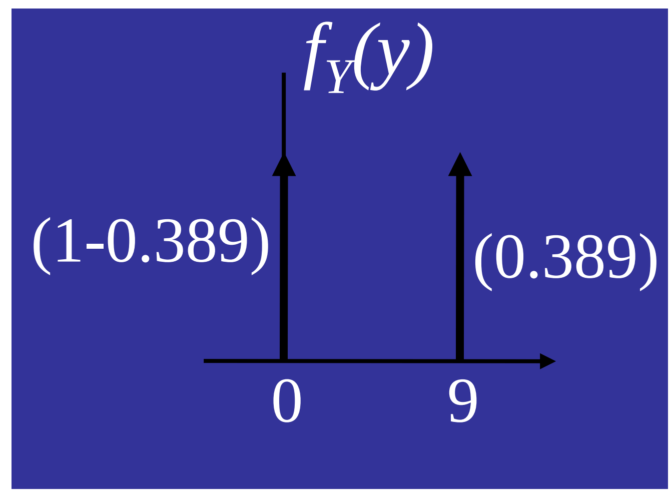
Let X be the height of children in a kindergarten class, and $E(X)=3.5$ *feet*. Find a bound on $P(X \geq 9$ *feet*$)$.

$$a=9 \quad \therefore P\left(X \geq 9\right) \leq \frac{3.5}{9}=0.389$$

very loose

This bound seems ridiculous, but it is possible to construct an RV for which the inequality is exact:

$$E\left(Y\right)=0\left(1-\ 0.389\right)+9\left(0.389\right)=3.5$$



Chebyshev's Inequality

Suppose X is any RV with finite mean μ and variance σ^2
Then for any $b>0$,

$$P\left(|X-\mu| \geq b\right) \leq \frac{\sigma^2}{b^2}$$

An upper bound on a double-tail probability

In words, the probability that X deviates from its mean by more than b is upper-bounded by s^2 / b^2 .

Chebyshev's Inequality Proof

$$P\left(\left|X-\mu\right|\geq b\right)\leq\frac{\sigma^2}{b^2}$$

An upper bound on a double-tail probability

Proof: Apply Markov inequality to $Y=\left[X-\mu\right]^2$ with $a=b^2$.

$$P\left(Y\geq a\right)=P\left(\left[X-\mu\right]^2\geq b^2\right)\leq\frac{E\left[Y\right]}{a}=\frac{\sigma^2}{b^2}$$

Example

The mean response time and the standard deviation in a multi-user computer network are known to be 0.5s and 2s, respectively.

Give an upper bound on the probability that the response time is more than 3s from the mean

$$P\left(\left|X-0.5\right|\geq 3\right)\leq \frac{\sigma^2}{9}=\frac{4}{9}$$

An interesting special case when $b = Ks$

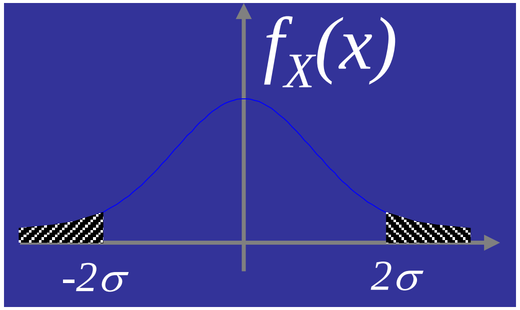
$$P\left(\left|X-\mu\right|\geq K\sigma\right)\leq \frac{\sigma^2}{(K\sigma)^2}=\frac{1}{K^2}$$

Compare to Exact Probability

From the previous slide:

$$P\left(\left|X-\mu\right|\geq K\sigma\right)\leq\frac{1}{K^2}$$

Ex: Suppose $X\sim N\left(\mu,\sigma^2\right),\quad K=2$



$$P\left(\left|X-\mu\right|\geq 2\sigma\right)=P\left(\left|\frac{X-\mu}{\sigma}\right|\geq 2\right)$$

$$=2\left[1-\Phi(2)\right]=0.0456$$

The bound gives 0.25.

The Chebyshev bound can also be quite loose, but it is useful in proving limit theorems.

The Chernoff bound

- The **chernoff bound** of a random variable X is given by

right hand (left side 是否也可用)

$$\Pr(X \geq a) \leq \min_{s \geq 0}[e^{-sa} M_X(s)]$$

$$\Pr(X \leq a) \leq \min_{s \leq 0}[e^{-sa} M_X(s)]$$

- where the minimum is over all s $\geq$ 0 for which $M_X(s)$ is finite.

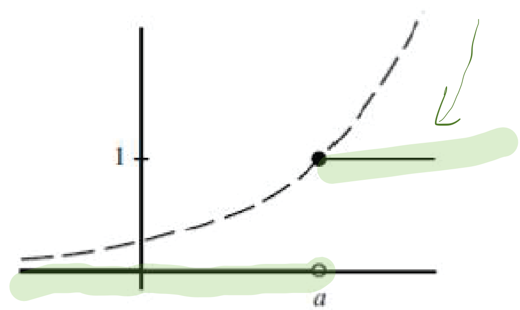
Proof

Every probability can be written as an expectation

$$P_r(X \geq a) = \int_a^{\infty} f_X(x)dx$$

$$= \int_{-\infty}^{\infty} I_{[a,\infty)}(X) f_X(x)dx = \mathrm{E}\left[I_{[a,\infty)}(X)\right] \quad (1)$$

$$I_{[a,\infty)}(X) \leq e^{s(X-a)} \quad (2)$$



(textbook) Figure 4.10 Graph showing that $I_{[a,\infty)}(X)$ (solid line) is upper bounded by $e^{s(X-a)}$ (dashed line) for any positive s. Note that the inequality (2) holds even if $s = 0$.

Proof

Taking expectations of (2)

$$\mathrm{E}\left[I_{[a,\infty)}(X)\right] \le \mathrm{E}\left[e^{s(X-a)}\right]$$

$$= e^{-sa}\mathrm{E}\left[e^{sX}\right] = e^{-sa}M_X(s)$$

Combining with (1),

$$\mathrm{P_r}(X \ge a) \le e^{-sa}M_X(s) \tag{3}$$

Inequality (3) is valid for all $s \ge 0$ *and the LHS of (3) does not depend on s. Consequently*

$$\boxed{\mathrm{P_r}(X \ge a) \le \min_{s \ge 0}\left[e^{-sa}M_X(s)\right]}$$

Short Summary

- The Markov inequality gives an upper bound on "tail probabilities" and applies only to non-negative RVs
- The Chebychev inequality gives an upper bound on "double-tail probabilities" and applies to any RV
- Both can be loose for certain RVs
- The Chernoff bound is usually tighter than the other two
  - For sufficiently large a, the bounds on $\Pr(X \geq a)$ have
    - the Chernoff bound < the Chebyshev bound < the Markov bound

Random Vectors

Random Vectors – Straight Forward Extensions

Random Vector (RVEC)

Straight forward extension of “Two Random Variables”

- Joint CDFs and PDFs
- Calculation of Probability
- Functions of Random Vector
- Independence
- Mean
- Correlation and Covariance Matrix (for real RVs)
- Jointly Gaussian
- Linear Transformations

This course proceeds: RVs → RVECs → Random sequences → Random processes

n-Dimensional Random Vectors

Random vectors (RVECs) are row vectors

$$\mathbf{X} = [X_1, X_2, \ldots, X_n]$$

Most fundamental description: Joint CDF

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\}$$

The joint PDF is the nth-order partial derivative of the CDF

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$$

Calculation of Probabilities

$$P(\mathbf{X} \in D) = \int_D f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

An n-dimensional integral of the PDF of $\mathbf{X}$ over the region D.

Some region in R^n

Vector

Multiple Functions of X

$$Y_1 = G_1(\mathbf{X}) \quad Y_2 = G_2(\mathbf{X}) \quad \cdots \quad Y_m = G_m(\mathbf{X})$$

Same procedure as before to get PDF of $\mathbf{Y}$ where

$$\mathbf{Y} = [Y_1, Y_2, \ldots, Y_m]$$

CDF approach:

Find $F_{\boldsymbol{Y}}(\boldsymbol{y})$, then differentiate to get $f_{\boldsymbol{Y}}(\boldsymbol{y})$

$$F_{\boldsymbol{Y}}(\boldsymbol{y}) = \Pr(Y_1 \le y_1 \cap \cdots \cap Y_m \le y_m)$$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\partial^m F_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_1 \ldots \partial y_m}$$

Jacobian Method

auxillary

1. Must have m=n (can use aux variables)
2. Find solutions $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_M$
3. Get Jacobian:

$$J(\mathbf{x}) = \det\begin{bmatrix} \frac{\partial G_1}{\partial x_1} & \frac{\partial G_1}{\partial x_2} & \cdots & \frac{\partial G_1}{\partial x_n} \\ \frac{\partial G_2}{\partial x_1} & \ddots & & \frac{\partial G_2}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial G_n}{\partial x_1} & & \cdots & \frac{\partial G_n}{\partial x_n} \end{bmatrix}$$

4. Plug into formula: $$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^{M} \frac{f_{\mathbf{X}}(\mathbf{x_i})}{|J(\mathbf{x_i})|}$$

Independent, Identically Distributed (iid) RVs

Elements of $\mathbf{X}$ are independent if CDF or PDF factors:

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} F_{X_i}(x_i) \qquad f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

A collection of RVs is called "iid" when they are independent and identically distributed.

Identically distributed means:

$$F_{X_i}(x) = F_X(x) \quad \forall i$$

X_i都与X
同分布

Same function

Mean Vector

$$\boldsymbol{\eta}_{\mathbf{X}} = E(\mathbf{X}) = [E(X_1), E(X_2), \ldots, E(X_n)]$$

Correlation matrix

X = row vector. 1xn.

$$\mathbf{R} = E\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) = E\left\{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}\begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}\right\}$$

correlation Cor[X₁, X₂]

$$= \begin{bmatrix} E\left(X_1^2\right) & E\left(X_1 X_2\right) & \cdots & E\left(X_1 X_n\right) \\ E\left(X_2 X_1\right) & E\left(X_2^2\right) & & E\left(X_2 X_n\right) \\ \vdots & & \ddots & \vdots \\ E\left(X_n X_1\right) & & \cdots & E\left(X_n^2\right) \end{bmatrix}$$

Cor[X₂]

Covariance matrix

$$\mathbf{C} = E\left\{\left[\mathbf{X} - \boldsymbol{\eta}_{\mathbf{X}}\right]^{\mathbf{T}}\left[\mathbf{X} - \boldsymbol{\eta}_{\mathbf{X}}\right]\right\}$$

$$= \begin{bmatrix} \sigma_{X_1}^2 & \operatorname{cov}(X_1 X_2) & \cdots & \operatorname{cov}(X_1 X_n) \\ \operatorname{cov}(X_2 X_1) & \sigma_{X_2}^2 & & \operatorname{cov}(X_2 X_n) \\ \vdots & & \ddots & \vdots \\ \operatorname{cov}(X_n X_1) & & \cdots & \sigma_{X_n}^2 \end{bmatrix}$$

$$= \mathbf{R} - \boldsymbol{\eta}_{\mathbf{X}}^{\mathbf{T}} \boldsymbol{\eta}_{\mathbf{X}}$$

$C = C^T$, symmetric

Expected Value of a Function

If Y=G(**X**) is a scalar-valued function, then

LOTUS

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} G(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

If **Y** is a vector-valued function of **X**, then

$$E(\mathbf{Y}) = [E(Y_1), E(Y_2), \ldots, E(Y_m)], \quad Y_i = G_i(\mathbf{X})$$

EACH ONE LOTUS

n Jointly Gaussian RVs

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=\frac{\exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\eta}_{\mathbf{X}})\mathbf{C}^{-1}(\mathbf{X}-\boldsymbol{\eta}_{\mathbf{X}})^{\mathbf{T}}\right\}}{(2\pi)^{n/2}\sqrt{\det\mathbf{C}}}$$

Linear Transformation of Jointly Gaussian RVs

$\underset{\uparrow}{}^{1\times n}\quad \underset{\uparrow}{}^{1\times n}\quad \underset{\uparrow}{}^{n\times n}$

Given that $\mathbf{Y} = \mathbf{X}\mathbf{A}$, $\mathbf{A}^{-1}$ exists, and $\mathbf{A} : n \times n$

若 X_i 全 $\sim N$

$Y_i \sim N \Longleftarrow$

$$J = \det \begin{bmatrix} \frac{\partial Y_1}{\partial x_1} & \cdots & \frac{\partial Y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_n}{\partial x_1} & \cdots & \frac{\partial Y_n}{\partial x_n} \end{bmatrix} = \det\left[\mathbf{A}\right]$$

$$\eta_{\mathbf{Y}} = \eta_{\mathbf{X}}\mathbf{A}$$

Linear Transformation of Jointly Gaussian RVs – Cont.

Cov of Y

$$\mathbf{C_Y} = E\left\{ \left[\mathbf{Y} - \eta_\mathbf{Y}\right]^\mathbf{T} \left[\mathbf{Y} - \eta_\mathbf{Y}\right] \right\}$$

$$= E\left\{ \left(\left[\mathbf{X} - \eta_\mathbf{X}\right]\mathbf{A}\right)^\mathbf{T} \left[\mathbf{X} - \eta_\mathbf{X}\right]\mathbf{A} \right\}$$

$$= \mathbf{A}^T \mathbf{C_X} \mathbf{A}$$

Cov of X

PDF of Y When Y=XA

1. There is just one solution: $\mathbf{X} = \mathbf{YA}^{-1}$

2. $$f_{\mathbf{Y}}(\mathbf{y}) = \frac{f_{\mathbf{X}}\left(\mathbf{YA}^{-1}\right)}{\left|\det \mathbf{A}\right|} = \frac{\exp\left\{-\frac{1}{2}\left(\mathbf{YA}^{-1} - h_{\mathbf{X}}\right)\mathbf{C}_{\mathbf{X}}^{-1}\left(\mathbf{YA}^{-1} - h_{\mathbf{X}}\right)^{\mathbf{T}}\right\}}{(2p)^{n/2}\sqrt{\det \mathbf{C}_{\mathbf{X}}}\left|\det \mathbf{A}\right|}$$

Note that $\boldsymbol{\eta}_{\mathbf{X}} = \boldsymbol{\eta}_{\mathbf{Y}}\mathbf{A}^{-1}$

$$\mathbf{A}^{-1}\mathbf{C}_{\mathbf{X}}^{-1}\left(\mathbf{A}^{-1}\right)^{\mathbf{T}} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{C}_{\mathbf{X}}\mathbf{A}\right)^{-1} = \mathbf{C}_{\mathbf{Y}}^{-1}$$

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\exp\left\{-\frac{1}{2}\left(\mathbf{Y} - \boldsymbol{\eta}_{\mathbf{Y}}\right)\boxed{\mathbf{A}^{-1}\mathbf{C}_{\mathbf{X}}^{-1}\left(\mathbf{A}^{-1}\right)^{\mathbf{T}}}\left(\mathbf{Y} - \boldsymbol{\eta}_{\mathbf{Y}}\right)^{\mathbf{T}}\right\}}{(2\pi)^{n/2}\sqrt{\det \mathbf{C}_{\mathbf{X}}}\left|\det \mathbf{A}\right|}$$

if $X \sim N$

Any linear transformation of X, $\sim N$

$Y = XA \sim N$

Random Vectors – MS Estimation

Recall the optimal linear mean square (MS) homogeneous estimate of the RV Y given an observation of the RV X,

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

where

$$a = \frac{E(XY)}{E(X^2)}$$

Now, we will consider estimating a RV Y from a row vector of observations (Rvec) $\mathbf{X}$

Estimator Form

The new estimator has the form

$$\hat{Y} = \mathbf{X}\mathbf{A}^T$$
$$= a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

$$\mathbf{X} = [X_1, X_2, \ldots, X_n] \quad \mathbf{A} = [a_1, a_2, \ldots, a_n]$$

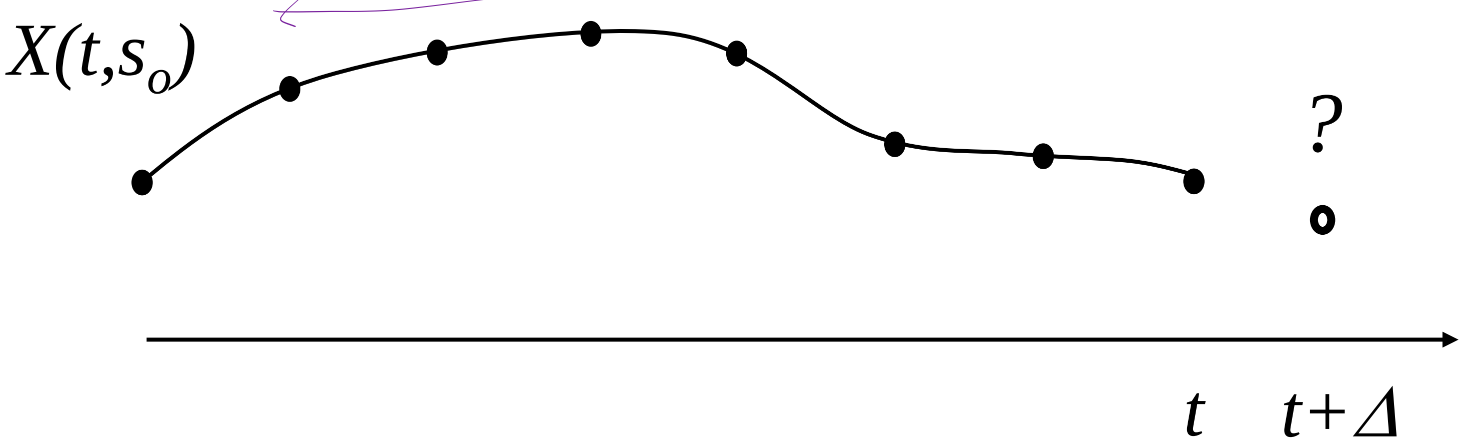
The observation data — The estimator coefficients

Applications

Prediction

A future value of a RP, $X(t+D)$ is estimated from the present and past measurements of the RP, $[X(t), X(t\text{-}D), \ldots, X(t\text{-}nD)]$

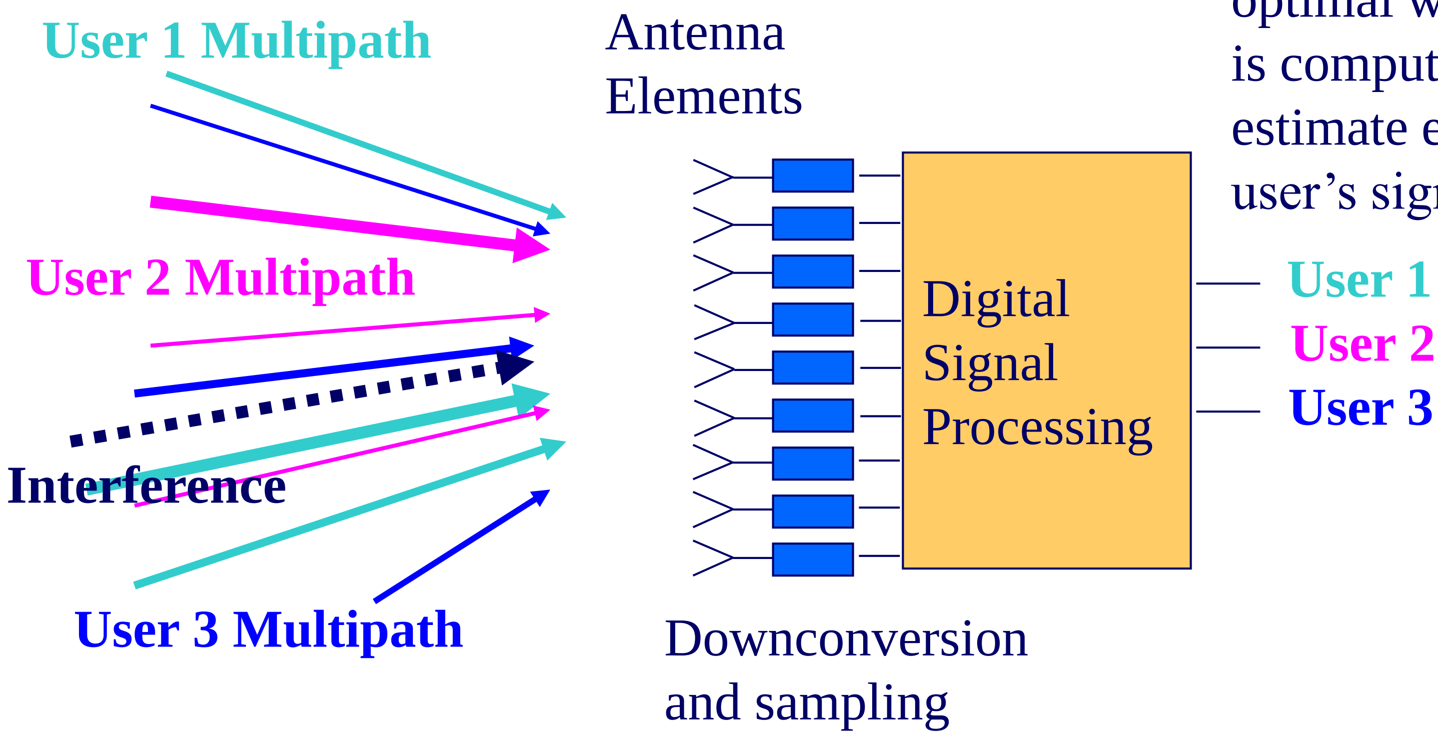
weight



Application: Multiuser Detecting Array Receiver

The observation vector comprises samples of the baseband signals at the outputs of the antenna elements

A different MSE-optimal weight vector is computed to estimate each desired user's signal



Hilbert Space

- Y and the elements of the vector X "span" a vector space
  - Space=all possible linear combinations of the elements of X and Y
  - A "point" in this space is a some linear combination

$$Z = b_0 Y + \sum_{i=1}^{n} b_i X_i$$

- A vector space with an inner product is a Hilbert space
  - For our Hilbert space, the inner product is the correlation
  - If Z and U are two points in this space, their inner product is E(ZU)

- The optimal MS estimator for Y is a point in the subspace spanned by the elements of X

MSE

Let U be some linear combination of the elements of X

$$U = \sum_{i=1}^{n} a_i X_i$$

Let $\mathbf{A} = [a_1, \ldots, a_n]$

To be the optimal estimator for Y, A must minimize the mean squared error (MSE)

$$E(\varepsilon^2) = E\left(\left[Y - \mathbf{X}\mathbf{A}^T\right]^2\right)$$

with respect to each element of A

Optimal Estimator

Setting each of the n partial derivatives

$$\frac{\partial E(\varepsilon^2)}{\partial a_i} \quad \text{for } i = 1,2,\ldots,n$$

equal to zero and solving for A yields the equation

$$\mathbf{A} = r_{XY}\mathbf{R}^{-1}$$

where R is the correlation matrix for $\mathbf{X}$ and r_{XY} is the cross correlation vector, $r_{XY} = E(\mathbf{X}Y)$

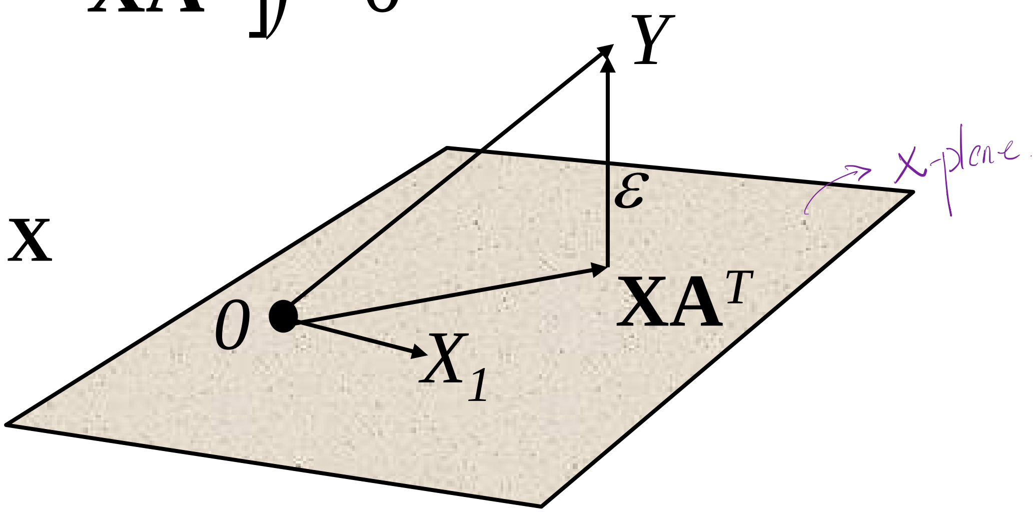
vector scalar

Orthogonality Principle

The partial derivative equations lead directly to the fact that the "data" are orthogonal to the "error"

$$E(\mathbf{X}\varepsilon) = E\left(\mathbf{X}\left[Y - \mathbf{X}\mathbf{A}^T\right]\right) = 0$$

Space spanned by **X**

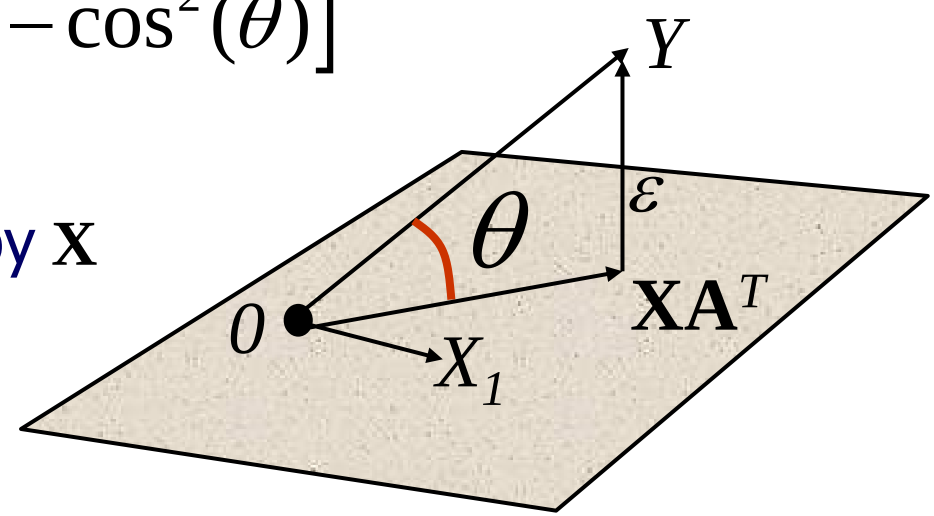


Performance

The MSE of the optimal estimator is the measure of its performance

$$MSE_{opt} = E(Y^2) - r_{XY} R^{-1} r_{XY}^T$$
$$= E(Y^2)\left[1 - \cos^2(\theta)\right]$$

Space spanned by **X**



Short Summary

The extension from two to n RVs is straightforward

A linear transformation on a Gaussian RVEC is another Gaussian RVEC

The optimal MSE estimator of a RV Y given observations of a Rvec X depends on

- The cross-correlation between X and Y
- The correlation matrix for X

Requires Only Second Order Statistics!

Thank You!