ELE 661 - ESTIMATION THEORY

INSTRUCTOR: STEVEN KAY

TEXT: FUNDAMENTALS OF STATISTICAL SIGNAL PROCESSING - ESTIMATION
S. KAY

REFERENCES: 1) ADVANCED THEORY OF STATISTICS VOL 2, KENDALL & STUART
2) PARAMETER ESTIMATION, SORENSON

COURSE OUTLINE: SEE TABLE OF CONTENTS

CLASSICAL { MLE, CRLB, LEAST SQUARES, SUFFICIENT STATISTICS, MVU, BLUE, METHOD OF MOMENTS

BAYESIAN { MMSE, MAP, KALMAN FILTERING

PREREQUISITES: 1) ELE 509 (SYSTEMS WITH RANDOM INPUTS)
2) LINEAR ALGEBRA + MATRIX THEORY

GRADING : 1) HOMEWORK ASSIGNED WEEKLY
DUE FOLLOWING WEEK - NOT
GRADED BUT REASONABLE
EFFORT <u>REQUIRED</u>

2) PROJECT (80%)
a) LITERATURE SURVEY
b) ANALYSIS (ORIGINAL RESEARCH)
c) COMPUTER SIMULATION

PROPOSAL, ORAL PRESENTATION AND
FINAL WRITTEN REPORT REQUIRED

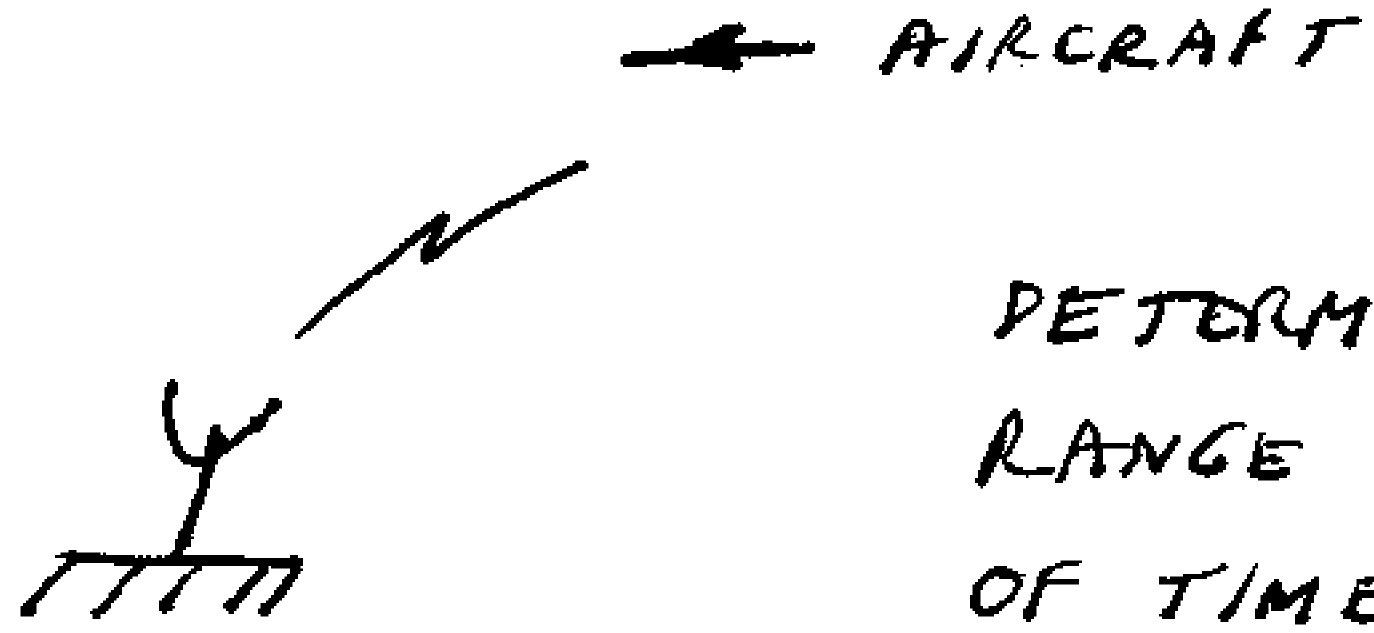
3) DATA ANALYSIS (20%)

OFFICE HOURS : M, W, F 9-11 874-5804
KELLEY ANNEX A123

INTRODUCTION

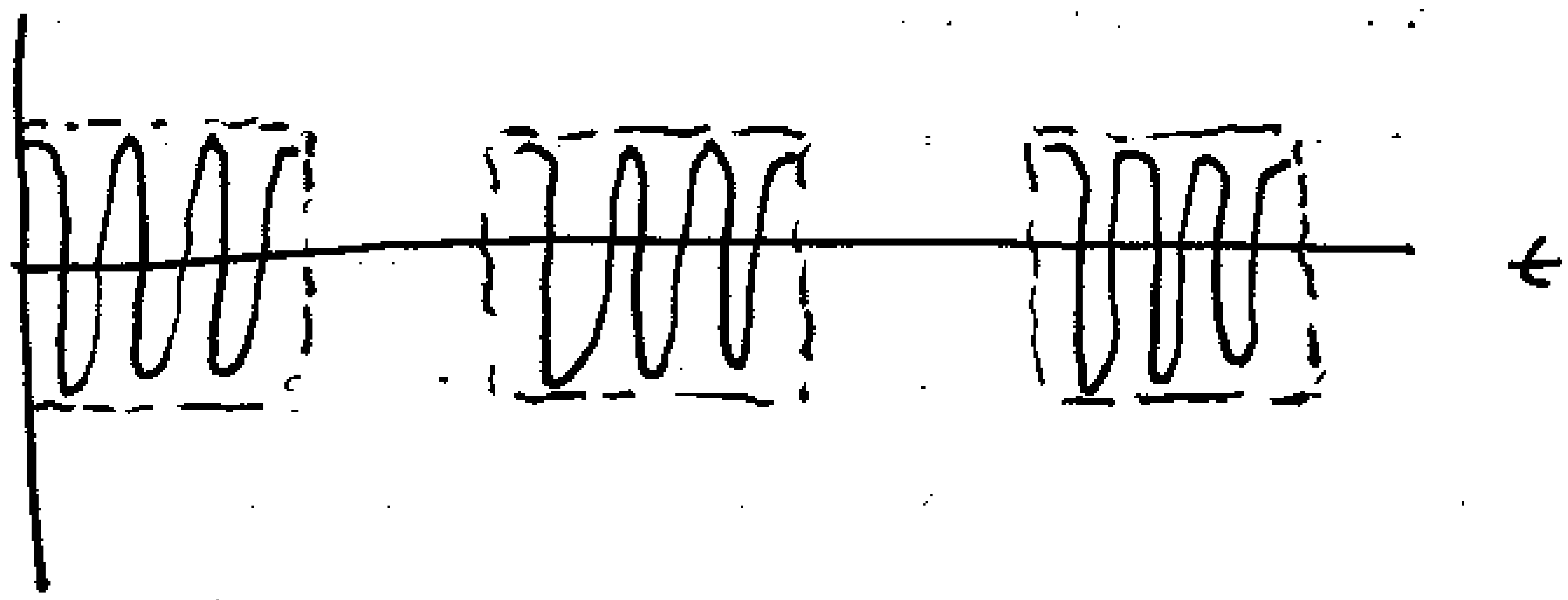
AREAS OF APPLICATION

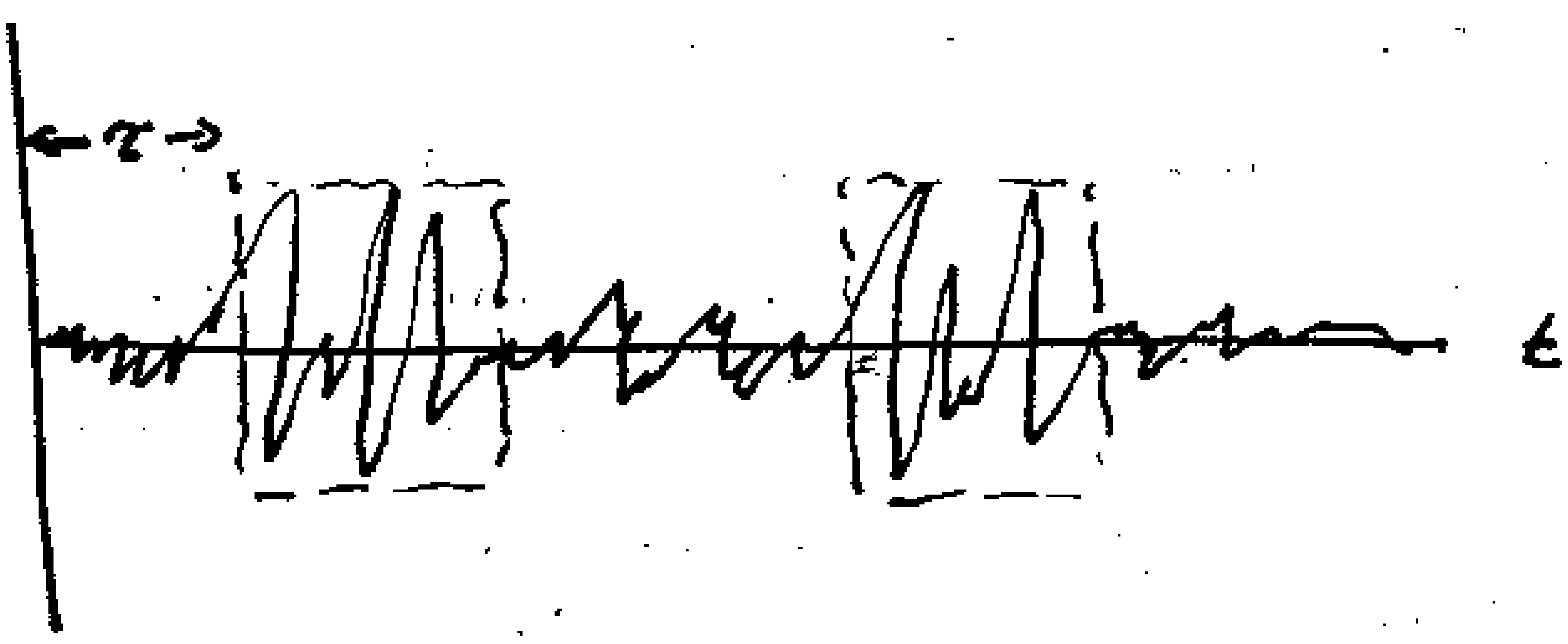
1) RADAR (AIRPORT SURVEILLANCE RADAR (ASR))



DETERMINE AIRCRAFT
RANGE AS A FUNCTION
OF TIME

METHOD: TRANSMIT SERIES OF PULSES, ANALYZE RETURNS





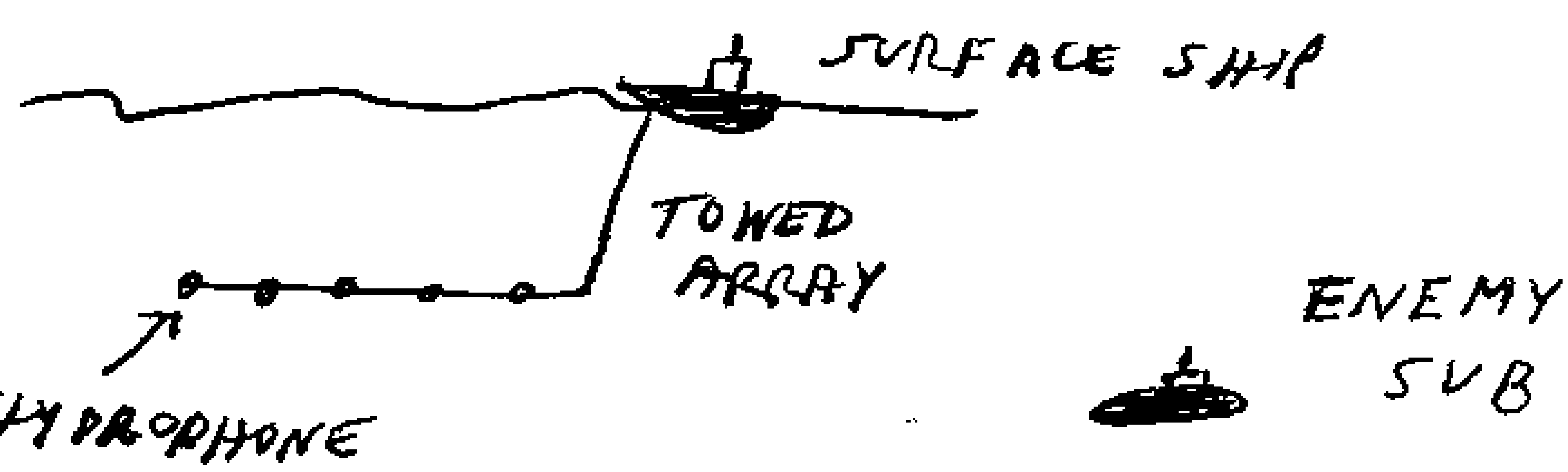
WISH TO FIND ROUND TRIP DELAY $= \tau = \frac{2R}{c}$

RANGE

SPEED OF LIGHT

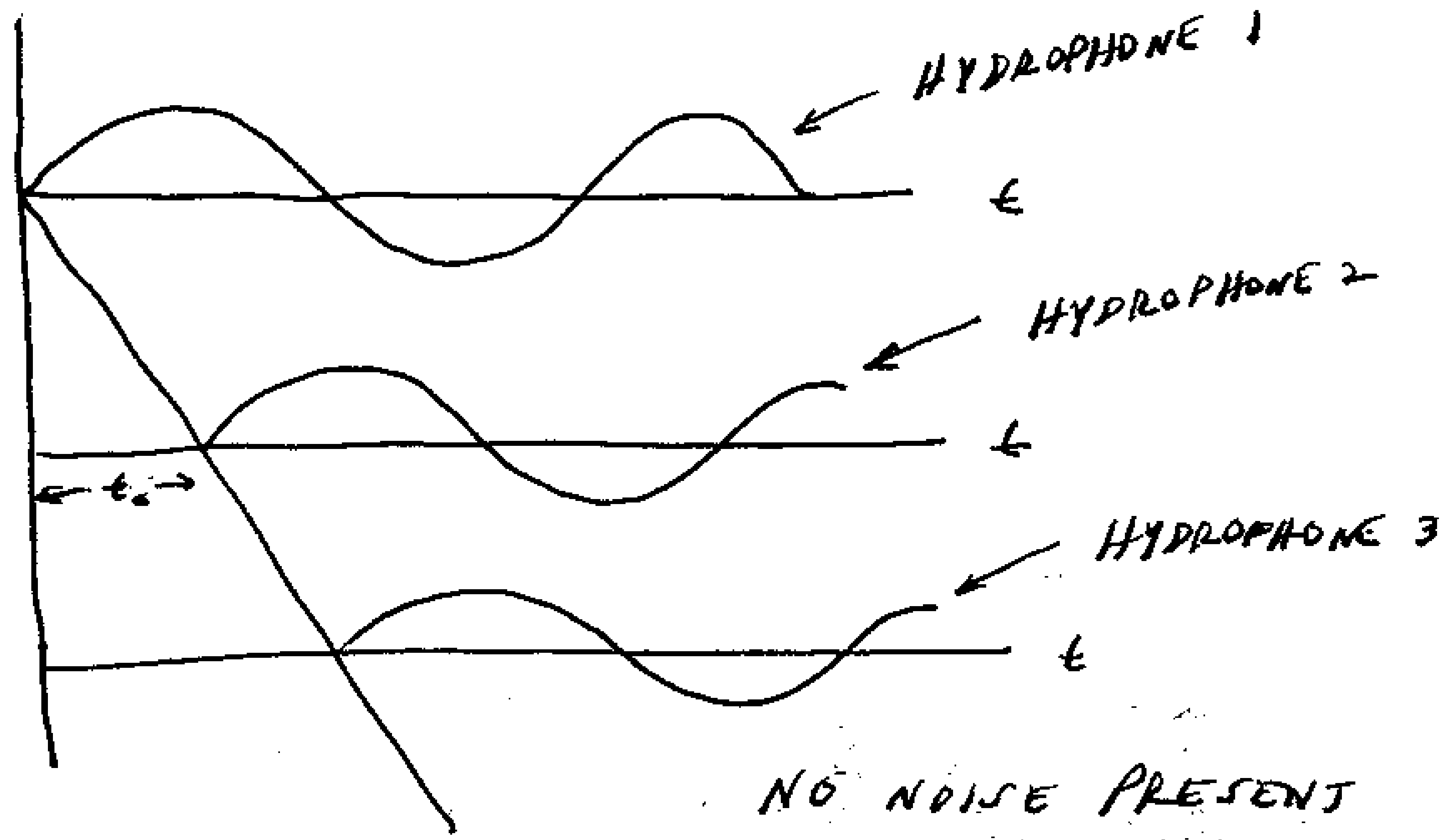
MUST <u>ESTIMATE</u> τ SINCE EACH RECEIVED PULSE IMBEDDED IN NOISE (AMBIENT, CLUTTER, ELECTRONIC)

2/ SONAR



WISH TO DETERMINE BEARING OF TARGET

METHOD: EXAMINE SIGNALS AT OUTPUT OF HYDROPHONE ARRAY



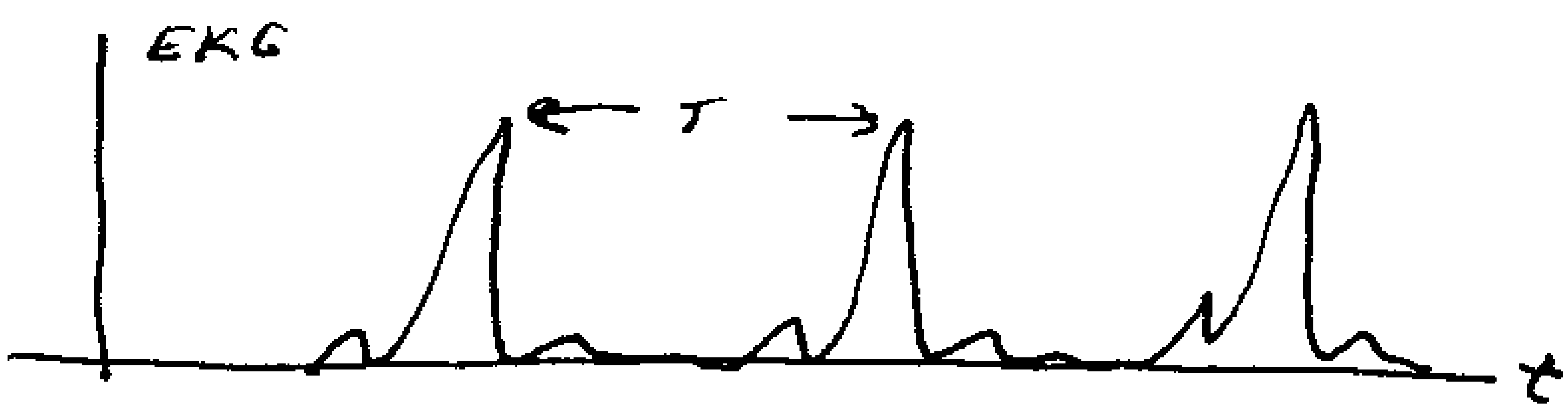
NO NOISE PRESENT

t_0 CAN BE USED TO FIND BEARING

SIGNALS WILL BE IMBEDDED IN NOISE

(AMBIENT, FLOW, ELECTRONIC, SHIPPING, ETC)

3) BIOMEDICAL



WISH TO DETERMINE HEART RATE ($1/T$).
SIGNAL CORRUPTED BY ELECTRONIC NOISE,
PATIENT MOVING, SENSOR NOISE, ETC.

4) COMMUNICATIONS
ESTIMATE NOISE LEVEL ON TELEPHONE
CHANNEL TO SET AMPLIFIER GAINS

5) ECONOMICS
ESTIMATE (GUESS) DOW JONES AVERAGE

6) RELIABILITY
ESTIMATE MEAN TIME BETWEEN FAILURES

7) ACTUARIAL
ESTIMATE AVERAGE LIFE SPAN OF
MALE IN USA

GENERIC EXAMPLE

DAMPED SINUSOID $s[n]$ OBSERVED BUT
CORRUPTED BY ADDITIVE NOISE $w[n]$.

$$x[n] = s[n] + w[n] \qquad n = 0, 1, \ldots, N-1$$

SINUSOID HAS UNKNOWN PARAMETERS

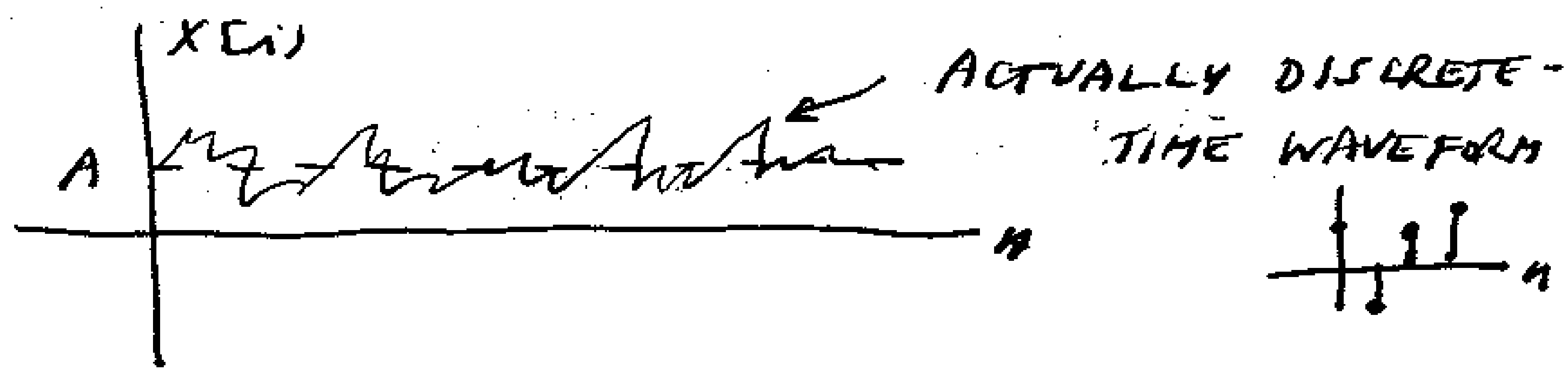
$$s[n] = A r^n \cos(2\pi f_0 n + \phi)$$

ESTIMATE $\{A, r, f_0, \phi\}$ OR SUBSET OF THESE. $w[n]$ IS ZERO MEAN WGN WITH VARIANCE σ^2.

EXAMPLE : $x[n] = A + w[n]$

"DC" LEVEL

$w[n]$ IS WHITE GAUSSIAN NOISE (WGN)



HOW DO WE ESTIMATE A?

INTUITIVELY, A IS AVERAGE LEVEL OF $x[n]$ ($w[n]$ IS ZERO MEAN)

REASONABLE ESTIMATOR :

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

DENOTES ESTIMATOR

ESTIMATOR DEPENDS ONLY ON OBSERVED DATA. CAN BE VIEWED AS REAL FUNCTION OF OBSERVATIONS OR $\hat{A} = g(x[0], x[1], \ldots, x[N-1])$

HOW ABOUT $\hat{A} = x[0]$?

ANY OTHER POSSIBLE ESTIMATORS?
WHICH IS THE BEST ONE?

EXPERIMENT 1

$N = 100 \quad A = 1$
MEASURE $\hat{A} = 0.9 \quad \hat{\hat{A}} = 0.95$

$\Rightarrow \hat{\hat{A}}$ IS BETTER?

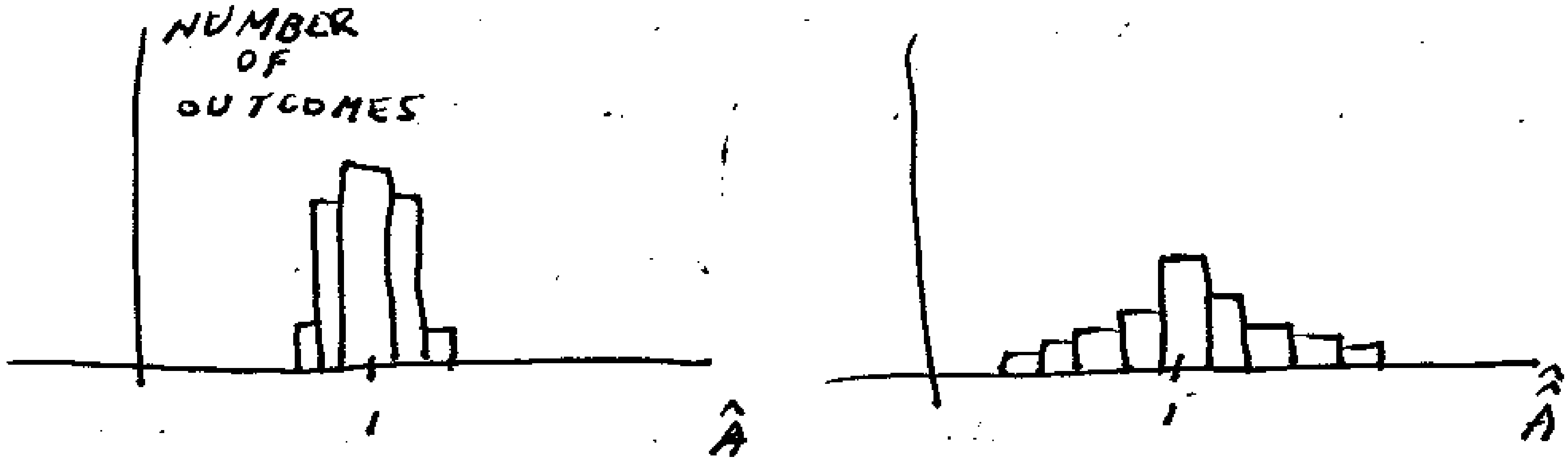
ESTIMATOR IS RANDOM VARIABLE (FUNCTION OF RANDOM VARIABLES) $\Rightarrow$ EACH OUTCOME IS DIFFERENT $\Rightarrow$ NEED TO DESCRIBE STATISTICALLY

EXPERIMENT 2

$N = 100 \quad A = 1$

REPEAT EXPERIMENT 1000 TIMES $\Rightarrow$ PLOT

HISTOGRAM OF RESULTS



BETTER ESTIMATOR? $\hat{A}$ BECAUSE PDF MORE CONCENTRATED ABOUT TRUE VALUE OR VARIANCE IS LESS. "ON THE AVERAGE $\hat{A}$ WILL PRODUCE VALUE CLOSER TO $A=1$".

"PROOF": COMPUTE MEAN AND VARIANCE OF ESTIMATORS

$$E(\hat{A}) = E\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N}\sum_{n=0}^{N-1} E(x[n]) = A$$

$$E(\hat{\hat{A}}) = E(x[0]) = A$$

$$\mathrm{VAR}(\hat{A}) = \mathrm{VAR}\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N^2}\sum_{n=0}^{N-1} \mathrm{VAR}(x[n]) = \sigma^2/N$$

$$\mathrm{VAR}(\hat{\hat{A}}) = \mathrm{VAR}(x[0]) = \sigma^2 > \mathrm{VAR}(\hat{A})$$

$\therefore \quad \hat{A}$ IS PREFERRED.

(PROBABILITY OF $\hat{A}$ BEING CLOSER TO A IS HIGHER — SEE HOMEWORK (HW?))

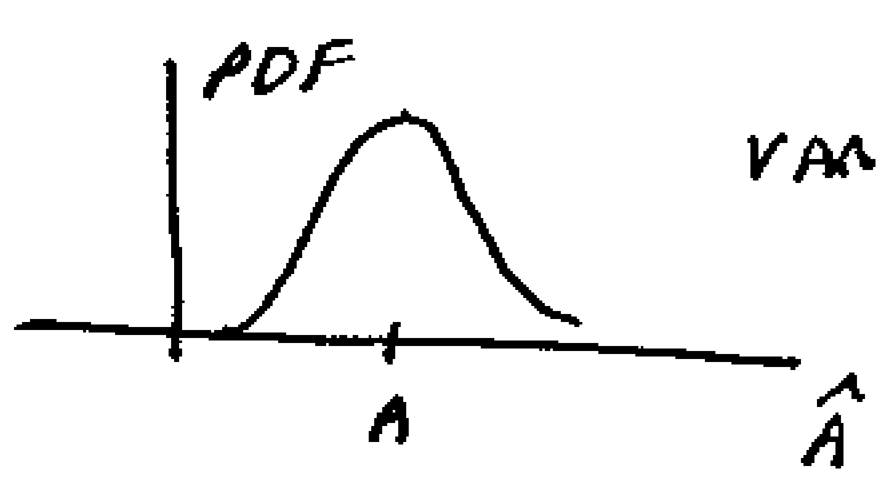
QUESTIONS:

1) GUESSED AT FORM OF ESTIMATOR — NOT POSSIBLE IN GENERAL — NEED A "TURN THE CRANK" METHOD.

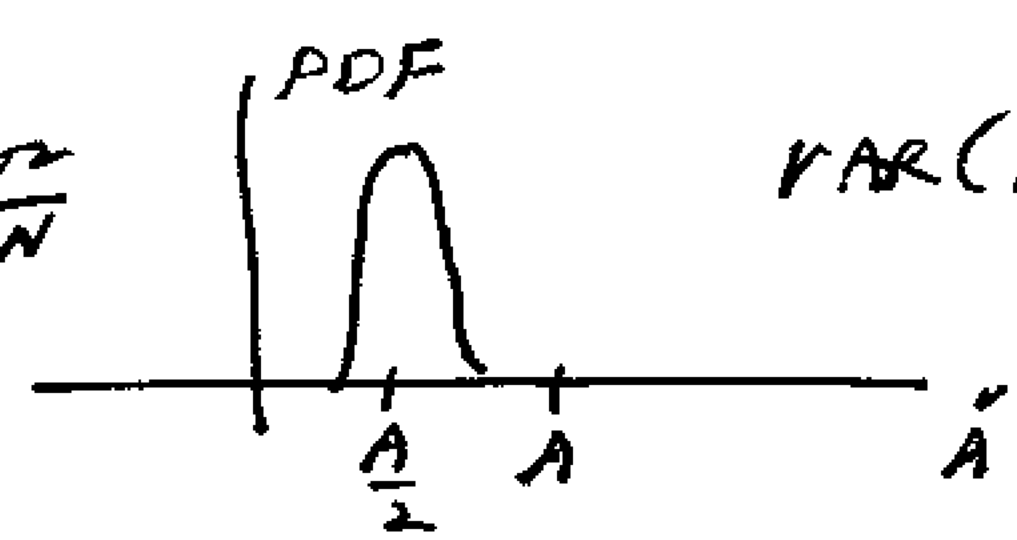
2) DOES AN OPTIMAL ESTIMATOR EXIST? ONE WITH MINIMUM VARIANCE?

IN PREVIOUS EXAMPLE $E(\hat{A}) = A$. WHAT HAPPENS IF

$$\check{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x[n] \; ?$$

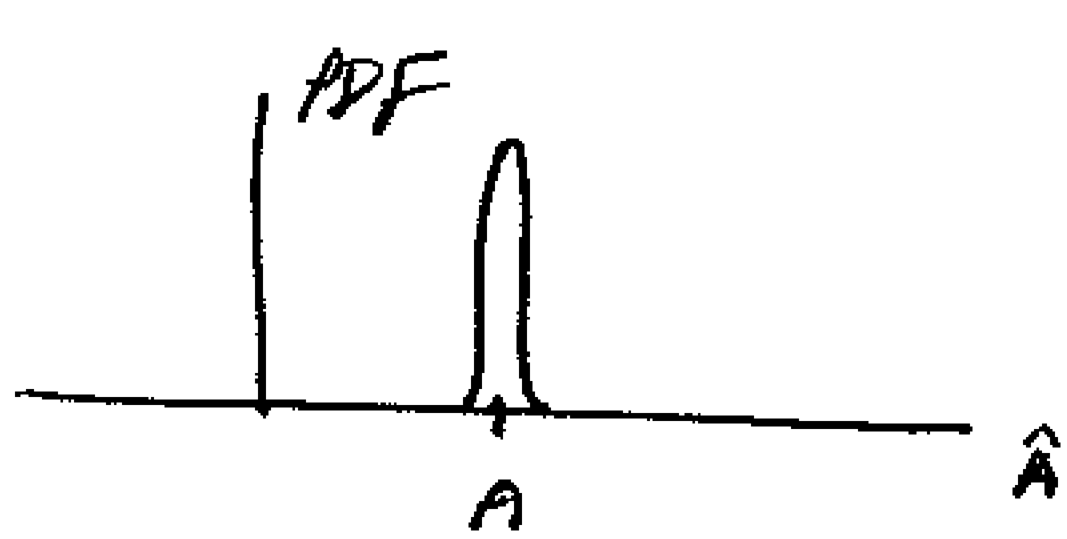


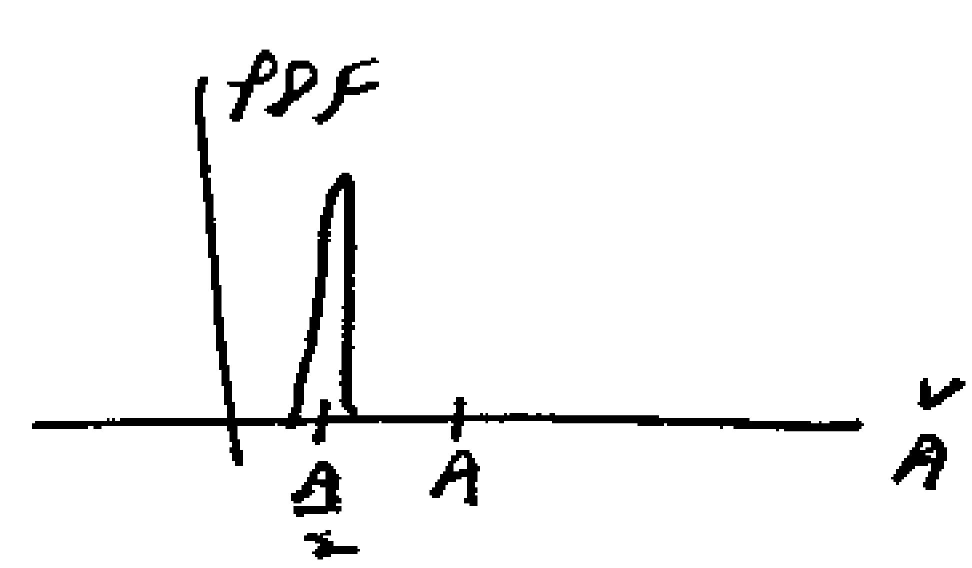
$\mathrm{VAR}(\hat{A}) = \frac{\sigma^2}{N}$



$\mathrm{VAR}(\check{A}) = \frac{1}{4}\frac{\sigma^2}{N}$

AS N INCREASES





IF WE ATTEMPT TO MINIMIZE VARIANCE, MUST CONSTRAIN ESTIMATOR TO HAVE $E(\hat{A}) = A \Rightarrow$ UNBIASED ESTIMATOR.

(IF $\hat{\hat{A}} = 0$, WHAT IS VARIANCE?)

CHAPTER 2 - MINIMUM VARIANCE UNBIASED ESTIMATION

UNBIASED - ON THE AVERAGE ESTIMATOR WILL YIELD TRUE VALUE OR

$$E(\hat{\theta}) = \underset{\substack{\uparrow \\ \text{TRUE VALUE}}}{\theta} \qquad \text{FOR ALL } \theta$$

EXAMPLE : DC LEVEL IN WGN

$$x[n] = A + \underset{\substack{\uparrow \\ \text{WGN}}}{w[n]} \qquad n = 0, 1, \ldots, N-1$$

A CAN TAKE ON VALUES $-\infty < A < \infty$

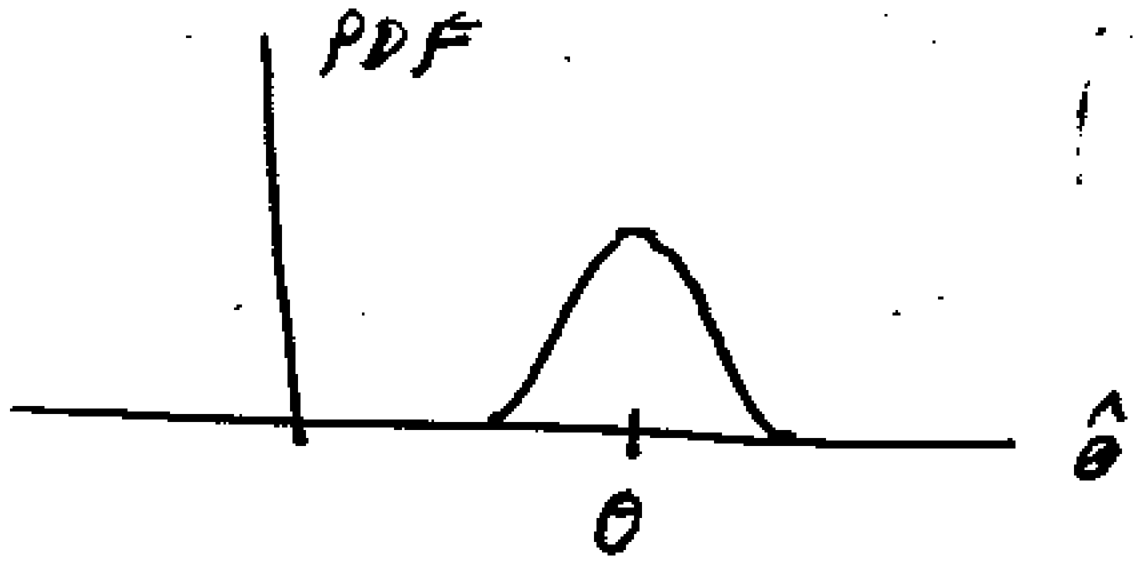
$$\hat{A} = 1/N \sum_{n=0}^{N-1} x[n]$$

$$E(\hat{A}) = 1/N \sum_{n=0}^{N-1} E(x[n])$$

$$= A \qquad \text{FOR ALL } A$$

$\Rightarrow$ UNBIASED

PDF OF UNBIASED $\hat{\theta}$ TENDS TO BE SYMMETRIC OR



NOT NECESSARY THOUGH. (EXAMPLE?)

TO BE UNBIASED $E(\hat{\theta}) = \theta$ FOR ALL θ

$$E(\hat{\theta}) = \int \underbrace{g(\underline{x})}_{\hat{\theta}} \underbrace{p(\underline{x};\theta)}_{\text{PDF OF DATA}} d\underline{x} = \theta$$

($\underline{x}$ DENOTES $N \times 1$ VECTOR $[x[0]\ x[1] \ldots x[N-1]]^T$.

$N \to$ EXAMPLE : SAME EXAMPLE

$$\check{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x[n]$$

UNBIASED ?

$$E(\check{A}) = \frac{1}{2} A$$
$$= A \qquad A = 0$$
$$\neq A \qquad A \neq 0$$

$\Rightarrow$ BIASED

BIAS OF ESTIMATOR $= E(\hat{\theta}) - \theta = b(\theta)$

GENERALLY SEEK UNBIASED ESTIMATORS (NECESSARY BUT NOT SUFFICIENT FOR GOOD ESTIMATOR)

MINIMUM VARIANCE CRITERION

NEED OPTIMALITY CRITERION TO ASSESS PERFORMANCE.

MEAN SQUARE ERROR (MSE) IS

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

GOOD ESTIMATOR HAS SMALL MSE OR AVERAGE SQUARED ERROR SMALL.

USE OF THIS LEADS TO <u>UNREALIZABLE</u> ESTIMATORS.

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

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

$$+ 2E\left[\underbrace{(\hat{\theta}-E(\hat{\theta}))}_{=0}\underbrace{(E(\hat{\theta})-\theta)}_{\text{CONSTANT}}\right]$$

$$= VAR(\hat{\theta}) + b(\theta)^2$$

= STRONG FUNCTION OF θ.

DIFFERENTIATION WILL PRODUCE A $\hat{\theta}$ DEPENDENT ON θ.

<u>EXAMPLE</u> : $\check{A} = a\bar{x}$ $\quad \bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

FIND BEST a.

$$E(\check{A}) = aA$$

$$VAR(\check{A}) = a^2 \sigma^2/N$$

$$MSE(\check{A}) = a^2\sigma^2/N + (aA - A)^2$$

$$\frac{dMSE}{da} = 2a\sigma^2/N + 2(aA-A)A = 0$$

$$\Rightarrow a_{OPT} = \frac{A^2}{A^2 + \sigma^2/N}$$

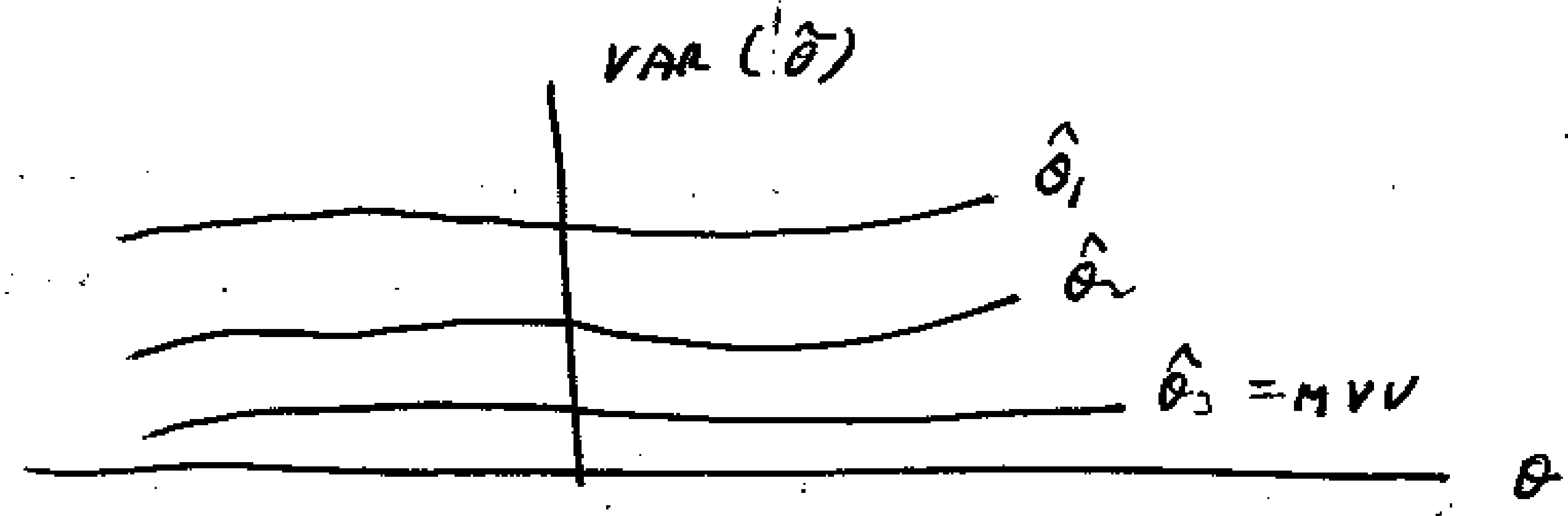
DEPENDS ON A.

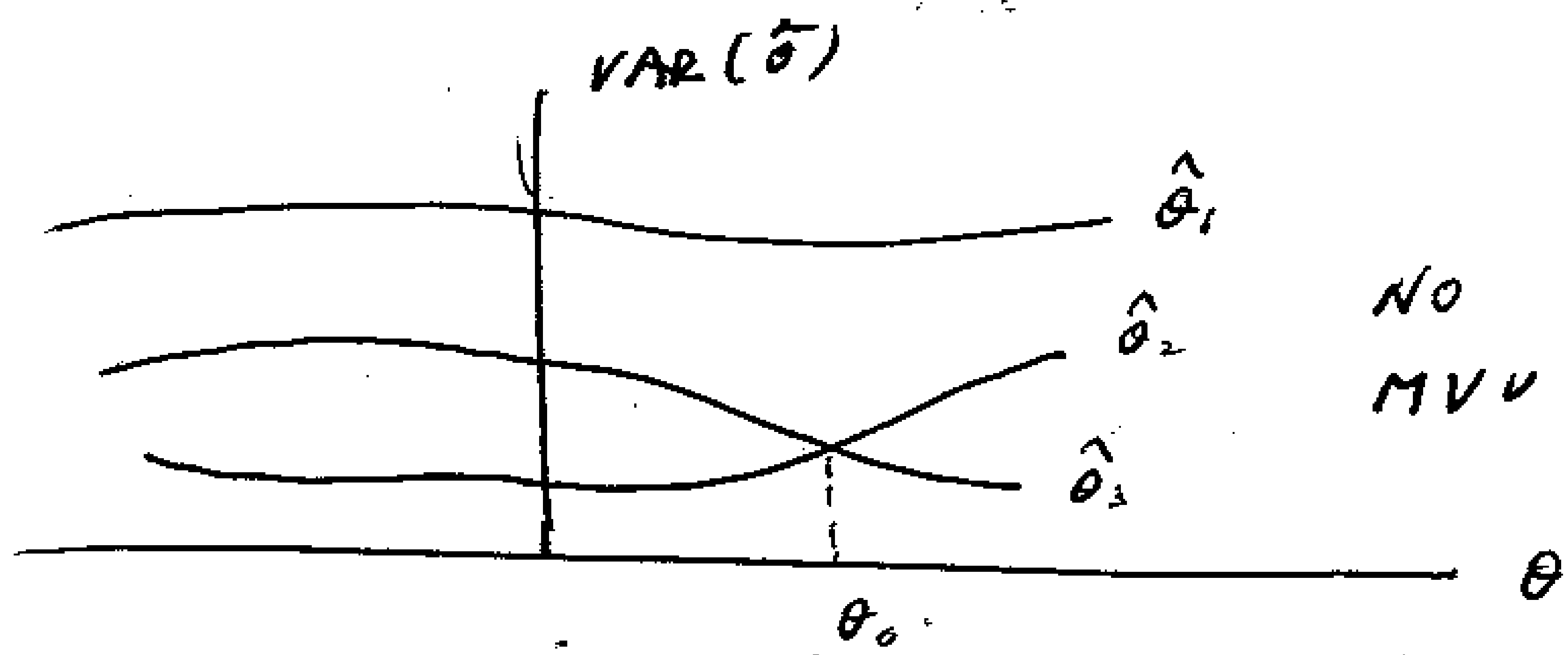
GENERALLY CONSTRAIN BIAS TO BE ZERO AND THEN MINIMIZE VARIANCE

$\Rightarrow$ MINIMUM VARIANCE UNBIASED (MVU) ESTIMATOR

EXISTENCE OF MVU ESTIMATOR

DOES MVU EXIST?





NO MVU

$\hat{\theta}$ MUST HAVE SMALLEST VARIANCE FOR ALL VALUES OF θ. ($\therefore$ UNIFORMLY MVU ESTIMATOR)

IN GENERAL, WILL NOT EXIST.

EXAMPLE :

$$x[0] \sim \underset{\text{NORMAL}}{N}(\underset{\text{MEAN}}{\theta}, \underset{\text{VARIANCE}}{1})$$

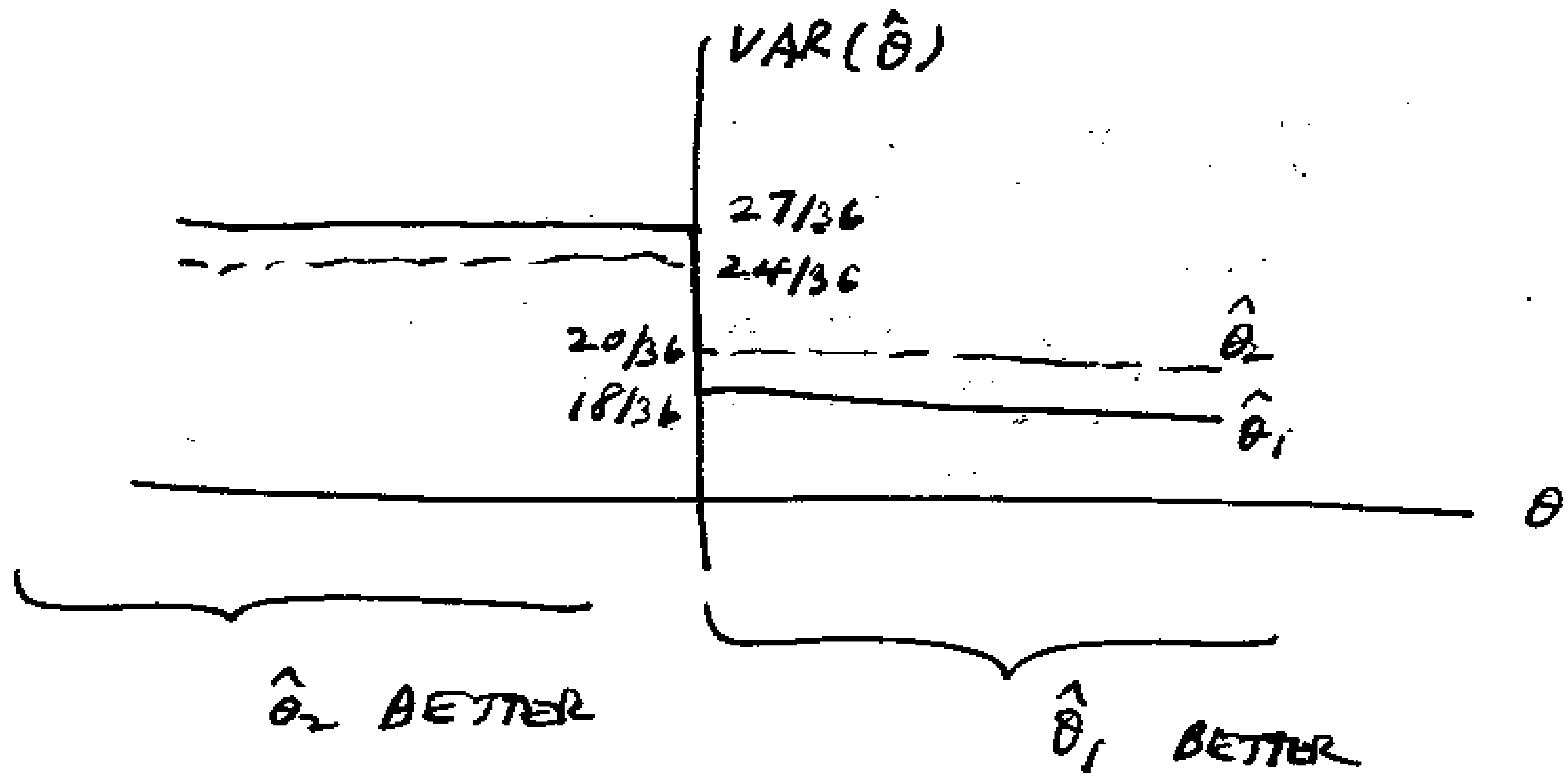
$$\left.\begin{array}{ll} x[0] \sim N(\theta, 1) & \\ x[1] \sim N(\theta, 1) & \theta \geq 0 \\ \qquad\;\; N(\theta, 2) & \theta < 0 \end{array}\right\} \text{INDEPENDENT}$$

$$\hat{\theta}_1 = \frac{1}{2}(x[0] + x[1])$$
$$\hat{\theta}_2 = \frac{2}{3}x[0] + \frac{1}{3}x[1]$$

(UNBIASED?)

$$\mathrm{VAR}(\hat{\theta}_1) = \frac{1}{4}\mathrm{VAR}(x[0]) + \frac{1}{4}\mathrm{VAR}(x[1])$$
$$\mathrm{VAR}(\hat{\theta}_2) = 4/9\,\mathrm{VAR}(x[0]) + \frac{1}{9}\mathrm{VAR}(x[1])$$

$$\mathrm{VAR}(x[0]) = 1$$
$$\mathrm{VAR}(x[1]) = \begin{cases} 1 & \theta \geq 0 \\ 2 & \theta < 0 \end{cases}$$



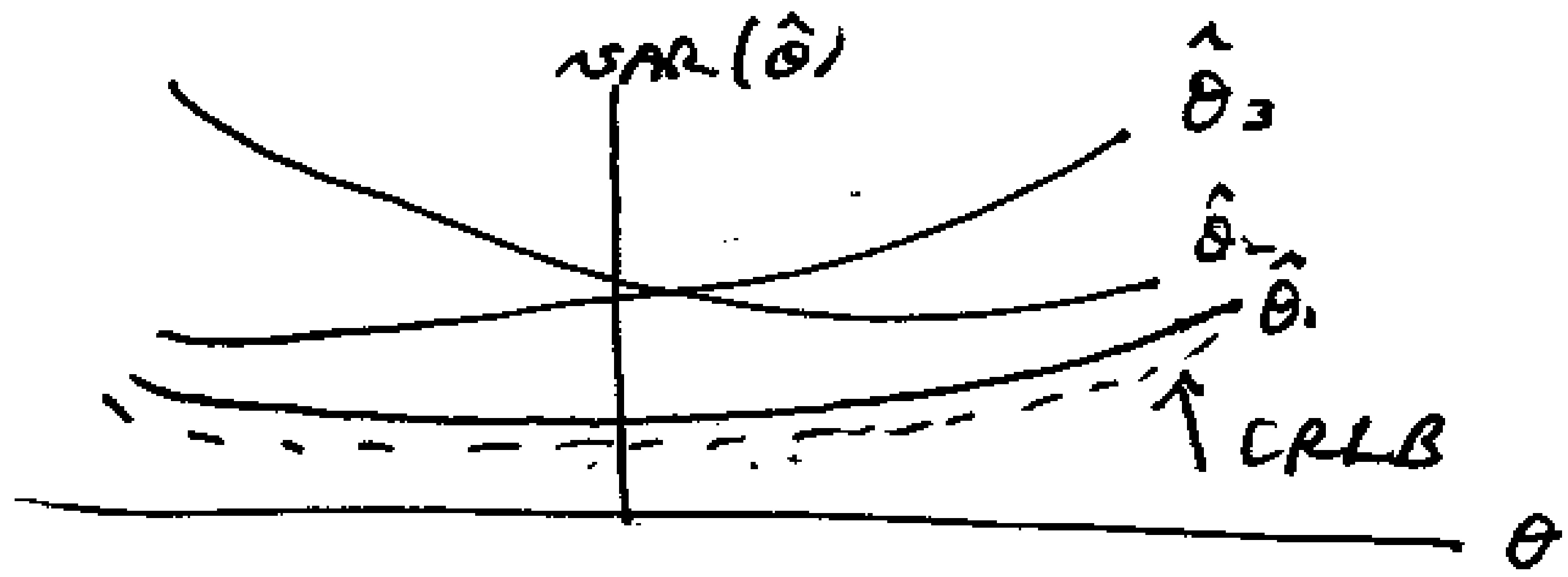
WHY NOT CHOOSE $\hat{\theta}_1$ FOR $\theta \geq 0$ AND $\hat{\theta}_2$ FOR $\theta < 0$?

(CAN ALSO SHOW THAT VAR($\hat{\theta}_1$) FOR $\theta \geq 0$ AND VAR($\hat{\theta}_2$) FOR $\theta < 0$ ARE SMALLEST POSSIBLE $\Rightarrow$ NO OTHER $\hat{\theta}$ WITH LOWER VARIANCE)

HOW DO WE FIND MVU?
(IF IT EXISTS)

APPROACHES:

CHAPTER 3 1) COMPUTE CRAMER-RAO LOWER BOUND ON VARIANCE. FIND ESTIMATOR THAT SATISFIES IT.



WHEN CRLB SATISFIED WITH EQUALITY FOR ALL θ $\Rightarrow$ MVU ESTIMATOR

CHAPTER 5 2) APPLY RAO-BLACKWELL-LEHMANN-SCHEFFE THEOREM

$\Rightarrow$ FIND SUFFICIENT STATISTIC AND MAKE IT UNBIASED $\Rightarrow$ MVU ESTIMATOR

CHAPTER 6 3) RESTRICT ESTIMATOR TO BE LINEAR

$\Rightarrow$ BEST LINEAR UNBIASED ESTIMATOR OR MVU FOR ALL LINEAR ESTIMATORS

MVU FOR VECTOR PARAMETER

$$\underline{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} \quad p \times 1 \text{ VECTOR OF PARAMETERS}$$

UNBIASED IF $E(\hat{\theta}_i) = \theta_i$ OR LETTING

$$E\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix}$$

$$\Rightarrow \quad E(\hat{\underline{\theta}}) = \begin{bmatrix} E(\hat{\theta}_1) \\ E(\hat{\theta}_2) \\ \vdots \\ E(\hat{\theta}_p) \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} = \underline{\theta} \quad \text{UNBIASED}$$

MINIMUM VARIANCE MEANS $VAR(\hat{\theta}_i)$ $i = 1, 2, \ldots, p$ IS MINIMUM

CRAMER-RAO LOWER BOUND (CRLB)

HOW ACCURATELY WE CAN ESTIMATE θ DEPENDS ON PDF. WHAT HAPPENS IF PDF DOES NOT DEPEND ON θ?

EXAMPLE – $x[0] = A + \underbrace{w[0]}_{\sim N(0, \sigma^2)}$

CHOOSE $\hat{A} = x[0]$

$= A + w[0]$

$\Rightarrow$ IF σ^2 LARGE, $\hat{A}$ POOR

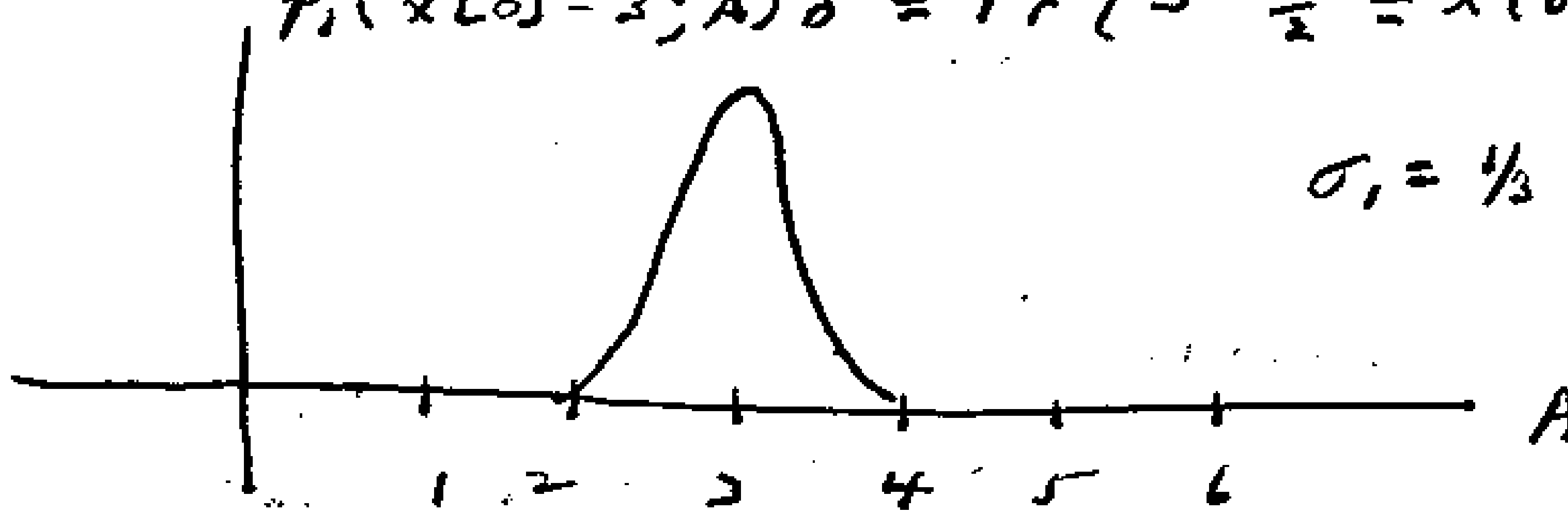
IF σ^2 SMALL, $\hat{A}$ GOOD

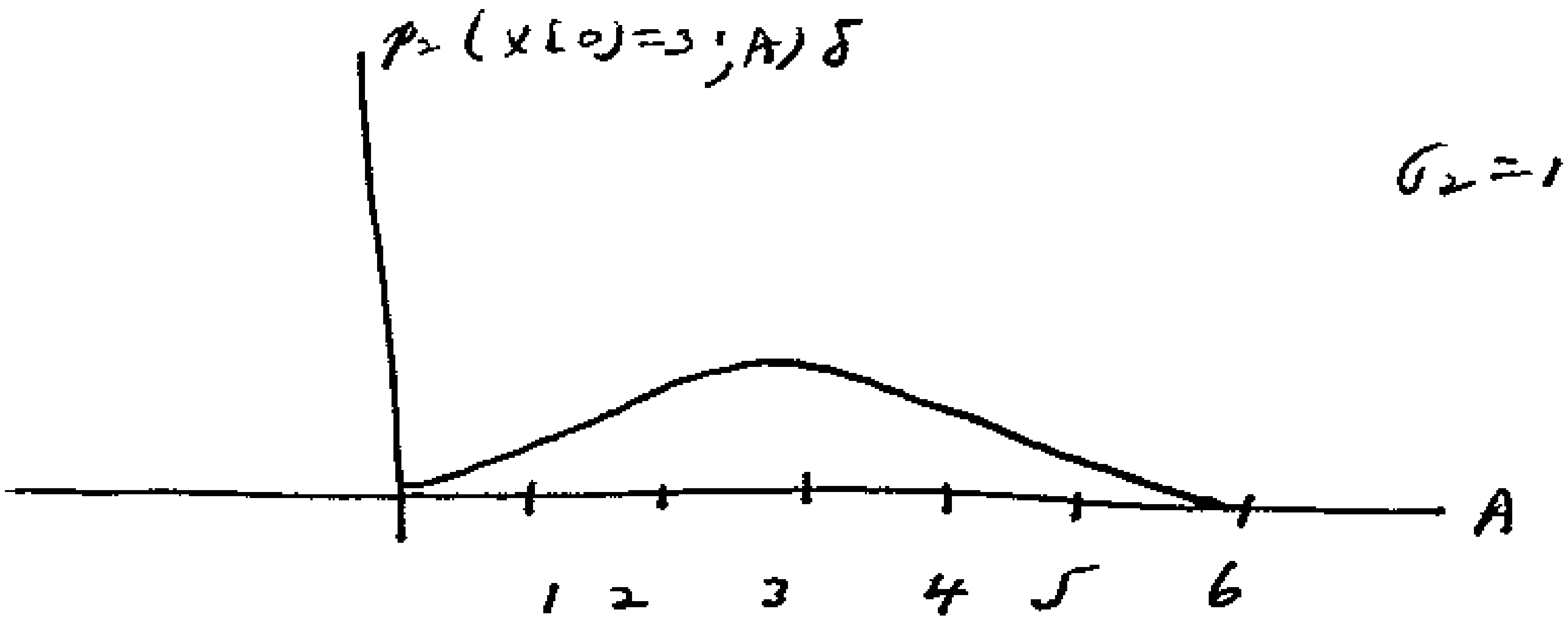
ALTERNATIVE VIEWPOINT!

$$p_i(x[0]; A) = \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2\sigma_i^2}(x[0]-A)^2}$$

$\uparrow$ PDF

$p_1(x[0]=3; A)\,\delta = P_r\left\{3 - \frac{\delta}{2} \le x[0] \le 3 + \frac{\delta}{2}\right\}$





$p_i(x[0]=3;A)$ VS A IS THE "PROBABILITY" OF OBSERVING $x[0]=3$ WHEN A IS THE TRUE VALUE.

IF $\sigma_1 = 1/2 \Rightarrow$ PROBABILITY $\simeq 0$ FOR $A > 4$
ONLY POSSIBLE VALUES OF A ARE
$3 \pm 3\sigma_1 = [2, 4]$

CONCENTRATION OF $p_i(x[0]=3;A)$ RESTRICTS POSSIBLE VALUES OF A

$\Rightarrow$ PDF CONCENTRATION $\uparrow$
PARAMETER ACCURACY $\uparrow$

TO MEASURE SHARPNESS OF PDF USE CURVATURE OR $-\dfrac{\partial^2 \ln p}{\partial A^2}$.

BUT

$$\ln p(x[0];A) = -\ln\sqrt{2\pi}\,\sigma - \frac{1}{2\sigma^2}(x[0]-A)^2$$

$$\frac{\partial \ln p}{\partial A} = \frac{1}{\sigma^2}(x[0]-A)$$

$$-\frac{\partial^2 \ln p}{\partial A^2} = \frac{1}{\sigma^2}$$

AS $\sigma^2 \downarrow$, CURVATURE INCREASES.

NOTE :

$$VAR(\hat{A}) = VAR(x[0]) = \sigma^2$$

$$= \frac{1}{-\frac{\partial^2 LNp}{\partial A^2}} = \frac{1}{CURVATURE}$$

IN GENERAL, $-\frac{\partial^2 LNp}{\partial \theta^2}$ IS RANDOM VARIABLE

TO MEASURE CURVATURE USE

$$E\left[-\frac{\partial^2 LNp}{\partial \theta^2}\right].$$

K →

CRLB

IF PDF $p(\underline{x};\theta)$ SATISFIES REGULARITY CONDITIONS, THEN VARIANCE OF UNBIASED ESTIMATOR $\hat{\theta}$ SATISFIES

$$VAR(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 LNp(\underline{x};\theta)}{\partial \theta^2}\right]\Bigg|_{\text{TRUE VALUE OF } \theta}}$$

ALSO, IF AND ONLY IF

$$\frac{\partial LNp(\underline{x};\theta)}{\partial \theta} = I(\theta)\left(g(\underline{x}) - \theta\right)$$