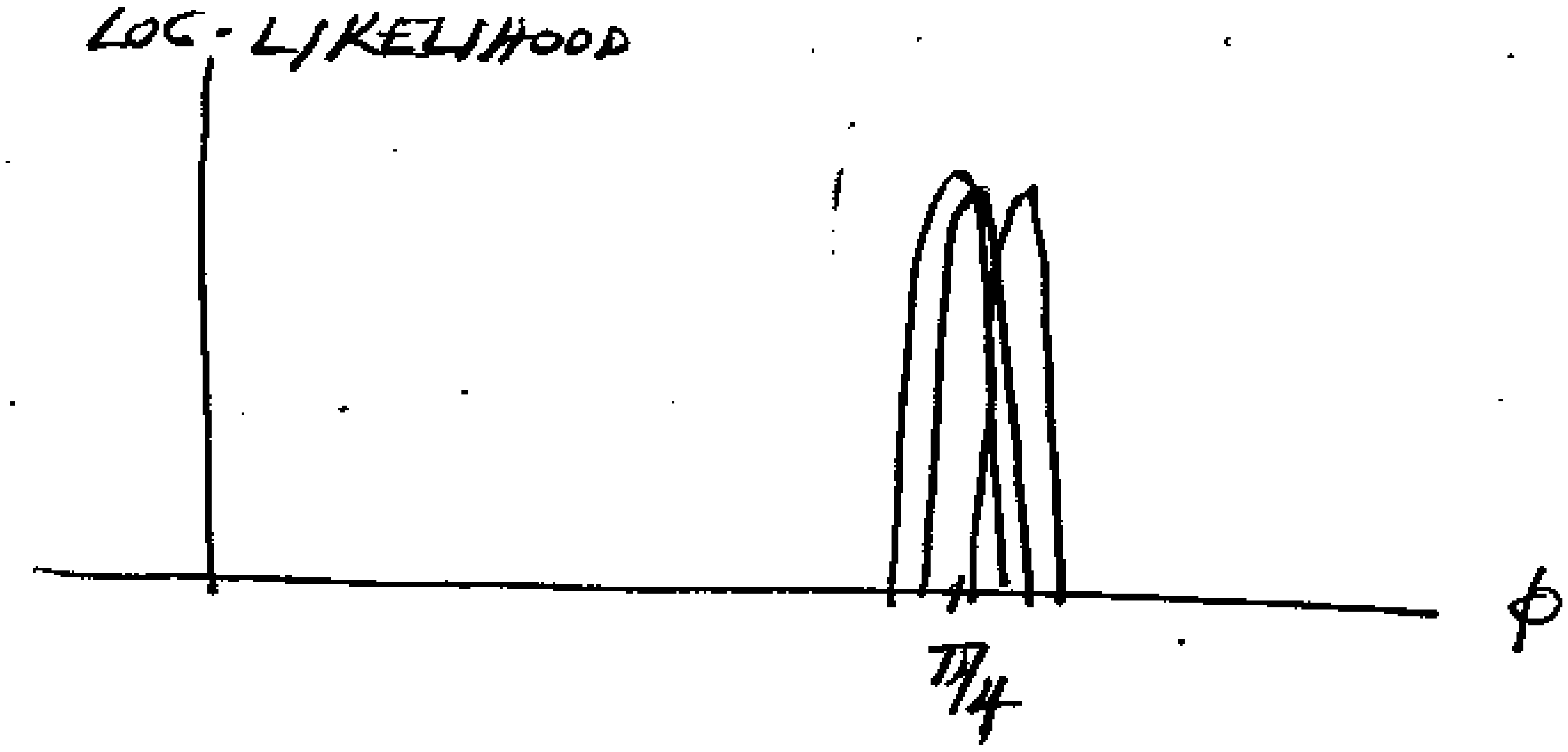
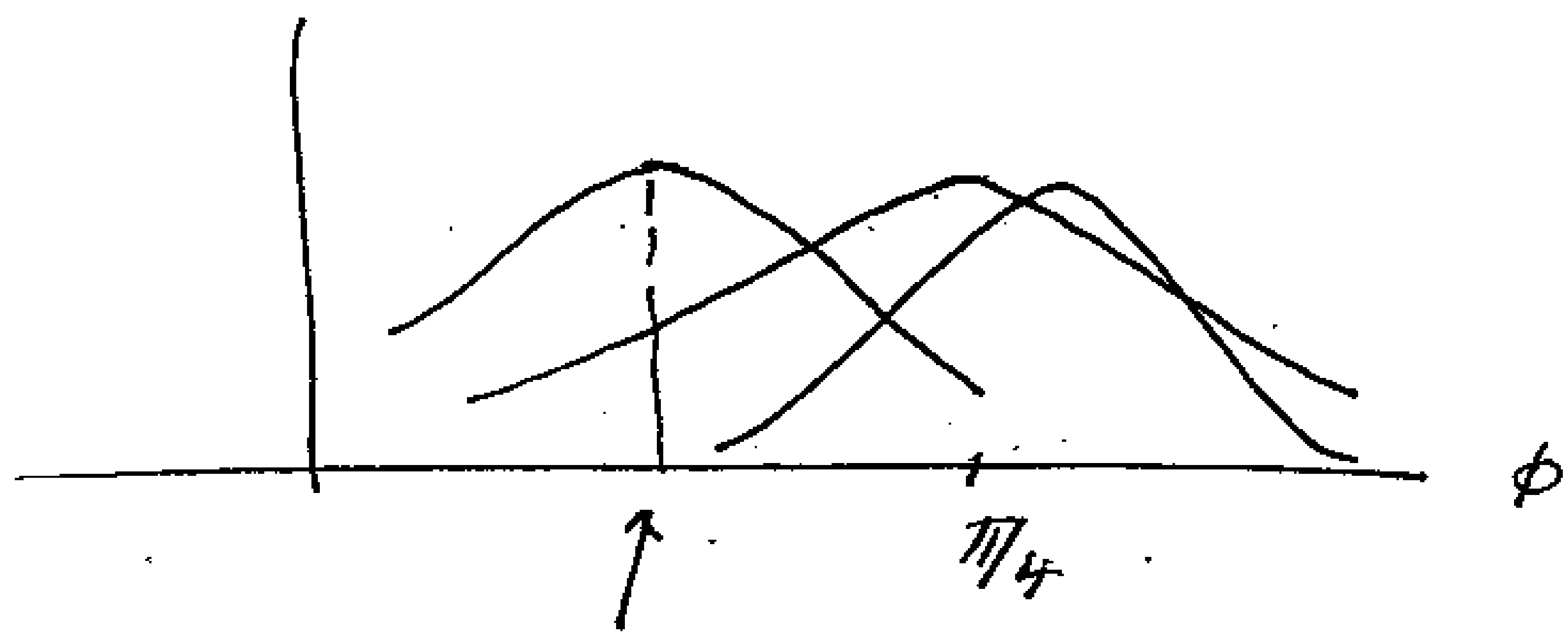
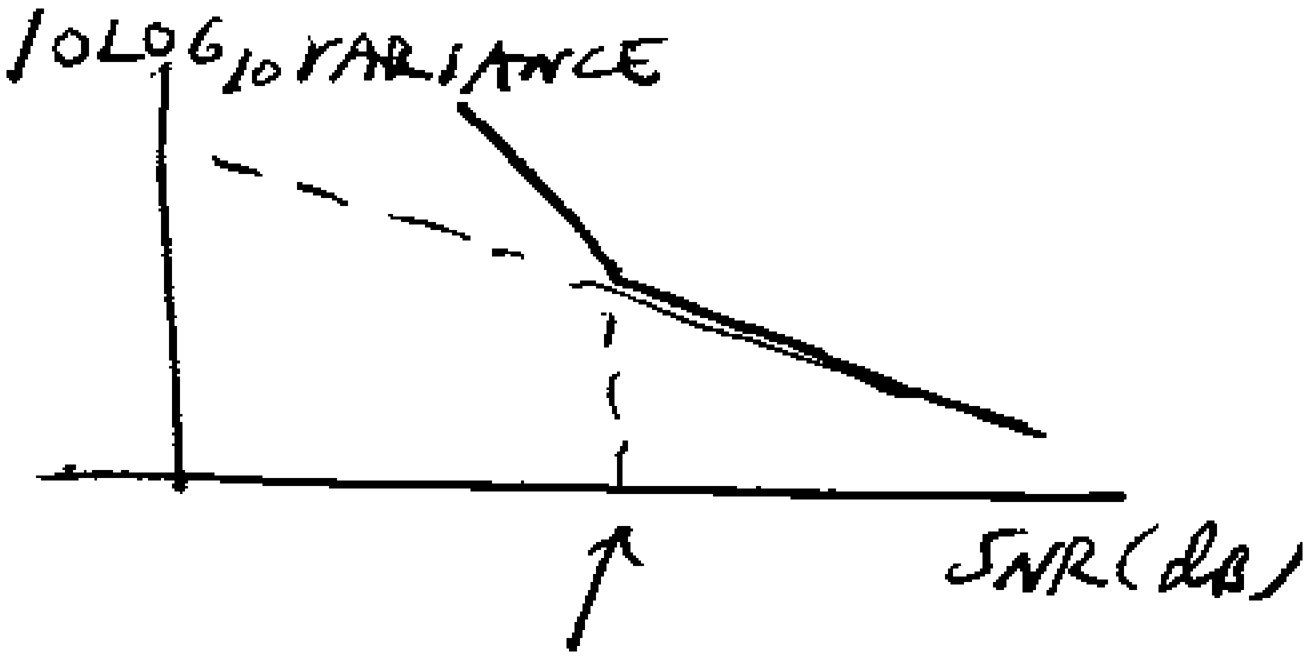
AT LOWER SNRs A THRESHOLD EFFECT OCCURS.





OUTLIER

OUTLIERS GIVE RISE TO INCREASED VARIANCE (MORE THAN PREDICTED BY CRLB)



THRESHOLD SNR

THIS BEHAVIOR IS CHARACTERISTIC OF ALL NONLINEAR ESTIMATORS

FOR SIGNAL IN NOISE PROBLEMS THE CRLB IS ATTAINED EVEN FOR SHORT DATA RECORDS IF SNR IS HIGH ENOUGH. WHY?

$$\hat{\phi} = -\arctan \frac{\sum_n (A\cos(2\pi f_0 n + \phi) + w[n]) \sin 2\pi f_0 n}{\sum_n (A\cos(2\pi f_0 n + \phi) + w[n]) \cos 2\pi f_0 n}$$

$$\approx -\arctan \frac{-\frac{NA}{2}\sin\phi + \sum_n w[n] \sin 2\pi f_0 n}{\frac{NA}{2}\cos\phi + \sum_n w[n]\cos 2\pi f_0 n}$$

$$= \arctan \frac{\sin\phi + \epsilon_s}{\cos\phi + \epsilon_c}$$

WHERE $\epsilon_s = -\frac{2}{NA} \sum_n w[n] \sin 2\pi f_0 n$

$$\epsilon_c = \frac{2}{NA} \sum_n w[n] \cos 2\pi f_0 n$$

IF $\epsilon_s = \epsilon_c = 0$ (NO NOISE), $\hat{\phi} = \phi$.

ASYMPTOTIC PDF WILL BE VALID WHEN ϵ_s, ϵ_c ARE SMALL. THEN, ARCTAN TRANSFORMATION $\simeq$ LINEAR

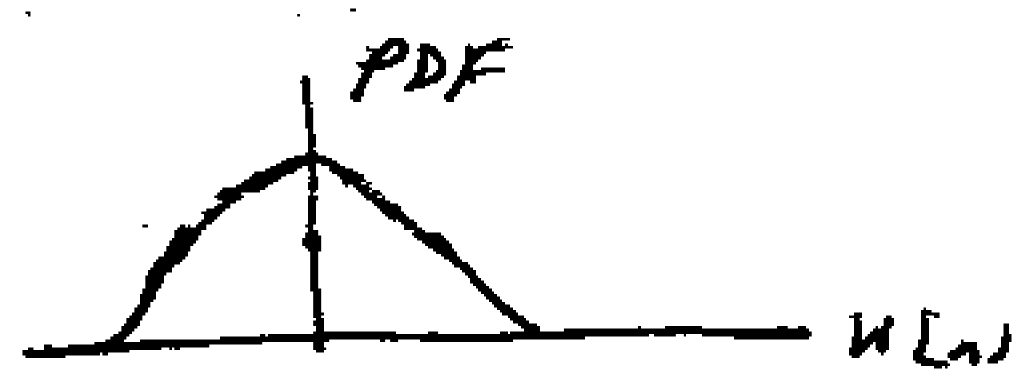
IN GENERAL, ASYMPTOTIC PDF HOLDS WHEN ERROR IS SMALL, EITHER $N \to \infty$

OR $A \to \infty$ (HIGH SNR) OR BOTH.

ASYMPTOTIC PDF INVALID WHEN ESTIMATION ERROR CANNOT BE MADE SMALL AS $N \to \infty$.

<u>EXAMPLE</u> : DC IN NONINDEPENDENT NON-GAUSSIAN NOISE

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

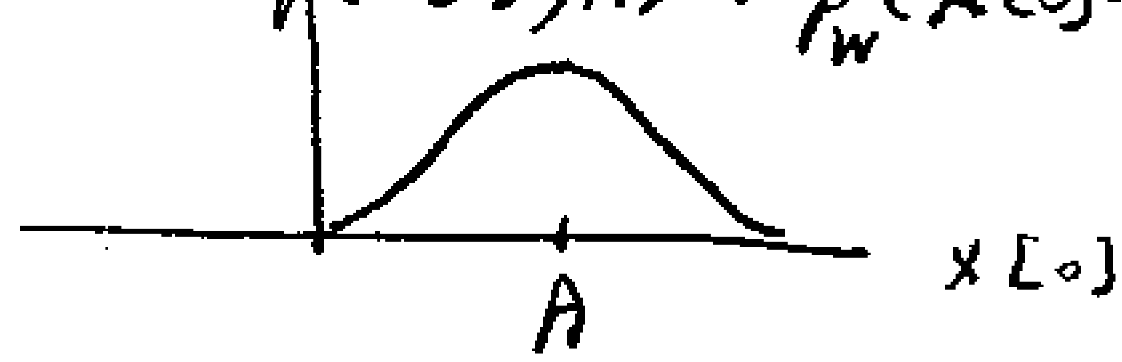


$E[w[n]] = 0$

OBSERVE $x[0], x[1], \ldots, x[N-1]$ BUT $w[0] = w[1] = \ldots = w[N-1] \Rightarrow$ ALL NOISE SAMPLES THE SAME : DISCARD $x[1], \ldots, x[N-1]$

MLE OF A FOUND BY MAXIMIZING $p(x[0]; A)$

$$p(x[0]; A) = p_w(x[0]-A; A)$$



$\Rightarrow \hat{A} = x[0] \Rightarrow$ PDF IS NONGAUSSIAN EVEN AS $N \to \infty$

ALSO NOT CONSISTENT SINCE $\text{VAR}(\hat{A}) = \text{VAR}(x[0]) \not\to 0$ AS $N \to \infty$

MLE FOR TRANSFORMED PARAMETERS

HOW DO WE FIND THE MLE FOR A FUNCTION OF θ, FOR EXAMPLE, A^2 INSTEAD OF A?

EXAMPLE : $x[n] = A + w[n]$ ← WGN

WANT MLE OF $\alpha = e^A$, WHICH IS A 1-1 TRANSFORMATION OF A.

$$p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]-A)^2} \qquad -\infty < A < \infty$$

THIS PDF IS PARAMETERIZED BY A. SINCE α IS 1-1 TRANSFORMATION, WE CAN EQUIVALENTLY USE

$$p_T(\underline{x}; \alpha) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]-\ln\alpha)^2} \qquad \alpha > 0$$

↑ TRANSFORMED

PDFS ARE EQUIVALENT. MLE OF α FOUND BY MAXIMIZING p_T OVER α.

$$\Rightarrow \frac{\partial}{\partial\alpha} \sum_n (x[n] - \ln\alpha)^2 = 0$$

$$-2 \sum_n (x[n] - \ln\hat{\alpha}) \frac{1}{\hat{\alpha}} = 0$$

$$\Rightarrow \quad \ln \hat{\alpha} = \frac{1}{N} \sum_n x[n] = \bar{x}$$

$$\tilde{\alpha} = e^{\bar{x}}$$

BUT $\hat{A} = \bar{x}$ IS MLE OF A.

$$\tilde{\alpha} = e^{\hat{A}} = e^{\hat{\theta}}$$

MLE IS FOUND BY SUBSTITUTING $\hat{\theta}$ INTO TRANFORMATION. INVARIANCE PROPERTY.

EXAMPLE: FOR SAME DATA LET $\alpha = A^2$
NOT 1-1 TRANSFORMATION.

$$A = \pm\sqrt{\alpha}$$

CANNOT PARAMETERIZE PDF BY α.
NEED TWO SETS OR

$$p_{T_1}(\underline{x};\alpha) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]-\sqrt{\alpha})^2} \qquad \alpha \geq 0$$

$$p_{T_2}(\underline{x};\alpha) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]+\sqrt{\alpha})^2} \qquad \alpha > 0$$

TO COVER ALL POSSIBILITIES.

MLE OF α IS FOUND BY FINDING α THAT

MAXIMIZES p_{T_1} AND p_{T_2} OR

$$\hat{\alpha} = \arg\max_{\alpha} \{ p_{T_1}(\underline{x};\alpha), p_{T_2}(\underline{x};\alpha) \}$$

OR

1) FOR GIVEN α, DETERMINE WHETHER p_{T_1} OR p_{T_2} IS LARGER $\Rightarrow \bar{p}_T$. REPEAT FOR ALL $\alpha > 0$ $\left(\bar{p}_T(\underline{x};\alpha=0) = p(\underline{x};A=0) \right)$

2) MLE GIVEN BY α THAT MAXIMIZES $\bar{p}_T$ OVER $\alpha \geq 0$.

FOR PREVIOUS EXAMPLE

$$\hat{\alpha} = \arg\max_{\alpha \geq 0} \{ \underbrace{p(\underline{x};\sqrt{\alpha})}_{p_{T_1}(\underline{x};\alpha)}, \underbrace{p(\underline{x};-\sqrt{\alpha})}_{p_{T_2}(\underline{x};\alpha)} \}$$

$$= \left[\arg\max_{\sqrt{\alpha} \geq 0} \{ p(\underline{x};\sqrt{\alpha}), p(\underline{x};-\sqrt{\alpha}) \} \right]^2$$

$$= \left(\arg\max_{-\infty < A < +\infty} p(\underline{x};A) \right]^2$$

$$= (\hat{A})^2 = (\bar{x})^2$$

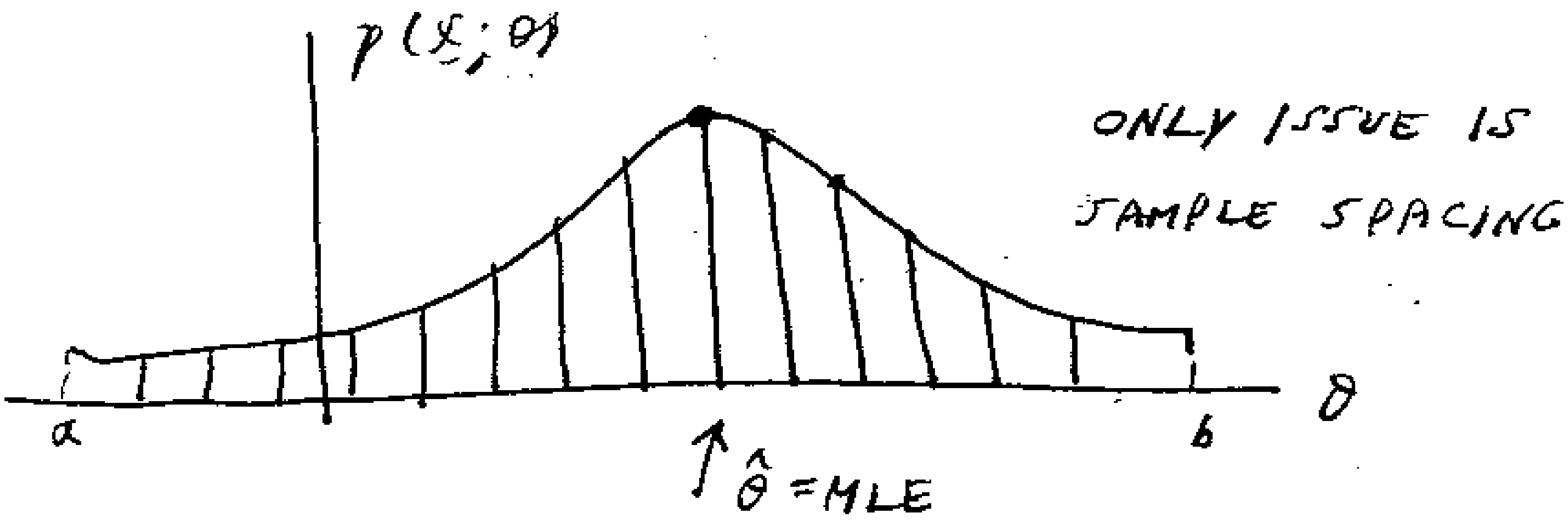
$\Rightarrow$ INVARIANCE PROPERTY STILL HOLDS, BUT $\hat{\alpha}$ MAXIMIZES <u>MODIFIED</u> LIKELIHOOD

FUNCTION OR $\bar{p}_T(\underline{x};\alpha)$ = MAXIMUM OF ALL $p(\underline{x};\theta)$, WHERE $\alpha = g(\theta)$ AND MAXIMUM IS TAKEN OVER ALL θ YIELDING SAME α. SEE FIGURE 7.7.

NUMERICAL DETERMINATION OF MLE

UNLIKE OTHER APPROACHES MLE CAN BE FOUND NUMERICALLY. ONLY NEED PDF.

1) $a < \theta < b \Rightarrow$ GRID SEARCH



MUST EVALUATE $p(\underline{x};\theta)$ FOR EACH NEW $\underline{x}$, SINCE $\hat{\theta} = g(\underline{x})$ AND g IS UNKNOWN.

2) θ LIES IN INFINITE INTERVAL

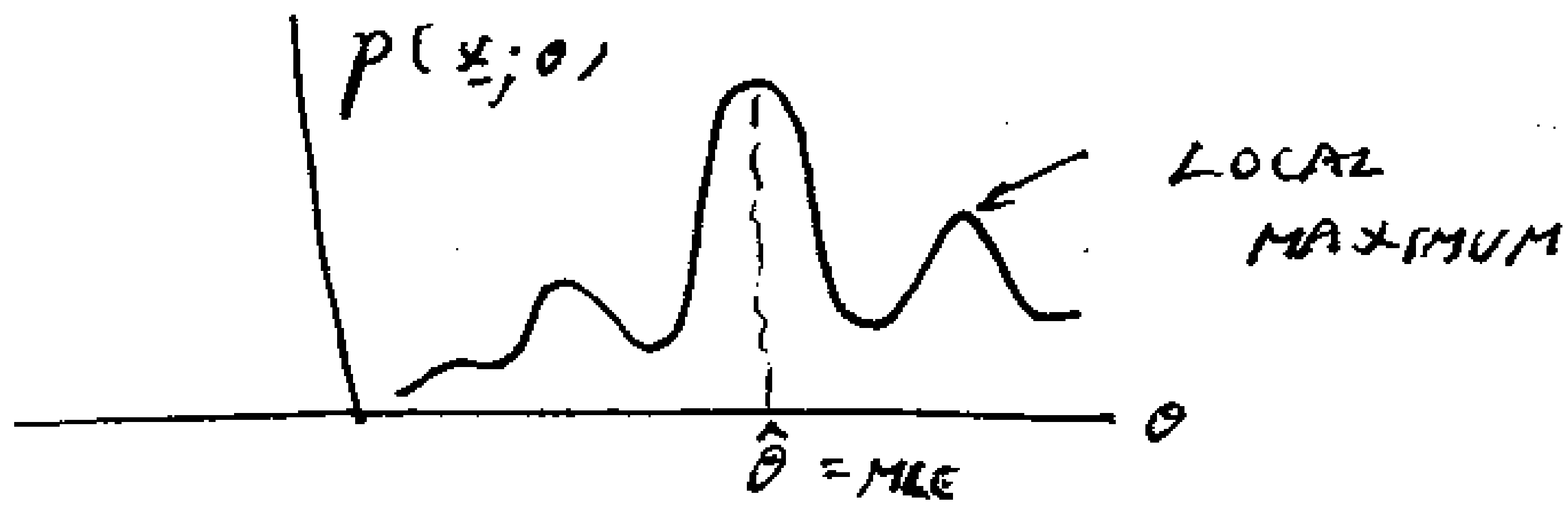
EXAMPLE: $\sigma^2 > 0$, WHERE σ^2 = VARIANCE OF $x[n]$

$\Rightarrow$ GRID SEARCH NO GOOD

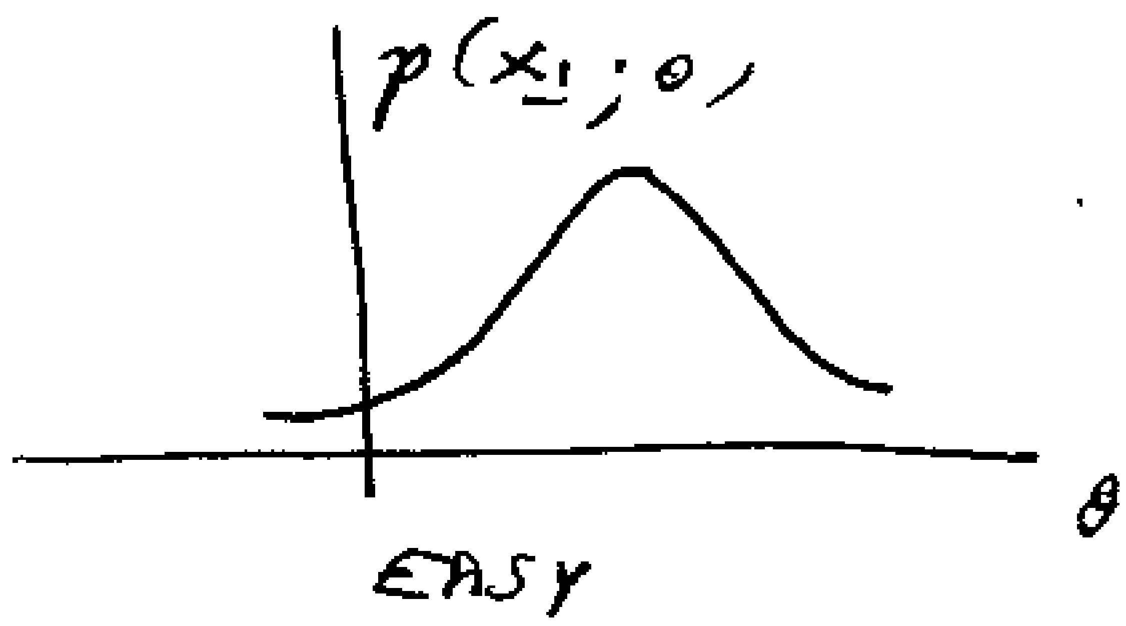
NEED NONLINEAR MAXIMIZATION ROUTINE –
NEWTON-RAPHSON, SCORING, EM

ALL THESE METHODS ARE ITERATIVE

1) MAY NOT CONVERGE
2) IF CONVERGENCE, LOCAL MAXIMUM?



NOTE: WE DO NOT KNOW FORM OF FUNCTION TO BE MAXIMIZED BEFOREHAND



USUAL METHODS OF NONLINEAR OPTIMIZATION MAY FAIL

EXAMPLE: $x[n] = r^n + w[n]$

↑ WGN WITH VARIANCE σ^2

ESTIMATE r, WHERE $r > 0$.

MLE: $p(\underline{x}; r) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x(n) - r^n)^2}$

$\Rightarrow$ MINIMIZE $J(r) = \sum_n (x(n) - r^n)^2$

$$\frac{dJ}{dr} = 0 \Rightarrow \sum_{n=0}^{N-1} (x[n] - r^n) n r^{n-1} = 0$$

NONLINEAR IN r (ALSO, MAY PRODUCE LOCAL MAXIMUM OR MINIMUM)

NEWTON - RAPHSON & SCORING TRY TO FIND ZERO OF dJ/dr OR IN GENERAL

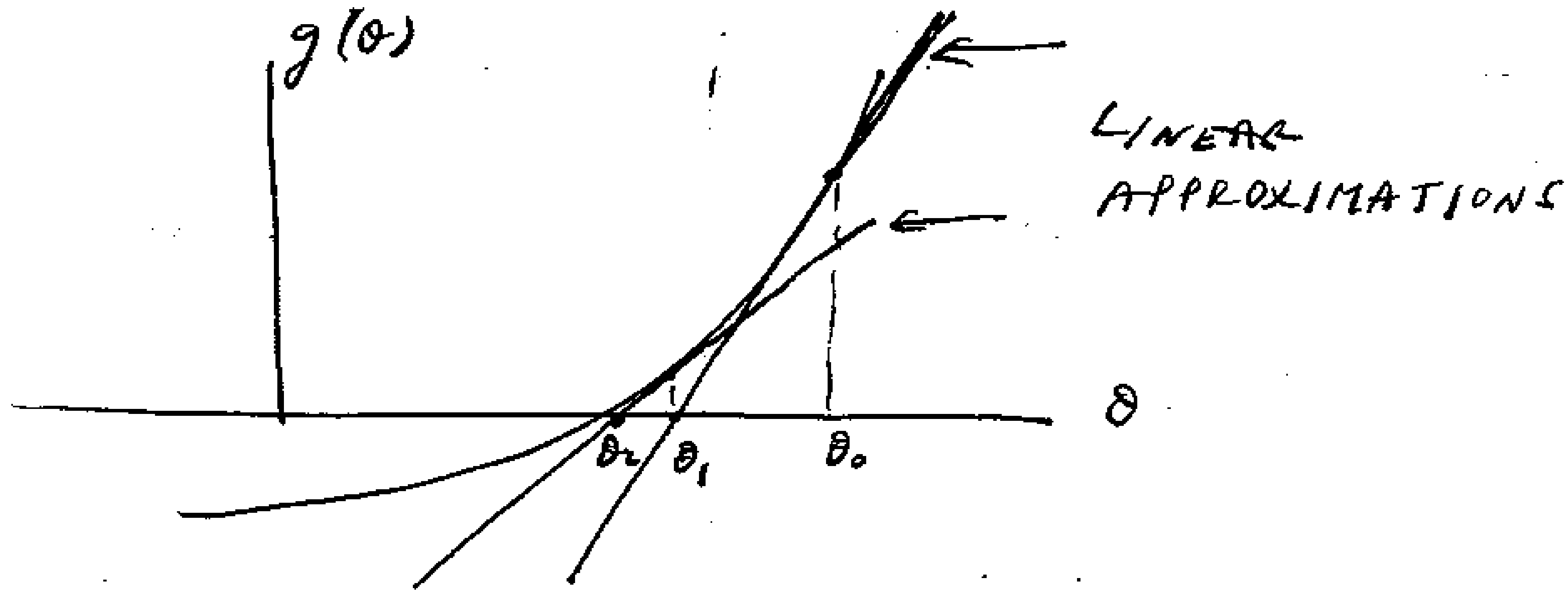
$$\frac{\partial \ln p(\underline{x}; \theta)}{\partial \theta} = 0 \qquad \text{SOLVE FOR } \theta.$$

LET $g(\theta) = \frac{\partial \ln p}{\partial \theta}$

ASSUME WE HAVE INITIAL GUESS OF SOLUTION, CALL IT θ_0. FOR g APPROXIMATELY LINEAR NEAR θ_0

$$g(\theta) \simeq g(\theta_0) + \frac{dg(\theta_0)}{d\theta_0}(\theta - \theta_0)$$

$$\left(\frac{dg(\theta_0)}{d\theta_0} = \left.\frac{dg}{d\theta}\right|_{\theta=\theta_0} \right)$$



SOLVE FOR NEW GUESS OR

$$0 = g(\theta_1) = g(\theta_0) + \frac{dg(\theta_0)}{d\theta_0}(\theta_1 - \theta_0)$$

$$\Rightarrow \quad \theta_1 = \theta_0 - \frac{g(\theta_0)}{\frac{dg(\theta_0)}{d\theta_0}}$$

NOW USE θ_1 TO LINEARIZE ABOUT, ETC.
SHOULD CONVERGE FOR g AS SHOWN.
IN GENERAL,

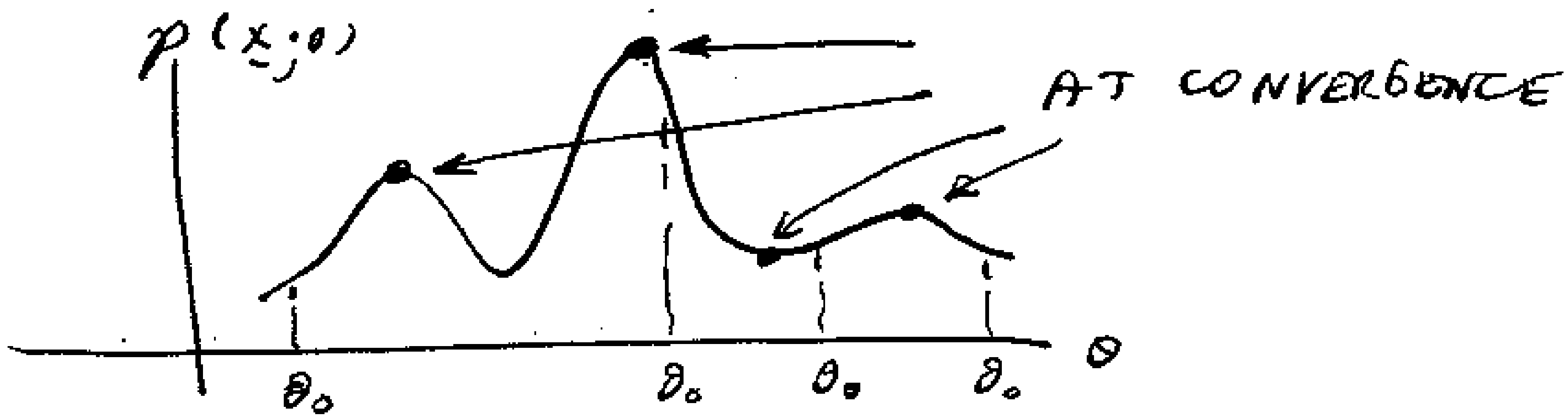
$$\theta_{k+1} = \theta_k - \frac{g(\theta_k)}{\frac{dg(\theta_k)}{d\theta_k}}$$

AT CONVERGENCE $\theta_{k+1} = \theta_k \Rightarrow g(\theta_k) = 0$

RECALL $g(\theta) = \frac{\partial \ln p(\underline{x};\theta)}{\partial \theta}$

$$\Rightarrow \quad \theta_{k+1} = \theta_k - \left[\frac{\partial^2 \ln p(\underline{x};\theta)}{\partial \theta^2} \right]^{-1} \frac{\partial \ln p(\underline{x};\theta)}{\partial \theta} \Bigg|_{\theta=\theta_k}$$

SHOULD USE SEVERAL STARTING POINTS



THEN COMPUTE $p(\underline{x};\theta)$ FOR EACH CONVERGED POINT AND CHOOSE ONE THAT MAXIMIZES p.

NEED GOOD INITIAL GUESS!

EXAMPLE : $x[n] = r^n + w[n]$

$\uparrow$ WGN

$$p(\underline{x};r) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]-r^n)^2}$$

$$\frac{\partial \ln p}{\partial r} = \frac{\partial}{\partial r}\left[-\frac{1}{2\sigma^2}\sum_n (x[n]-r^n)^2\right]$$

$$= \frac{1}{\sigma^2}\sum_n (x[n]-r^n)\, n r^{n-1}$$

$$\frac{\partial^2 \ln p}{\partial r^2} = \frac{1}{\sigma^2}\Big[\sum_n n(n-1)x[n]r^{n-2} - \sum_n n(2n-1)r^{2n-2}\Big]$$

$$= \frac{1}{\sigma^2}\sum_n n r^{n-2}\left[(n-1)x[n] - (2n-1)r^n\right]$$

$$r_{k+1} = r_k - \frac{\sum_{n=0}^{N-1}(x[n]-r_k^n)\,n\,r_k^{n-1}}{\sum_{n=0}^{N-1} n\,r_k^{n-2}\left[(n-1)x[n]-(2n-1)r_k^n\right]}$$

SCORING METHOD NOTES THAT

$$\left.\frac{\partial^2 \ln p(\underline{x};\theta)}{\partial\theta^2}\right|_{\theta=\theta_k} \approx -I(\theta_k)$$

SINCE FOR IID SAMPLES

INDEPENDENT

$$\frac{\partial^2 \ln p(\underline{x};\theta)}{\partial\theta^2} = \frac{\partial^2}{\partial\theta^2}\ln\prod_{n=0}^{N-1} p(x[n];\theta)$$

$$= \frac{\partial^2}{\partial\theta^2}\sum_n \ln p(x[n];\theta)$$

$$= \sum_n \frac{\partial^2 \ln p(x[n];\theta)}{\partial\theta^2}$$

$$= N\,\frac{1}{N}\sum_n \frac{\partial^2 \ln p(x[n];\theta)}{\partial\theta^2}$$

$$\approx N E\left[\frac{\partial^2 \ln p(x[n];\theta)}{\partial\theta^2}\right]$$

BY LAW OF LARGE NUMBERS ↑ IDENTICALLY DISTRIBUTED

(FOR $x_1, x_2, \ldots, x_n$ IID WITH $E(x_i) = \mu$
$\frac{1}{N}\sum_{i=1}^{n} x_i \to \mu$ AS $n \to \infty$)

$$= -N i(\theta)$$
$$= -I(\theta)$$

$\Rightarrow$ SCORING METHOD IS

$$\theta_{k+1} = \theta_k + I^{-1}(\theta) \left.\frac{\partial \ln p(\underline{x};\theta)}{\partial\theta}\right|_{\theta=\theta_k}$$

HAS SAME GENERAL CONVERGENCE PROBLEMS BUT MAY WORK BETTER THAN N-R.

<u>EXAMPLE</u> : $x[n] = r^n + w[n]$ ↑ WGN

SEE EXAMPLE 7.11

$$I(\theta) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} n^2 r^{2n-2}$$

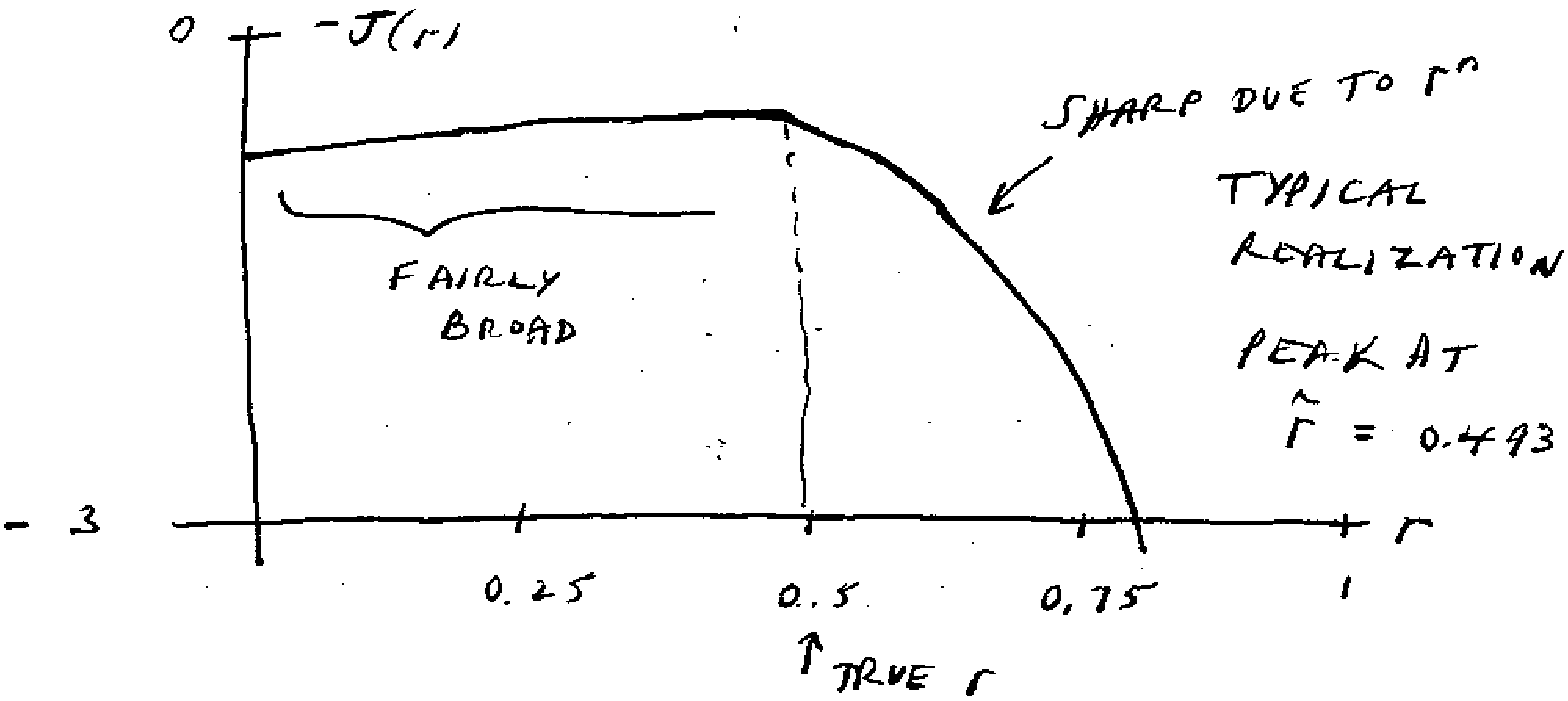
$$\Rightarrow \quad r_{k+1} = r_k + \frac{\sum_{n=0}^{N-1}(x[n] - r_k^n)\, n r_k^{n-1}}{\sum_{n=0}^{N-1} n^2 r_k^{2n-2}}$$

COMPUTER EXAMPLE : N-R FOR
$N = 50, \quad r = 0.5, \quad \sigma^2 = 0.01.$

WISH TO MAXIMIZE

$$\ln p(\underline{x}; r) = \ln \frac{1}{(2\pi\sigma^2)^{N/2}} - \frac{1}{2\sigma^2} \sum_n (x[n] - r^n)^2$$

OVER r OR $-\sum_{n=0}^{N-1} (x[n] - r^n)^2 = -J(r)$



USING A N-R FOR SEVERAL r_0.
FOR $r_0 > 1.2$ OR $r_0 < 0.18$ ITERATION "BLEW UP" DUE TO OVERFLOW (r^n TERM).

ITERATION	INITIAL GUESS, r_0		
0	0.8	0.2	1.2
1	0.723	0.799	1.187
2	0.638	0.722	1.174
3	0.561	0.637	1.161
4	0.510	0.560	1.148
5	0.494	0.510	1.136
6	0.493	0.494	1.123
7		0.493	1.111
8			1.098
9			1.086
10			1.074
⋮			⋮
29			0.493

SLOWER CONVERGENCE

THIRD METHOD IS EM - MOST USEFUL FOR VECTOR PARAMETER.

VECTOR PARAMETER

MLE IS THAT VALUE OF $\underline{\theta}$ THAT MAXIMIZES $p(\underline{x};\underline{\theta})$. CAN TRY

$$\frac{\partial \ln p(\underline{x};\underline{\theta})}{\partial \underline{\theta}} = \underline{0}$$

AND PICK SOLUTION THAT PRODUCES MAXIMUM.

EXAMPLE : $x[n] = A + w[n] \quad n = 0, 1, \ldots, N-1$

$\uparrow$ WGN

FIND MLE OF $\underline{\theta} = [A \ \sigma^2]^T$

$$p(\underline{x}; \underline{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_n (x[n]-A)^2}$$

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

$$\Rightarrow \frac{1}{\sigma^2}\sum_n (x[n]-A) = 0 \Rightarrow \hat{A} = \bar{x}$$

$$\frac{\partial \ln p}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_n (x[n]-A)^2\right] = 0$$

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (x[n]-A)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = 1/N \sum_n (x[n]-\hat{A})^2$$

OR

$$\hat{\underline{\theta}} = \begin{bmatrix} \bar{x} \\ \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-\bar{x})^2 \end{bmatrix} \qquad \text{UNBIASED?}$$

SHOULD CHECK TO BE SURE THIS MAXIMIZES $p(\underline{x};\theta)$. HOW?

ASYMPTOTIC PROPERTIES OF MLE :

$$\hat{\underline{\theta}} \overset{a}{\sim} N(\underline{\theta}, \underline{I}^{-1}(\underline{\theta}))$$

AS $N \to \infty$

EXAMPLE : SEE PREVIOUS ONE

$$\hat{A} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2$$

ACCORDING TO THEOREM

$$\begin{bmatrix} \hat{A} \\ \hat{\sigma}^2 \end{bmatrix} \overset{a}{\sim} N\left(\begin{bmatrix} A \\ \sigma^2 \end{bmatrix}, \begin{bmatrix} \sigma^2/N & 0 \\ 0 & 2\sigma^4/N \end{bmatrix} \right)$$

OR $\hat{A}, \hat{\sigma}^2$ ARE UNCORRELATED $\Rightarrow$ INDEPENDENT
TO VERIFY THIS RECALL

$$\hat{A} = \bar{x} \sim N(A, \sigma^2/N)$$

IT CAN BE SHOWN THAT (FOR FINITE N)

$$T = \frac{\sum_{n=0}^{N-1} (x[n] - \bar{x})^2}{\sigma^2} \sim \chi^2_{N-1}$$

CHI SQUARED PDF WITH N-1 DEGREES OF FREEDOM

AND $T, \bar{x}$ ARE INDEPENDENT

χ^2_ν HAS PDF $p(u) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} u^{\frac{\nu}{2}-1} e^{-u/2} & u > 0 \\ 0 & u < 0 \end{cases}$

ALSO, IF $y = \sum_{i=1}^{n} x_i^2$ $\quad x_i \sim N(0,1)$ IID

$$\Rightarrow \quad y \sim \chi^2_n .$$

AS $n \to \infty$, χ_n^2 PDF BECOMES GAUSSIAN DUE TO CENTRAL LIMIT THEOREM.

ALSO $E(\chi_n^2) = n$

$VAR(\chi_n^2) = 2n$.

$$\Rightarrow \quad T \sim \chi^2_{N-1} \to N(N-1,\ 2(N-1))$$

BUT $\hat{\sigma}^2 = \frac{1}{N} \sum_n (x[n] - \bar{x})^2$

$$= \frac{\sigma^2}{N} T \overset{a}{\sim} N\left(\frac{N-1}{N}\sigma^2,\ \frac{\sigma^4}{N^2} 2(N-1)\right)$$

$$\overset{a}{\sim} N\left(\sigma^2,\ 2\sigma^4/N\right)$$

FURTHERMORE, $\underline{\hat{\theta}}$ IS ASYMPTOTICALLY JOINTLY GAUSSIAN SINCE T, $\bar{x}$ ARE INDEPENDENT AND INDIVIDUALLY GAUSSIAN.

COUNTEREXAMPLE : ASYMPTOTIC PROPERTIES GENERALLY DON'T HOLD IF WE ATTEMPT TO ESTIMATE TOO MANY PARAMETERS.

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

$\uparrow$ IID

$$p(w[n]) = \frac{1}{4} e^{-\frac{1}{2}|w[n]|}$$

LAPLACIAN NOISE

ESTIMATE $\underline{\theta} = (s[0]\; s[1] \ldots s[N-1])^T$.

$x[n]$ HAS PDF $\frac{1}{4} e^{-\frac{1}{2}|x[n]-s[n]|}$

$$\Rightarrow p(\underline{x}; \underline{\theta}) = \prod_{n=0}^{N-1} \frac{1}{4} e^{-\frac{1}{2}|x[n]-s[n]|}$$

MLE IS $\hat{s}[n] = x[n] \qquad n = 0, 1, \ldots, N-1$

OR $\hat{\underline{\theta}} = \underline{x}$

PDF OF $\hat{\underline{\theta}}$ NOT GAUSSIAN. IT IS

$$p(\hat{\underline{\theta}}) = \prod_{n=0}^{N-1} \frac{1}{4} e^{-\frac{1}{2}|\hat{\theta}_n - s[n]|}$$

PROBLEM IS NO AVERAGING POSSIBLE DUE TO TOO MANY PARAMETERS $\Rightarrow$ CENTRAL LIMIT THEOREM VIOLATED (SEE APPENDIX 7B).

INVARIANCE PROPERTY

IF $\underline{\alpha} = \underline{g}(\underline{\theta})$ IS r DIMENSIONAL FUNCTION, THEN MLE OF $\underline{\alpha}$ IS

$$\hat{\underline{\alpha}} = \underline{g}(\hat{\underline{\theta}})$$

MLE OF $\underline{\theta}$

IF $\underline{g}$ IS NOT INVERTIBLE, THEN $\hat{\underline{\alpha}}$ MAXIMIZES THE MODIFIED LIKELIHOOD FUNCTION

$$\bar{p}_T(\underline{x};\underline{\alpha}) = \max_{\{\underline{\theta}\,:\,\underline{\alpha}=\underline{g}(\underline{\theta})\}} p(\underline{x};\underline{\theta})$$

ALL $\underline{\theta}$ THAT MAP INTO SAME $\underline{\alpha}$

TO FIND MLE FOR GENERAL GAUSSIAN CASE OR $\underline{x} \sim N(\underline{\mu}(\underline{\theta}), \underline{C}(\underline{\theta}))$ WE SET PARTIALS EQUAL TO ZERO OR

FROM APPENDIX 3C

$$\frac{\partial \ln p(\underline{x};\underline{\theta})}{\partial \theta_k} = -\frac{1}{2}\,\mathrm{tr}\left[\underline{C}^{-1}(\underline{\theta})\,\frac{\partial \underline{C}(\underline{\theta})}{\partial \theta_k}\right]$$