At lower SNR's a threshold effect occurs. Log-likelihood

Outliers give rise to increased variance (more than predicted by CRLB)

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?

\[ \hat{\phi} = -\arctan \frac{\sum \left( A \cos(2\pi f_0 + \phi) + WN \right) \sin 2\pi f_0}{\sum \left( A \cos(2\pi f_0 + \phi) + WN \right) \cos 2\pi f_0} \]

\[ \hat{\sigma} = -\arctan \frac{-\frac{N A}{2} \sin \phi + \sum WN \sin 2\pi f_0}{\frac{N A}{2} \cos \phi + \sum WN \cos 2\pi f_0} \]

\[ = \arctan \frac{\sin \phi + \frac{E_s}{\cos \phi + E_c}}{\cos \phi + E_c} \]

Where

\[ E_s = -\frac{1}{N A} \sum WN \sin 2\pi f_0 \]

\[ E_c = \frac{1}{N A} \sum WN \cos 2\pi f_0 \]

If \( E_s = E_c = 0 \) (no noise), \( \hat{\phi} = \phi \).

Asymptotic PDF will be valid when \( E_s, E_c \) are small. Then, \( \arctan \) transformation \( \approx \) 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 $N \to \infty$.

**Example**: 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 $W[0] = W[1] = \cdots = W[N-1] \equiv$ all noise samples the same. Discard $X[1], \ldots, X[N-1]$

MLE of $A$ found by maximizing $p(X[0] \mid A)$.

$$p(X[0] \mid A) = p_w(X[0] - A \mid A)$$

$$\Rightarrow \hat{A} = X[0] \Rightarrow\text{PDF is non-Gaussian even as } N \to \infty$$

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 $\theta$, for example, $A^2$ instead of $A$?

Example: $X(n) = A + WN$

Want MLE of $\alpha = e^A$, which is a 1-1 transformation of $A$.

\[
p(X; A) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum (x(n) - A)^2}
\]

\[-\infty < A < \infty\]

This PDF is parameterized by $A$. Since $\alpha$ is 1-1 transformation, we can equivalently use

\[
p_T(X; \alpha) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum (x(n) - \ln(\alpha))^2}
\]

\[\alpha > 0\]

\[
p_{T}(X; \alpha) \uparrow \text{TRANSFORMED}
\]

PDFs are equivalent. MLE of $\alpha$ found by maximizing $p_T$ over $\alpha$.

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

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

\[ \hat{\alpha} = e^{\bar{x}} \]

But \( \hat{\alpha} = \bar{x} \) is MLE of \( \alpha \).

\[ \bar{x} = e^{\hat{\alpha}} = e^\hat{\theta} \]

MLE is found by substituting \( \hat{\theta} \) into transformation. Invariance property.

Example: For same data let \( \alpha = a^2 \)

Not 1-1 transformation.

\[ a = \pm \sqrt{\alpha} \]

Cannot parameterize PDF by \( \alpha \).

Need two sets or

\[ p_{T_1}(x; \alpha) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum (x(n) - \mu_2)^2} \quad \alpha > 0 \]

\[ p_{T_2}(x; \alpha) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum (x(n) + \mu_2)^2} \quad \alpha > 0 \]

To cover all possibilities.

MLE of \( \alpha \) is found by finding \( \alpha \) that
MAXIMIZES $\rho T_1$ AND $\rho T_2$ OR

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

OR

1) FOR GIVEN $\alpha$, DETERMINE WHETHER $\rho T_1$ OR $\rho T_2$ IS LARGER $\Rightarrow \bar{\rho}_T$. REPEAT FOR ALL $\alpha > 0$ ($\bar{\rho}_T(x;\alpha < 0) = \rho(x;\alpha = 0)$

2) MLE GIVEN BY $\hat{\alpha}$ THAT MAXIMIZES $\bar{\rho}_T$ OVER $\alpha > 0$.

FOR PREVIOUS EXAMPLE

$$\hat{\alpha} = \arg \max_{\alpha \geq 0} \{ \rho(x; \Gamma \alpha), \rho(x; -\Gamma \alpha) \}$$

$$\uparrow_{\rho T_1} \quad \uparrow_{\rho T_2}$$

$$= \left( \arg \max_{\alpha \geq 0} \{ \rho(x; \Gamma \alpha), \rho(x; -\Gamma \alpha) \} \right)^2$$

$$= \left( \arg \max_{-\infty < \alpha < \infty} \rho(x; \alpha) \right)^2$$

$$= (\hat{\alpha})^2 = (x)^2$$

$\Rightarrow$ INVARIANCE PROPERTY STILL HOLDS, BUT $\hat{\alpha}$ MAXIMIZES MODIFIED LIKELIHOOD.
FUNCTION OR \( P_T(x; \alpha) = \text{MAXIMUM OF ALL } p(x; \theta), \text{ WHERE } \alpha = g(\theta) \text{ AND MAXIMUM IS TAKEN OVER ALL } \theta \text{ YIELDING SAME } \alpha. \text{ SEE FIGURE 7.7.} \)

**NUMERICAL DETERMINATION OF MLE**

Unlike other approaches MLE can be found numerically, only need PDF.

1) \( a < \theta < b \) \( \Rightarrow \) Grid search

Only issue is sample spacing

\[ p(x; \theta) \]

\[ \alpha \leq \theta = \text{MLE} \leq b \]

Must evaluate \( p(x; \theta) \) for each new \( x \), since \( \hat{\theta} = g(x) \) and \( g \) is unknown.

2) \( \theta \) LIES IN INFINITE INTERVAL

**EXAMPLE:** \( \theta^2 > 0 \), WHERE \( \theta^2 = \text{VARIANCE OF } X(\theta) \)

\( \Rightarrow \) 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: $x \in \mathbb{R} \rightarrow \langle \omega, x \rangle + \text{wgn}$

\text{wgn with variance } \sigma^2
Estimate \( \theta \), where \( n > 0 \).

MLE: \( p(x; \theta) = \frac{1}{(2\pi \sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum (x_i - \theta)^2} \)

\( \Rightarrow \) Minimize \( J(\theta) = \sum \frac{1}{\sigma^2} (x_i - \theta)^2 \)

\( \frac{dJ}{d\theta} = 0 \Rightarrow \sum (x_i - \theta) \ln \sigma + \sigma^{-1} = 0 \)

Nonlinear in \( \theta \) (Also, may produce local maximum or minimum)

Newton-Raphson & Scoring try to find zero of \( dJ/d\theta \) or in general

\[ \frac{d \ln p(x; \theta)}{d\theta} = 0 \quad \text{solve for } \theta. \]

Let \( g(\theta) = \frac{d \ln p}{d\theta} \)

Assume we have initial guess of solution, call it \( \theta_0 \). For \( g \) approximately linear near \( \theta_0 \)

\[ g(\theta) \approx g(\theta_0) + \frac{dg(\theta_0)}{d\theta} (\theta - \theta_0) \]
\[
\left( \frac{dg(\theta)}{d\theta} \right)_{\theta = \theta_0} = \frac{dg}{d\theta} \bigg|_{\theta = \theta_0}
\]

Linear approximations

Solve for new guess \( \theta_1 \):

\[0 = g(\theta_1) = g(\theta_0) + \frac{dg(\theta)}{d\theta} \bigg|_{\theta = \theta_0} (\theta_1 - \theta_0) \]

\[\Rightarrow \theta_1 = \theta_0 - \frac{g(\theta_0)}{\frac{dg(\theta_0)}{d\theta}} \]

Now use \( \theta_1 \) to linearize about, etc. Should converge for \( g \) as shown.

In general:

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

At convergence \( \theta_{k+1} = \theta_k \Rightarrow g(\theta_k) = 0 \)
RECALL $q(\theta) = \frac{\partial \ln p(x; \theta)}{\partial \theta}$

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

**Should use several starting points**

**At convergence**

Then compute $p(x; \theta)$ for each converged point and choose one that maximizes $p$.

*Need good initial guess!*

**Example:** $x(n) = c^n + W[n]

$p(x; \theta) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2} \sum (x(n) - c^n)^2}$

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

$$= \frac{1}{\sigma^2} \sum (x(n) - c^n) = n-1$$
\[
\frac{\partial^2 \ln p}{\partial \theta^2} = \frac{1}{2^2} \left[ \frac{2}{n} \sum \left( \frac{1}{n(n-1)x \theta} \right)^{n-2} \right. \\
\left. - \frac{2}{n} \sum \left( \frac{1}{n(2n-1) \theta} \right)^{2n-2} \right] \\
= \frac{1}{2} \sum \frac{n}{n(n-1)} \left[ \frac{1}{(n-1)x \theta} - \frac{1}{2n-1} \frac{1}{\theta} \right]
\]

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

SCORING METHOD NOTES THAT

\[
\frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} \bigg|_{\theta = \theta_0}
\]

SINCE FOR IID SAMPLES

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

\[
= N \left\{ \frac{1}{N} \sum \frac{\partial^2 \ln p(x[n]; \theta)}{\partial \theta^2} \right\}
\]
\[ N \sim \mathcal{N} \left( \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2}, \frac{1}{n} \right) \]

By Law of Large Numbers \( \text{IDENTICALLY DISTRIBUTED} \)

\[ \left( \text{for } x_1, x_2, \ldots, x_n \text{ i.i.d with } \mathbb{E}(x_i) = \mu, \right. \]
\[ \frac{1}{n} \sum_{i=1}^{n} x_i \to \mu \text{ as } n \to \infty \]

\[ = -N \ell(\theta) \]
\[ = -I(\theta) \]

\( \Rightarrow \text{ SCORING METHOD IS} \)

\[ \theta_{k+1} = \theta_k + I^{-1}(\theta_k) \frac{\partial \ln p(x; \theta)}{\partial \theta} \bigg|_{\theta = \theta_k} \]

\( \text{GENERAL HAS SAME CONVERGENCE PROBLEMS BUT MAY WORK BETTER THAN } N-R. \)

**EXAMPLE** \( x(t) = r(t) + \epsilon(t) \)

\[ \uparrow \text{ WGN} \]

\( \text{SEE EXAMPLE 7.11} \)

\[ I(\theta) = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} h^{n+2n+2} \]

\[ \Rightarrow \theta_{k+1} = \theta_k + \frac{\sum_{n=0}^{N-1} (x(t_n) - \theta_k h^{n}) h^{n+2n+2}}{\sum_{n=0}^{N-1} h^{n+2n+2}} \]
COMPUTER EXAMPLE: N-R FOR

\[ N = 50, \quad r = 0.5, \quad \sigma^2 = 0.01. \]

WISH TO MAXIMIZE

\[ \ln p(x; r) = \ln \left( \frac{1}{(2\pi\sigma^2)^{N/2}} \right) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x - \ln(r))^2 \]

\[ \text{over } r \quad \text{or} \quad -\sum_{n=0}^{N-1} (x - \ln(r))^2 = -J(r) \]

\[ -J(r) \]

SHARP DUE TO $r^n$

FAIRLY BROAD

TYPICAL REALIZATION

PEAK AT

\[ \hat{r} = 0.493 \]

USING A N-R FOR SEVERAL $r_0$. OR $r_0$ IF

FOR $r_0 > 1.2$ AN ITERATION "BLEW UP" DUE TO OVERFLOW ($r^n$ TERM).
<table>
<thead>
<tr>
<th>Iteration</th>
<th>Initial Guess, $\theta_0$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.8</td>
<td>0.2</td>
<td>1.2</td>
</tr>
<tr>
<td>1</td>
<td>0.723</td>
<td>0.799</td>
<td>1.187</td>
</tr>
<tr>
<td>2</td>
<td>0.638</td>
<td>0.722</td>
<td>1.174</td>
</tr>
<tr>
<td>3</td>
<td>0.561</td>
<td>0.637</td>
<td>1.161</td>
</tr>
<tr>
<td>4</td>
<td>0.510</td>
<td>0.560</td>
<td>1.148</td>
</tr>
<tr>
<td>5</td>
<td>0.494</td>
<td>0.510</td>
<td>1.136</td>
</tr>
<tr>
<td>6</td>
<td>0.493</td>
<td>0.494</td>
<td>1.123</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.493</td>
<td>1.111</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>1.109</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>1.098</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>1.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.493</td>
</tr>
</tbody>
</table>

*Third method is EM - most useful for vector parameter.*

**Vector Parameter**

MLE is that value of $\theta$ that maximizes $p(x; \theta)$. Can try

$$\frac{\partial \ln p(x; \theta)}{\partial \theta} = 0$$
AND PICK SOLUTION THAT PRODUCES MAXIMUM.

**Example:** \( X(n) = A + W(n) \quad n = 0, 1, \ldots, N-1 \) \( WGN \)

Find MLE of \( \theta = (A, \sigma^2)^T \)

\[
\rho(x ; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum \left( x(n) - A \right)^2}
\]

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

\[
\Rightarrow \quad \frac{1}{\sigma^2} \sum \left( x(n) - A \right) = 0 \quad \Rightarrow \quad \hat{A} = \bar{x}
\]

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

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

\[
\Rightarrow \quad \sigma^2 = \frac{1}{N} \sum \left( x(n) - \hat{A} \right)^2
\]

OR

\[
\hat{\theta} = \begin{bmatrix} \hat{A} \\ \frac{1}{N} \sum \frac{X(n) - \hat{A}}{\sigma^2} \end{bmatrix}
\]

UNBIASED?

Should check to be sure this maximizes \( \rho(x ; \theta) \). How?
Asymptotic properties of MLE:

\[
\hat{\theta} \xrightarrow{\mathcal{D}} N(\theta, \frac{1}{N} I^{-1}(\theta)) \quad \text{as } N \to \infty
\]

**Example:** See previous one

\[
\hat{A} = \bar{X} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (X_n - \bar{X})^2
\]

According to theorem

\[
\begin{bmatrix}
\hat{A} \\
\hat{\sigma}^2
\end{bmatrix} \xrightarrow{\mathcal{D}} N \left( \begin{bmatrix}
A \\
\sigma^2
\end{bmatrix}, \begin{bmatrix}
\sigma^2 / N & 0 \\
0 & 2\sigma^4 / N
\end{bmatrix} \right)
\]

OR \( \hat{A}, \hat{\sigma}^2 \) ARE UNCORRELATED \( \Rightarrow \) INDEPENDENT

To verify this recall

\[
\hat{A} = \bar{X} \sim N(\theta, \frac{\sigma^2}{N})
\]

It can be shown that (for finite \( N \))

\[
T = \frac{1}{\hat{\sigma}^2} \sum_{n=0}^{N-1} (X_n - \bar{X})^2 \quad \xrightarrow{\text{cht sqared}} \chi^2_{N-1} \quad \text{PDF with } \ N-1 \text{ degrees of freedom}
\]

And \( T, \bar{X} \) ARE INDEPENDENT
$\chi^2_n$ has PDF

\[ p(u) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} u^{n/2-1} e^{-u/2} & u > 0 \\ 0 & u \leq 0 \end{cases} \]

Also, if $y = \frac{1}{2} \sum_{i=1}^{n} x_i^2$, $x_i \sim N(0,1)$ i.i.d.

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

As $n \to \infty$, $\chi^2_n$ PDF becomes Gaussian due to Central Limit Theorem.

Also $E(\chi^2_n) = n$

$\text{VAR}(\chi^2_n) = 2n$.

$\Rightarrow T \sim \chi^2_{n-1} \Rightarrow N\left( \frac{n-1}{2}, \frac{2}{n} \right)$

But $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{n} (X_i - \bar{X})^2$

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

\sim N\left( \sigma^2, 2\sigma^4/n \right)$

Furthermore, $\hat{\sigma}$ 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.

\[ x(n) = s(n) + w(n) \quad n = 0, 1, \ldots, N-1 \]

\[ w(n) \text{ i.i.d.} \]

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

Laplacian noise

Estimate \( \hat{\theta} = (s(0), s(1), \ldots, s(N-1))^T \).

\( x(n) \) has pdf

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

MLE is \( \hat{s}(n) = x(n) \quad n = 0, 1, \ldots, N-1 \)

or \( \hat{\theta} = x \)

Pdf of \( \hat{\theta} \) not Gaussian. It is

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

Problem is no averaging possible due to too many parameters \( \Rightarrow \) central limit theorem violated (see Appendix 7B).
INVARINANCE PROPERTY

If \( \mathbf{x} = q(\mathbf{\theta}) \) is a dimensional function, then MLE of \( \mathbf{x} \) is

\[
\hat{\mathbf{x}} = q(\hat{\mathbf{\theta}})
\]

MLE of \( \mathbf{\theta} \)

If \( q \) is not invertible, then \( \hat{\mathbf{x}} \) maximizes the modified likelihood function

\[
\hat{\mathbf{\theta}}_{\mathbf{x}}(\mathbf{x}; \mathbf{x}) = \max_{\mathbf{\theta} \colon \mathbf{x} = q(\mathbf{\theta})} p(x; \mathbf{\theta})
\]

\( \forall \mathbf{\theta} \) that map into same \( \mathbf{x} \)

To find MLE for general Gaussian case or \( \mathbf{x} \sim N(\mathbf{m}(\mathbf{\theta}), \mathbf{C}(\mathbf{\theta})) \) we set partials equal to zero or

From Appendix 3C

\[
\frac{\partial \ln p(x; \mathbf{\theta})}{\partial \mathbf{\theta}_k} = -\frac{1}{2} \text{tr} \left[ \mathbf{C}^{-1}(\mathbf{\theta}) \frac{\partial \mathbf{C}(\mathbf{\theta})}{\partial \mathbf{\theta}_k} \right]
\]