Chapter 7
Maximum Likelihood Estimate (MLE)

Motivation for MLE

Problems: **1. MVUE often does not exist or can't be found**

<See Ex. 7.1 in the textbook for such a case>

2. BLUE may not be applicable ($x \neq H\theta + w$)

Solution: *If* the PDF is known, then MLE can always be used!!!

This makes the MLE one of the most popular practical methods

- Advantages:
  1. It is a "Turn-The-Crank" method
  2. "Optimal" for large enough data size

- Disadvantages:
  1. Not optimal for small data size
  2. Can be computationally complex
     - may require numerical methods

Rationale for MLE

Choose the parameter value that:

makes the data you did observe…

the most likely data to have been observed!!!

Consider 2 possible parameter values: θ_1 & θ_2

Ask the following: If θ_i were really the true value, what is the probability that I would get the data set I really got ?

Let this probability be P_i

So if P_i is small… it says you actually got a data set that was unlikely to occur! Not a good guess for θ_i!!!

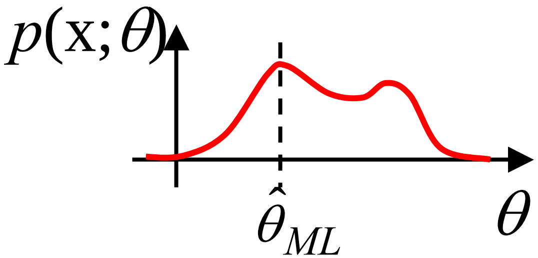
But $p_1 = p(\mathbf{x};\theta_1)\, d\mathbf{x}$

$p_2 = p(\mathbf{x};\theta_2)\, d\mathbf{x}$

$\Rightarrow$ **pick** $\hat{\theta}_{ML}$ **so that** $p(\mathbf{x};\hat{\theta}_{ML})$ **is largest**

Definition of the MLE

$\hat{\theta}_{ML}$ is the value of θ that <u>maximizes</u> the "Likelihood Function" $p(\mathrm{x};\theta)$ for the <u>specific measured</u> data x



$\hat{\theta}_{ML}$ maximizes the likelihood function

<u>Note</u>: Because ln(z) is a monotonically increasing function…

$\hat{\theta}_{ML}$ maximizes the <u>log</u> likelihood function $\ln\{p(\mathbf{x};\theta)\}$

General Analytical Procedure to Find the MLE

1. Find log-likelihood function: $\ln p(\mathbf{x};\theta)$
2. Differentiate w.r.t θ and set to 0: $\partial \ln p(\mathbf{x};\theta)/\partial\theta = 0$
3. Solve for θ value that satisfies the equation

Ex. 7.3: Ex. of MLE When MVUE Non-Existent

$$x[n] = A + w[n] \qquad \Rightarrow x[n] \sim N(A,A)$$

$A > 0$

WGN $\sim N(0,A)$

Likelihood Function: $p(\mathbf{x};A) = \frac{1}{(2\pi A)^{\frac{N}{2}}} \exp\left[-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$

To take *ln* of this… use log properties:

Take $\partial/\partial A$, set = 0, and change A to $\hat{A}$

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

Expand this:

$$-\frac{N}{2\hat{A}} + \frac{1}{\hat{A}}\sum x[n] - \frac{1}{\hat{A}}N\hat{A} + \frac{1}{2\hat{A}}\sum x^2[n] - \frac{1}{2\hat{A}^2}2\hat{A}\sum x[n] + \frac{\hat{A}^2 N}{2\hat{A}^2} = 0$$

Cancel

Manipulate to get: $\hat{A}^2 + \hat{A} - \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = 0$

Solve quadratic equation to get MLE:

$$\hat{A}_{ML} = -\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}$$

Can show this estimator biased (see bottom of p. 160)

But it is asymptotically unbiased…

Use the "Law of Large Numbers":

Sample Mean → True Mean $\quad \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \xrightarrow[as\ N\to\infty]{} E\{x^2[n]\}$

So can use this to show:

$$E\{\hat{A}_{ML}\} \to E\left\{-\frac{1}{2} + \sqrt{E\{x^2[n]\} + \frac{1}{4}}\right\} = -\frac{1}{2} + \sqrt{\underbrace{E\{x^2[n]\}}_{=A^2+A} + \frac{1}{4}} = A$$

$$\text{var}(\hat{A}) \to \frac{A^2}{N\left(A + \frac{1}{2}\right)} = CRLB$$

Asymptotically…Unbiased & Efficient

7.5 Properties of the MLE (or… "Why We Love MLE")

The MLE is **asymptotically**:

1. unbiased
2. efficient (i.e. achieves CRLB)
3. Gaussian PDF

Also, if a truly efficient estimator exists, then the ML procedure finds it !

The **asymptotic properties** are captured in **Theorem 7.1**:

If $p(\mathbf{x};\theta)$ satisfies some "regularity" conditions, then the MLE is asymptotically distributed according to

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

where $\mathbf{I}(\theta)$ = Fisher Information Matrix

Size of *N* to Achieve Asymptotic

This Theorem only states what happens asymptotically…
when N is small there is no guarantee how the MLE behaves

Q: How large must N be to achieve the asymptotic properties?

A: In practice: use “Monte Carlo Simulations” to answer this

Monte Carlo Simulations: see Appendix 7A

A methodology for doing computer simulations to evaluate performance of any estimation method

Not just for the MLE!!!

Illustrate for deterministic signal $s[n; \theta]$ in AWGN

Monte Carlo Simulation:

Data Collection:

1. Select a particular true parameter value, θ_{true}
   - you are often interested in doing this for a variety of values of θ so you would run one MC simulation for each θ value of interest
2. Generate signal having true θ: $s[n;\theta_t]$ (call it s in matlab)
3. Generate WGN having unit variance

   w = randn (size(s));
4. Form measured data: x = s + sigma*w;
   - choose σ to get the desired SNR
   - usually want to run at many SNR values

     → do one MC simulation for each SNR value

Data Collection (Continued):

5. Compute estimate from data **x**

6. Repeat steps 3-5 *M* times

   - (call *M* "# of MC runs" or just "# of runs")

7. Store all M estimates in a vector EST (assumes scalar θ)

Statistical Evaluation:

1. Compute bias

$$b = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta_{true}\right)$$

2. Compute error RMS

$$RMS = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta_t\right)^2}$$

3. Compute the error Variance

$$VAR = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \left(\frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_i\right)\right)^2$$

4. Plot Histogram or Scatter Plot (if desired)

Now explore (via plots) how: Bias, RMS, and VAR vary with:

θ value, SNR value, *N* value, Etc.

Is B ≈ 0 ?

Is RMS ≈ $(\text{CRLB})^{1/2}$?

Ex. 7.6: Phase Estimation for a Sinusoid

Some Applications:

1. Demodulation of phase coherent modulations (e.g., DSB, SSB, PSK, QAM, etc.)
2. Phase-Based Bearing Estimation

Signal Model: $x[n] = A\cos(2\pi f_o n + \phi) + w[n], \qquad n = 0, 1, \ldots, N-1$

White $\sim N(0,\sigma^2)$

A and f_o known, ϕ unknown

Recall CRLB:

$$\operatorname{var}\left(\hat{\phi}\right) \geq \frac{2\sigma^2}{NA^2} = \frac{1}{N \cdot SNR}$$

For this problem… all methods for finding the MVUE will fail!!
$\Rightarrow$ So… try MLE!!

So first we write the **likelihood function**:

$$p(\mathbf{x};\phi) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \underbrace{\sum_{n=0}^{N-1}\left[x[n] - A\cos\left(2\pi f_o n + \phi\right)\right]^2} \right\}$$

GOAL: Find ϕ that maximizes this → … equivalent to minimizing this

End up in same place if we maximize LLF

So, minimize: $J(\phi) \triangleq \sum_{n=0}^{N-1}\left[x[n] - A\cos\left(2\pi f_o n + \phi\right)\right]^2$ Setting $\frac{\partial J(\phi)}{\partial \phi} = 0$ gives

$$\sum_{n=0}^{N-1} x[n]\sin\left(2\pi f_o n + \hat{\phi}\right) = A \underbrace{\sum_{n=0}^{N-1} \sin\left(2\pi f_o n + \hat{\phi}\right)\cos\left(2\pi f_o n + \phi\right)}_{\approx 0}$$

sin and cos are ⊥ when summed over full cycles

So… MLE Phase Estimate satisfies:

$$\sum_{n=0}^{N-1} x[n]\sin\left(2\pi f_o n + \hat{\phi}\right) = 0$$

Interpret via inner product or correlation

Now…using a Trig Identity and then re-arranging gives:

$$\cos(\hat{\phi})\left[\sum_n x[n]\sin(2\pi f_o n)\right] = -\sin(\hat{\phi})\left[\sum_n x[n]\cos(2\pi f_o n)\right]$$

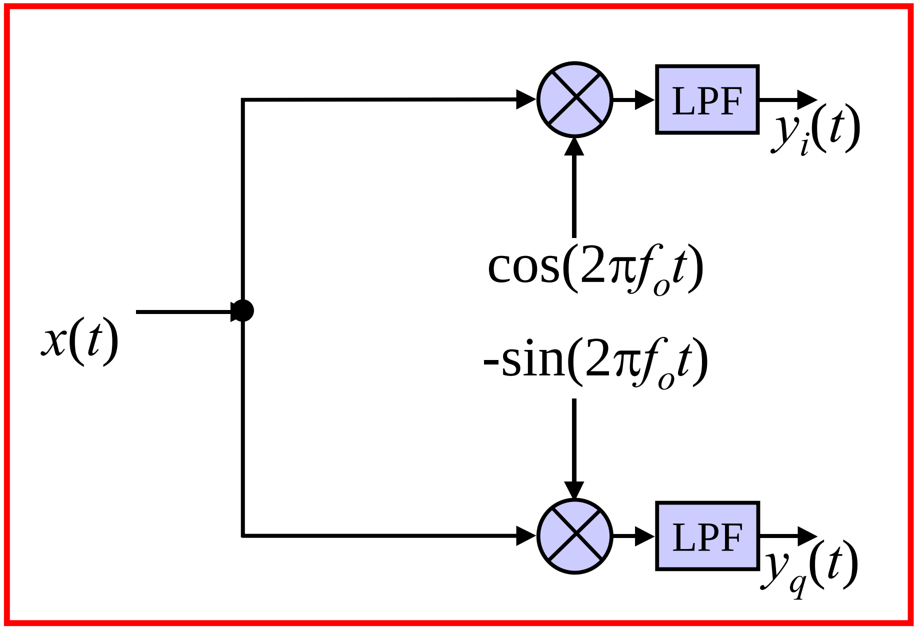
Or…

$$\hat{\phi}_{ML} = -\tan^{-1}\left[\frac{\sum_n x[n]\sin(2\pi f_o n)}{\sum_n x[n]\cos(2\pi f_o n)}\right]$$

Recall: This is the *approximate* MLE

Don't need to know A or σ^2 but do need to know f_o

Recall: I-Q Signal Generation



The "sums" in the above equation play the role of the LPF's in the figure (why?)
Thus, ML phase estimator can be viewed as: atan of ratio of Q/I

Monte Carlo Results for ML Phase Estimation

See figures 7.3 & 7.4 in text book