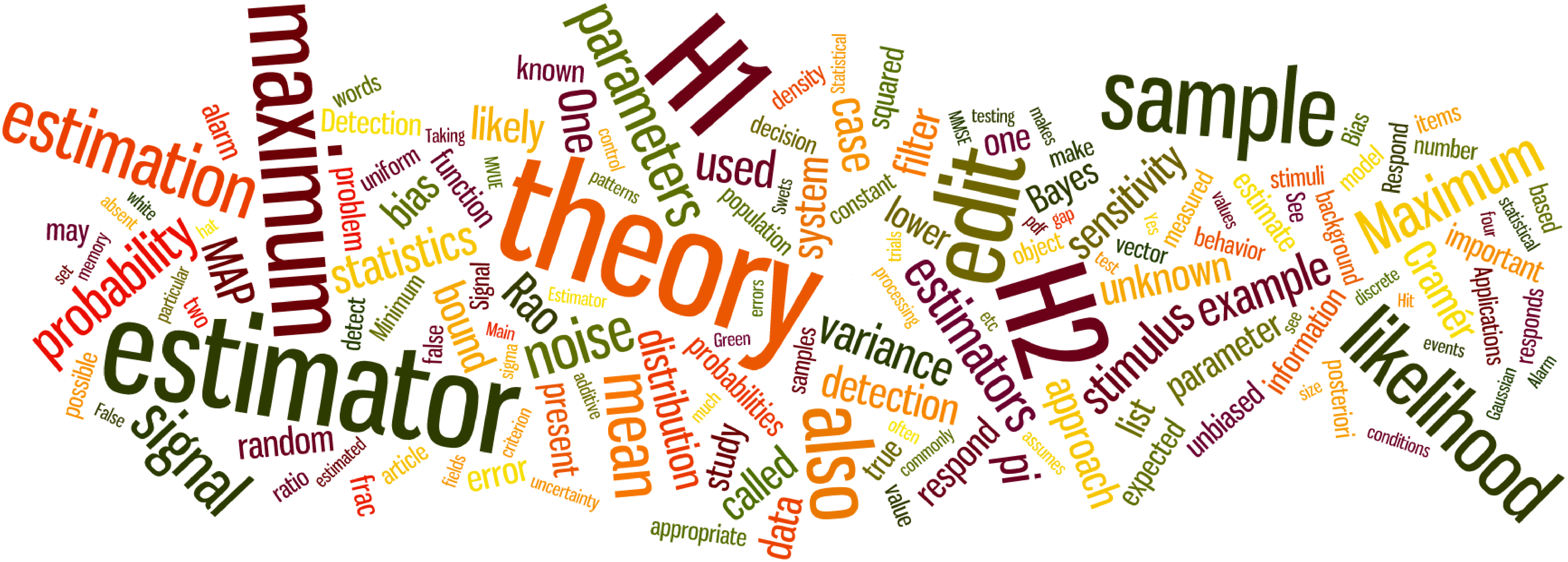


Detection and Estimation Theory
Lecture 10

Mojtaba Soltanalian- UIC
msol@uic.edu
http://msol.people.uic.edu

Based on ECE 531 Slides- 2011 (Prof. Natasha Devroye)

Maximum-Likelihood (ML) Estimation

Minimum variance unbiased estimators are often difficult to find, in which case we can resort to sub-optimal, but tractable estimators such as the Maximum Likelihood Estimator (MLE). This estimator is widely used in part because of its simplicity (you need to find the maximum of a function which may be done either analytically or numerically) as well as its asymptotic efficiency (as the number of observations $\to \infty$).

Maximum-Likelihood (ML) Estimation

"If the probability of an event x dependent on model parameters p is written

$$p (x \mid \theta)$$

then we would talk about the likelihood

$$L = p (\theta \mid x)$$

that is, the likelihood of the parameters given the data."

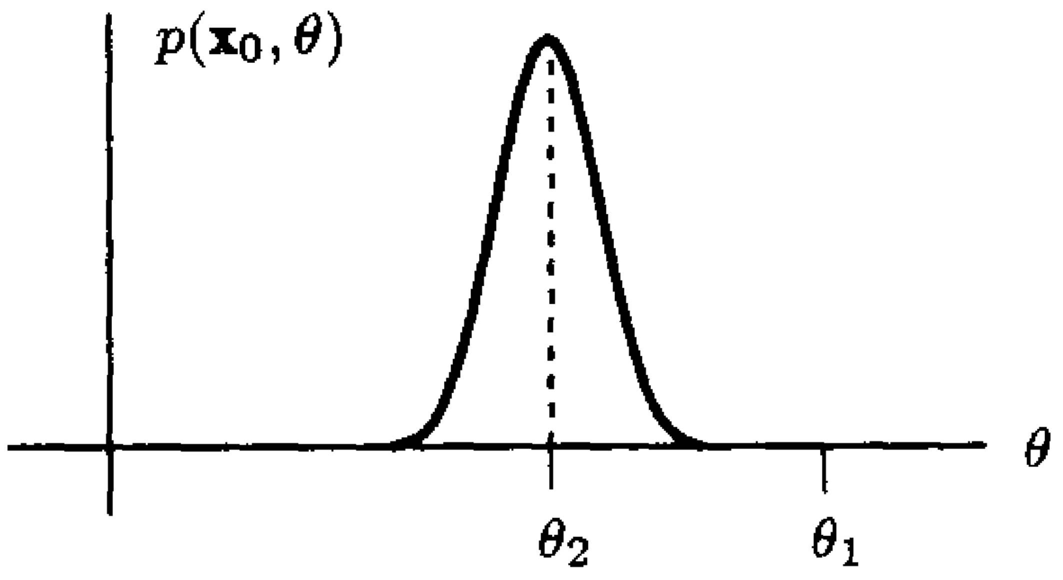
Maximum-Likelihood (ML) Estimation

Consider estimating θ from a set of observations $\mathbf{x}$. The *likelihood function* $p(\mathbf{x};\theta)$ is a function of θ for a given, fixed $\mathbf{x}$. The MLE is the value of θ that maximizes the likelihood function, that is

$$\text{Maximum Likelihood Estimator:} \qquad \hat{\theta} = \arg\max_{\theta} p(\mathbf{x};\theta)$$

Maximum-Likelihood (ML) Estimation

Note: The 2D nature of the $p(x;\theta)$.



Maximum-Likelihood (ML) Estimation
A motivational example

Consider estimating A from the data $x[n] = A + w[n], \quad w[n] \sim \mathcal{N}(0, A), \quad A > 0.$
Try finding the MVUE and the ML estimator.

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

$$\Downarrow$$

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} &= -\frac{N}{2A} + \frac{1}{A}\sum_{n=0}^{N-1}(x[n] - A) + \frac{1}{2A^2}\sum_{n=0}^{N-1}(x[n] - A)^2 \\ &\stackrel{?}{=} I(A)(\hat{A} - A).\end{aligned}$$

Maximum-Likelihood (ML) Estimation
A motivational example

Consider estimating A from the data $x[n] = A + w[n], \quad w[n] \sim \mathcal{N}(0, A), \quad A > 0.$
Try finding the MVUE and the ML estimator.

$$p(\mathbf{x}; A) = \underbrace{\frac{1}{(2\pi A)^{\frac{N}{2}}} \exp\left[-\frac{1}{2}\left(\frac{1}{A}\sum_{n=0}^{N-1} x^2[n] + NA\right)\right]}_{g\left(\sum_{n=0}^{N-1} x^2[n],\, A\right)} \underbrace{\exp(N\bar{x})}_{h(\mathbf{x})}$$

$$\Downarrow$$

$$E\left[g\left(\sum_{n=0}^{N-1} x^2[n]\right)\right] = A \qquad \text{for all } A > 0.$$

Since

$$\begin{aligned} E\left[\sum_{n=0}^{N-1} x^2[n]\right] &= NE[x^2[n]] \\ &= N\left[\operatorname{var}(x[n]) + E^2(x[n])\right] \\ &= N(A + A^2) \end{aligned}$$

it is not obvious how to choose g.

Maximum-Likelihood (ML) Estimation
A motivational example

Consider estimating A from the data $x[n] = A + w[n], \quad w[n] \sim \mathcal{N}(0, A), \quad A > 0.$
Try finding the MVUE and the ML estimator.

Faced with our inability to find the MVU estimator we will propose an estimator that is *approximately optimal*. We claim that for large data records or as $N \to \infty$, the proposed estimator is efficient. This means that as $N \to \infty$

$$E(\hat{A}) \to A, \quad \Longrightarrow \quad \textit{asymptotically unbiased}$$

$$\text{var}(\hat{A}) \to \text{CRLB}, \quad \Longrightarrow \quad \textit{asymptotically efficient}$$

For finite data records, however, we can say nothing about its optimality.

Better estimators may exist, but finding them may not be easy!

Maximum-Likelihood (ML) Estimation

- the MLE is asymptotically unbiased (i.e. consistent: converges in probability to the true parameter value)
- the MLE is asymptotically efficient (meets the CRLB)
- the MLE is asymptotically distributed as $\mathcal{N}(\theta, I^{-1}(\theta))$, where $I(\theta)$ is the Fisher information.
- if an efficient estimator exists, the MLE finds it

The above statements hold under the regularity conditions:

- The existence of $\frac{\partial \ln p(x;\theta)}{\partial \theta}$ and $\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}$
- $E\left[\frac{\partial \ln p(x;\theta)}{\partial \theta}\right] = 0$

PROOF!

Maximum-Likelihood (ML) Estimation
Example for the last property

When we have a linear model, meaning $\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$, where $\mathbf{H}$ is a known $N \times p$ matrix with $N > p$ and rank p, and $\mathbf{w} \sim \mathcal{N}(0, \mathbf{C})$, then the MLE of θ is

$$\hat{\theta} = \left(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{\mathbf{T}} \mathbf{C}^{-1} \mathbf{x}, \quad \hat{\theta} \sim \mathcal{N}(\mathbf{0}, (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1})$$

Thus, in the linear model, the MLE is efficient and MVUE!

Maximum-Likelihood (ML) Estimation

These results are asymptotic as the number of observations $N \to \infty$. What finite N is large enough for these nice asymptotic results to start to hold is an important question. There is unfortunately no general answer. An important tool for determining how large N should be (or rather for getting a feel for it) is Monte Carlo.

ML Invariance Property

The MLE also has what's known as an *invariance property*, i.e. the MLE of the parameter $\alpha = g(\theta)$ where pdf $p(\mathbf{x};\theta)$ parameterized by θ is

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

where $\hat{\theta}$ is the MLE of θ. The MLE of $\hat{\theta}$ is obtained by maximizing $p(\mathbf{x};\theta)$. If $g(\cdot)$ is not a 1:1 function, then $\hat{\alpha}$ maximizes the modified likelihood function

$$p_T(\mathbf{x};\alpha) = \max_{\theta:\alpha=g(\theta)} p(\mathbf{x};\theta)$$

ML Invariance Property

– Given the data

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

where $w[n]$ is WGN with variance σ^2, find the MLE of $\alpha = exp(A)$.

– Given the data

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

where $w[n]$ is WGN with variance σ^2, find the MLE of $\alpha = A^2$.