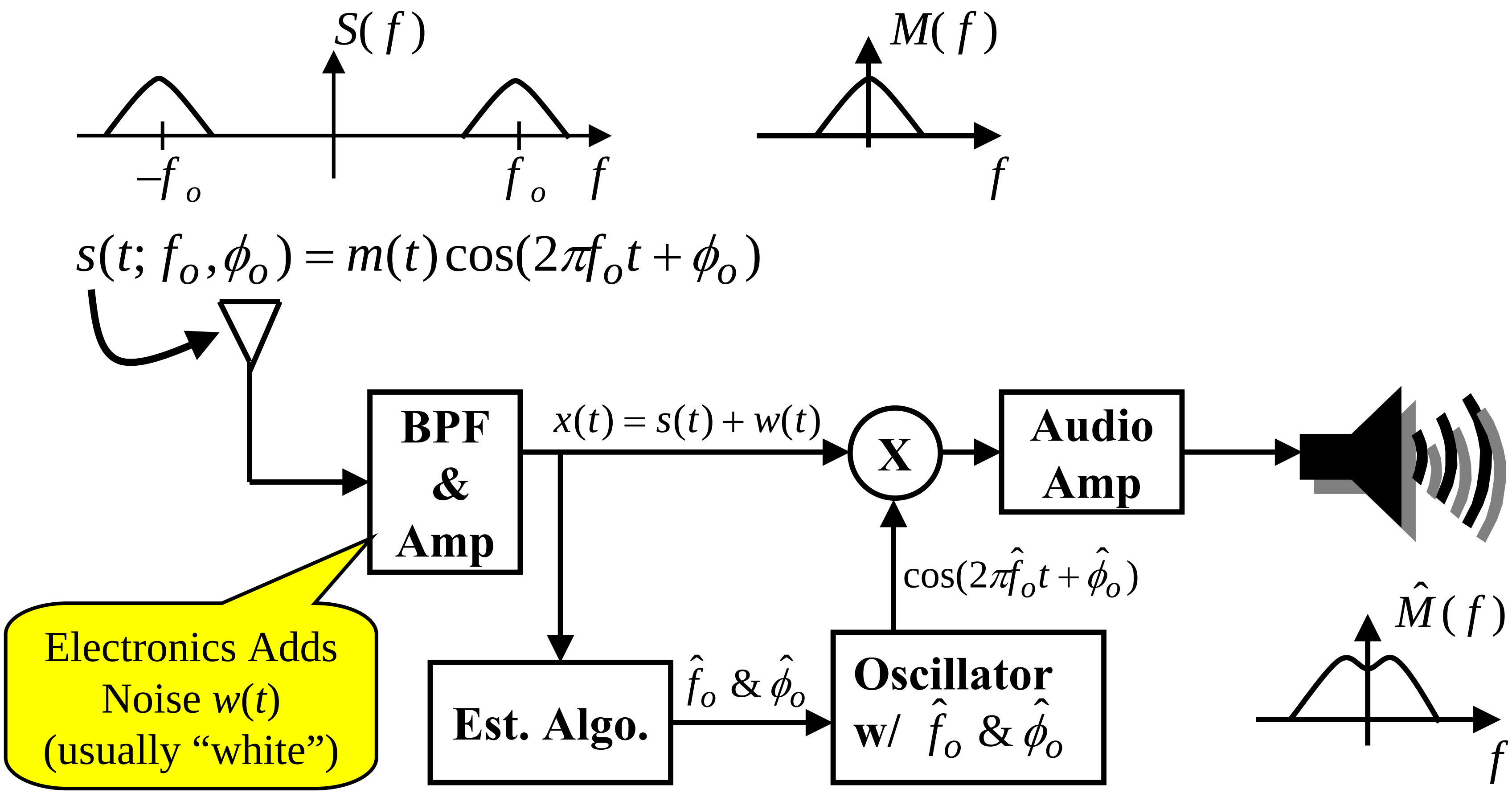
Ch. 1 Introduction to Estimation

An Example Estimation Problem: DSB Rx



Goal: Given $x(t) = s(t; f_o, \phi_o) + w(t)$

Find Estimates
(that are optimal in some sense)

Describe with Probability Model: PDF & Correlation

Discrete-Time Estimation Problem

These days, almost always work with samples of the observed signal (signal plus noise):

$$\boxed{x[n] = s[n; f_o, \phi_o] + w[n]}$$

Our "Thought" Model: Each time you "observe" $x[n]$ it contains same $s[n]$ but different "realization" of noise $w[n]$, so the estimate is different each time. ⇨ $\hat{f}_o$ & $\hat{\phi}_o$ are RVs

Our Job: Given finite data set $x[0]$, $x[1]$, … $x[N$-1]
Find estimator functions that map data into estimates:

$$\hat{f}_o = g_1(x[0], x[1], \ \ldots, x[N-1]) = g_1(\mathbf{x})$$

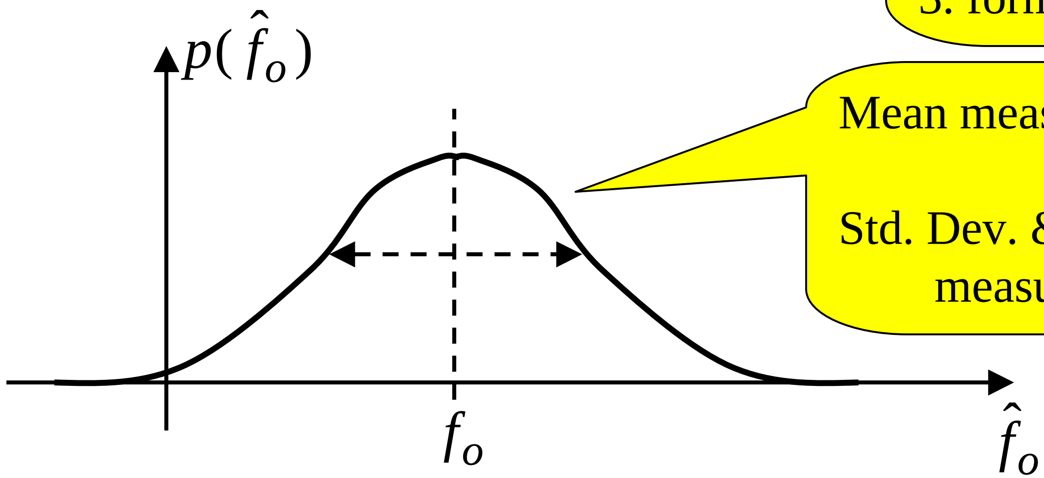
$$\hat{\phi}_o = g_2(x[0], x[1], \ \ldots, x[N-1]) = g_2(\mathbf{x})$$

These are RVs… Need to describe w/ probability model

PDF of Estimate

Because estimates are RVs we describe them with a PDF…

Will depend on:
1. structure of $s[n]$
2. probability model of $w[n]$
3. form of est. function $g(\mathbf{x})$



Mean measures centroid

Std. Dev. & Variance measure spread

Desire: $E\{\hat{f}_o\} = f_o$

$$\sigma^2_{\hat{f}_o} = E\left\{\left(\hat{f}_o - E\{\hat{f}_o\}\right)^2\right\} = \text{small}$$

1.2 Mathematical Estimation Problem

General Mathematical Statement of Estimation Problem:

For… Measured Data $\mathbf{x} = [\ x[0]\ x[1]\ \ldots\ x[N\text{-}1]\ ]$

Unknown Parameter $\theta = [\theta_1\ \ \theta_2\ \ \ldots\ \ \theta_p\ ]$

θ is Not Random

$\mathbf{x}$ is an N-dimensional random data vector

Q: What captures all the statistical information needed for an estimation problem ?

A: Need the N-dimensional PDF of the data, parameterized by θ

In practice, not given PDF!!!

Choose a suitable model

- **Captures Essence of Reality**
- **Leads to Tractable Answer**

$$\boxed{p(\mathbf{x};\boldsymbol{\theta})}$$

We'll use $p(\mathbf{x};\theta)$ to find $\hat{\boldsymbol{\theta}} = g(\mathbf{x})$

Ex. Estimating a DC Level in Zero Mean AWGN

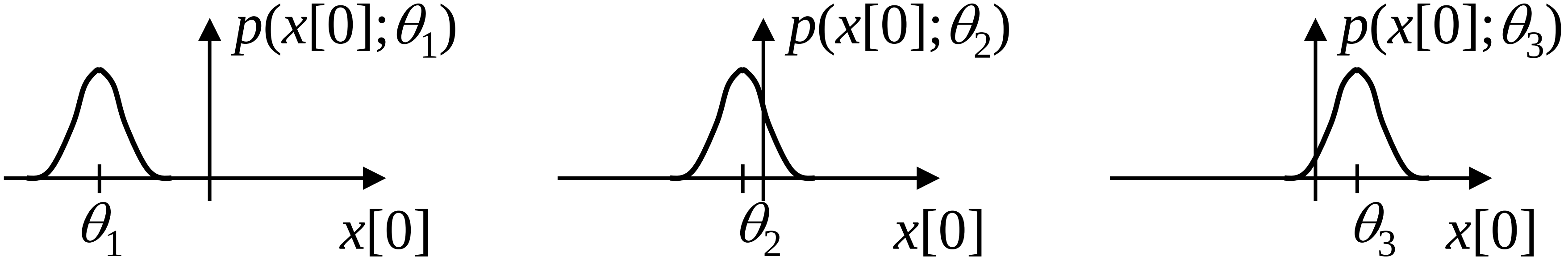
Consider a single data point is observed $x[0] = \underbrace{\theta + w[0]}_{\boxed{\sim N(\theta,\, \sigma^2)}}$

Gaussian
zero mean
variance σ^2

So… the needed parameterized PDF is:

$p(x[0];\theta)$ which is Gaussian with mean of θ

So… in this case the parameterization changes the data PDF mean:



Ex. Modeling Data with Linear Trend

See Fig. 1.6 in Text

Looking at the figure we see what looks like a linear trend perturbed by some noise…

So the engineer proposes signal and noise models:

$$x[n] = \underbrace{\left[A + Bn\right]}_{s[n;A,B]} + w[n]$$

Signal Model: Linear Trend **Noise Model: AWGN w/ zero mean**

AWGN = "Additive White Gaussian Noise"
"White" = $x[n]$ and $x[m]$ are uncorrelated for $n \neq m$ $\rightarrow$ $E\left\{(\mathbf{w} - \overline{\mathbf{w}})(\mathbf{w} - \overline{\mathbf{w}})^T\right\} = \sigma^2 \mathbf{I}$

Typical Assumptions for Noise Model

- W and G is always easiest to analyze
  - Usually assumed unless you have reason to believe otherwise
  - Whiteness is usually first assumption removed
  - Gaussian is less often removed due to the validity of Central Limit Thm
- Zero Mean is a nearly universal assumption
  - Most practical cases have zero mean
  - But if not… $w[n] = w_{zm}[n] + \mu$

Non-Zero Mean of μ

Zero Mean

Now group into signal model

- Variance of noise doesn't always have to be known to make an estimate
  - BUT, must know to assess expected "goodness" of the estimate
  - Usually perform "goodness" analysis as a function of noise variance (or SNR = Signal-to-Noise Ratio)
  - Noise variance sets the SNR level of the problem

Classical vs. Bayesian Estimation Approaches

If we view θ (parameter to estimate) as Non-Random

$\rightarrow$ Classical Estimation

Provides no way to include *a priori* information about θ

If we view θ (parameter to estimate) as Random

$\rightarrow$ Bayesian Estimation

Allows use of some *a priori* PDF on θ

The first part of the course: Classical Methods

- Minimum Variance, Maximum Likelihood, Least Squares

Last part of the course: Bayesian Methods

- MMSE, MAP, Wiener filter, Kalman Filter

1.3 Assessing Estimator Performance

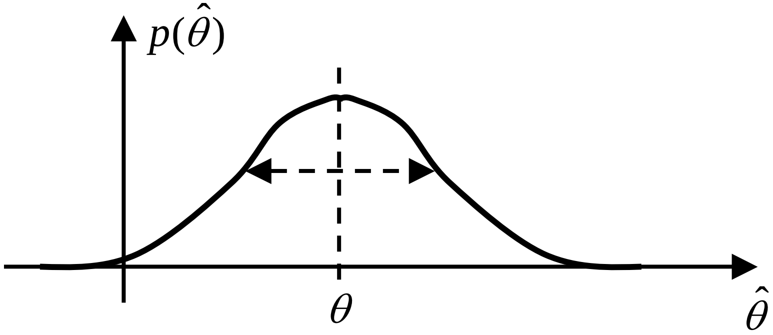
Can only do this when the value of θ is known:

- Theoretical Analysis, Simulations, Field Tests, etc.

Recall that the estimate $\hat{\theta} = g(\mathbf{x})$ is a random variable

Thus it has a PDF of its own… and that PDF <u>completely</u> displays the quality of the estimate.

Illustrate with 1-D parameter case



Often just capture quality through <u>mean</u> and <u>variance</u> of $\hat{\theta} = g(\mathbf{x})$

Desire: $m_{\hat{\theta}} = E\{\hat{\theta}\} = \theta$

$$\sigma^2_{\hat{\theta}} = E\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\} = \text{small}$$

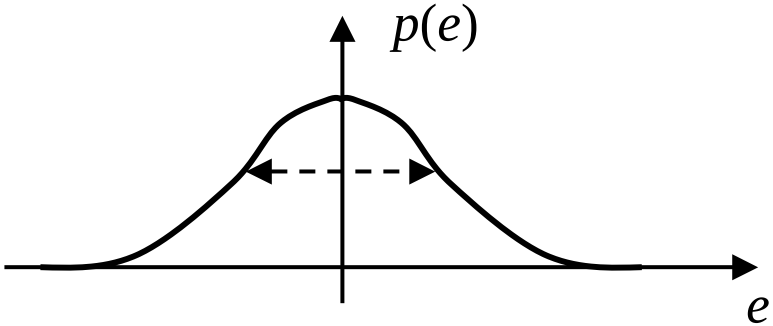
If this is true: say estimate is “unbiased

Equivalent View of Assessing Performance

Define estimation error: $e = \hat{\theta} - \theta$ $\qquad (\hat{\theta} = \theta + e)$

RV RV Not RV

Completely describe estimator quality with error PDF: $p(e)$



Desire:

$$m_e = E\{e\} = 0$$

$$\sigma_e^2 = E\left\{\left(e - E\{e\}\right)^2\right\} = \text{small}$$

If this is true: say estimate is "unbiased

Example: DC Level in AWGN

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

Gaussian, zero mean, variance σ^2
White (uncorrelated sample-to-sample)

PDF of an individual data sample:

$$\boxed{p(x[i]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[i]-A)^2}{2\sigma^2} \right]}$$

Uncorrelated Gaussian RVs are Independent…
so joint PDF is the product of the individual PDFs:

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

(***property: prod of exp's gives sum inside exp***)

Prob. Theory

Each data sample has the same mean (A), which is the thing we are trying to estimate… so, we can imagine trying to estimate A by finding the sample mean of the data:

Statistics

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

Let's analyze the quality of this estimator…

- Is it unbiased?

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

$$\Rightarrow \quad E\{\hat{A}\} = A$$

Yes! Unbiased!

- Can we get a small variance?

Due to Indep.
(white & Gauss.
$\Rightarrow$ Indep.)

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

$$\Rightarrow \quad \mathrm{var}(\hat{A}) = \frac{\sigma^2}{N}$$

Can make var small by increasing *N*!!!

Theoretical Analysis vs. Simulations

- Ideally we'd like to be always be able to theoretically analyze the problem to find the bias and variance of the estimator
  - Theoretical results show how performance depends on the problem specifications
- But sometimes we make use of simulations
  - to verify that our theoretical analysis is correct
  - sometimes can't find theoretical results

Course Goal = Find "Optimal" Estimators

- There are several different definitions or criteria for optimality!
- Most Logical: Minimum MSE (Mean-Square-Error)
  - See Sect. 2.4
  - To see this result:

$$mse(\hat{\theta}) = E\left\{\left(\hat{\theta} - \theta\right)^2\right\}$$

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

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

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

$$mse(\hat{\theta}) = E\left\{\left(\hat{\theta} - \theta\right)^2\right\}$$

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

Bias

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

Although MSE makes sense, estimates usually rely on θ