ECE 830 Fall 2011 Statistical Signal Processing

instructor: R. Nowak

Lecture 13: Parameter Estimation

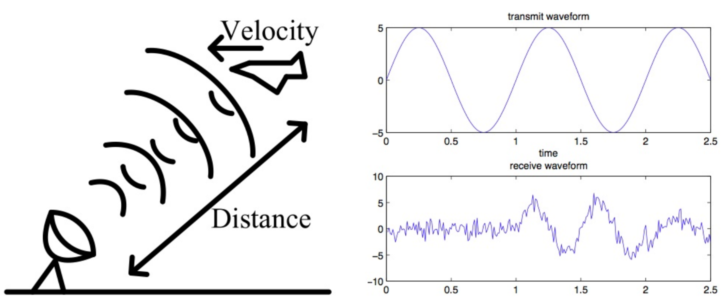
We observe $X \sim p(x|\theta), \theta \in \Theta$, and the goal is to determine the θ that produced X.

1 Generative Model:

Select $\theta \in \Theta \;\rightarrow$ randomly generate $X \sim p(x|\theta)$, i.e. "draw X from $p(x|\theta)$"

- Estimator
  Observe $X \sim p(x|\theta) \xrightarrow{infer} \theta \in \Theta$
  Estimation is a sort of "inverse" problem

2 Radar Example:



Radar is used to estimate the distance of an object

The received waveform is time-dilated and shifted version of original waveform $g(t)$ plus noise

$$x(t) = g(\alpha t - \tau) + w(t) \ .$$

The parameter of interest is $\theta = \begin{bmatrix} \alpha \\ \tau \end{bmatrix}$, where α is related to velocity / Doppler shift, τ is related to distance

$\tau = \frac{2d}{c}, c =$ speed of light

$x \rightarrow \widehat{\tau} \rightarrow \widehat{d} = \frac{c\widehat{\tau}}{2}$

3 Imaging Example:

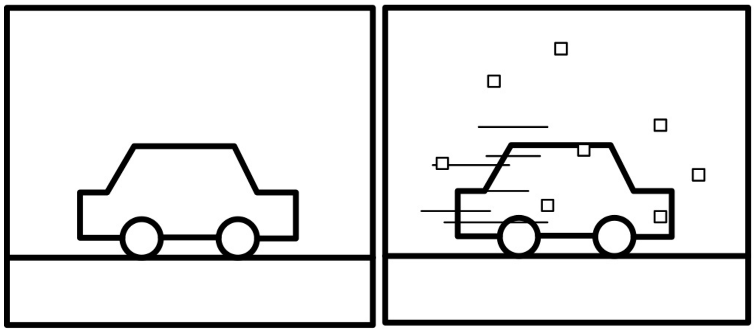


Image processing can involve complicated estimation problems. For example, suppose we observe a moving object with noise. This image is blurry and noisy and our goal is to "restore" this image by deblurring and denoising.

We can model the observed image as

$$x = \underbrace{h * \theta}_{\text{motion blur (convolution)}} + \underbrace{w}_{noise}$$

where the parameter of interest θ is the ideal image.

- Vector model:
  $X = H\theta + w, w \sim N(0, \sigma^2 I)$
  $X \sim N(H\theta, \sigma^2 I)$

- Estimator:
  $\widehat{\theta} = f(x)$, a function of x

4 Basic Ingredients of Estimation Theory

- Observation model:
  $X \sim p(x|\theta), \theta \in \Theta, x \in \mathcal{X}$

- Estimator:
  $\widehat{\theta} : \mathcal{X} \rightarrow \Theta$, a mapping from $\mathcal{X}$ to Θ

- Loss/Error Function:
  $\ell : \Theta \times \Theta \rightarrow \mathbb{R}^+$
  $\ell(\theta, \widehat{\theta})$ measures proximity of $\widehat{\theta}$ to θ

- Risk (Average/Expected Loss:)

$$\begin{aligned} R(\theta, \widehat{\theta}) &= \mathbb{E}[\ell(\theta, \widehat{\theta}(x))] \\ &= \int_{\mathcal{X}} \ell(\theta, \widehat{\theta}(x))\, p(x|\theta) dx \end{aligned}$$

5 Optimal Estimator:

$\widehat{\theta}_{opt} = \arg\min_{\widehat{\theta}:\mathcal{X}\to\Theta} R(\theta, \widehat{\theta})$
$\widehat{\theta}_{opt}$ is optimal with respect to chosen loss function.

- Squared Error (ℓ_2 loss)
  $\ell(\theta, \widehat{\theta}) = \|\theta - \widehat{\theta}\|_2^2 = \sum_{i=1}^{n}(\theta_i - \widehat{\theta}_i)^2$
- Absolute Error (ℓ_1 loss)
  $\ell(\theta, \widehat{\theta}) = \|\theta - \widehat{\theta}\|_1 = \sum_{i=1}^{n}|\theta_i - \widehat{\theta}_i|$
  penalize large errors less than ℓ_2
- 0/1 loss
  $\ell(\theta, \widehat{\theta}) = \begin{cases} 1 & \widehat{\theta} \neq \theta \\ 0 & \text{otherwise} \end{cases}$
  $R(\theta, \widehat{\theta}) = \mathbb{E}[1_{\{\widehat{\theta}(x)\neq\theta\}}] = P(\widehat{\theta} \neq \theta)$

5.1 Special Case - Hypothesis Testing:

Hypothesis testing can be viewed as a special case in which the parameter θ takes one of two possible values, 0 or 1, and the loss is 0/1 loss or some weighted version of it.

6 Basic Concepts:

- Estimator:
  $\widehat{\theta} : \mathcal{X} \to \Theta$ a function of x
  $\Rightarrow \widehat{\theta}(x)$ is a statistic
- Estimate:
  Given a particular observation of X, say x, $\widehat{\theta}(x)$ is called the estimate of θ given observation x.
- Bias:
  $\text{bias}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}(x)] - \theta$

- Variance:

$$\begin{aligned}\mathrm{var}(\widehat{\theta}) &= tr(\mathbb{E}[(\widehat{\theta}(x) - \mathbb{E}[\widehat{\theta}(x)])(\widehat{\theta}(x) - \mathbb{E}[\widehat{\theta}(x)])^T]) \\ &= \mathbb{E}[(\widehat{\theta}(x) - \mathbb{E}[\widehat{\theta}(x)])^T(\widehat{\theta}(x) - \mathbb{E}[\widehat{\theta}(x)])]\end{aligned}$$

- Mean Squared Error (MSE):

$$\begin{aligned}MSE(\widehat{\theta}) &= \mathbb{E}[\|\theta - \widehat{\theta}(x)\|_2^2] \\ &= \mathbb{E}[\|\theta - \mathbb{E}[\widehat{\theta}(x)] + \mathbb{E}[\widehat{\theta}(x)] - \widehat{\theta}(x)\|_2^2] \\ &= \|\theta - \mathbb{E}[\widehat{\theta}(x)]\|_2^2 + \mathbb{E}[\|\widehat{\theta}(x) - \mathbb{E}[\widehat{\theta}(x)]\|_2^2] + 2(\theta - \mathbb{E}[\widehat{\theta}(x)])^T\mathbb{E}[\underbrace{\widehat{\theta}(x) - \mathbb{E}[\widehat{\theta}(x)]}_{=0}] \\ &= \underbrace{\mathrm{bias}(\widehat{\theta})^T\mathrm{bias}(\widehat{\theta})}_{\text{squared bias}} + \underbrace{\mathrm{var}(\widehat{\theta})}_{\text{variance}}\end{aligned}$$

- Asymptotics
  Suppose $X_1, X_2, \cdots, X_n \overset{iid}{\sim} p(x|\theta), \theta \in \Theta$, and consider an estimator $\widehat{\theta}_n = \widehat{\theta}(X_1, \cdots, X_n)$. How does $\widehat{\theta}_n$ behave as $n \to \infty$

Definition 1 $\widehat{\theta}_n$ *is asymptotically unbiased if* $\lim_{x\to\infty} \mathbb{E}[\widehat{\theta}_n] - \theta = 0$, *for all* $\theta \in \Theta$

Definition 2 $\widehat{\theta}_n$ *is consistent (w.r.t. chosen loss/risk) if* $R(\theta, \widehat{\theta}_n) \to 0$, *as* $n \to \infty$, *for all* $\theta \in \Theta$

6.1 A Simple Example

$X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu, 1)$
$\widehat{\mu}_n = \widehat{\mu}(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^n X_i$

- *loss:* (ℓ_2)
  $\ell(\mu, \widehat{\mu}) = \|\mu - \widehat{\mu}_n\|_2^2$

- *risk: (MSE)*

$$\begin{aligned}R(\mu, \widehat{\mu}) &= \mathbb{E}[\|\mu - \widehat{\mu}_n\|_2^2] \\ &= \underbrace{\mathbb{E}[\|\mu - \mathbb{E}[\widehat{\mu}_n]\|_2^2]}_{bias^2} + \underbrace{\mathbb{E}[\|\widehat{\mu}_n - \mathbb{E}[\widehat{\mu}_n]\|_2^2]}_{variance}\end{aligned}$$

- *bias:*
  $\mathbb{E}[\widehat{\mu}_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$ *(unbiased estimator)*

- *variance:*

$$
\begin{aligned}
\mathbb{E}[\|\widehat{\mu}_n - \mathbb{E}[\widehat{\mu}_n]\|_2^2] &= \mathbb{E}[(n\sum_i X_i - \mu)^2] \\
&= \mathbb{E}[\frac{1}{n^2}\left(\sum_i (X_i - \mu)\right)^2] \\
&= \mathbb{E}[\frac{1}{n^2}\sum_{ij}(X_i - \mu)(X_j - \mu)] \\
&= \frac{1}{n^2}\sum_i \mathbb{E}[(X_i - \mu)^2] \\
&= \frac{1}{n}
\end{aligned}
$$

- *consistency*
  $R(\mu, \widehat{\mu}) = \frac{1}{n} \to 0 \ \textit{as} \ n \to \infty$