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

1 Generator Model:

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

• Estimator

Observe $X \sim p(x|\theta) \xrightarrow{\text{infer}} \theta \in \Theta$

Estimation is a sort of “inverse” problem

2 Radar Example:

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 $\alpha$ is related to velocity / Doppler shift, $\tau$ is related to distance

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

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

\[ x = h \ast \theta + w \]

where the parameter of interest \( \theta \) is the ideal image.

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

- Estimator:
  \( \hat{\theta} = f(x) \), 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:
  \( \hat{\theta} : \mathcal{X} \rightarrow \Theta \), a mapping from \( \mathcal{X} \) to \( \Theta \)

- Loss/Error Function:
  \( \ell : \Theta \times \Theta \rightarrow \mathbb{R}^+ \)
  \( \ell(\theta, \hat{\theta}) \) measures proximity of \( \hat{\theta} \) to \( \theta \)
Lecture 13: Parameter Estimation

- Risk (Average/Expected Loss:)

\[ R(\theta, \hat{\theta}) = \mathbb{E}[\ell(\theta, \hat{\theta}(x))] = \int X \ell(\theta, \hat{\theta}(x)) p(x|\theta) dx \]

5 Optimal Estimator:

\( \hat{\theta}_{opt} = \arg\min_{\hat{\theta} : \mathcal{X} \to \Theta} R(\theta, \hat{\theta}) \)

\( \hat{\theta}_{opt} \) is optimal with respect to chosen loss function.

- Squared Error (\( \ell_2 \) loss)

\[ \ell(\theta, \hat{\theta}) = \| \theta - \hat{\theta} \|_2^2 = \sum_{i=1}^n (\theta_i - \hat{\theta}_i)^2 \]

- Absolute Error (\( \ell_1 \) loss)

\[ \ell(\theta, \hat{\theta}) = \| \theta - \hat{\theta} \|_1 = \sum_{i=1}^n |\theta_i - \hat{\theta}_i| \]

penalize large errors less than \( \ell_2 \)

- 0/1 loss

\[ \ell(\theta, \hat{\theta}) = \begin{cases} 1 & \hat{\theta} \neq \theta \\ 0 & \text{otherwise} \end{cases} \]

\[ R(\theta, \hat{\theta}) = \mathbb{E}[1_{\{\hat{\theta}(x) \neq \theta\}}] = P(\hat{\theta} \neq \theta) \]

5.1 Special Case - Hypothesis Testing:

Hypothesis testing can be viewed as a special case in which the parameter \( \theta \) 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:

\( \hat{\theta} : \mathcal{X} \to \Theta \) a function of \( x \)

\( \Rightarrow \hat{\theta}(x) \) is a statistic

- Estimate:

Given a particular observation of \( X \), say \( x \), \( \hat{\theta}(x) \) is called the estimate of \( \theta \) given observation \( x \).

- Bias:

\( \text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}(x)] - \theta \)
• Variance:

\[
\text{var}(\hat{\theta}) = tr(\mathbb{E}[(\hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)])(\hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)])^T]) = \mathbb{E}[(\hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)])(\hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)])^T]
\]

• Mean Squared Error (MSE):

\[
\text{MSE}(\hat{\theta}) = \mathbb{E}[(\theta - \hat{\theta}(x))^2] = \mathbb{E}[(\theta - \mathbb{E}[\hat{\theta}(x)])^2 + (\theta - \mathbb{E}[\hat{\theta}(x)])\mathbb{E}[\hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)]] = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

\[
\text{bias} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

\[
\text{variance} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

• Asymptotics

Suppose \(X_1, X_2, \ldots, X_n \sim p(x|\theta), \theta \in \Theta\), and consider an estimator \(\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)\).

How does \(\hat{\theta}_n\) behave as \(n \rightarrow \infty\)

**Definition 1** \(\hat{\theta}_n\) is asymptotically unbiased if \(\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}_n] - \theta = 0\), for all \(\theta \in \Theta\)

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

### 6.1 A Simple Example

\(X_1, X_2, \ldots, X_n \sim N(\mu, 1)\)

\(\hat{\mu}_n = \hat{\mu}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i\)

• loss: \(l_2\)

\(\ell(\mu, \hat{\mu}) = ||\mu - \hat{\mu}_n||_2^2\)

• risk: \(MSE\)

\[
R(\mu, \hat{\mu}) = \mathbb{E}[||\mu - \hat{\mu}_n||_2^2] = \mathbb{E}[||\mu - \mathbb{E}[\hat{\mu}_n]||_2^2] + \mathbb{E}[||\hat{\mu}_n - \mathbb{E}[\hat{\mu}_n]||_2^2]
\]

• bias:

\(\mathbb{E}[\hat{\mu}_n] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu \) (unbiased estimator)
• variance:

\[
\mathbb{E}[\|\hat{\mu}_n - \mathbb{E}[\hat{\mu}_n]\|_2^2] = \mathbb{E}[(n \sum_i X_i - \mu)^2]
\]

\[
= \mathbb{E}[\frac{1}{n^2} \left( \sum_i (X_i - \mu) \right)^2]
\]

\[
= \mathbb{E}[\frac{1}{n^2} \sum_{ij} (X_i - \mu)(X_j - \mu)]
\]

\[
= \frac{1}{n^2} \sum_i \mathbb{E}[(X_i - \mu)^2]
\]

\[
= \frac{1}{n}
\]

• consistency

\[R(\mu, \hat{\mu}) = \frac{1}{n} \to 0 \ \text{as} \ n \to \infty\]