ESTIMATION THEORY

What is Estimation?

The main actors in the story of estimation theory are

$\mathcal{X} \subseteq \mathbb{R}^N$	the sample space
$\underline{X} \in \mathcal{X}$	a random vector of measurements
$\Theta \subseteq \mathbb{R}^P$	the parameter space
$\underline{\theta} \in \Theta$	the unknown parameter
$f_{\underline{\theta}}(\underline{x})$	the pdf/pmf of $\underline{X}$

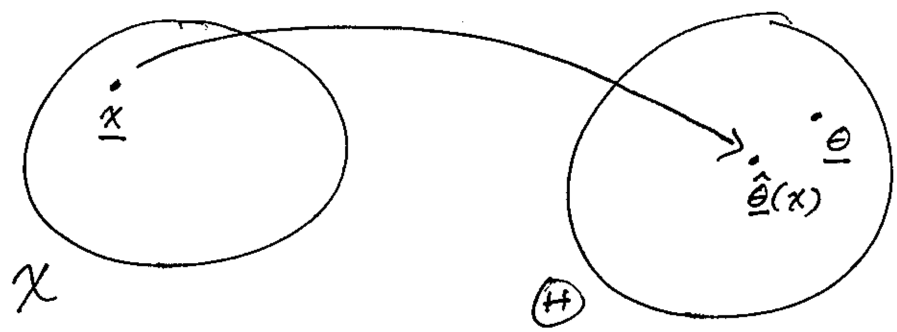
The basic plot is a follows:

A realization $\underline{x} \in \mathcal{X}$ of $\underline{X}$ is observed. The distribution of $\underline{X}$ is specified entirely by the unknown parameter $\underline{\theta}$. $\underline{x}$ must be used to estimate $\underline{\theta}$.

The protagonist of our story is a function

$$\underline{\hat{\theta}} : \mathcal{X} \longrightarrow \Theta$$

called an <u>estimator</u>. Estimation theory studies various kinds of estimators and their properties.



The notation $\underline{\hat{\theta}}$ will be used in several ways

- $\underline{\hat{\theta}}(\underline{X})$ is a random variable
- $\underline{\hat{\theta}}$ may also denote the same random variable
- for a particular $\underline{x}$, $\underline{\hat{\theta}}(\underline{x})$ is the <u>estimate</u> of $\underline{\theta}$

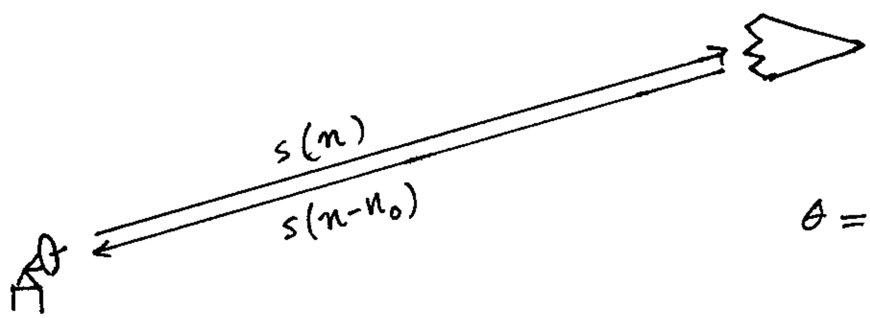
The meaning will be clear from the context.

The methods we will study will be quite general, but we will emphasize applications in signal processing.

The methods to be covered might also be covered in a statistics course on "parameter estimation" or "point estimation."

$\boxed{\text{Examples}}$

1. Signal delay estimation

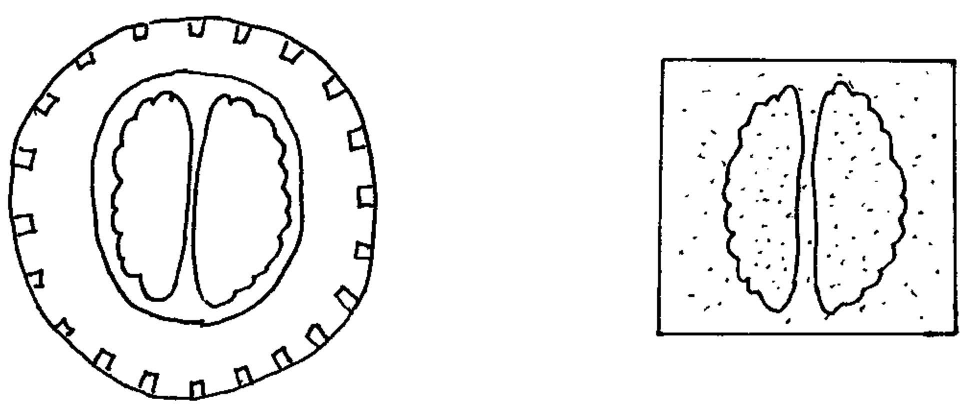


$$\theta = n_0$$

$$x(n) = s(n-n_0) + w(n), \quad n = 0, 1, \ldots, N-1$$

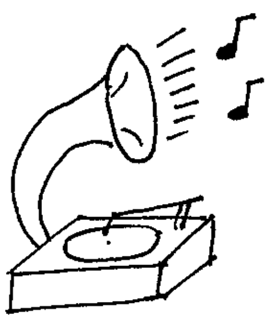
2. Image reconstruction

$\underline{\theta}$ = image of brain



radial projections $\Longrightarrow$ CT brain image

3. Signal / Image Denoising

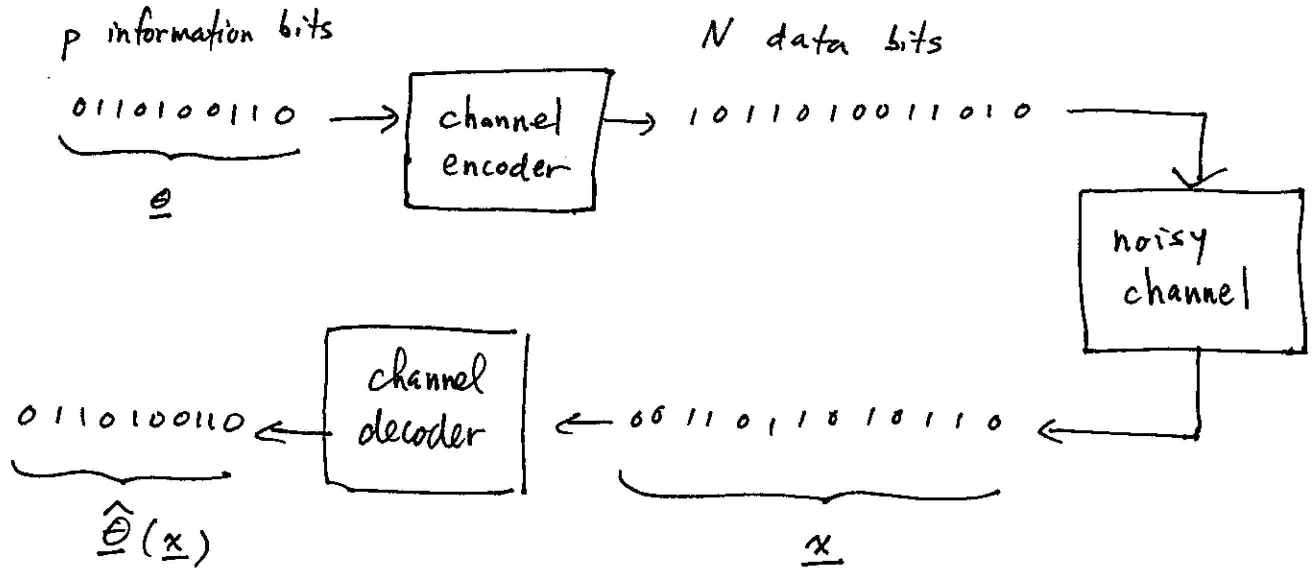


old record with static $\Longrightarrow$ clean recording with crackles and pops removed

$$x(n) = s(n) + w(n) \Longrightarrow \hat{s}(n)$$

$$\underline{\theta} = [s(0) \;\cdots\; s(N-1)]^T.$$

4. Communication / Transmission over noisy channels



5. Many, many others

Estimation Categories

There are several ways to classify estimators and estimation strategies. Here are some of the main ones.

I. Optimality Criterion

The primary element in an optimality criterion is whether the unknown parameter is viewed as <u>random</u> or <u>nonrandom</u>.

$\boxed{A}$ $\underline{\theta}$ nonrandom : <u>classical</u> (frequentist) estimation

- minimum variance unbiased estimation
- maximum likelihood estimation
- method of moments
- least squares (nonstatistical)

$\boxed{B}$ $\underline{\theta}$ random : <u>Bayesian</u> estimation

- minimum mean squared error
- minimum absolute deviation
- maximum a posteriori

Depending on the problem, sometimes different criteria are optimized by the same estimator.

II. Form

The primary distinction regarding the form of an estimator is whether it is linear or nonlinear.

A linear estimator has the form

$$\underline{\hat{\theta}}(\underline{x}) = C\underline{x}, \qquad C \in \mathbb{R}^{p \times N}$$

Such estimators arise frequently in conjunction with the multivariate Gaussian distribution.

Because of their simplicity, sometimes we will optimize our criteria while restricting the estimator to be linear.

III. Offline vs. Online

When we study filtering we will focus on estimators that can efficiently update their estimate as data streams in.

Classical Estimation: Basic Notions

Our study of estimation will begin with classical estimators. Thus, assume $\underline{\theta}$ is nonrandom, that is, unknown but fixed.

Definitions Let $\hat{\underline{\theta}}$ be an estimator of $\underline{\theta}$.

The mean squared error of $\hat{\underline{\theta}}$ is

$$\text{MSE}_{\underline{\theta}}(\hat{\underline{\theta}}) = E\left[\|\hat{\underline{\theta}} - \underline{\theta}\|^2\right]$$

$$= E\left[\|\hat{\theta}(\underline{x}) - \underline{\theta}\|^2\right]$$

The variance of $\hat{\underline{\theta}}$ is

$$\text{Var}_{\underline{\theta}}(\hat{\underline{\theta}}) = E\left[\|\hat{\underline{\theta}} - E[\hat{\underline{\theta}}]\|^2\right]$$

The bias of $\hat{\underline{\theta}}$ is

$$\text{Bias}_{\underline{\theta}}(\hat{\underline{\theta}}) = E\left[\hat{\underline{\theta}}\right] - \underline{\theta}$$

We say $\hat{\underline{\theta}}$ is unbiased if

$$\text{Bias}_{\underline{\theta}}(\hat{\underline{\theta}}) = \underline{0} \qquad \forall \theta \in \Theta$$

Otherwise, we say $\hat{\underline{\theta}}$ is biased.

Let $\{\hat{\underline{\theta}}_N\}_{N=1}^{\infty}$ be a family of estimators.

We say $\{\hat{\underline{\theta}}_N\}$ is <u>asymptotically unbiased</u> if

$$\text{Bias}_{\underline{\theta}}(\hat{\underline{\theta}}_N) \to \underline{0} \quad \text{as} \quad N \to \infty \qquad \forall \underline{\theta} \in \Theta$$

and <u>consistent</u> if

$$\text{MSE}_{\underline{\theta}}(\hat{\underline{\theta}}_N) \to 0 \quad \text{as} \quad N \to \infty \qquad \forall \theta \in \Theta$$

<u>Example</u> Suppose $\underline{X} = [X_1, \ldots, X_N]^T$ where

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2), \qquad i = 1, \ldots, N$$

Consider the estimator of μ given by

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i .$$

What is the bias of $\hat{\mu}$?

$$E\{\hat{\mu}\} = E\left\{\frac{1}{N}\sum_{i=1}^{N} X_i\right\}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\{X_i\}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu$$

$$= \mu$$

$\Longrightarrow$ $\hat{\mu}$ is unbiased.

<u>Exercise</u> Find the variance of $\hat{\mu}$. Is $\hat{\mu}$ consistent?

Solution 1

Approach 1 :

$$
\begin{aligned}
Var_\mu(\hat{\mu}) &= E\left\{(\hat{\mu}-\mu)^2\right\} \\
&= E\left\{\left(\frac{1}{N}\sum X_i - \mu\right)^2\right\} \\
&= E\left\{\sum_{i=1}^{N}\left(\frac{X_i-\mu}{N}\right)^2\right\} \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E\left\{(X_i-\mu)^2\right\} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2 \\
&= \frac{\sigma^2}{N}
\end{aligned}
$$

Approach 2

Since $\underline{X} \sim N(\underline{1}\mu, \sigma^2\underline{I})$ and $\bar{X} = A \cdot \underline{X}$,

$A = \left[\frac{1}{N} \cdots \frac{1}{N}\right]$, we deduce

$$
\begin{aligned}
\hat{\mu} &\sim N(A \cdot \underline{1}\mu,\ A \cdot \sigma^2 \underline{I} \cdot A^T) \\
&\sim N\left(\mu, \frac{\sigma^2}{N}\right)
\end{aligned}
$$

As for consistency, note

$$
\begin{aligned}
MSE_\mu(\hat{\mu}) &= E\{(\hat{\mu}-\mu)^2\} \\
&= E\{(\hat{\mu}-E\hat{\mu})^2\} \\
&= Var_\mu(\hat{\mu}) \\
&= \frac{\sigma^2}{N} \longrightarrow 0 \quad \text{as } N \to \infty.
\end{aligned}
$$

<u>Example</u> Note that for an estimator to be unbiased, its expected value must be the true value for <u>all $\underline{\theta} \in \Theta$</u>.

In the previous example, suppose we take

$$\hat{\mu} = \frac{2}{N}\sum_{i=1}^{N} X_i .$$

Then

$$
\begin{aligned}
E\{\hat{\mu}\} &= \mu \qquad \text{if } \mu = 0 \\
E\{\hat{\mu}\} &\neq \mu \qquad \text{if } \mu \neq 0.
\end{aligned}
$$

$\Longrightarrow \hat{\mu}$ is biased.