Detection and Estimation Theory

Sau-Hsuan Wu
Institute of Communications Engineering
National Chiao Tung University
Detection theory
- Determine whether or not an event of interest occurs
- E.g. Whether or not an aircraft occurs

Estimation theory
- Determine the values of parameters pertaining to an event of interest
- E.g. The altitude and position of an aircraft
- Under what basis do we do detection and estimation
  - The signal source model and the observations

Figure 1.1. Radar system (a) Radar (b) Radar waveforms.
Another application of detection and estimation theory

- Determine the existence of a submarine

- Estimate the bearing

(a) Passive sonar

(b) Received signals at array sensors
Many more applications

- Biomedicine: estimate the heart rate or even heart diseases
- Image analysis: estimate the size, position and orientation of an object in an image
- Seismology: estimate the underground distance of an oil deposit
- Communications: estimate the signal distortion and determine the transmitted symbols
- Economics: estimate the Don-Jones industrial average
The mathematical estimation problem

- We want to know (or estimate) an unknown quantity $\theta$
- This quantity, $\theta$, may not be obtained directly, or it is a notion (quantity) derived from a set of observations
- E.g. the Don-Jones average

We seem to have a model $x[n] = A + Bn + w[n]$, $n = 0, \ldots, N-1$
The observations are corrupted by noise $w[n]$.

The noise may be modeled as a white Gaussian noise $\mathcal{N}(0,\sigma^2)$.

We want to determine $\theta = [A, B]$ from the set of corrupted observations $x = x[0], \ldots, x[N-1]$ which can be modeled as

$$P(x; A, B) = \frac{1}{N} \exp\left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)^2 \right]$$
Our goal is to determine $\theta$ according to the observations, namely:

$$\hat{\theta} = g(x[0], \ldots, x[N - 1])$$

Examples:

- $x[n] = A_1 + w[n]$, $n=0,\ldots,N-1$
- $x[n] = A_2 + B_2 n + w[n]$, $n=0,\ldots,N-1$

- $\hat{A}_1 = \sum x[n]/N$
- $\hat{A}_2$ and $\hat{B}_2$? Can we try the least square method

How many other kinds of methods can we use?

- Minimum variance unbiased estimator (MVU)
- Best linear unbiased estimator (BLUE)
- Maximum likelihood estimator (ML)
- Bayesian estimator: MMSE, MAP, and …
The mathematical detection problem

- We want to determine the presence or absence of an object
- Similarly, we have some observations \( x[n], n=0, \ldots, N-1 \)
- Based on our problem setting, \( x[n] \) can be modeled as
  \[ H_0 : x[n] = s_0 + w[n] \]
  \[ H_1 : x[n] = s_1 + w[n] \]
In this special case, \( s_0 = 0 \) and \( s_1 = 1 \).

Then, we have two hypotheses for \( x[n] \) in this case:

- \( \mathcal{H}_0 : x[n] = w[n] \)
- \( \mathcal{H}_1 : x[n] = 1 + w[n] \)

Suppose the noise \( w[n] \sim \mathcal{N}(0, \sigma^2) \).

And we have one observation only.

\[
p(x[0]; H_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{1}{2\sigma^2} x^2[0] \right)
\]

\[
p(x[0]; H_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{1}{2\sigma^2} (x[0] - 1)^2 \right)
\]
A reasonable approach is to decide $\mathcal{H}_1$ if $x[0] > 1/2$

Type I error: We decide $\mathcal{H}_1$ when $\mathcal{H}_0$ is true (False Alarm)

Type II error: We decide $\mathcal{H}_0$ when $\mathcal{H}_1$ is true (Miss)
As the threshold $\gamma$ increases, the missing probability $P_M$ increases, and $P_D$ and $P_{FA} = 1 - P_M$ decrease.

$$P_{FA} = P(\mathcal{H}_1; \mathcal{H}_0)$$
$$= \Pr\{x[0] > \gamma; \mathcal{H}_0\}$$
$$= \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t^2\right) dt$$
$$= Q(\gamma).$$

$$P(\mathcal{H}_0; \mathcal{H}_1) = 1 - P(\mathcal{H}_1; \mathcal{H}_1)$$

$$P_D = P(\mathcal{H}_1; \mathcal{H}_1)$$
$$= \Pr\{x[0] > \gamma; \mathcal{H}_1\}$$
$$= \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t - 1)^2\right] dt$$
In general, we have multiple observations $x[n]$, $n=0,\ldots,N-1$

$\mathcal{H}_0 : x[n] = w[n]$
$\mathcal{H}_1 : x[n] = A + w[n]$

- We would decide $\mathcal{H}_1$ if
  
  $T = \frac{\sum x[n]}{N} > \gamma$

- The detection performance increases with $d^2 = \frac{NA^2}{\sigma^2}$

Other approach for detection

- Neyman-Pearson approach

When $\mathcal{H}_0$ and $\mathcal{H}_1$ are thought of random variables

- Maximum likelihood (ML) detector
- Maximum a posterior (MAP) detector
- Composite hypothesis testing
Detection Theory
Neyman-Pearson (NP) approach

- We have observations $x[n]$, $n=0,\ldots, N-1$, $x\equiv\{x[0],\ldots,x[N-1]\}$
- Given $P(x; \mathcal{H}_1)$, we decide $\mathcal{H}_1$ if $P(x; \mathcal{H}_1) / P(x; \mathcal{H}_0) > \gamma$
- Goal: maximize $P_D = P(\mathcal{H}_1; \mathcal{H}_1)$ subject to $P_{FA} = P(\mathcal{H}_1; \mathcal{H}_0) \leq \alpha$
- $\alpha$ is called the significance level

For example, $\mathcal{H}_0: x[n] = w[n]$  
$\mathcal{H}_1: x[n] = A + w[n]$, $n=0,\ldots, N-1$, 

$$\frac{p(x; \mathcal{H}_1)}{p(x; \mathcal{H}_0)} = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] > \gamma.$$
which results in

\[ \frac{1}{2\sigma^2} \left( -2A \sum_{n=0}^{N-1} x[n] + NA^2 \right) > \ln \gamma \]

and, hence,

\[ T(x) \equiv \frac{1}{N} \sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA} \ln \gamma + \frac{A}{2} = \gamma'. \]

The decision statistic \( T(x) \) is Gaussian distributed

\[ T(x) \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{under } \mathcal{H}_1. \end{cases} \]

Thus, \( P_D = P(\mathcal{H}_1 ; \mathcal{H}_1) = \Pr\{T(x) > \gamma'; \mathcal{H}_1\} = Q\left(\gamma' - \frac{A}{\sqrt{\sigma^2/N}}\right) \)

And \( P_{FA} = P(\mathcal{H}_1 ; \mathcal{H}_0) = \Pr\{T(x) > \gamma'; \mathcal{H}_0\} = Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) \)
Notice that both $P_D$ and $P_{FA}$ increase when $\gamma'$ decreases.

Therefore to maximize $P_D$ subject to $P_{FA} \leq \alpha$, the optimal threshold $\gamma'$ is found by setting $P_{FA} = \alpha$.

This results in $Q\left(\frac{\gamma' / \sqrt{\sigma^2/N}}{\alpha}\right) = \alpha$, and, hence, $\gamma' = \sqrt{\frac{\sigma^2}{N}} Q^{-1}(\alpha)$.
Bayesian approaches

In some detection problems, it makes sense to assign probabilities to the various hypotheses $\mathcal{H}_i$

Define a detector $\delta(x) = P(\text{decide } \mathcal{H}_1|x) \in [0, 1]$

The error probability of detection becomes

$$P_e = P(\mathcal{H}_1 | \mathcal{H}_0) P(\mathcal{H}_0) + P(\mathcal{H}_0 | \mathcal{H}_1) P(\mathcal{H}_1)$$

$$= P(\mathcal{H}_0) \int \delta(x) P(x | \mathcal{H}_0) dx + P(\mathcal{H}_1) \int [1 - \delta(x)] P(x | \mathcal{H}_1) dx$$

$$= P(\mathcal{H}_1) + \int \delta(x) [P(x | \mathcal{H}_0) P(\mathcal{H}_0) - P(x | \mathcal{H}_1) P(\mathcal{H}_1)] dx$$

We want to design a $\delta(x) \in [0, 1]$ to minimize $P_e$

- $\delta(x) = 0$ if $P(x | \mathcal{H}_0) P(\mathcal{H}_0) - P(x | \mathcal{H}_1) P(\mathcal{H}_1) > 0$
- $\delta(x) = 1$ if $P(x | \mathcal{H}_0) P(\mathcal{H}_0) - P(x | \mathcal{H}_1) P(\mathcal{H}_1) \leq 0$

Or, $\delta(x) = 1$ if $\frac{P(x | \mathcal{H}_1)}{P(x | \mathcal{H}_0)} = \Lambda(x) \geq \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = \gamma$

$= 0$ if $\Lambda(x) < \gamma$
The function \( \Lambda(x) = \frac{P(x | \mathcal{H}_1)}{P(x | \mathcal{H}_0)} \) is the likelihood ratio. And the detection process is called the likelihood ratio test.

- **Maximum a posteriori (MAP) detector**
  - We may rewrite:
    - \( \delta(x) = 0 \) if \( P(x | \mathcal{H}_0) P(\mathcal{H}_0) - P(x | \mathcal{H}_1) P(\mathcal{H}_1) > 0 \)
    - \( \delta(x) = 1 \) if \( P(x | \mathcal{H}_0) P(\mathcal{H}_0) - P(x | \mathcal{H}_1) P(\mathcal{H}_1) \leq 0 \)
  - Into:
    - \( \delta(x) = \arg \max_{\theta \in \{0, 1\}} P(x | \mathcal{H}_\theta) P(\mathcal{H}_\theta) \)
    - \[ = \arg \max_{\theta \in \{0, 1\}} P(\mathcal{H}_\theta | x) P(x) = \arg \max_{\theta \in \{0, 1\}} P(\mathcal{H}_\theta | x) \]

- **Maximum likelihood (ML) detector**
  - If \( P(\mathcal{H}_0) = P(\mathcal{H}_1) \)
    - \( \delta(x) = \arg \max_{\theta \in \{0, 1\}} P(x | \mathcal{H}_\theta) \)
For example, $\mathcal{H}_0 : x[n] = w[n]$  
$\mathcal{H}_1 : x[n] = A + w[n]$, $n = 0, \ldots, N-1$

- $P(\mathcal{H}_0) = P(\mathcal{H}_1) = \frac{1}{2}$ and $w[n] \sim \mathcal{N}(0, \sigma^2)$
- As a result, $\gamma = P(\mathcal{H}_0)/P(\mathcal{H}_1) = 1$, and
- $\Lambda(x) = P(x|\mathcal{H}_1)/P(x|\mathcal{H}_0) > \gamma = 1$ yields

\[
\frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] > 1.
\]

- Or equivalently

\[
-\frac{1}{2\sigma^2} \left( -2A \sum_{n=0}^{N-1} x[n] + NA^2 \right) > 0
\]
Thus, we decide $\mathcal{H}_1$ if $\bar{x} > A/2$.

Given that

$$\bar{x} \sim \begin{cases} 
\mathcal{N}(0, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_0 \\
\mathcal{N}(A, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_1.
\end{cases}$$

As a result

$$P_e = \frac{1}{2} [P(\mathcal{H}_0|\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)]$$

$$= \frac{1}{2} \left[ \Pr\{\bar{x} < A/2|\mathcal{H}_1\} + \Pr\{\bar{x} > A/2|\mathcal{H}_0\} \right]$$

$$= \frac{1}{2} \left[ \left( 1 - Q \left( \frac{A/2 - A}{\sqrt{\sigma^2/N}} \right) \right) + Q \left( \frac{A/2}{\sqrt{\sigma^2/N}} \right) \right]$$

Since $Q(-x) = 1 - Q(x)$, we have

$$P_e = Q \left( \sqrt{\frac{NA^2}{4\sigma^2}} \right)$$
For MAP detector: \( \arg \max_{\theta \in \{0, 1\}} P(\mathcal{H}_\theta | x) \)

- Thus, we decide \( \mathcal{H}_1 \) if \( P(\mathcal{H}_1 | x) > P(\mathcal{H}_0 | x) \)
- Suppose we have only one observation \( x[0] \)
- The decision region varies with the prior probability

\[
\begin{align*}
(a) & \quad P(\mathcal{H}_0) = P(\mathcal{H}_1) = \frac{1}{2} \\
(b) & \quad P(\mathcal{H}_0) = \frac{1}{4}, \quad P(\mathcal{H}_1) = \frac{3}{4}
\end{align*}
\]
In more general cases, we want to design a detector $\delta(x)$ to minimize the average cost of decision errors

- Consider a cost function $C_{ab}$ of deciding $H_a$ when $H_b$ is true
- Given the general set of costs $\{C_{00}, C_{01}, C_{10}, C_{11}\}$
- The average cost, or also called Bayesian Risk, is given by

$$R = C_{00} P(H_0 | H_0) P(H_0) + C_{01} P(H_0 | H_1) P(H_1) + C_{10} P(H_1 | H_0) P(H_0) + C_{11} P(H_1 | H_1) P(H_1)$$

$$= C_{00} P(H_0) \int [1 - \delta(x)] P(x | H_0) dx + C_{01} P(H_0) \int \delta(x) P(x | H_0) dx + C_{10} P(H_1) \int [1 - \delta(x)] P(x | H_1) dx + C_{11} P(H_1) \int \delta(x) P(x | H_1) dx$$

- $\delta(x) = 0$ if $\delta(x) > 0$
- $\delta(x) = 1$ if $\delta(x) \leq 0$
An alternative form

\[
\begin{align*}
& [(C_{10} - C_{00})P(H_0) P(x | H_0) + (C_{11} - C_{01})P(H_1) P(x | H_1)] \\
& = \left[ C_{10} P(H_0 | x) + C_{11} P(H_1 | x) - C_{00} P(H_0 | x) - C_{01} P(H_1 | x) \right] P(x) \\
& \equiv [C_1(x) - C_0(x)] P(x)
\end{align*}
\]

where \( C_i(x) = C_{i0} P(H_0 | x) + C_{i1} P(H_1 | x) \) is the cost of deciding \( H_i \)

Therefore,

\[
\delta(x) = 0 \text{ if } C_1(x) - C_0(x) > 0 \text{ or } C_1(x) > C_0(x) \\
\delta(x) = 1 \text{ if } C_1(x) - C_0(x) \leq 0
\]

Or, we can define two decision regions

\[
R_1 \equiv \{ x | C_1(x) \leq C_0(x) \} \\
R_0 \equiv \{ x | C_1(x) > C_0(x) \}
\]

\[
\delta(x) = 1 \text{ if } x \in R_1 \\
\delta(x) = 0 \text{ if } x \in R_0
\]
Multiple hypothesis testing

Suppose we must decide among $M$ hypotheses for which

\[ H_0 : x \sim P(x | H_0), \quad H_1 : x \sim P(x | H_1), \ldots, H_{M-1} : x \sim P(x | H_{M-1}) \]

The average cost is given by

\[ R = \Sigma_i \Sigma_j C_{ij} P(H_i | H_j) P(H_j) \]

Redefine the decision rule

\[ \delta_i(x) = P(\text{decide } H_i | x) \in [0, 1], \text{ and } \Sigma_i \delta_i(x) = 1 \]

\[ R = \Sigma_i \Sigma_j C_{ij} \int \delta_i(x) P(H_j | x) P(x) dx \]

Redefine the cost of $C_i(x) = \Sigma_j C_{ij} P(H_j | x)$

And define the decision region for $H_i$ as

\[ R_i \equiv \{ x | C_i(x) \leq C_j(x), \forall j \neq i \} \]

The decision rule $\delta_i(x)$ that minimizes $R$ is

\[ \delta_i(x) = 1 \text{ if } x \in R_i, \text{ and } \delta_i(x) = 0 \text{ otherwise} \]
Estimation Theory
Let us recall the problems we talked about:

- \( x[n] = A_1 + w[n], \quad n=0,\ldots,N-1 \)
- \( x[n] = A_2 + B_2 n + w[n], \quad n=0,\ldots,N-1 \)

We define a parameter \( \theta \) to represent the quantities to be estimated, e.g. \( \theta = A_1 \) and \( \theta = [A_2, B_2] \) in the above cases.

We model the data by its probability density function (PDF), assuming that the data are inherently random.

As an example:

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

We have a class of PDFs where each one is different due to a different value of \( \theta \), i.e. the PDFs are parameterized by \( \theta \).

The parameter \( \theta \) is assumed deterministic but unknown.
We conclude for the time being that we are hoping to have an estimator that gives

- An unbiased mean of the estimate: \( E\{\hat{\theta}\} = \int g(x)p(x; \theta) = \theta \)

- A minimum variance for the estimate: \( Var(\theta) \)

Is an unbiased estimator always the optimal estimator?

Consider a widely used optimality criterion: the minimum mean squared error (MSE) criterion

\[
\text{mse}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]
\]

\[
= E\{[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2\}
\]

\[
= \text{var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2
\]

\(\triangleq \text{var}(\hat{\theta}) + b^2\)
As an example, consider the modified estimator

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

We attempt to find the ‘a’ which yields the minimum MSE

- Since $E(\tilde{A}) = aA$ and $\text{var}(\tilde{A}) = a^2 \sigma^2 / N$, we have
  $$\text{mse}(\tilde{A}) = a^2 \sigma^2 / N + (a-1)^2 A^2$$

- Taking the derivative w.r.t. to $a$ and setting it to zero leads to
  $$a_{\text{opt}} = \frac{A^2}{(A^2 + \sigma^2 / N)}$$

- The optimal value of $a$ depends upon the unknown parameter $A \Rightarrow$ The estimator is not realizable

- How do we resolve this dilemma?
  - Since, $\text{mse}(\tilde{A}) = \text{var}(\tilde{A}) + b^2$, as an alternative,
    we set $b=0$ and search for the estimator that minimizes $\text{var}(\tilde{A})$
    $\Rightarrow$ minimum variance unbiased (MVU) estimator
However, does a MVU estimator always exist for all $\theta$?

- Example:
  - $x[0] \sim \mathcal{N}(\theta, 1)$
  - $x[1] \sim \mathcal{N}(\theta, 1)$ if $\theta \geq 0$
    $\sim \mathcal{N}(\theta, 2)$ if $\theta \leq 0$
  - Both of the two estimators are unbiased
    $$\hat{\theta}_1 = \frac{1}{2}(x[0] + x[1]) \quad \Rightarrow \quad \text{var} (\hat{\theta}_1) = \frac{1}{4} (\text{var}(x[0]) + \text{var}(x[1]))$$
  $$\hat{\theta}_2 = \frac{2}{3} x[0] + \frac{1}{3} x[1] \quad \Rightarrow \quad \text{var} (\hat{\theta}_2) = \frac{4}{9} \text{var}(x[0]) + \frac{1}{9} \text{var}(x[1])$$
  - Therefore
    $$\text{var}(\hat{\theta}_1) = \begin{cases} \frac{18}{36} & \text{if } \theta \geq 0 \\ \frac{27}{36} & \text{if } \theta < 0 \end{cases} \quad \text{var}(\hat{\theta}_2) = \begin{cases} \frac{20}{36} & \text{if } \theta \geq 0 \\ \frac{24}{36} & \text{if } \theta < 0 \end{cases}$$
  - None of the estimator has a variance uniformly less than or equal to $18/36$
Is there a systematic way to find the MVU if it exists?

We start by defining the set of data that is sufficient for estimation?

What do we mean by sufficiency in estimation?

- We want to have a set of data \( T(x) \) such that given \( T(x) \), any individual data \( x(n) \) is statistically independent of \( A \)

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

Suppose \( \hat{A}_1 = \sum x[n]/N \), then are the followings sufficient?

- \( S_1 = \{x[0], x[1], \ldots, x[N-1]\} \)
- \( S_2 = \sum x[n] \)
- Given \( T_0 = \sum x[n] \), do we still need the individual data?
We say the conditional PDF \( p(\mathbf{x} | \sum x[n] = T_0 ; A) \) should not depend on \( A \) if \( T_0 \) is sufficient.

E.g.

- For (a), a value of \( A \) near \( A_0 \) is more likely even given \( T_0 \).
- For (b), however, \( p(\mathbf{x} | \sum x[n] = T_0 ; A) \) is a constant.

(a) Observations provide information after \( T(x) \) observed—\( T(x) \) is not sufficient
(b) No information from observations after \( T(x) \) observed—\( T(x) \) is sufficient
Now, we need to determine \( p( x \mid \sum x[n] = T_0 \mid A) \) to show \( \sum x[n] = T_0 \) is sufficient

By Baye’s rule

\[
p(x\mid T(x) = T_0; A) = \frac{p(x, T(x) = T_0; A)}{p(T(x) = T_0; A)}
\]

Since \( T(x) \) is a direct function of \( x \),

\[
p(x\mid T(x) = T_0; A) = \frac{p(x; A)\delta(T(x) - T_0)}{p(T(x) = T_0; A)}
\]

Clearly, we have

\[
p(x; A)\delta(T(x) - T_0)) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \delta(T(x) - T_0))
\]
Thus, we have

\[ p(x; A) \delta(T(x) - T_0) = \frac{1}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \delta(T(x) - T_0) \]

\[ = \frac{1}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} x^2[n] - 2AT(x) + NA^2 \right) \right] \delta(T(x) - T_0) \]

\[ = \frac{1}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} x^2[n] - 2AT_0 + NA^2 \right) \right] \delta(T(x) - T_0). \]
Since \( T(x) = \sum x[n] \sim \mathcal{N}(NA, N\sigma^2) \)

\[
p(T(x) = T_0; A) = \frac{1}{\sqrt{(2\pi N\sigma^2)}} \exp \left[ -\frac{1}{2N\sigma^2}(T_0 - NA)^2 \right]
\]

Thus

\[
p(x|T(x) = T_0; A) = \frac{\frac{1}{(2\pi\sigma^2)^\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[ -\frac{1}{2\sigma^2}(-2AT_0 + NA^2) \right]}{\sqrt{2\pi N\sigma^2} \exp \left[ -\frac{1}{2N\sigma^2}(T_0 - NA)^2 \right]} \delta(T(x) - T_0)
\]

\[
= \frac{\sqrt{N}}{(2\pi\sigma^2)^{\frac{N-1}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[ \frac{T_0^2}{2N\sigma^2} \right] \delta(T(x) - T_0)
\]

which does not depend on A

Institute of Communications Engineering, EE, NCTU

35
In general, to identify potential sufficient statistics is difficult

An efficient procedure for finding the sufficient statistics is to employ the **Neyman-Fisher factorization** theorem

Observe that

\[ p(x|T(x) = T_0; A) = \frac{p(x; A) \delta(T(x) - T_0)}{p(T(x) = T_0; A)} \]

If we can factor \( p(x; \theta) \) into \( p(x; \theta) = g(T(x), \theta) \cdot h(x) \)

where

- \( g \) is a function depends only on \( x \) through \( T(x) \)
- \( h \) is a function depends only on \( x \)
- \( \Rightarrow T(x) \) is a sufficient statistics for \( \theta \)
- The converse is also true
  - If \( T(x) \) is a sufficient statistics \( \Rightarrow p(x; \theta) = g(T(x), \theta) \cdot h(x) \)
Recall \( p(x;A) \)

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

On the other hand, we want to estimate \( \sigma^2 \) of \( y[n]=A+x[n] \)

- Suppose \( A \) is given, then define \( x[n] = y[n]-A \)

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

- Clearly, \( T(x) = \sum x^2[n] \) is a sufficient statistics for \( \sigma^2 \)
Ex. we want to estimate the phase of a sinusoid
\[ x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad n=0,1,\ldots,N-1 \]

- Suppose \( A \) and \( f_0 \) are given

\[
p(x; \phi) = \frac{1}{(2\pi \sigma^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x[n] - A \cos(2\pi f_0 n + \phi)]^2 \right\}
\]

- Expand the exponent

\[
\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] \cos(2\pi f_0 n + \phi) + \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi)
\]

\[
= \sum_{n=0}^{N-1} x^2[n] - 2A \left( \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right) \cos \phi
\]

\[
+ 2A \left( \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \right) \sin \phi + \sum_{n=0}^{N-1} A^2 \cos^2(2\pi f_0 n + \phi)
\]
In this case, no single sufficient statistics exists, however

\[
p(x; \phi) = \frac{1}{(2\pi \sigma^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{n=0}^{N-1} A^2 \cos^2 (2\pi f_0 n + \phi) - 2AT_1(x) \cos \phi + 2AT_2(x) \sin \phi \right] \right\}
\]

\[g(T_1(x), T_2(x), \phi)\]

\[\times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x[n]^2 \right\}\]

\[h(x)\]

\[T_1(x) = \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n\]

\[T_2(x) = \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n.\]
The \( r \) statistics \( T_1(x), T_2(x), \ldots, T_r(x) \) are jointly sufficient if \( p(x \mid T_1(x), T_2(x), \ldots, T_r(x); \theta) \) does not depend on \( \theta \).

If \( p(x; \theta) = g(T_1(x), T_2(x), \ldots, T_r(x), \theta) \cdot h(x) \)

\[ \iff \{T_1(x), T_2(x), \ldots, T_r(x)\} \text{ are sufficient statistics for } \theta \]

Now, we know how to obtain the sufficient statistics.

How do we apply them to help obtain the MVU estimator?

The **Rao-Blackwell-Lehmann-Scheffe** Theorem

- If \( \tilde{\theta} \) is an unbiased estimator of \( \theta \) and \( T(x) \) is a sufficient statistic for \( \theta \), then \( \hat{\theta} = E\{\tilde{\theta} \mid T(x)\} \) is unbiased and
  - A valid estimator for \( \theta \) (not dependent on \( \theta \))
  - Of lesser or equal variance than that of \( \tilde{\theta} \) for all \( \theta \)
  - If \( T(x) \) is complete, then \( \hat{\theta} \) is the MVU estimator
Finally, a statistic is **complete** if there is **only one** function, say \( g \), of the statistic that is **unbiased**

\[ \hat{\theta} = g(T(x)) = E\{\hat{\theta}|T(x)\} \]

is solely a function of \( T(x) \)

\( \Rightarrow \) If \( T(x) \) is complete \( \Rightarrow \hat{\theta} \) is unique and unbiased

Besides, \( \text{var}(\hat{\theta}) \leq \text{var}(\tilde{\theta}) \) for any unbiased estimator \( \tilde{\theta} \)

Then, \( \hat{\theta} = g(T(x)) = E\{\hat{\theta}|T(x)\} \) must be the MVU

In summary, the MVU can be found by

- Taking any unbiased \( \tilde{\theta} \) and carrying out \( \hat{\theta} = E\{\tilde{\theta}|T(x)\} \)
- Alternatively, since there is only one function of \( T(x) \) that leads to an unbiased estimator
  \( \Rightarrow \) find the unique \( g(T(x)) \) that makes \( \hat{\theta} = g(T(x)) \) unbiased
The Best Linear Unbiased Estimator (BLUE)

The constraints or limitations on finding the MVU

- Do not know the PDF
- Not able to produce the MVU estimator even if the PDF is given

Faced with our inability to determine the optimal MVU estimator, it is reasonable to resort to a suboptimal one

- An estimator which is linear in the data
- The linear estimator is unbiased as well and has minimum variance

The estimator is termed the best linear unbiased estimator

- Can be determined with the first and the second moments of PDF, thus complete knowledge of the PDF is not necessary
The BLUE formulation

- Linear in data ⇒ \( \hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] \)

- Unbiased ⇒ \( E(\hat{\theta}) = \sum_{n=0}^{N-1} a_n E(x[n]) = \theta \)

- As a result, the variance is given by

\[
\text{Var}(\hat{\theta}) = E \left[ \left( \sum_{n=0}^{N-1} a_n x[n] - \sum_{n=0}^{N-1} a_n E(x[n]) \right)^2 \right] \\
= E \left[ \left( a^T x - a^T E(x) \right)^2 \right] = E \left[ \left( a^T (x - E(x)) \right)^2 \right] \\
= E \left[ a^T (x - E(x))(x - E(x))^T a \right] = a^T Ca
\]
To satisfy the unbiased constraint, $E(x[n])$ must be linear in $\theta$, namely

$$E(x[n]) = s[n] \theta$$

where $s[n]$’s are known.

Rewrite $x[n]$ as

$$x[n] = E(x[n]) + [x[n] - E(x[n])] = s[n] \theta + w[n]$$

This means that the BLUES is applicable to amplitude estimation of known signals in noise.

Let $s = [s[0], s[1], \ldots, s[N-1]]^T$. Based on the above assumption, we reformulate the BLUE as

$$\hat{\theta} = \arg \min_a a^T Ca \quad \text{subject to } a^T s = 1$$
Using the method of Lagrangian multiplier, the Lagrangian function becomes

$$J = a^T Ca + \lambda (a^T s - 1)$$

Taking the gradient with respect to $a$ gives

$$\frac{\partial J}{\partial a} = 2Ca + \lambda s$$

Setting this equal to the zero vector produces

$$a = -\frac{1}{2} \lambda C^{-1} s$$

Substituting this result back into the constraint yields

$$-\frac{1}{2} \lambda s^T C^{-1} s = 1, \quad \Rightarrow \lambda = \frac{-2}{s^T C^{-1} s}, \quad \Rightarrow a_{opt} = \frac{C^{-1} s}{s^T C^{-1} s}$$
The corresponding variance is given by

$$a_{opt}^T C a_{opt} = \frac{s^T C^{-1}}{s^T C^{-1} s} C \frac{C^{-1} s}{s^T C^{-1} s} = \frac{1}{s^T C^{-1} s}$$

The resultant estimator is

$$\hat{\theta} = \frac{s^T C^{-1} x}{s^T C^{-1} s}$$

since $E(x) = s\theta$

To determine the BLUE, we only require knowledge of

- $s$ or the scaled mean
- $C$, the covariance

which are the first and second moments, but not the entire PDF.
Ex. $x[n] = A + w[n]$, $n=0,1,\ldots,N-1$ with $\text{var}(w[n]) = \sigma_n^2$

Since $E(x[n]) = A \Rightarrow s[n]=1 \Rightarrow s = I$

Then

$$\hat{A} = \frac{I^T C^{-1} x}{I^T C^{-1} I}$$

with

$$C = \begin{bmatrix}
\sigma_0^2 & 0 & \ldots & 0 \\
0 & \sigma_1^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_{N-1}^2
\end{bmatrix}$$

$$\Rightarrow \hat{A} = \frac{\sum_{n=0}^{N-1} x[n]}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}$$

The min variance is

$$\text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}$$
When do we use the maximum likelihood estimator (MLE)

- Often there is no obvious way to find the MVU estimator
- The MLE is approximately efficient, thus approximately the MVU estimator

We demonstrate this through an example

- \( x[n] = A + w[n] \), \( n=0,1\ldots,N-1 \), with \( w[n] \sim \mathcal{N}(A, A) \), \( A > 0 \)
- The PDF is

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

- Can we find the MVU estimator?
- We first search for the sufficient statistic
By the Neyman-Fisher factorization theorem, we attempt to factorize the PDF into

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

\[
g \left( \sum_{n=0}^{N-1} x^2[n], A \right)
\]

\[h(x)\]

Assume T(x) is a complete sufficient statistic

We may try to find the MVU estimator if

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

not obvious how to choose \( g \)

- Or we may try \( E(\hat{A} | \sum_{n=0}^{N-1} x^2[n]) \) if \( \hat{A} \) is an unbiased estimator
- Suppose \( \hat{A} = x[0] \)

\[ E \left( x[0] \left| \sum_{n=0}^{N-1} x^2[n] \right. \right) \]

We now have exhausted our possible optimal approaches

- We try to find an estimator that maximizes the likelihood
Differentiating the LLK

\[
\frac{\partial \ln p(x; A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A} \sum_{n=0}^{N-1} (x[n] - A) + \frac{1}{2A^2} \sum_{n=0}^{N-1} (x[n] - A)^2
\]

and setting it to zero produces

\[
\hat{A}^2 + \hat{A} - \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] = 0.
\]

Solving for \(\hat{A}\) gives

\[
\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}} > 0
\]

Is it biased
Now, can we say anything about the MLE estimator?

\[ E(\hat{A}) = E \left( -\frac{1}{2} + \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}} \right) \]

\[ \neq -\frac{1}{2} + \sqrt{E \left( \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \right) + \frac{1}{4}} \]

\[ = -\frac{1}{2} + \sqrt{A + A^2 + \frac{1}{4}} \]

\[ = A. \]

But, \( \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \rightarrow E(x^2[n]) = A + A^2 \)

Therefore, \( \hat{A} \rightarrow A \) as \( N \rightarrow \infty \), i.e. asymptotically unbiased.
The Bayesian Philosophy and General Estimators
Note: In an actual problem, we are not given a PDF but must choose one that is not only consistent with the problem constraints and any prior knowledge, but one that is also mathematically tractable.

Sometimes, we might want to constraint the estimator to produce values in a certain range.

- To incorporate this prior knowledge, we can assume that $\theta$ is no longer deterministic but a *random variable* having a uniform distribution over the $[-U, U]$ interval for instance.
- Assign a PDF to $\theta$, then the data are described by the joint PDF

$$p(x, \theta) = p(x|\theta)p(\theta)$$

- Any estimator that yields estimates according to the prior knowledge of $\theta$ is termed a *Bayesian* estimator.
For example, the MVU estimator is a sample mean if 
\[ x[n] = A + w[n] \]
- This assumes that \( \hat{A} \) could take on any value in \([-\infty, \infty]\).
- However, due to physical constraints, it may be more reasonable to assume that \( A \) can take on values in \([-A_0, A_0]\).
- We would expect to improve our estimation if we used
\[
\hat{A} = \begin{cases} 
-A_0 & \bar{x} < -A_0 \\
\bar{x} & -A_0 \leq \bar{x} \leq A_0 \\
A_0 & \bar{x} > A_0 
\end{cases}
\]
- Such an estimator would have the PDF
\[
p_{\hat{A}}(\xi; A) = \Pr\{\bar{x} \leq -A_0\} \delta(\xi + A_0) \\
+ p_{\hat{A}}(\xi; A)[u(\xi + A_0) - u(\xi - A_0)] \\
+ \Pr\{\bar{x} \geq A_0\} \delta(\xi - A_0)
\]
If we compare the MSE of the two estimators

\[
\text{mse}(\hat{A}) = \int_{-\infty}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A) \, d\xi
\]

\[
= \int_{-\infty}^{-A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A) \, d\xi + \int_{A_0}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A) \, d\xi
\]

\[
+ \int_{A_0}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A) \, d\xi
\]

\[
> \int_{-\infty}^{-A_0} (-A_0 - A)^2 p_{\hat{A}}(\xi; A) \, d\xi + \int_{A_0}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A) \, d\xi
\]

\[
+ \int_{A_0}^{\infty} (A_0 - A)^2 p_{\hat{A}}(\xi; A) \, d\xi
\]

\[
= \text{mse}(\ddot{A}).
\]

Hence, \( \ddot{A} \) is better than \( \hat{A} \) although \( \hat{A} \) is still the MVUE.

We can reduce the MSE by allowing \( \ddot{A} \) to be biased.
Recall that the MMSE criterion
\[ \text{mse}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \]
\[ = E\{[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)]^2\} \]
\[ = \text{var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 \]
\[ \triangleq \text{var}(\hat{\theta}) + b^2 \]

- We consider the modified estimator \( \tilde{A} = \frac{a}{N} \sum_{n=0}^{N-1} x[n] \) and attempt to find ‘a’ that min the MSE
- Since \( E(\tilde{A}) = aA \) and \( \text{var}(\tilde{A}) = a^2 \sigma^2 / N \), we have \( \text{mse}(\tilde{A}) = a^2 \sigma^2 / N + (a-1)^2A \)
- Taking the derivative w.r.t. to ‘a’ and setting to zero leads to \( a_{opt} = A^2 / (A^2 + \sigma^2 / N) \) \( \Rightarrow \) The estimator is not realizable
- We instead find \( \hat{A} \) that min \( \text{var}(\hat{A}) \) while setting \( b=0 \)
We can resolve this paradox in the Bayesian approach.

To this end, we need to reformulate the data model.

In the previous example, knowing ‘A’ lying in a known interval, while no inclination as to whether A should be nearer any particular value, we may have $A \sim U[-A_0, A_0]$.

Since A is a random variable, we have

$$B\text{mse}(\hat{A}) = \int \int (A - \hat{A})^2 p(x, A) \, dx \, dA.$$  

in stead of the classic one with

$$\text{mse}(\hat{A}) = \int (\hat{A} - A)^2 p(x; A) \, dx$$

Thus

$$B\text{mse}(\hat{A}) = \int \left[ \int (A - \hat{A})^2 p(A|x) \, dA \right] p(x) \, dx$$
In fact, we have integrated the parameter dependence away.

Now, go back to

\[ B_{\text{mse}}(\hat{A}) = \int \left[ \int (A - \hat{A})^2 p(A|x) \, dA \right] p(x) \, dx \]

Since \( P(x) \geq 0 \) for all \( x \), if the integral in brackets can be minimized for each \( x \), then the Bayesian MSE will be minimized.

We have

\[ \frac{\partial}{\partial \hat{A}} \int (A - \hat{A})^2 p(A|x) \, dA = \int \frac{\partial}{\partial \hat{A}} (A - \hat{A})^2 p(A|x) \, dA \]

\[ = \int -2(A - \hat{A}) p(A|x) \, dA \]

\[ = -2 \int A p(A|x) \, dA + 2\hat{A} \int p(A|x) \, dA \]

Which results in

the MMSE estimator: \( \hat{A} = \int A p(A|x) \, dA \)
In determining the MMSE estimator, we require the \( p(A|x) \).

By Bayes' rule, we have

\[
p(A|x) = \frac{p(x|A)p(A)}{p(x)} = \frac{\int p(x|A)p(A) \, dA}{p(x)}
\]

Recall that \( p(A) = \mathcal{U}[-A_0, A_0] \) and \( x[n] = A + w[n] \) with \( w[n] \sim \mathcal{N}(0, \sigma^2) \). we have

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

Thus

\[
p(A|x) = \begin{cases} 
\frac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] & |A| \leq A_0 \\
\int_{-A_0}^{A_0} \frac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \, dA & |A| > A_0.
\end{cases}
\]
But, \[
\sum_{n=0}^{N-1} (x[n] - A)^2 = \sum_{n=0}^{N-1} x^2[n] - 2NA\bar{x} + NA^2
\]

\[
= N(A - \bar{x})^2 + \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2
\]

so that we have

\[
p(A|x) = \begin{cases} 
\frac{1}{c\sqrt{2\pi}\sigma^2_N} \exp \left[ -\frac{1}{2\sigma^2_N} (A - \bar{x})^2 \right] & |A| \leq A_0 \\
0 & |A| > A_0
\end{cases}
\]

where \[
c = \int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi}\sigma^2_N} \exp \left[ -\frac{1}{2\sigma^2_N} (A - \bar{x})^2 \right] dA
\]

- The PDF is a truncated Gaussian
The MMSE estimator is
\[ \hat{A} = E(A|x) = \int_{-\infty}^{\infty} A p(A|x) \, dA \]
\[ = \int_{-A_0}^{A_0} A \frac{1}{\sqrt{2\pi \frac{\sigma^2}{N}}} \exp \left[ -\frac{1}{2 \frac{\sigma^2}{N}} (A - \bar{x})^2 \right] \, dA \]

The estimator relies less on \( p(A) \) and more on data as \( N \) increases.
The Bayesian risk functions
- Let $\varepsilon = \theta - \hat{\theta}$ and the cost function $C(\varepsilon)$
- MSE: $C(\varepsilon) = \varepsilon^2$
- Absolute error: $C(\varepsilon) = |\varepsilon|$ (penalizing errors proportionally)
- Hit-Or-Miss: $C(\varepsilon) = \begin{cases} 0 & |\varepsilon| < \delta \\ 1 & |\varepsilon| \geq \delta \end{cases}$

The Bayesian risk $\mathcal{R} = E[C(\varepsilon)]$
$$= \int \int C(\theta - \hat{\theta})p(x, \theta) \, dx \, d\theta$$
$$= \int \left[ \int C(\theta - \hat{\theta})p(\theta|x) \, d\theta \right] p(x) \, dx.$$
Considering the absolute error cost function, we have

\[ g(\hat{\theta}) = \int |\theta - \hat{\theta}| p(\theta|x) \, d\theta \]

\[ = \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|x) \, d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta|x) \, d\theta. \]

To minimize \( g(\hat{\theta}) \) w.r.t. \( \hat{\theta} \), we differentiate \( g(\hat{\theta}) \) w.r.t. \( \hat{\theta} \)

Recall the Leibnitz’s rule

\[
\frac{\partial}{\partial u} \int_{\phi_1(u)}^{\phi_2(u)} h(u, v) \, dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u, v)}{\partial u} \, dv + \frac{d\phi_2(u)}{du} h(u, \phi_2(u)) - \frac{d\phi_1(u)}{du} h(u, \phi_1(u)).
\]

For the first integral, we have \( d\phi_1(u) / du = 0 \) and

\[ h(u, \phi_2(u)) = h(\hat{\theta}, \hat{\theta}) = (\hat{\theta} - \hat{\theta}) p(\hat{\theta}|x) = 0 \]
For the second integral, the corresponding term are zero

Hence, we obtain

$$\frac{dg(\hat{\theta})}{d\hat{\theta}} = \int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|x) d\theta = 0$$

Or

$$\int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta = \int_{\hat{\theta}}^{\infty} p(\theta|x) d\theta$$

Therefore, $\hat{\theta}$ is the median of the posterior PDF or the point for which $\Pr\{\theta \leq \hat{\theta} | x\} = 1/2$

For the MSE cost function, we already have $\hat{\theta} = E[\theta | x]$

We next determine $\hat{\theta}$ for the hit-and-miss cost function
For the hit-and-miss cost function, we have
\[ C(\varepsilon) = 1 \text{ for } \varepsilon > \delta \text{ and } \varepsilon < -\delta \text{ or for } \theta > \hat{\theta} + \delta \text{ and } \theta < \hat{\theta} - \delta \]

- Thus,
  \[ g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta} - \delta} 1 \cdot p(\theta|x) \, d\theta + \int_{\hat{\theta} + \delta}^{\infty} 1 \cdot p(\theta|x) \, d\theta \]

- But \( \int_{-\infty}^{\infty} p(\theta|x) \, d\theta = 1 \), yielding
  \[ g(\hat{\theta}) = 1 - \int_{\hat{\theta} - \delta}^{\hat{\theta} + \delta} p(\theta|x) \, d\theta \]

- This is minimized by maximizing \( \int_{\hat{\theta} - \delta}^{\hat{\theta} + \delta} p(\theta|x) \, d\theta \)

- For \( \delta \) arbitrary small, this is maximized by choosing \( \hat{\theta} \) to the location of the maximum of \( p(\theta|x) \), or the mode of \( p(\theta|x) \)

- The estimator is termed the \textit{maximum a posteriori} estimator
In summary, the estimators that minimize the Bayes risk for the MSE, absolute and the hit-and-miss cost functions are the mean, median and mode of the posterior PDF.

For some posterior PDFs, these estimators are the same. A notable example is the Gaussian posterior PDF.
Summary

Detection theory
- Neyman-Pearson detector
- Bayesian detector, including
  - MAP estimator
  - ML detector
  - Multiple hypothesis testing

Estimation theory
- MUV estimator
- Sufficient statistics
- The BLUE estimator
- The ML estimator
- The general Bayesian estimators