Parameter Estimation

Saravanan Vijayakumaran
sarva@ee.iitb.ac.in

Department of Electrical Engineering
Indian Institute of Technology Bombay

October 25, 2012

Motivation

System Model used to Derive Optimal Receivers

$$y(t) = s(t) + n(t)$$

$s(t)$ Transmitted Signal

$y(t)$ Received Signal

$n(t)$ Noise

Simplified System Model. Does Not Account For

- Propagation Delay
- Carrier Frequency Mismatch Between Transmitter and Receiver
- Clock Frequency Mismatch Between Transmitter and Receiver

In short, Lies! Why?


A Few Good Men © 1992
You want answers?
© 2008 moviescreenshots.blogspot.com


A Few Good Men © 1992
I want the truth!
© 2008 moviescreenshots.blogspot.com

. . . right at the beginning of the course. Now you can.

Why Study the Simplified System Model?

$$y(t) = s(t) + n(t)$$

- Receivers estimate propagation delay, carrier frequency and clock frequency before demodulation
- Once these unknown parameters are estimated, the simplified system model is valid
- Then why not study parameter estimation first?
  - Hypothesis testing is easier to learn than parameter estimation
  - Historical reasons

Unsimplifying the System Model

Effect of Propagation Delay

- Consider a complex baseband signal

$$s(t) = \sum_{n=-\infty}^{\infty} b_n p(t - nT)$$

and the corresponding passband signal

$$s_p(t) = \text{Re}\left[\sqrt{2} s(t) e^{j2\pi f_c t}\right].$$

- After passing through a noisy channel which causes amplitude scaling and delay, we have

$$y_p(t) = A s_p(t - \tau) + n_p(t)$$

where A is an unknown amplitude, τ is an unknown delay and $n_p(t)$ is passband noise

Unsimplifying the System Model

Effect of Propagation Delay

- The delayed passband signal is

$$
\begin{aligned}
s_p(t-\tau) &= \text{Re}\left[\sqrt{2}s(t-\tau)e^{j2\pi f_c(t-\tau)}\right] \\
&= \text{Re}\left[\sqrt{2}s(t-\tau)e^{j\theta}e^{j2\pi f_c t}\right]
\end{aligned}
$$

  where $\theta = -2\pi f_c \tau \mod 2\pi$. For large f_c, θ is modeled as uniformly distributed over $[0, 2\pi]$.

- The complex baseband representation of the received signal is then

$$
y(t) = Ae^{j\theta}s(t-\tau) + n(t)
$$

  where $n(t)$ is complex Gaussian noise.

Unsimplifying the System Model

Effect of Carrier Offset

- Frequency of the local oscillator (LO) at the receiver differs from that of the transmitter
- Suppose the LO frequency at the transmitter is f_c

$$s_p(t) = \text{Re}\left[\sqrt{2}s(t)e^{j2\pi f_c t}\right].$$

- Suppose that the LO frequency at the receiver is $f_c - \Delta f$
- The received passband signal is

$$y_p(t) = As_p(t-\tau) + n_p(t)$$

- The complex baseband representation of the received signal is then

$$y(t) = Ae^{j(2\pi\Delta f t + \theta)}s(t-\tau) + n(t)$$

Unsimplifying the System Model

Effect of Clock Offset

- Frequency of the clock at the receiver differs from that of the transmitter
- The clock frequency determines the sampling instants at the matched filter output
- Suppose the symbol rate at the transmitter is $\frac{1}{T}$ symbols per second
- Suppose the receiver sampling rate is $\frac{1+\delta}{T}$ symbols per second where $|\delta| \ll 1$ and δ may be positive or negative
- The actual sampling instants and ideal sampling instants will drift apart over time

The Solution

Estimate the unknown parameters τ, θ, Δf and δ

Timing Synchronization Estimation of τ

Carrier Synchronization Estimation of θ and Δf

Clock Synchronization Estimation of δ

Perform demodulation after synchronization

Parameter Estimation

Parameter Estimation

- Hypothesis testing was about making a choice between discrete states of nature
- Parameter or point estimation is about choosing from a continuum of possible states

Example

Consider the complex baseband signal below

$$y(t) = Ae^{j\theta} s(t - \tau) + n(t)$$

- The phase θ can take any real value in the interval $[0, 2\pi)$
- The amplitude A can be any real number
- The delay τ can be any real number

System Model for Parameter Estimation

- Consider a family of distributions

$$\mathbf{Y} \sim P_\theta, \quad \theta \in \Lambda$$

  where the observation vector $\mathbf{Y} \in \Gamma \subseteq \mathbb{R}^n$ for $n \in \mathbb{N}$ and $\Lambda \subseteq \mathbb{R}^m$ is the parameter space
- Example:

$$Y = A + N$$

  where A is an unknown parameter and N is a standard Gaussian RV
- The goal of parameter estimation is to find θ given $\mathbf{Y}$
- An estimator is a function from the observation space to the parameter space

$$\hat{\theta} : \Gamma \to \Lambda$$

Which is the Optimal Estimator?

- Assume there is a cost function C which quantifies the estimation error

$$C : \Lambda \times \Lambda \to \mathbb{R}$$

such that $C[a, \theta]$ is the cost of estimating the true value of θ as a

- Examples of cost functions

Squared Error $C[a, \theta] = (a - \theta)^2$

Absolute Error $C[a, \theta] = |a - \theta|$

Threshold Error $C[a, \theta] = \begin{cases} 0 & \text{if } |a - \theta| \leq \Delta \\ 1 & \text{if } |a - \theta| > \Delta \end{cases}$

Which is the Optimal Estimator?

- With an estimator $\hat{\theta}$ we associate a conditional cost or risk conditioned on θ

$$R_\theta(\hat{\theta}) = E_\theta \left\{ C \left[\hat{\theta}(\mathbf{Y}), \theta \right] \right\}$$

- Suppose that the parameter θ is the realization of a random variable Θ
- The average risk or Bayes risk is given by

$$r(\hat{\theta}) = E \left\{ R_\Theta(\hat{\theta}) \right\}$$

- The optimal estimator is the one which minimizes the Bayes risk

Which is the Optimal Estimator?

- Given that

$$R_\theta(\hat{\theta}) = E_\theta\left\{C\left[\hat{\theta}(\mathbf{Y}), \theta\right]\right\} = E\left\{C\left[\hat{\theta}(\mathbf{Y}), \Theta\right] \middle| \Theta = \theta\right\}$$

the average risk or Bayes risk is given by

$$\begin{aligned} r(\hat{\theta}) &= E\left\{C\left[\hat{\theta}(\mathbf{Y}), \Theta\right]\right\} \\ &= E\left\{E\left\{C\left[\hat{\theta}(\mathbf{Y}), \Theta\right] \middle| \mathbf{Y}\right\}\right\} \end{aligned}$$

- The optimal estimate for θ can be found by minimizing for each $\mathbf{Y} = \mathbf{y}$ the posterior cost

$$E\left\{C\left[\hat{\theta}(\mathbf{y}), \Theta\right] \middle| \mathbf{Y} = \mathbf{y}\right\}$$

Minimum-Mean-Squared-Error (MMSE) Estimation

- $C[a, \theta] = (a - \theta)^2$
- The posterior cost is given by

$$
\begin{aligned}
E\left\{(\hat{\theta}(\mathbf{y}) - \Theta)^2 \middle| \mathbf{Y} = \mathbf{y}\right\} &= \left[\hat{\theta}(\mathbf{y})\right]^2 \\
&\quad -2\hat{\theta}(\mathbf{y}) E\left\{\Theta \middle| \mathbf{Y} = \mathbf{y}\right\} \\
&\quad +E\left\{\Theta^2 \middle| \mathbf{Y} = \mathbf{y}\right\}
\end{aligned}
$$

- The Bayes estimate is given by

$$
\hat{\theta}_{MMSE}(\mathbf{y}) = E\left\{\Theta \middle| \mathbf{Y} = \mathbf{y}\right\}
$$

Example 1: MMSE Estimation

- Suppose X and Y are jointly Gaussian random variables
- Let the joint pdf be given by

$$p_{XY}(x, y) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{s} - \mu)^T \Sigma^{-1} (\mathbf{s} - \mu)\right)$$

where $\mathbf{s} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$

- Suppose Y is observed and we want to estimate X
- The MMSE estimate of X is

$$\hat{X}_{MMSE}(y) = E\left[X \middle| Y = y\right]$$

Example 1: MMSE Estimation

- The conditional distribution of X given $Y = y$ is a Gaussian RV with mean

$$\mu_{X|y} = \mu_x + \frac{\sigma_x}{\sigma_y}\rho(y - \mu_y)$$

and variance

$$\sigma^2_{X|y} = (1 - \rho^2)\sigma^2_x$$

- Thus the MMSE estimate of X given $Y = y$ is

$$\hat{X}_{MMSE}(y) = \mu_x + \frac{\sigma_x}{\sigma_y}\rho(y - \mu_y)$$

Example 2: MMSE Estimation

- Suppose A is a Gaussian RV with mean μ and known variance v^2
- Suppose we observe $Y_i,\ i = 1, 2, \ldots, M$ such that

$$Y_i = A + N_i$$

  where N_i's are independent Gaussian RVs with mean 0 and known variance σ^2
- Suppose A is independent of the N_i's
- The MMSE estimate is given by

$$\hat{A}_{MMSE}(\mathbf{y}) = \frac{\frac{Mv^2}{\sigma^2}\hat{A}_1(\mathbf{y}) + \mu}{\frac{Mv^2}{\sigma^2} + 1}$$

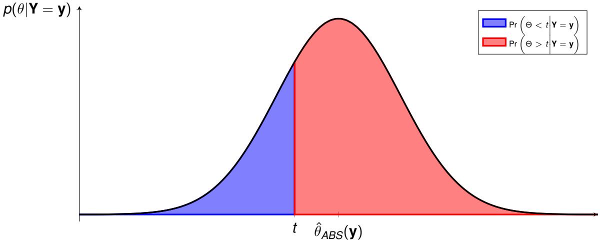
  where $\hat{A}_1(\mathbf{y}) = \frac{1}{M}\sum_{i=1}^{M} y_i$

Minimum-Mean-Absolute-Error (MMAE) Estimation

- $C[a, \theta] = |a - \theta|$
- The Bayes estimate $\hat{\theta}_{ABS}$ is given by the median of the posterior density $p(\Theta|\mathbf{Y} = \mathbf{y})$

$$\Pr\left(\Theta < t \middle| \mathbf{Y} = \mathbf{y}\right) \leq \Pr\left(\Theta > t \middle| \mathbf{Y} = \mathbf{y}\right), \quad t < \hat{\theta}_{ABS}(\mathbf{y})$$

$$\Pr\left(\Theta < t \middle| \mathbf{Y} = \mathbf{y}\right) \geq \Pr\left(\Theta > t \middle| \mathbf{Y} = \mathbf{y}\right), \quad t > \hat{\theta}_{ABS}(\mathbf{y})$$



Minimum-Mean-Absolute-Error (MMAE) Estimation

- For $\Pr[X \geq 0] = 1$, $E[X] = \int_0^\infty \Pr[X > x]\, dx$
- Since $|\hat{\theta}(\mathbf{y}) - \Theta| \geq 0$

$$E\left\{ |\hat{\theta}(\mathbf{y}) - \Theta| \middle| \mathbf{Y} = \mathbf{y} \right\}$$

$$\begin{aligned}
&= \int_0^\infty \Pr\left[|\hat{\theta}(\mathbf{y}) - \Theta| > x \middle| \mathbf{Y} = \mathbf{y} \right] dx \\
&= \int_0^\infty \Pr\left[\Theta > x + \hat{\theta}(\mathbf{y}) \middle| \mathbf{Y} = \mathbf{y} \right] dx \\
&\quad + \int_0^\infty \Pr\left[\Theta < -x + \hat{\theta}(\mathbf{y}) \middle| \mathbf{Y} = \mathbf{y} \right] dx \\
&= \int_{\hat{\theta}(\mathbf{y})}^\infty \Pr\left[\Theta > t \middle| \mathbf{Y} = \mathbf{y} \right] dt \\
&\quad + \int_{-\infty}^{\hat{\theta}(\mathbf{y})} \Pr\left[\Theta < t \middle| \mathbf{Y} = \mathbf{y} \right] dt
\end{aligned}$$

Minimum-Mean-Absolute-Error (MMAE) Estimation

Differentiating $E\left\{|\hat{\theta}(\mathbf{y}) - \Theta|\middle|\mathbf{Y} = \mathbf{y}\right\}$ wrt to $\hat{\theta}(\mathbf{y})$

$$
\begin{aligned}
&\frac{\partial}{\partial\hat{\theta}(\mathbf{y})} E\left\{|\hat{\theta}(\mathbf{y}) - \Theta|\middle|\mathbf{Y} = \mathbf{y}\right\} \\
&= \frac{\partial}{\partial\hat{\theta}(\mathbf{y})} \int_{\hat{\theta}(\mathbf{y})}^{\infty} \Pr\left[\Theta > t\middle|\mathbf{Y} = \mathbf{y}\right] dt \\
&\quad + \frac{\partial}{\partial\hat{\theta}(\mathbf{y})} \int_{-\infty}^{\hat{\theta}(\mathbf{y})} \Pr\left[\Theta < t\middle|\mathbf{Y} = \mathbf{y}\right] dt \\
&= \Pr\left[\Theta < \hat{\theta}(\mathbf{y})\middle|\mathbf{Y} = \mathbf{y}\right] - \Pr\left[\Theta > \hat{\theta}(\mathbf{y})\middle|\mathbf{Y} = \mathbf{y}\right]
\end{aligned}
$$

- The derivative is nondecreasing tending to -1 as $\hat{\theta}(\mathbf{y}) \to -\infty$ and $+1$ as $\hat{\theta}(\mathbf{y}) \to \infty$
- The minimum risk is achieved at the point the derivative changes sign

Minimum-Mean-Absolute-Error (MMAE) Estimation

- Thus the MMAE $\hat{\theta}_{ABS}$ is given by any value θ such that

$$\Pr\left(\Theta < t \middle| \mathbf{Y} = \mathbf{y}\right) \leq \Pr\left(\Theta > t \middle| \mathbf{Y} = \mathbf{y}\right), \quad t < \hat{\theta}_{ABS}(\mathbf{y})$$
$$\Pr\left(\Theta < t \middle| \mathbf{Y} = \mathbf{y}\right) \geq \Pr\left(\Theta > t \middle| \mathbf{Y} = \mathbf{y}\right), \quad t > \hat{\theta}_{ABS}(\mathbf{y})$$

- Why not the following expression?

$$\Pr\left(\Theta < \hat{\theta}_{ABS}(\mathbf{y}) \middle| \mathbf{Y} = \mathbf{y}\right) = \Pr\left(\Theta \geq \hat{\theta}_{ABS}(\mathbf{y}) \middle| \mathbf{Y} = \mathbf{y}\right)$$

- Why not the following expression?

$$\Pr\left(\Theta < \hat{\theta}_{ABS}(\mathbf{y}) \middle| \mathbf{Y} = \mathbf{y}\right) = \Pr\left(\Theta > \hat{\theta}_{ABS}(\mathbf{y}) \middle| \mathbf{Y} = \mathbf{y}\right)$$

- MMAE estimation for discrete distributions requires the more general expression above

Maximum A Posteriori (MAP) Estimation

- The MAP estimator is given by

$$\hat{\theta}_{MAP}(\mathbf{y}) = \underset{\theta}{\operatorname{argmax}}\, p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right)$$

- It can be obtained as the optimal estimator for the threshold cost function

$$C[a, \theta] = \begin{cases} 0 & \text{if } |a - \theta| \le \Delta \\ 1 & \text{if } |a - \theta| > \Delta \end{cases}$$

  for small $\Delta > 0$

Maximum A Posteriori (MAP) Estimation

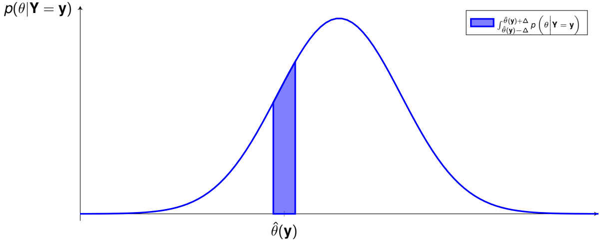
- For the threshold cost function, we have[1]

$$
\begin{aligned}
& E\left\{C\left[\hat{\theta}(\mathbf{y}), \Theta\right] \middle| \mathbf{Y}=\mathbf{y}\right\} \\
= & \int_{-\infty}^{\infty} C[\hat{\theta}(\mathbf{y}), \theta] p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right) \, d\theta \\
= & \int_{-\infty}^{\hat{\theta}(\mathbf{y})-\Delta} p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right) \, d\theta + \int_{\hat{\theta}(\mathbf{y})+\Delta}^{\infty} p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right) \, d\theta \\
= & \int_{-\infty}^{\infty} p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right) \, d\theta - \int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right) \, d\theta \\
= & \; 1 - \int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right) \, d\theta
\end{aligned}
$$

- The Bayes estimate is obtained by maximizing the integral in the last equality

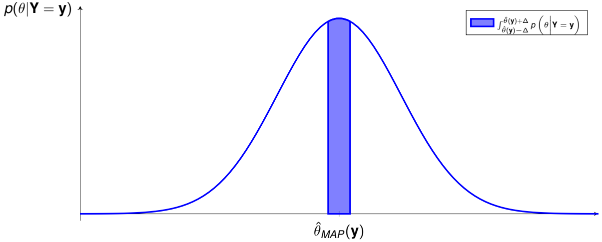
[1] Assume a scalar parameter θ for illustration

Maximum A Posteriori (MAP) Estimation



- The shaded area is the integral $\int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta$
- To maximize this integral, the location of $\hat{\theta}(\mathbf{y})$ should be chosen to be the value of θ which maximizes $p(\theta|\mathbf{Y} = \mathbf{y})$

Maximum A Posteriori (MAP) Estimation



- This argument is not airtight as $p(\theta|\mathbf{Y}=\mathbf{y})$ may not be symmetric at the maximum
- But the MAP estimator is widely used as it is easier to compute than the MMSE or MMAE estimators

Maximum Likelihood (ML) Estimation

- The ML estimator is given by

$$\hat{\theta}_{ML}(\mathbf{y}) = \underset{\theta}{\operatorname{argmax}}\, p\left(\mathbf{Y} = \mathbf{y} \middle| \theta\right)$$

- It is the same as the MAP estimator when the prior probability distribution of Θ is uniform
- It is also used when the prior distribution is not known

Example 1: ML Estimation

- Suppose we observe $Y_i, \; i = 1, 2, \ldots, M$ such that

$$Y_i \sim \mathcal{N}(\mu, \sigma^2)$$

  where Y_i's are independent, μ is unknown and σ^2 is known
- The ML estimate is given by

$$\hat{\mu}_{ML}(\mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} y_i$$

Assignment 5

Example 2: ML Estimation

- Suppose we observe $Y_i, \ i = 1, 2, \ldots, M$ such that

$$Y_i \sim \mathcal{N}(\mu, \sigma^2)$$

  where Y_i's are independent, both μ and σ^2 are unknown
- The ML estimates are given by

$$\begin{aligned} \hat{\mu}_{ML}(\mathbf{y}) &= \frac{1}{M} \sum_{i=1}^{M} y_i \\ \hat{\sigma}^2_{ML}(\mathbf{y}) &= \frac{1}{M} \sum_{i=1}^{M} (y_i - \hat{\mu}_{ML}(\mathbf{y}))^2 \end{aligned}$$

Assignment 5

Example 3: ML Estimation

- Suppose we observe $Y_i,\ i = 1, 2, \ldots, M$ such that

$$Y_i \sim \text{Bernoulli}(p)$$

  where Y_i's are independent and p is unknown
- The ML estimate of p is given by

$$\hat{p}_{ML}(\mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} y_i$$

Assignment 5

Example 4: ML Estimation

- Suppose we observe $Y_i, \; i = 1, 2, \ldots, M$ such that

$$Y_i \sim \text{Uniform}[0, \theta]$$

  where Y_i's are independent and θ is unknown
- The ML estimate of θ is given by

$$\hat{\theta}_{ML}(\mathbf{y}) = \max\left(y_1, y_2, \ldots, y_{M-1}, y_M\right)$$

Assignment 5

Reference

- Chapter 4, *An Introduction to Signal Detection and Estimation*, H. V. Poor, Second Edition, Springer Verlag, 1994.

Parameter Estimation of Random Processes

ML Estimation Requires Conditional Densities

- ML estimation involves maximizing the conditional density wrt unknown parameters
- Example: $Y \sim \mathcal{N}(\theta, \sigma^2)$ where θ is known and σ^2 is unknown

$$p\left(Y = y \middle| \theta\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}$$

- Suppose the observation is the realization of a random process

$$y(t) = Ae^{j\theta} s(t - \tau) + n(t)$$

- What is the conditional density of $y(t)$ given A, θ and τ?

Maximizing Likelihood Ratio for ML Estimation

- Consider $Y \sim \mathcal{N}(\theta, \sigma^2)$ where θ is unknown and σ^2 is known

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}$$

- Let $q(y)$ be the density of a Gaussian with distribution $\mathcal{N}(0, \sigma^2)$

$$q(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}}$$

- The ML estimate of θ is obtained as

$$\begin{aligned} \hat{\theta}_{ML}(y) &= \operatorname*{argmax}_{\theta} p(y|\theta) = \operatorname*{argmax}_{\theta} \frac{p(y|\theta)}{q(y)} \\ &= \operatorname*{argmax}_{\theta} L(y|\theta) \end{aligned}$$

where $L(y|\theta)$ is called the likelihood ratio

Likelihood Ratio and Hypothesis Testing

- The likelihood ratio $L(y|\theta)$ is the ML decision statistic for the following binary hypothesis testing problem

$$
\begin{aligned}
H_1 &: \quad Y \sim \mathcal{N}(\theta, \sigma^2) \\
H_0 &: \quad Y \sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

  where θ is assumed to be known
- H_0 is a dummy hypothesis which makes calculation of the ML estimator easy for random processes

Likelihood Ratio of a Signal in AWGN

- Let $H_s(\theta)$ be the hypothesis corresponding the following received signal

$$H_s(\theta) \quad : \quad y(t) = s_\theta(t) + n(t)$$

  where θ can be a vector parameter

- Define a noise-only dummy hypothesis H_0

$$H_0 \quad : \quad y(t) = n(t)$$

- Define Z and $y^\perp(t)$ as follows

$$\begin{aligned} Z &= \langle y, s_\theta \rangle \\ y^\perp(t) &= y(t) - \langle y, s_\theta \rangle \frac{s_\theta(t)}{\|s_\theta\|^2} \end{aligned}$$

- Z and $y^\perp(t)$ completely characterize $y(t)$

Likelihood Ratio of a Signal in AWGN

- Under both hypotheses $y^{\perp}(t)$ is equal to $n^{\perp}(t)$ where

$$n^{\perp}(t) = n(t) - \langle n, s_\theta \rangle \frac{s_\theta(t)}{\|s_\theta\|^2}$$

- $n^{\perp}(t)$ is independent of the noise component in Z and has the same distribution under both hypotheses
- $n^{\perp}(t)$ is irrelevant for this binary hypothesis testing problem
- The likelihood ratio of $y(t)$ equals the likelihood ratio of Z under the following hypothesis testing problem

$$\begin{aligned} H_s(\theta) &: \quad Z \sim \mathcal{N}(\|s_\theta\|^2, \sigma^2\|s_\theta\|^2) \\ H_0(\theta) &: \quad Z \sim \mathcal{N}(0, \sigma^2\|s_\theta\|^2) \end{aligned}$$

Likelihood Ratio of Signals in AWGN

- The likelihood ratio of signals in real AWGN is

$$L(y|s_\theta) = \exp\left(\frac{1}{\sigma^2}\left[\langle y, s_\theta\rangle - \frac{\|s_\theta\|^2}{2}\right]\right)$$

- The likelihood ratio of signals in complex AWGN is

$$L(y|s_\theta) = \exp\left(\frac{1}{\sigma^2}\left[\mathrm{Re}(\langle y, s_\theta\rangle) - \frac{\|s_\theta\|^2}{2}\right]\right)$$

- Maximizing these likelihood ratios as functions of θ results in the ML estimator

Thanks for your attention