Statistical Signal Processing

Lecture 5

chapter 1: parameter estimation: deterministic parameters

- some optimality properties

- Maximum Likelihood estimation, examples

- Fischer Information Matrix

- Cramer-Rao lower bound on the MSE, example

- linear model

- asymptotic (large sample) properties

- recap: estimator properties and estimators

- simplified estimators: BLUE, (W)LS, method of moments

Asymptotic (Large Sample) Properties

- asymptotic: $n \rightarrow \infty$

- *asymptotically unbiased* : $\lim_{n \to \infty} b_n(\theta) = 0$, $\forall \theta \in \Theta$

- Example (mean and variance of Gaussian i.i.d. variables):

$$E[\widehat{\sigma^2}_{ML}|\mu, \sigma^2] = \frac{n-1}{n}\sigma^2$$

$$b_n = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \xrightarrow{n \to \infty} 0$$

$\widehat{\sigma^2}_{ML}$: biased but asymptotically unbiased

- *consistency* : convergence of (a series of random vectors:) $\widehat{\theta}_n \rightarrow \theta$

  – convergence in probability

  – mean square convergence

  – convergence with probability one

  – convergence in distribution

Consistency

the sequence of estimates $\widehat{\theta}(Y_n)$ is said to be

- *simply* or *weakly consistent* if

$$\lim_{n\to\infty} \mathrm{Pr}_{Y_n|\theta}\left\{\|\widehat{\theta}(Y_n) - \theta\| < \epsilon\right\} = 1\,, \quad \forall \epsilon > 0, \ \forall \theta \in \Theta$$

- *mean-square consistent* if

$$\lim_{n\to\infty} \mathrm{MSE}_n = \lim_{n\to\infty} E_{Y_n|\theta}\|\widehat{\theta}(Y_n) - \theta\|^2 = 0\,, \quad \forall \theta \in \Theta$$

- *strongly consistent* if

$$\mathrm{Pr}_{Y_\infty|\theta}\left\{\lim_{n\to\infty} \widehat{\theta}(Y_n) = \theta\right\} = 1\,, \quad \forall \theta \in \Theta$$

- Any of these 3 consistencies implies asymptotic unbiasedness. E.g. for mean-square:

$$\underbrace{E_{Y_n|\theta}\|\widehat{\theta}(Y_n) - \theta\|^2}_{\text{MSE}} = \|\underbrace{E_{Y_n|\theta}\widehat{\theta}(Y_n) - \theta}_{\text{bias}}\|^2 + \underbrace{E_{Y_n|\theta}\|\widehat{\theta}(Y_n) - E_{Y_n|\theta}\widehat{\theta}\|^2}_{\text{variance}} \to 0$$

$$\Rightarrow \ \lim_{n\to\infty} E_{Y_n|\theta}\widehat{\theta}(Y_n) = \theta$$

Consistency (2)

- Strong and mean-square consistency do not imply each other in general. Either implies weak consistency (e.g. use the Chebyshev inequality to show that mean-square consistency implies weak consistency), but not conversely. Except when Θ is bounded: then weak consistency implies mean-square consistency.

- example: i.i.d. $y_i \sim \mathcal{N}(\mu, \sigma^2)$, $\theta = \mu$, σ^2 known. $\widehat{\mu}_{ML} = \overline{y}$

$$Var(\widehat{\mu}_{ML}) = \frac{\sigma^2}{n} \xrightarrow{n \to \infty} 0 \ \ \text{mean-square consistent}$$

- example: i.i.d. $y_i \sim U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, $\widehat{\theta}_{ML} = \frac{y_{min} + y_{max}}{2}$

$$\begin{cases} y_{min} \ \to \ \theta - \frac{1}{2} \ \ \text{in probability} \\ y_{max} \ \to \ \theta + \frac{1}{2} \ \ \text{in probability} \end{cases} \quad \text{weak consistency}$$

$$\widehat{\theta}_{ML} \ \to \ \theta \ \ \text{in probability}$$

mean-square consistency can also be shown

Asymptotic Normality

- if $\widehat{\theta}_n$ consistent, then $\widetilde{\theta} \to 0$ in some sense

- introduce a magnifying glass: $d_n(\widehat{\theta}_n - \theta)$ where $0 < d_{n-1} \leq d_n \to \infty$

- *convergence in distribution*: weaker than the 3 forms of convergence of sequences of random vectors mentioned before

- if $d_n(\widehat{\theta}_n - \theta) \overset{in\,dist.}{\longrightarrow} \xi$, some random vector, then the distribution of ξ useful as a measure for the limiting behavior of $\widehat{\theta}_n$

- usually $d_n = \sqrt{n}$

- $\widehat{\theta}_n$ *consistent asymptotically normal* (CAN) :
  if $\widehat{\theta}_n$ simply consistent and $d_n(\widehat{\theta}_n - \theta) \overset{in\,dist.}{\longrightarrow} \mathcal{N}(0, \Xi(\theta))$
  CAN implies asympt. unbiased (which requires that bias $\longrightarrow 0$ faster than $\frac{1}{d_n}$),
  $\Xi =$ asymptotic normalized covariance of $\widehat{\theta}_n$. We say that $\widehat{\theta}_n = \theta + \mathcal{O}_p(\frac{1}{d_n})$

- distinguish $\Xi(\theta)$ from $V(\theta) = \lim_{n \to \infty} d_n^2 C_{\widetilde{\theta}\widetilde{\theta}}(\theta)$ which may not even exist for a CAN estimate (if $\widehat{\theta}_n$ is simply but not mean-square consistent). $V(\theta)$ exists for a mean-square consistent $\widehat{\theta}_n$, but is not necessarily $= \Xi(\theta)$.

- Hence CAN can be used to formulate *interval estimators* on the basis of *point estimators*.

Asymptotic Optimality of ML

- *asymptotic normalized information matrix* : $J_0(\theta) = \lim\limits_{n \to \infty} \dfrac{1}{d_n^2} J_n(\theta)$ if it exists

  ($J_0(\theta)$ = asymptotic average information per data sample y_n if $d_n = \sqrt{n}$)

- *best asymptotically normal* (BAN): CAN and $\Xi(\theta) = J_0^{-1}(\theta)$

  also called *asymptotically efficient*

- under some regularity conditions (maximum of the likelihood function unique, y_i given θ i.i.d.,...) the ML estimate is strongly consistent and BAN with $d_n = \sqrt{n}$ ($\Rightarrow$ another use of the CRB). In particular, the ML estimate is

  - asymptotically unbiased
  - asymptotically efficient (i.i.d.: $J_n = nJ_1 \Rightarrow J_0 = J_1$)
  - asymptotically normal

- example: i.i.d. $y_i \sim \mathcal{N}(\mu, \sigma^2)\,,\; \theta = \mu,\; \sigma^2$ known. $\widehat{\mu}_{ML} = \overline{y}$

$$\widehat{\mu}_{ML} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \;\longrightarrow\; \sqrt{n}(\widehat{\mu}_{ML} - \mu) \sim \mathcal{N}(0, \sigma^2)\,,\; J_n = \frac{n}{\sigma^2} \Rightarrow J_0^{-1} = \sigma^2 = \Xi(\theta)$$

Recap: Properties of Estimators $\widehat{\theta}(Y)$

small sample (finite n):

- *bias*: $b_{\widehat{\theta}}(\theta) = E_{Y|\theta}\widehat{\theta}(Y) - \theta \quad (= 0, \ \forall \theta \in \Theta : \text{unbiased})$

- *error correlation*: $R_{\widetilde{\theta}\widetilde{\theta}} = E_{Y|\theta}\left(\widehat{\theta}(Y) - \theta\right)\left(\widehat{\theta}(Y) - \theta\right)^T$

  Cramer-Rao Bound : $\widehat{\theta}$ unbiased: $R_{\widetilde{\theta}\widetilde{\theta}} = C_{\widetilde{\theta}\widetilde{\theta}} = C_{\widehat{\theta}\widehat{\theta}}$

  $$C_{\widetilde{\theta}\widetilde{\theta}} \geq J^{-1}(\theta) \ , \quad J(\theta) = -E_{Y|\theta}\frac{\partial}{\partial\theta}\left(\frac{\partial \ln f(Y|\theta)}{\partial\theta}\right)^T \ \text{ information matrix}$$

  efficient: $C_{\widetilde{\theta}\widetilde{\theta}} = J^{-1}(\theta) \ , \ \forall \theta \in \Theta \quad \Rightarrow \quad \widehat{\theta}(Y) \text{ is UMVUE}$

large sample ($n \to \infty$):

- *asymptotically unbiased*: $\lim\limits_{n\to\infty} b_{\widehat{\theta}}(\theta) = 0, \ \forall \theta \in \Theta$

- *consistency* (weak, in mean square, strong): $\Rightarrow$ asymptotically unbiased

- *asymptotic normality*:

$$\text{BAN} \begin{cases} \diamond \text{ weakly consistent} \\ \diamond \text{ asymptotically normal} \\ \diamond \text{ asymptotically efficient} \end{cases} \left. \vphantom{\begin{cases} \\ \\ \end{cases}} \right\} \text{CAN}$$

Recap: Estimation Techniques

- *Uniformly Minimum Variance Unbiased Estimator* (UMVUE):
  complicated (via "sufficient statistics")

- *Maximum likelihood* (ML): $\widehat{\theta}_{ML} = \arg\max\limits_{\theta} f(Y|\theta)$

  Qualities:

  $\diamond$ if $\exists$ efficient $\widehat{\theta} = \widehat{\theta}_{eff}$ and $\widehat{\theta}_{ML}$ is obtained from $\dfrac{\partial \ln f(Y|\theta)}{\partial \theta} = 0$

      $\Rightarrow\ \widehat{\theta}_{eff} = \widehat{\theta}_{ML} = \widehat{\theta}_{UMVUE}$

  $\diamond$ $\widehat{\theta}_{ML}$ = BAN

  Problems:

  $\diamond$ what if $f(Y|\theta)$ is unknown?

  $\diamond$ if $f(Y|\theta)$ is not concave (local maxima)

- simplified estimators:

  $\diamond$ *Best Linear Unbiased Estimator* (BLUE) $\rightarrow$ linear model

  $\diamond$ *Method of Moments*

  $\diamond$ *Least-Squares* (LS) $\rightarrow$ linear model

Best Linear Unbiased Estimator (BLUE)

- deterministic analog of LMMSE in the Bayesian case

- *linear*: $\widehat{\theta}(Y) = FY \quad (F : m \times n)$

- *unbiased*: $E_{Y|\theta}\widehat{\theta} = F\,E\,(Y|\theta) = \theta$

- *best = minimum variance*: $\quad \min C_{\widetilde{\theta}\widetilde{\theta}}$

- remarks:

  – BLUE inferior to UMVUE unless UMVUE is linear

  – generalizations: $X = g(Y) \; : \; \widehat{\theta}(Y) = F\,X = F\,g(Y) \quad$ (linear in X)

      e.g.: linear in Y inappropriate if $\theta \neq 0$ and $E\,(Y|\theta) = 0$

Example of $X = g(Y)$

- $y_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d., $\theta = \sigma^2$, $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

- linear: $\widehat{\sigma^2} = F\,Y \;\Rightarrow\; E_{Y|\sigma^2}\widehat{\sigma^2} = F\,E\left(Y|\sigma^2\right) = 0 \neq \sigma^2$
  no linear unbiased estimator $\widehat{\sigma^2}$ exists

- however, let $x_i = y_i^2$, $\; X = \begin{bmatrix} y_1^2 \\ \vdots \\ y_n^2 \end{bmatrix}$

- $\widehat{\sigma^2} = F\,X \;\Rightarrow\; E_{Y|\sigma^2}\widehat{\sigma^2} = F\,E\left(X|\sigma^2\right) = \sigma^2\,F\,\mathbf{1} = \sigma^2 \;\Rightarrow\; F\,\mathbf{1} = 1$

- for this problem: $\widehat{\sigma^2}_{UMVUE} = \frac{1}{n}\mathbf{1}^T X = \widehat{\sigma^2}_{BLUE} \;\; (F = \frac{1}{n}\mathbf{1}^T)$

BLUE Assumptions

- unbiased: $\quad F E\left(Y|\theta\right) = \theta\,, \quad \forall\theta \in \Theta$

  unbiasedness and the requirement that a large class of linear unbiased estimators (many F satisfying $F\,E\left(Y|\theta\right) = \theta$) should exist naturally lead to:

- *assumption 1* : $\quad E\left(Y|\theta\right) = H\,\theta\,, \quad (H : n \times m)$

  unbiasedness $\quad \rightarrow \quad F\,H = I_m \quad (\Rightarrow\ n \geq m)$

- variance:

$$C_{\widetilde{\theta}\widetilde{\theta}} = C_{\widehat{\theta}\widehat{\theta}} = E_{Y|\theta}\left(\widehat{\theta} - E_{Y|\theta}\widehat{\theta}\right)\left(\widehat{\theta} - E_{Y|\theta}\widehat{\theta}\right)^T$$
$$= E_{Y|\theta}\left(F\,Y - F\,E\left(Y|\theta\right)\right)\left(F\,Y - F\,E\left(Y|\theta\right)\right)^T$$
$$= F\,E_{Y|\theta}\left(Y - E\left(Y|\theta\right)\right)\left(Y - E\left(Y|\theta\right)\right)^T F^T = F\,C_{YY}(\theta)\,F^T$$

- *assumption 2* : $\quad C_{YY}(\theta) = c(\theta)\,C$

  $c(\theta)\ (> 0, \forall\theta)$ is a scalar function of θ, $C > 0$ is constant w.r.t. θ

BLUE Optimization Problem

- $$\min_{\hat{\theta}: E_{Y|\theta}\hat{\theta}(Y)=\theta} C_{\tilde{\theta}\tilde{\theta}} \quad \rightarrow \quad \min_{F: FH=I} F\,C\,F^T$$

- introduce matrix square root $B\,(n \times n)$ of $C = C^T > 0\,(n \times n)$: $C = B\,B^T$
  notation: $B = C^{1/2}$, $C^{T/2} = (C^{1/2})^T$, $C = C^{1/2}C^{T/2}$, $C^{-1} = C^{-T/2}C^{-1/2}$

- Consider a vector space of $m \times n$ matrices with matrix inner product $\langle X_1, X_2 \rangle = X_1 X_2^T$. Take $X_1 = H^T C^{-T/2}$, $X_2 = F\,C^{1/2}$. With $FH = I$:

$$\left\langle \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right\rangle = \begin{bmatrix} H^T C^{-T/2} \\ F\,C^{1/2} \end{bmatrix} \begin{bmatrix} H^T C^{-T/2} \\ F\,C^{1/2} \end{bmatrix}^T = \begin{bmatrix} H^T C^{-1} H & I \\ I & FCF^T \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \geq 0$$

- From the Schur Complements Lemma, $R_{22} \geq R_{21}R_{11}^{-1}R_{12}$ with equality iff $X_2 = R_{21}R_{11}^{-1}X_1$.

- Hence $\min_{F: FH=I} F\,C\,F^T = \left(H^T C^{-1} H\right)^{-1}$ for $F = \left(H^T C^{-1} H\right)^{-1} H^T C^{-1}$.

- Or $\hat{\theta}_{BLUE} = \left(H^T C^{-1} H\right)^{-1} H^T C^{-1} Y = \left(H^T C_{YY}^{-1} H\right)^{-1} H^T C_{YY}^{-1} Y$
  with $C_{\tilde{\theta}\tilde{\theta}} = F\,C_{YY}\,F^T = c(\theta)\,F\,C\,F^T = c(\theta)\left(H^T C^{-1} H\right)^{-1} = \left(H^T C_{YY}^{-1} H\right)^{-1}$

BLUE: Example Cont'd and Recap

Example Cont'd:

- $y_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d., $\theta = \sigma^2$, $x_i = y_i^2$, $\widehat{\sigma^2} = F\,X$

- BLUE assumptions OK: $E\left(X|\sigma^2\right) = \mathbf{1}\,\sigma^2 = H\,\theta$, $C_{XX} = 2\sigma^4\,I = c(\theta)\,C$

- $\widehat{\sigma^2}_{BLUE} = \left(H^T C^{-1} H\right)^{-1} H^T C^{-1} X = \frac{1}{n}\mathbf{1}^T X$

  $$C_{\widehat{\sigma^2}\widehat{\sigma^2}}(\sigma^2) = \left(H^T C_{XX}^{-1} H\right)^{-1} = \frac{2\sigma^4}{n}$$

- note: this example is not a linear model!

Recap: BLUE assumptions:

- $\begin{cases} (1)\ E(Y|\theta) = H\,\theta \\ (2)\ C_{YY}(\theta) = c(\theta)\,C \end{cases}$

  Only need to know the first two moments of $f(Y|\theta)$ which need to satisfy these assumptions. The higher-order moments of $f(Y|\theta)$: don't need to know, can be arbitrary functions of θ. So the problem should more or less look like a linear model problem, up to the second-order moments.

BLUE: Linear Model

- $Y = H\,\theta + V$, $EV = 0$, $EVV^T = C_{VV}$

  (EV and C_{VV} independent of θ, only first two moments of V specified)

- BLUE assumptions satisfied:

$$\begin{cases} E\left(Y|\theta\right) = H\,\theta \\ C_{YY}(\theta) = E_{Y|\theta}\left(Y - E\left(Y|\theta\right)\right)\left(Y - E\left(Y|\theta\right)\right)^T = E_V VV^T = C_{VV} = C\ (c(\theta) = 1) \end{cases}$$

- $\widehat{\theta}_{BLUE} = \left(H^T C_{VV}^{-1} H\right)^{-1} H^T C_{VV}^{-1} Y$ with $C_{\tilde{\theta}\tilde{\theta}} = \left(H^T C_{VV}^{-1} H\right)^{-1}$

- If $V \sim \mathcal{N}(0, C_{VV})$ then $\widehat{\theta}_{BLUE} = \widehat{\theta}_{ML} =$ efficient $\Rightarrow = \widehat{\theta}_{UMVUE}$

Method of Moments

Principle:

- m unknown parameters $\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}$

- $f(Y|\theta)$ depends on $\theta \quad \Rightarrow \quad$ its moments also

- take m moments $\mu = g(\theta) = \begin{bmatrix} g_1(\theta) \\ \vdots \\ g_m(\theta) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$

  such that $g(.)$ is invertible, i.e. $\theta = g^{-1}(\mu)$: can determine θ from μ.

- estimate the moments: $\widehat{\mu}$ (e.g. sample moments)

- method of moments: $\widehat{\theta}_{MM} = g^{-1}(\widehat{\mu})$

Method of Moments: Example 1

- $y_i, i = 1, \dots, n$ i.i.d., $f(y|\theta)$ mixture distribution, θ mixture parameter

$$f(y|\theta) = (1-\theta)\phi_1(y) + \theta\phi_2(y) , \quad \phi_k(y) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{y^2}{2\sigma_k^2}}, k = 1, 2$$

- $\mu = E\left(y^2|\theta\right) = (1-\theta)\,\sigma_1^2 + \theta\,\sigma_2^2 = g(\theta) \;\Rightarrow\; \theta = g^{-1}(\mu) = \dfrac{\mu - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}$

- $\widehat{\theta}_{MM} = g^{-1}(\widehat{\mu}) = \dfrac{\widehat{\mu} - \sigma_1^2}{\sigma_2^2 - \sigma_1^2} , \quad \widehat{\mu} = \dfrac{1}{n} \sum\limits_{i=1}^{n} y_i^2$ sample mean squared value

- bias: $E\widehat{\theta} = \frac{1}{\sigma_2^2-\sigma_1^2}E\widehat{\mu} - \frac{\sigma_1^2}{\sigma_2^2-\sigma_1^2} = \frac{1}{\sigma_2^2-\sigma_1^2}\mu - \frac{\sigma_1^2}{\sigma_2^2-\sigma_1^2} = \theta$: unbiased

Method of Moments: Example 1 (cont'd)

-

$$Var\left(\widehat{\theta}\right) = Var\left(\frac{1}{\sigma_2^2-\sigma_1^2}\widehat{\mu} - \frac{\sigma_1^2}{\sigma_2^2-\sigma_1^2}\right) = \frac{1}{\left(\sigma_2^2-\sigma_1^2\right)^2}Var\left(\widehat{\mu}\right) = \frac{1}{\left(\sigma_2^2-\sigma_1^2\right)^2}Var\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2\right)$$

$$= \frac{1}{\left(\sigma_2^2-\sigma_1^2\right)^2}\sum_{i=1}^{n}Var\left(\frac{1}{n}y_i^2\right) = \frac{1}{\left(\sigma_2^2-\sigma_1^2\right)^2}\sum_{i=1}^{n}\frac{1}{n^2}Var\left(y_i^2\right) = \frac{1}{\left(\sigma_2^2-\sigma_1^2\right)^2}\frac{1}{n}Var\left(y^2\right)$$

$$f(y|\theta) = (1-\theta)\,\phi_1(y) \; +\theta\,\phi_2(y)$$

- $Var(y^2) = Ey^4 - (Ey^2)^2 \,,$

$$Ey^2 = (1-\theta)\,\sigma_1^2 \quad +\theta\,\sigma_2^2$$

$$Ey^4 = (1-\theta)\,3\sigma_1^4 \quad +\theta\,3\sigma_2^4$$

- $\Rightarrow \;\; Var(\widehat{\theta}_{MM}) = \dfrac{3(1-\theta)\sigma_1^4 + 3\theta\sigma_2^4 - \left[(1-\theta)\sigma_1^2 + \theta\sigma_2^2\right]^2}{n\,(\sigma_1^2 - \sigma_2^2)^2} \;\; \overset{n\to\infty}{\Longrightarrow} \;\; 0$

$\Rightarrow \;\; \widehat{\theta}_{MM}$ = mean-square consistent

MM Example 2: Sinusoid in White Noise

- $y_k = s_k + v_k = A\cos(\omega k + \phi) + v_k\,, \quad k = 1, \ldots, n$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ S = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix},\ V = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix},\ \theta = \begin{bmatrix} A \\ \omega \\ \sigma_v^2 \end{bmatrix},\ \Theta : A > 0, \omega \in [0, \pi], \sigma_v^2 > 0$$

- distributions: $\phi \sim \mathcal{U}[0, 2\pi]$ independent of $\theta, V;\ EV = 0,\ EVV^T = \sigma_v^2 I_n$

  randomness: $f(Y, \phi | \theta) = f(\phi | \theta)\, f(Y | \theta, \phi) = f(\phi)\, f_{\mathbf{V}|\sigma_v^2}(Y - S(A, \omega, \phi) | \sigma_v^2)$

  in what follows: only first and second moments of V needed

- mean: $E_{Y,\phi|\theta}\, y_k = AE\cos(\omega k + \phi) + Ev_k = 0$

  covariance sequence:

$$\begin{aligned}
r_{yy}(i) &=\ Ey_k y_{k+i} = A^2 E\cos(\omega k + \phi)\cos(\omega k + \phi + \omega i) \\
&\quad + AE\cos(\omega k + \phi)Ev_{k+i} + AE\cos(\omega k + \phi + \omega i)Ev_k + Ev_k v_{k+i} \\
&=\ \frac{A^2}{2}E\cos(2\omega k + 2\phi + \omega i) + \frac{A^2}{2}E\cos(\omega i) + \sigma_v^2 \delta_{i0} \\
&=\ \frac{A^2}{2}\cos(\omega i) + \sigma_v^2 \delta_{i0}
\end{aligned}$$

MM Example 2: Sinusoid in White Noise (2)

- moments:
$$\mu = \begin{bmatrix} r_{yy}(0) \\ r_{yy}(1) \\ r_{yy}(2) \end{bmatrix} = \begin{bmatrix} \dfrac{A^2}{2} + \sigma_v^2 \\ \dfrac{A^2}{2}\cos(\omega) \\ \dfrac{A^2}{2}\cos(2\omega) \end{bmatrix} = g(\theta)$$

- $\theta = g^{-1}(\mu)$:

$$\omega = \begin{cases} \arccos\left(\dfrac{r_{yy}(2) + \sqrt{r_{yy}^2(2) + 8r_{yy}^2(1)}}{4\,r_{yy}(1)}\right) & ,\ r_{yy}(1) \neq 0 \\[4mm] \dfrac{\pi}{2} & ,\ r_{yy}(1) = 0 \end{cases}$$

$$A = \begin{cases} \sqrt{\dfrac{2\,r_{yy}(1)}{\cos(\omega)}} & ,\ r_{yy}(1) \neq 0 \\[4mm] \sqrt{-2\,r_{yy}(2)} & ,\ r_{yy}(1) = 0 \end{cases} \quad,\quad \sigma_v^2 = r_{yy}(0) - \dfrac{A^2}{2}$$

- sample moments $\widehat{\mu}$: $\hat{r}_{yy}(i) = \dfrac{1}{n}\sum_{k=1}^{n-i} y_k y_{k+i}\,,\ i = 0,1,2$

Method of Moments: Properties

- $\widehat{\mu}$ easy to compute, $\widehat{\theta}_{MM} = g^{-1}(\widehat{\mu})$ straightforward if μ chosen well, hence $\widehat{\theta}_{MM}$ easy to determine and easy to implement

- no optimality properties but usually consistent (since $\widehat{\mu}$ consistent)

- if performance of $\widehat{\theta}_{MM}$ not satisfactory, can use $\widehat{\theta}_{MM}$ as initialization in an iterative optimization procedure that finds $\widehat{\theta}_{ML}$