

$$17) \quad a) \quad p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x(n) - A \cos 2\pi f_0 n)^2}$$

$$\text{But } \sum_n (x(n) - A \cos 2\pi f_0 n)^2 =$$

$$\sum_n x^2(n) - 2A \sum_n x(n) \cos 2\pi f_0 n + A^2 \sum_n \cos^2 2\pi f_0 n$$

$$p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} (A^2 \sum_n \cos^2 2\pi f_0 n - 2A \sum_n x(n) \cos 2\pi f_0 n)}$$

$$= \underbrace{e^{-\frac{1}{2\sigma^2} \sum_n x^2(n)}}_{h(\underline{x})} \underbrace{e^{-\frac{1}{2\sigma^2} (A^2 \sum_n \cos^2 2\pi f_0 n - 2A \sum_n x(n) \cos 2\pi f_0 n)}}_{g(T(\underline{x}), A)}$$

where $T(\underline{x}) = \sum_{n=0}^{N-1} x(n) \cos 2\pi f_0 n$ is the sufficient statistic

$$E(T(\underline{x})) = \sum_{n=0}^{N-1} A \cos^2 2\pi f_0 n$$

$$\Rightarrow \hat{A} = \frac{\sum_{n=0}^{N-1} x(n) \cos 2\pi f_0 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n}$$

$$b) \quad \text{Let } \underline{\theta} = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}$$

From part a we have

$$T(\underline{x}) = \begin{bmatrix} \sum_n x(n) \cos 2\pi f_0 n \\ \sum_n x^2(n) \end{bmatrix}$$

is a sufficient statistic

To make $T_1(x)$ unbiased let

$$\hat{A} = \frac{\sum_n x[n] \cos 2\pi f_0 n}{\sum_n \cos^2 2\pi f_0 n} \quad \text{as before.}$$

To make $T_2(x)$ unbiased:

$$\begin{aligned} E(T_2(x)) &= \sum_n E(x^2[n]) \\ &= \sum_n E[(A \cos 2\pi f_0 n + w[n])^2] \end{aligned}$$

$$= \sum_n A^2 \cos^2 2\pi f_0 n + N \sigma^2$$

From Example 5.11 we expect that we will have to subtract out the squared mean to generate an unbiased estimator of σ^2 .

But

$$\begin{aligned} E(\hat{A}^2) &= \frac{E\left[\left(\sum_n x[n] \cos 2\pi f_0 n\right)^2\right]}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2} \\ &= \frac{\sum_{m,n} E(x[m] x[n]) \cos 2\pi f_0 m \cos 2\pi f_0 n}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2} \\ &= \frac{\sum_{m,n} \left(A^2 \cos 2\pi f_0 m \cos 2\pi f_0 n + \sigma^2 \delta_{mn} \right) \cdot \begin{matrix} (\cos 2\pi f_0 m) \\ (\cos 2\pi f_0 n) \end{matrix}}{\left(\sum_n \cos^2 2\pi f_0 n\right)^2} \end{aligned}$$

$$= \frac{A^2 \left(\sum_n \cos^2 2\pi f_0 n \right)^2 + \sigma^2 \sum_n \cos^2 2\pi f_0 n}{\left(\sum_n \cos^2 2\pi f_0 n \right)^2}$$

$$= A^2 + \frac{\sigma^2}{\sum_n \cos^2 2\pi f_0 n}$$

so that if $T_2'(x) = T_2(x) - \sum_n \cos^2 2\pi f_0 n \hat{A}^2$

$$\begin{aligned} E(T_2'(x)) &= A^2 \sum_n \cos^2 2\pi f_0 n + N \sigma^2 \\ &\quad - A^2 \sum_n \cos^2 2\pi f_0 n \cdot \sigma^2 \\ &= (N-1) \sigma^2 \end{aligned}$$

\Rightarrow Let $T_2''(x) = \frac{1}{N-1} T_2'(x)$

$$= \frac{1}{N-1} \left[\sum_n x^2(n) - \sum_n \cos^2 2\pi f_0 n \hat{A}^2 \right]$$

$\therefore \hat{\theta} = \begin{bmatrix} \hat{A} \\ \sigma^2 \end{bmatrix}$

$$= \begin{bmatrix} \frac{\sum_{n=0}^{N-1} x(n) \cos 2\pi f_0 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \\ \frac{1}{N-1} \left[\sum_{n=0}^{N-1} x^2(n) - \hat{A}^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right] \end{bmatrix}$$

$$18) \quad p(x(n)) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \theta_1 \leq x(n) \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$$p(\underline{x}; \underline{\theta}) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^N} & \text{all } x(n) \text{ satisfy} \\ & \theta_1 \leq x(n) \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

Alternatively, for the PDF to be nonzero

$\min x(n) \geq \theta_1, \max x(n) \leq \theta_2$ so that

$$p(\underline{x}; \underline{\theta}) = \underbrace{\frac{1}{(\theta_2 - \theta_1)^N} u(\min x(n) - \theta_1) u(\max x(n) - \theta_2)}_{g(T(\underline{x}), \underline{\theta})}$$

$$\underbrace{\quad}_{h(\underline{x})}$$

$$\Rightarrow T(\underline{x}) = \begin{bmatrix} \min x(n) \\ \max x(n) \end{bmatrix} \text{ is a sufficient statistic}$$

$$\begin{aligned} 19) \quad & (\underline{X} - \underline{H}\hat{\underline{\theta}})^T (\underline{X} - \underline{H}\hat{\underline{\theta}}) + (\hat{\underline{\theta}} - \underline{\theta})^T \underline{H}^T \underline{H} (\hat{\underline{\theta}} - \underline{\theta}) \\ &= (\underline{X} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X})^T (\underline{X} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X}) \\ &\quad + (\hat{\underline{\theta}} - (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X})^T \underline{H}^T \underline{H} (\hat{\underline{\theta}} - (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X}) \\ &= \underline{X}^T (\underline{I} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T) (\underline{I} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T) \underline{X} \\ &\quad + \hat{\underline{\theta}}^T \underline{H}^T \underline{H} \hat{\underline{\theta}} - \hat{\underline{\theta}}^T \underline{H}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X} \\ &\quad - \underline{X}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} \hat{\underline{\theta}} + \underline{X}^T \underline{H}^T (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X} \\ &= \underline{X}^T \underline{X} - \underline{X}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X} + \hat{\underline{\theta}}^T \underline{H}^T \underline{H} \hat{\underline{\theta}} - 2 \hat{\underline{\theta}}^T \underline{H}^T \underline{X} \\ &\quad + \underline{X}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{X} = (\underline{X} - \underline{H}\hat{\underline{\theta}})^T (\underline{X} - \underline{H}\hat{\underline{\theta}}) \end{aligned}$$

$$\begin{aligned}
 p(\underline{x}; \underline{\theta}) &= \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} (\underline{x} - H\underline{\theta})^T (\underline{x} - H\underline{\theta})} \\
 &= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} (\underline{\theta} - \hat{\underline{\theta}})^T H^T H (\underline{\theta} - \hat{\underline{\theta}})}}_{g(T(\underline{x}), \underline{\theta})} \underbrace{e^{-\frac{1}{2\sigma^2} (\underline{x} - H\hat{\underline{\theta}})^T (\underline{x} - H\hat{\underline{\theta}})}}_{h(\underline{x})}
 \end{aligned}$$

where $T(\underline{x}) = \hat{\underline{\theta}}$ = sufficient statistic
 Since we already know that $\hat{\underline{\theta}}$ is unbiased,
 it is the MVU estimator. This is not
 unexpected since we saw in Chapter 3
 that $\hat{\underline{\theta}}$ is efficient.

Chapter 6

$$1) \quad \hat{A} = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{x}$$

$$\text{where } \underline{H} = \begin{bmatrix} 1 \\ r \\ \vdots \\ r^{N-1} \end{bmatrix} \quad \underline{C} = \sigma^2 \underline{I}$$

$$\begin{aligned} \hat{A} &= \left(\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} \right)^{-1} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} x[n] r^n \\ &= \frac{\sum_{n=0}^{N-1} x[n] r^n}{\sum_{n=0}^{N-1} r^{2n}} \end{aligned}$$

$$\text{var}(\hat{A}) = \frac{1}{\underline{H}^T \underline{C}^{-1} \underline{H}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}}$$

$$\text{var}(\hat{A}) \rightarrow 0 \quad \text{if } |r| \geq 1.$$

$$2) \quad \text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} 1/\sigma_n^2}$$

$$\text{If } \sigma_n^2 = n+1, \quad \text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} 1/(n+1)}$$

As $N \rightarrow \infty$, $\sum_{n=0}^{N-1} \frac{1}{n+1} \rightarrow \infty$ since this is a harmonic series $\Rightarrow \text{var}(\hat{A}) \rightarrow 0$

$$\text{If } \sigma_n^2 = (n+1)^2, \quad \text{as } N \rightarrow \infty \quad \sum_{n=0}^{N-1} \frac{1}{(n+1)^2} \rightarrow \text{Constant} \\ \Rightarrow \text{var}(\hat{A}) \not\rightarrow 0.$$

In this case the noise samples have such a large variance that the estimator

variance does not go to zero.

$$3) \quad \hat{A} = \frac{\underline{1}^T \underline{C}^{-1} \underline{x}}{\underline{1}^T \underline{C}^{-1} \underline{1}}$$

$$\underline{C}^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} \underline{B} & & 0 \\ & \underline{B} & \\ 0 & \ddots & \underline{B} \end{bmatrix} \quad \text{where } \underline{B} = \frac{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}{1-\rho^2}$$

$\underline{1}^T \underline{C}^{-1} \underline{1} =$ Sum of all elements in \underline{C}^{-1}

$$= \frac{N/2}{\sigma^2} \frac{2-2\rho}{1-\rho^2} = \frac{N}{\sigma^2(1+\rho)}$$

Let $\underline{x} = [\underline{x}_1^T \underline{x}_2^T \dots \underline{x}_{N/2}^T]^T$ where each \underline{x}_i is 2×1 .

$$\underline{1}^T \underline{C}^{-1} \underline{x} = \frac{1}{\sigma^2} \underline{1}^T \begin{bmatrix} \underline{B} \underline{x}_1 \\ \vdots \\ \underline{B} \underline{x}_{N/2} \end{bmatrix}$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{N/2} \underset{\substack{\uparrow \\ 2 \times 1}}{\underline{1}^T \underline{B} \underline{x}_i}$$

$$\text{But } \underline{B}^T \underline{1} = \frac{\begin{bmatrix} 1-\rho \\ 1-\rho \end{bmatrix}}{1-\rho^2} \Rightarrow \underline{1}^T \underline{B} \underline{x}_i = \underline{x}_i^T \underline{B}^T \underline{1} = \frac{(1-\rho)[\underline{x}_i]_1 + (1-\rho)[\underline{x}_i]_2}{1-\rho^2}$$

$$= \frac{[\underline{x}_i]_1 + [\underline{x}_i]_2}{1+\rho}$$

$$\underline{1}^T \underline{C}^{-1} \underline{x} = \frac{1}{\sigma^2} \frac{1}{1+\rho} \sum_{i=1}^{N/2} ([\underline{x}_i]_1 + [\underline{x}_i]_2)$$

$$= \frac{1}{\sigma^2} \frac{1}{1+\rho} \sum_{n=0}^{N-1} x(n)$$

$$\hat{A} = \frac{\frac{1}{\sigma^2} \frac{N \bar{x}}{1+p}}{\frac{N}{\sigma^2(1+p)}} = \bar{x}$$

$$\text{var}(\hat{A}) = \frac{1}{\frac{N}{\sigma^2(1+p)}} = \frac{\sigma^2(1+p)}{N}$$

Since the subvectors $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{N/2}$ are uncorrelated, we average them. For a single subvector we also average the samples (see Probs. 3.9 and 4.11). Hence, we obtain \bar{x} . The variance is σ^2/N for $p=0$ (our usual case), $2\sigma^2/N$ for $p \rightarrow 1$ since the samples of each subvector are equal and hence we have only $N/2$ uncorrelated samples, and $\rightarrow 0$ for $p \rightarrow -1$ since then the noise samples cancel (see Probs. 3.9 and 4.11).

4) In either case we have the model

$$\underline{x} = \underline{1}u + \underline{w} \quad \text{where } E(\underline{w}) = \underline{0}$$

$$\text{and } E(\underline{w}\underline{w}^T) = \text{var}(\underline{w}) \underline{I}$$

The BLUE for each case is

$$\hat{u} = \frac{\underline{1}^T \underline{C}^{-1} \underline{x}}{\underline{1}^T \underline{C}^{-1} \underline{1}} = \frac{\underline{1}^T \underline{x}}{\underline{1}^T \underline{1}} = \bar{x}$$

But in the Gaussian case the BLUE is also the MVU estimator - not so for the Laplacian PDF.

$$(5) \quad E(x) = \int_0^{\infty} x \frac{1}{\sqrt{2\pi} x} e^{-\frac{1}{2}(\ln x - \theta)^2} dx$$

$$\text{Let } y = \ln x \quad dy/dx = 1/x \Rightarrow dx = e^y dy$$

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2} e^y dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2 - 2y(\theta+1) + (\theta+1)^2)} \\ &\quad \cdot e^{-\frac{1}{2}(\theta^2 - (\theta+1)^2)} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - (\theta+1))^2} dy \\ &\quad \cdot e^{-\frac{1}{2}(-2\theta-1)} \\ &= e^{\theta+1/2} \end{aligned}$$

$$\text{Now let } y = \ln x, \quad dy/dx = 1/x = e^{-y}$$

$$\begin{aligned} p(y) &= \frac{p(x(y))}{|dy/dx|} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}}{e^{-y}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2} \end{aligned}$$

$$\Rightarrow y \sim N(\theta, 1)$$

The BLUE is $\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} y[n]$

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

b) For $\hat{\theta}$ to be unbiased

$$\begin{aligned} E(\hat{\theta}) &= E\left(\sum_n a_n x[n] + b\right) \\ &= \sum_n a_n (0.5) + \beta + b \\ &= \theta \end{aligned}$$

$$\Rightarrow \sum_n a_n = 1, \quad \beta \sum_n a_n = -b$$

$$\begin{aligned} \text{or } \hat{\theta} &= \sum_n a_n x[n] - \beta \sum_n a_n \\ &= \sum_n a_n (x[n] - \beta) \end{aligned}$$

Let $x'[n] = x[n] - \beta$. Then, we have the same problem as before. Thus,

$$\hat{\theta} = \frac{\underline{y}^T \underline{C}^{-1} \underline{x}'}{\underline{y}^T \underline{C}^{-1} \underline{y}} = \frac{\underline{y}^T \underline{C}^{-1} (\underline{x} - \beta \underline{1})}{\underline{y}^T \underline{C}^{-1} \underline{y}}$$

and since the covariance for $\underline{x} - \beta \underline{1}$ is \underline{C} ,

$$\text{var}(\hat{\theta}) = \frac{1}{\underline{y}^T \underline{C}^{-1} \underline{y}}$$

7) $\hat{A} = \frac{\underline{y}^T \underline{C}^{-1} \underline{x}}{\underline{y}^T \underline{C}^{-1} \underline{y}}$

Assume $\underline{C} \underline{y} = \lambda \underline{y}$. Then, $\underline{C}^{-1} \underline{y} = 1/\lambda \underline{y}$

$$\hat{A} = \frac{\underline{x}^T \underline{C}^{-1} \underline{y}}{\underline{y}^T \underline{C}^{-1} \underline{y}} = \frac{\underline{x}^T \underline{1}/\lambda \underline{y}}{\underline{y}^T \underline{1}/\lambda \underline{y}} = \frac{\underline{x}^T \underline{y}}{\underline{y}^T \underline{y}}$$

$$\text{var}(\hat{A}) = \frac{1}{\underline{y}^T \underline{C}^{-1} \underline{y}} = \frac{\lambda}{\underline{y}^T \underline{y}}$$

In this case we obtain same result as if $\underline{C} = \sigma^2 \underline{I}$. We do not need a prewhitener.

$$\begin{aligned} 8) \quad \text{var}(\hat{A}) &= \frac{1}{\underline{y}^T \underline{C}^{-1} \underline{y}} \\ &= \frac{1}{\left(\sum_i \alpha_i \underline{v}_i \right)^T \underline{C}^{-1} \left(\sum_j \alpha_j \underline{v}_j \right)} \\ &= \frac{1}{\sum_i \sum_j \alpha_i \alpha_j \underbrace{\underline{v}_i^T \underline{C}^{-1} \underline{v}_j}_{\underline{v}_i^T \underline{1}/\lambda_j \underline{v}_j}} \\ &= \frac{1}{\sum_i \alpha_i^2 / \lambda_i} \end{aligned}$$

$$\begin{aligned} \Sigma = \underline{y}^T \underline{y} &= \left(\sum_i \alpha_i \underline{v}_i \right)^T \left(\sum_j \alpha_j \underline{v}_j \right) \\ &= \sum_i \sum_j \alpha_i \alpha_j \underbrace{\underline{v}_i^T \underline{v}_j}_{\delta_{ij}} = \sum_i \alpha_i^2 \end{aligned}$$

Must minimize $\frac{1}{\sum_i \alpha_i^2 / \lambda_i}$ subject to constraint $\sum_i \alpha_i^2 = \epsilon_0$. Equivalently, we must maximize $\sum_i \alpha_i^2 / \lambda_i$. Using Lagrangian multipliers

$$F = \sum_i \alpha_i^2 / \lambda_i + \lambda (\sum_i \alpha_i^2 - \epsilon_0)$$

$$\frac{\partial F}{\partial \alpha_k} = \frac{2\alpha_k}{\lambda_k} + \lambda 2\alpha_k = 0$$

$$\Rightarrow \alpha_k = 0 \text{ or}$$

$$\lambda = -1/\lambda_k \text{ all } k$$

Clearly, we cannot have $\alpha_k = 0$ for all k , since then constraint could not be satisfied. Since the eigenvalues are distinct, we also cannot have $\lambda = -1/\lambda_k$ for all k . Thus, we must have

$$\alpha_k = 0 \text{ except for } k=j$$

$$\lambda = -1/\lambda_j \text{ and } \alpha_j \neq 0.$$

Hence, $\underline{s} = \alpha_j \underline{v}_j$. To determine which eigenvector to use

$$\text{var}(\hat{A}) = \frac{1}{\alpha_j^2 / \lambda_j} = \lambda_j / \alpha_j^2$$

And since $\epsilon_0 = \alpha_j^2$, $\text{var}(\hat{A}) = \lambda_j / \epsilon_0$ so that λ_j should be the minimum eigenvalue. Hence, the optimal signal is

$$\underline{s} = c \underline{v}_{\text{MIN}} = \sqrt{\epsilon_0} \underline{v}_{\text{MIN}}$$

where $\underline{v}_{\text{MIN}}$ is the eigenvector associated with the smallest eigenvalue. Intuitively, we place signal along direction where there is the least amount of noise.

9) $\theta = A$

$$\underline{s} = [1 \cos 2\pi f_1 \dots \cos 2\pi f_1 (N-1)]^T$$

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

$$= \frac{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_1 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n}$$

This is a scaled Fourier coefficient since

$$\hat{A} = \frac{N/2}{\sum_n \cos^2 2\pi f_1 n} \underbrace{\frac{2}{N} \sum_n x[n] \cos 2\pi f_1 n}_{\text{Fourier coefficient}}$$

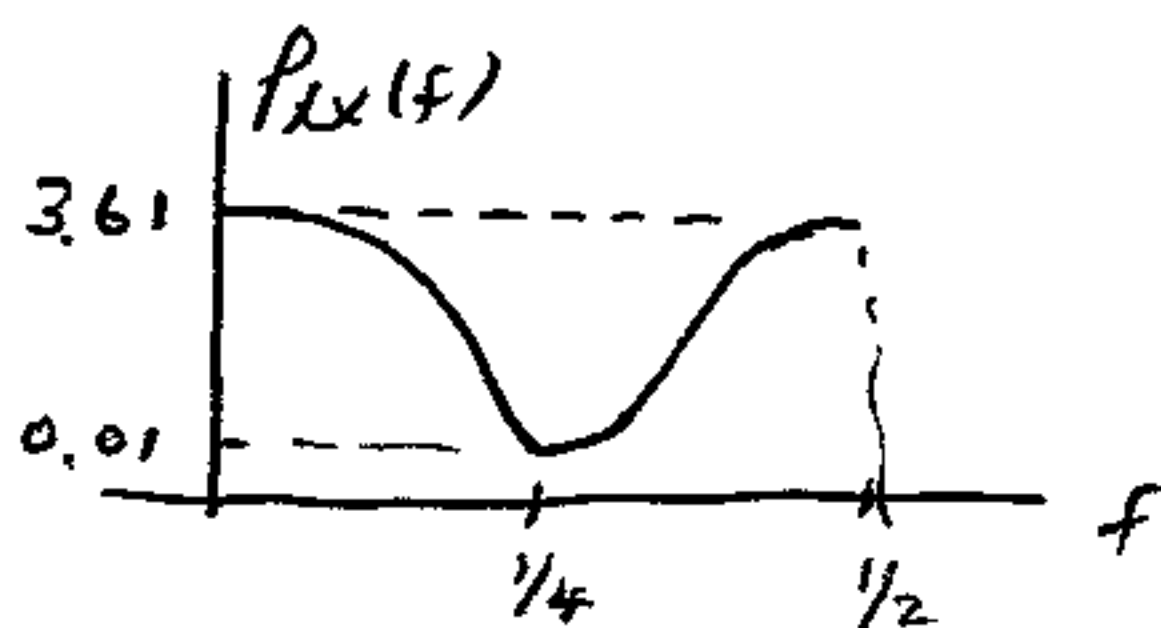
and for large N and f_1 not near 0 or $1/2$ the scale factor is one.

$$\begin{aligned} \text{var}(\hat{A}) &= \frac{1}{\underline{S}^T \underline{C}^{-1} \underline{S}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} S^2[n]} \\ &= \frac{\sigma^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n} \geq \frac{\sigma^2}{N} \end{aligned}$$

Since $\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n \leq N$

\Rightarrow variance is minimized if $f_1 = 0$.
Note that this choice maximizes the signal energy.

$$\begin{aligned} 10) \quad P_{xx}(f) &= r_{xx}(0) + r_{xx}(2) e^{-j4\pi f} + r_{xx}(2) e^{j4\pi f} \\ &= 1.81 + 2(0.9) \cos 4\pi f \\ &= 1.81 + 1.8 \cos 4\pi f \end{aligned}$$

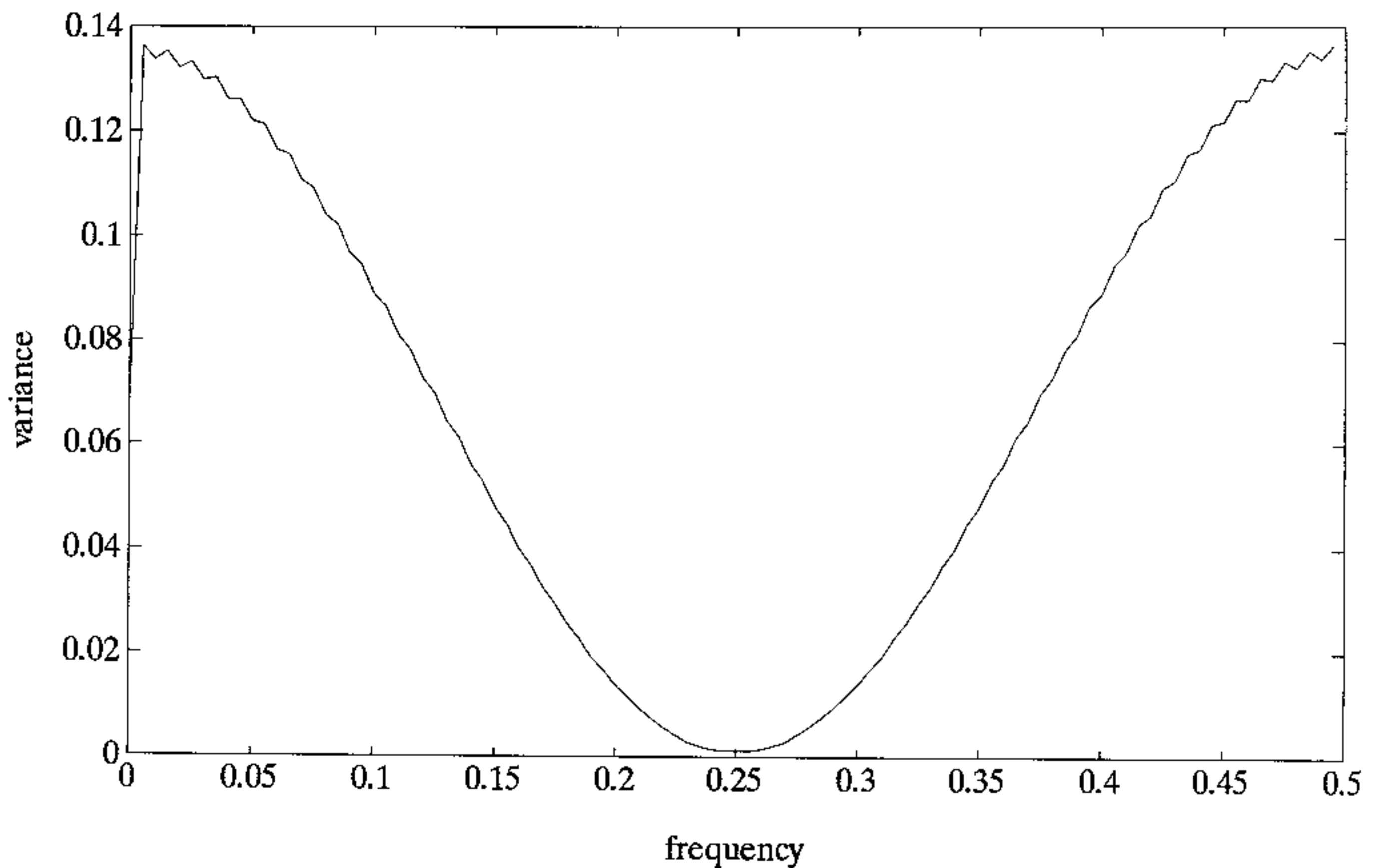


Need to minimize $\frac{1}{\underline{S}^T \underline{C}^{-1} \underline{S}} = \text{var}(\hat{A})$
where $\underline{S} = [1 \cos 2\pi f_1 \dots \cos 2\pi f_1 (N-1)]^T$
and

$$\underline{C} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \dots & r_{xx}(N-1) \\ r_{xx}(1) & r_{xx}(0) & & \\ & & \ddots & \\ & & & r_{xx}(0) \end{bmatrix}$$

$$= \begin{bmatrix} 1.81 & 0 & 0.9 & 0 & \dots & 0 \\ 0 & 1.81 & 0 & 0.9 & \dots & 0 \\ & & \ddots & & & \\ & & & & & 1.81 \end{bmatrix}$$

As seen in the following graph, the variance is minimized for $f_1 = 0.25$ or where the PSD is minimum.



$$11) \quad H(e^{j2\pi f}) = \sum_{n=0}^{N-1} h[n] e^{-j2\pi f n}$$

$$H(e^{j0}) = \sum_{n=0}^{N-1} h[n] = 1$$

The noise power at the output at $n = N-1$ is

$$\begin{aligned} & E \left[\left(\sum_k h[k] w[N-1-k] \right)^2 \right] \\ &= \sum_k \sum_l h[k] h[l] \underbrace{E[w[N-1-k] w[N-1-l]]}_{r_{ww}[k-l]} \\ &= \underline{h}^T \underline{C} \underline{h} \quad \text{where } (C)_{kl} = r_{ww}[k-l] \end{aligned}$$

Thus, to find the FIR filter coefficients we need to minimize $\underline{h}^T \underline{C} \underline{h}$ subject to the constraint $\underline{h}^T \underline{1} = 1$. This is just the BLUE setup so that

$$\underline{h}_{opt} = \frac{\underline{C}^{-1} \underline{1}}{\underline{1}^T \underline{C}^{-1} \underline{1}}$$

The noise power at the output is

$$\underline{h}_{opt}^T \underline{C} \underline{h}_{opt} = \frac{\underline{1}^T \underline{C}^{-1} \underline{C} \underline{C}^{-1} \underline{1}}{(\underline{1}^T \underline{C}^{-1} \underline{1})^2}$$

$$= \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}$$

which is just the variance of \hat{A} since

$$\begin{aligned} \text{var}(\hat{A}) &= E[(\hat{A} - E(\hat{A}))^2] \\ &= E\left[\left(\sum_k h[k] x[N-1-k] - \sum_k h[k] A\right)^2\right] \end{aligned}$$

Since $\sum_k h[k] = 1$

$$\begin{aligned} &= E\left[\left(\sum_k h[k] (x[N-1-k] - A)\right)^2\right] \\ &= E\left[\left(\sum_k h[k] w[N-1-k]\right)^2\right] \end{aligned}$$

The BLUE may be viewed as the output of a linear filter constrained to pass the DC signal and whose coefficients are chosen to minimize the noise at the filter output.

12) Since $\underline{x} = \underline{H}\underline{\theta} + \underline{w}$ and \underline{B} is invertible,

$$\underline{0} = \underline{B}^{-1}(\underline{x} - \underline{b}) \Rightarrow \underline{x} = \underline{H}\underline{B}^{-1}(\underline{x} - \underline{b}) + \underline{w}$$

$$\text{or } \underline{x} = \underline{H}\underline{B}^{-1}\underline{x} - \underline{H}\underline{B}^{-1}\underline{b} + \underline{w}$$

$$\underbrace{\underline{x} + \underline{H}\underline{B}^{-1}\underline{b}}_{\underline{x}'} = \underbrace{\underline{H}\underline{B}^{-1}\underline{x}}_{\underline{u}'} + \underline{w}$$

$$\begin{aligned}
\Rightarrow \hat{\underline{\alpha}} &= (\underline{H}'^T \underline{C}^{-1} \underline{H}')^{-1} \underline{H}'^T \underline{C}^{-1} \underline{x}' \\
&= (\underline{B}^{-1T} \underline{H}^T \underline{C}^{-1} \underline{H} \underline{B}^{-1})^{-1} \underline{B}^{-1T} \underline{H}^T \underline{C}^{-1} (\underline{x} + \underline{H} \underline{B}^{-1} \underline{b}) \\
&= \underline{B} (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} (\underline{x} + \underline{H} \underline{B}^{-1} \underline{b}) \\
&= \underline{B} \hat{\underline{\theta}} + \underline{B} \underline{B}^{-1} \underline{b} = \underline{B} \hat{\underline{\theta}} + \underline{b}
\end{aligned}$$

$$\begin{aligned}
13) \quad J &= \underline{x}^T \underline{C}^{-1} \underline{x} - \underline{x}^T \underline{C}^{-1} \underline{H} \underline{\theta} - \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{x} \\
&\quad + \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{H} \underline{\theta} \\
&= \underline{x}^T \underline{C}^{-1} \underline{x} - 2 \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{x} + \underline{\theta}^T \underline{H}^T \underline{C}^{-1} \underline{H} \underline{\theta}
\end{aligned}$$

Using (4.3) we have

$$\frac{\partial J}{\partial \underline{\theta}} = -2 \underline{H}^T \underline{C}^{-1} \underline{x} + 2 \underline{H}^T \underline{C}^{-1} \underline{H} \underline{\theta} = \underline{0}$$

$$\Rightarrow \hat{\underline{\theta}} = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{x}$$

14) Since $p(W(n))$ is even, $E(W(n)) = 0$.

$$\text{var}(W(n)) = E(W^2(n))$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} W^2 \left(\frac{1-E}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2} W^2/\sigma_0^2} \right. \\
&\quad \left. + \frac{E}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2} W^2/\sigma_1^2} \right) dW
\end{aligned}$$

$$\begin{aligned}
&= (1-\epsilon) \int_{-\infty}^{\infty} W^2 \frac{1}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{1}{2}W^2/\sigma_B^2} dW \\
&\quad + \epsilon \int_{-\infty}^{\infty} W^2 \frac{1}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{1}{2}W^2/\sigma_I^2} dW \\
&= (1-\epsilon)\sigma_B^2 + \epsilon\sigma_I^2
\end{aligned}$$

The BLUE of σ^2 is $\hat{\sigma}^2 \equiv \frac{1}{N} \sum_{n=0}^{N-1} W^2(n)$ (see Section 6.3). This is because the $W(n)$'s are independent and thus $y(n) = W^2(n)$'s are independent. The mean is $E(y(n)) = \sigma^2$ and covariance matrix is $\underline{C} = \text{var}(y(n)) \underline{I}$ so that

$$\hat{\sigma}^2 = \frac{\underline{y}^T \underline{C}^{-1} \underline{y}}{\underline{y}^T \underline{C}^{-1} \underline{y}} = \frac{\underline{1}^T \underline{y}}{\underline{1}^T \underline{1}} = \frac{1}{N} \sum_{n=0}^{N-1} y(n)$$

Now using results from Prob. 6.12

$$\sigma_I^2 = \frac{\sigma^2 - (1-\epsilon)\sigma_B^2}{\epsilon}$$

$$\begin{aligned}
\Rightarrow \hat{\sigma}_I^2 &= \frac{\hat{\sigma}^2 - (1-\epsilon)\sigma_B^2}{\epsilon} \\
&= \frac{\frac{1}{N} \sum_{n=0}^{N-1} W^2(n) - (1-\epsilon)\sigma_B^2}{\epsilon}
\end{aligned}$$

$$15) \quad \underbrace{\underline{X} - \underline{\varepsilon}}_{\underline{X}'} = \underline{H} \underline{\theta} + \underline{W}$$

$$\Rightarrow \hat{\underline{\theta}} = (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} \underline{X}'$$

$$= (\underline{H}^T \underline{C}^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}^{-1} (\underline{X} - \underline{\varepsilon})$$

$$16) \quad \hat{A} = (\underline{1}^T \hat{\underline{C}}^{-1} \underline{1})^{-1} \underline{1}^T \hat{\underline{C}}^{-1} \underline{X}$$

$$E(\hat{A}) = (\underline{1}^T \hat{\underline{C}}^{-1} \underline{1})^{-1} \underline{1}^T \hat{\underline{C}}^{-1} \underbrace{E(\underline{X})}_{\underline{A}} = A$$

\Rightarrow unbiased for any $\hat{\underline{C}}$

$$\text{var}(\hat{A}) = E[(\hat{A} - A)^2] = E[(\underline{1}^T \hat{\underline{C}}^{-1} \underline{1})^{-1} \underline{1}^T \hat{\underline{C}}^{-1} \underline{W}]^2]$$

$$= \frac{\underline{1}^T \hat{\underline{C}}^{-1} E(\underline{W} \underline{W}^T) \hat{\underline{C}}^{-1} \underline{1}}{(\underline{1}^T \hat{\underline{C}}^{-1} \underline{1})^2}$$

$$= \frac{\underline{1}^T \hat{\underline{C}}^{-1} \underline{C} \hat{\underline{C}}^{-1} \underline{1}}{(\underline{1}^T \hat{\underline{C}}^{-1} \underline{1})^2}$$

$$\text{If } \underline{C} = \underline{I}, \quad \hat{\underline{C}} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} \Rightarrow \hat{\underline{C}}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix}$$

$$\text{var}(\hat{A}) = \frac{\underline{1}^T \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix} \underline{1}}{(\underline{1}^T \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix} \underline{1})^2}$$

$$= \frac{1 + 1/\alpha^2}{(1 + 1/\alpha)^2}$$

Clearly, by the minimum variance property of the BLUE

$$\text{var}(\hat{A})_{\text{MIN}} = 1/2 \quad \text{for } \alpha = 1$$

$$\text{Let } J = \frac{\text{var}(\hat{A})}{\text{var}(\hat{A})_{\text{MIN}}} = \frac{2(1 + 1/\alpha^2)}{(1 + 1/\alpha)^2}$$

$$= \frac{2(1 + \alpha^2)}{(1 + \alpha)^2}$$



For $\alpha \rightarrow 0$ or $\alpha \rightarrow \infty$ we discard one data sample and thus the variance doubles. For $\alpha = 1$ we have the BLUE and hence the minimum variance.

Chapter 7

$$1) \quad p(\underline{x}; A) = \frac{1}{(2\pi A)^{N/2}} e^{-\frac{1}{2A} \sum_n (x[n] - A)^2}$$

$$\frac{\partial \log p}{\partial A} = -\frac{N}{2A} + \frac{1}{A} \sum_n (x[n] - A) + \frac{1}{2A^2} \sum_n (x[n] - A)^2$$

$$\begin{aligned} \frac{\partial^2 \log p}{\partial A^2} &= \frac{N}{2A^2} - \frac{1}{A^2} \sum_n x[n] - \frac{1}{A^2} \sum_n (x[n] - A) \\ &\quad - \frac{1}{A^3} \sum_n (x[n] - A)^2 \end{aligned}$$

$$\begin{aligned} E \left[\frac{\partial^2 \log p}{\partial A^2} \right] &= \frac{N}{2A^2} - \frac{NA}{A^2} - 0 - \frac{1}{A^3} (NA) \\ &= -\frac{N}{2A^2} - \frac{N}{A} \end{aligned}$$

$$I(A) = \frac{N}{2A^2} + \frac{N}{A} \Rightarrow \text{var}(\hat{A}) \geq \frac{1}{\frac{N}{2A^2} + \frac{N}{A}}$$

$$\text{or } \text{var}(\hat{A}) \geq \frac{A^2}{N(A + \frac{1}{2})}$$

$$2) \quad \text{var}(\bar{x}) = \sigma^2/N = A/N$$

$$\text{But } \text{var}(\hat{A}) \geq \frac{A}{N} \left(\frac{A}{A + 1/2} \right) < A/N$$

Even as $N \rightarrow \infty$, \bar{x} does not attain CRLB. Thus, MLE is better (at least for large data records). For finite data records we would need to determine the exact mean and variance of \hat{A} and compare them to \bar{x} .

$$\begin{aligned}
 3) \quad a) \quad p(\underline{x}; \mu) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x[n] - \mu)^2} \\
 &= \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2} \sum_{n=0}^{N-1} (x[n] - \mu)^2}
 \end{aligned}$$

To maximize p , we minimize $\sum_n (x[n] - \mu)^2$. Since it is a quadratic in μ , differentiation produces a global minimum.

$$\Rightarrow \sum_n (x[n] - \mu) = 0 \Rightarrow \hat{\mu} = \bar{x}$$

(This is just a DC level, μ , in WGN).

$$b) \quad p(\underline{x}; \lambda) = \lambda^N e^{-\lambda \sum_n x[n]} \quad \begin{array}{l} \text{all } x[n] > 0 \\ 0 \quad \text{otherwise} \end{array}$$

Assuming all $x[n] > 0$ we have

$$p = \lambda^N e^{-\lambda N \bar{x}}$$

$$\frac{dp}{d\lambda} = N \lambda^{N-1} e^{-\lambda N \bar{x}} + \lambda^N (-N \bar{x}) e^{-\lambda N \bar{x}} = 0$$

$$\Rightarrow \hat{\lambda} = 1/\bar{x}$$

To verify that $\hat{\lambda}$ yields the global maximum we can consider $\ln p$, which is a monotonic function of p .

$$\ln p = N \ln \lambda - \lambda N \bar{x}$$

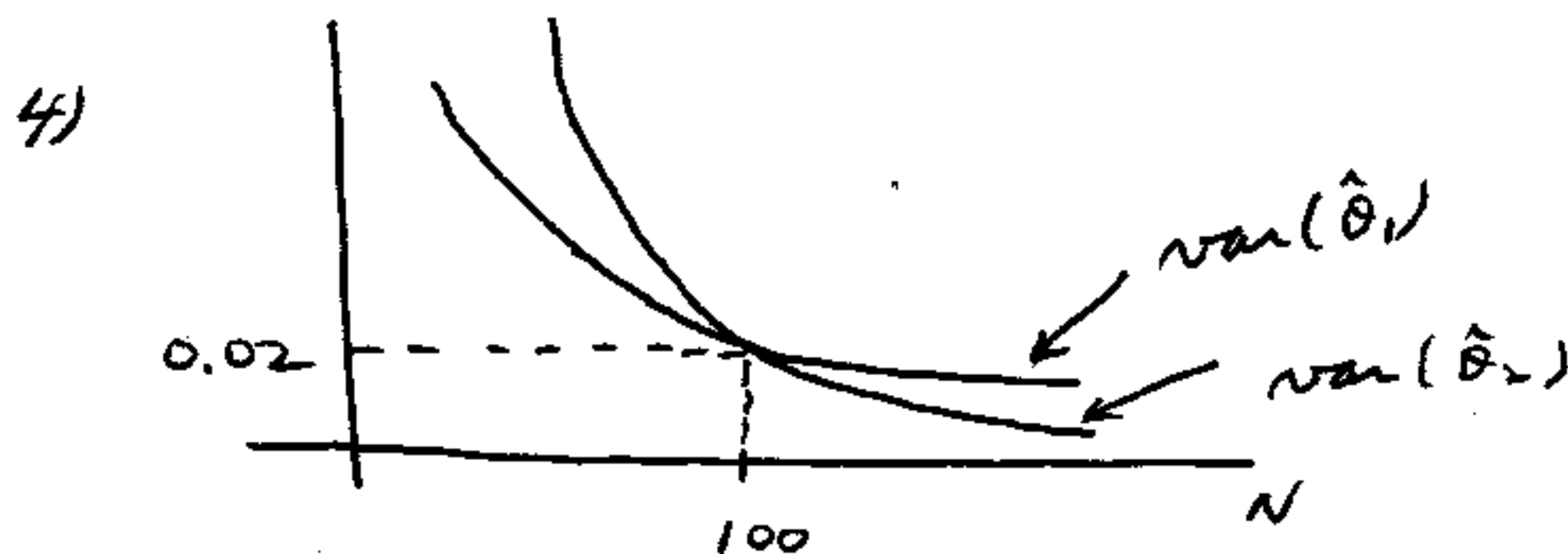
$$\frac{d \ln p}{d \lambda} = N/\lambda - N \bar{x}$$

$$\frac{d^2 \ln p}{d \lambda^2} = -N/\lambda^2 < 0 \quad \text{for all } \lambda$$

$\Rightarrow \ln p$ is concave function

$\Rightarrow \hat{\lambda}$ is global maximum solution

This result is reasonable since the mean of x is $1/\lambda$.



$\hat{\theta}_1$ better for $N < 100$

$\hat{\theta}_2$ better for $N > 100$

$$5) \quad P\{|\hat{A} - A| > \epsilon\} = P\left\{\left|\frac{\bar{x} - A}{\sigma/\sqrt{N}}\right| > \frac{\epsilon}{\sigma/\sqrt{N}}\right\}$$

$$\leq \frac{1}{(\epsilon/\sigma/\sqrt{N})^2}$$

$$= \frac{\sigma^2}{N \epsilon^2} \rightarrow 0 \text{ as } N \rightarrow \infty$$

$\Rightarrow \bar{x}$ is consistent

- 6) Linearizing about the true value of θ , i.e., θ_0 yields

$$\alpha = g(\theta) = g(\theta_0) + \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\theta - \theta_0)$$

$$\begin{aligned} \text{But } \hat{\alpha} - \alpha &= g(\hat{\theta}) - g(\theta_0) \\ &\approx \left[g(\theta_0) + \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \right] - g(\theta_0) \\ &= \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \end{aligned}$$

$$\begin{aligned} P_r \{ |\hat{\alpha} - \alpha| > \epsilon \} &= P_r \left\{ \left| \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \right| > \epsilon \right\} \\ &= P_r \left\{ |\hat{\theta} - \theta_0| > \frac{\epsilon}{\left| \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} \right|} \right\} \rightarrow 0 \text{ as } N \rightarrow \infty \end{aligned}$$

since $\hat{\theta}$ is consistent for θ (as long as $dg/d\theta$ is bounded).

$$\begin{aligned} 7) \quad p(\underline{x}; \theta) &= \prod_{n=0}^{N-1} e^{A(\theta)B(x(n)) + C(x(n)) + D(\theta)} \\ &= e^{A(\theta) \sum_n B(x(n)) + \sum_n C(x(n)) + ND(\theta)} \end{aligned}$$

To maximize p we must minimize

$$A(\theta) \sum_n B(x(n)) + ND(\theta)$$

Differentiating produces the necessary condition

$$\frac{dA(\theta)}{d\theta} \sum_{n=0}^N B(x(n)) + N \frac{dD(\theta)}{d\theta} = 0$$

For the PDFs of Prob. 7.2

$$a) p(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2\mu x + \mu^2)}$$

$$\Rightarrow A(\theta) = \mu \quad B(x) = x \quad D(\theta) = -\frac{1}{2} \mu^2$$

$$(i) \sum_n x(n) + N(-\mu) = 0 \Rightarrow \hat{\mu} = \bar{x}$$

$$b) p(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

$$= \begin{cases} e^{-\lambda x + \ln \lambda} & x > 0 \\ 0 & x < 0 \end{cases}$$

$$\Rightarrow A(\theta) = -\lambda \quad B(x) = x \quad D(\theta) = \ln \lambda$$

$$(-i) \sum_n x(n) + N(1/\lambda) = 0 \Rightarrow \hat{\lambda} = 1/\bar{x}$$

8) For discrete random variables the ML principle applies ^{to} the probability function

$$\begin{aligned} p_r\{x\} &= \prod_{n=0}^{N-1} p^{x(n)} (1-p)^{1-x(n)} \\ &= p^{\sum_n x(n)} (1-p)^{N - \sum_n x(n)} \end{aligned}$$

$$= p^{N\bar{x}} (1-p)^{N-N\bar{x}}$$

Maximizing $\ln P\{x\}$ over p

$$\frac{d \ln P\{x\}}{dp} = \frac{d}{dp} [N\bar{x} \ln p + (N-N\bar{x}) \ln (1-p)]$$

$$= \frac{N\bar{x}}{p} + \frac{N-N\bar{x}}{1-p} (-1) = 0$$

$$N\bar{x}(1-p) - (N-N\bar{x})p = 0 \Rightarrow \hat{p} = \bar{x}$$

$$9) \quad p(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

$$p(\underline{x}) = \prod_{n=0}^{N-1} p(x(n)) = \begin{cases} \frac{1}{\theta^N} & 0 < \text{all } x(n) < \theta \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $p(\underline{x})$ is maximized over θ when θ is as small as possible. But $\theta > x(n)$ for all n . Thus, $\theta_{\min} = \max x(n)$ and thus $\hat{\theta} = \max x(n)$.

$$10) \quad p(\underline{x}; A) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x(n) - AS(n))^2}$$

Minimize $\sum_n (x(n) - AS(n))^2$ to find MLE.

$$-2 \sum_n (x[n] - AS[n]) S[n] = 0$$

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

$$\begin{aligned} E(\hat{A}) &= \frac{\sum_n E(x[n]) S[n]}{\sum_n S^2[n]} = \frac{\sum_n AS[n] S[n]}{\sum_n S^2[n]} \\ &= A \end{aligned}$$

$$\begin{aligned} \text{var}(\hat{A}) &= \frac{\sum_n \text{var}(x[n]) S^2[n]}{(\sum_n S^2[n])^2} \\ &= \sigma^2 \frac{\sum_n S^2[n]}{(\sum_n S^2[n])^2} = \frac{\sigma^2}{\sum_{n=0}^{N-1} S^2[n]} \\ &= \mathbf{I}^{-1}(A) \quad (\text{see Theorem 4.1}) \end{aligned}$$

\hat{A} is Gaussian since it is linear in the data. Hence, $\hat{A} \sim N(A, \mathbf{I}^{-1}(A))$ and thus asymptotic PDF holds for finite data records. This problem is special case of linear model.

$$\begin{aligned}
 \text{ii) } p(\underline{x}; \rho) &= \prod_{n=0}^{N-1} \frac{1}{2\pi \sqrt{\det(\underline{C})}} e^{-\frac{1}{2} \underline{x}^T(n) \underline{C}^{-1} \underline{x}(n)} \\
 &= \frac{1}{(2\pi)^N [\det(\underline{C})]^{N/2}} e^{-\frac{1}{2} \sum_n \underline{x}^T(n) \underline{C}^{-1} \underline{x}(n)}
 \end{aligned}$$

$$\text{But } \underline{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \Rightarrow \underline{C}^{-1} = \frac{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}{1-\rho^2}$$

$$\det(\underline{C}) = 1-\rho^2$$

$$\begin{aligned}
 \ln p &= -N \ln 2\pi - \frac{N}{2} \ln(1-\rho^2) \\
 &\quad - \frac{1}{2(1-\rho^2)} \underbrace{\sum_n \underline{x}^T(n) \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \underline{x}(n)}_Q
 \end{aligned}$$

$$\begin{aligned}
 Q &= \sum_n (x_1(n) \ x_2(n)) \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} \\
 &= \sum_n [x_1^2(n) + x_2^2(n) - 2\rho x_1(n) x_2(n)]
 \end{aligned}$$

$$\text{Let } C_{ij} = \sum_n x_i(n) x_j(n)$$

$$\Rightarrow Q = C_{11} + C_{22} - 2\rho C_{12}$$

$$\begin{aligned}
 \ln p &= -N \ln 2\pi - \frac{N}{2} \ln(1-\rho^2) \\
 &\quad - \frac{1}{2(1-\rho^2)} (C_{11} + C_{22} - 2\rho C_{12})
 \end{aligned}$$

$$\frac{d \ln p}{d\rho} = \frac{-N/2(-2\rho)}{1-\rho^2} - \frac{1}{2(1-\rho^2)} (-2C_{12})$$

$$- \frac{1}{2} (C_{11} + C_{22} - 2\rho C_{12}) \left[\frac{2\rho}{(1-\rho^2)^2} \right] = 0$$

$$N\rho(1-\rho^2) + C_{12}(1-\rho^2) - \rho(C_{11} + C_{22} - 2\rho C_{12}) = 0$$

$$\rho^3 - \frac{1}{N} C_{12} \rho^2 + \left(\frac{1}{N} C_{11} + \frac{1}{N} C_{22} - 1 \right) \rho - \frac{1}{N} C_{12} = 0$$

As $N \rightarrow \infty$, $\frac{C_{11}}{N} \rightarrow 1$, $\frac{C_{22}}{N} \rightarrow 1$

$$\rho^3 - \frac{1}{N} C_{12} \rho^2 + \rho - \frac{1}{N} C_{12} = 0$$

for which a solution is $\hat{\rho} = \frac{1}{N} C_{12}$
 $= \frac{1}{N} \sum_{n=0}^{N-1} x_1(n) x_2(n)$

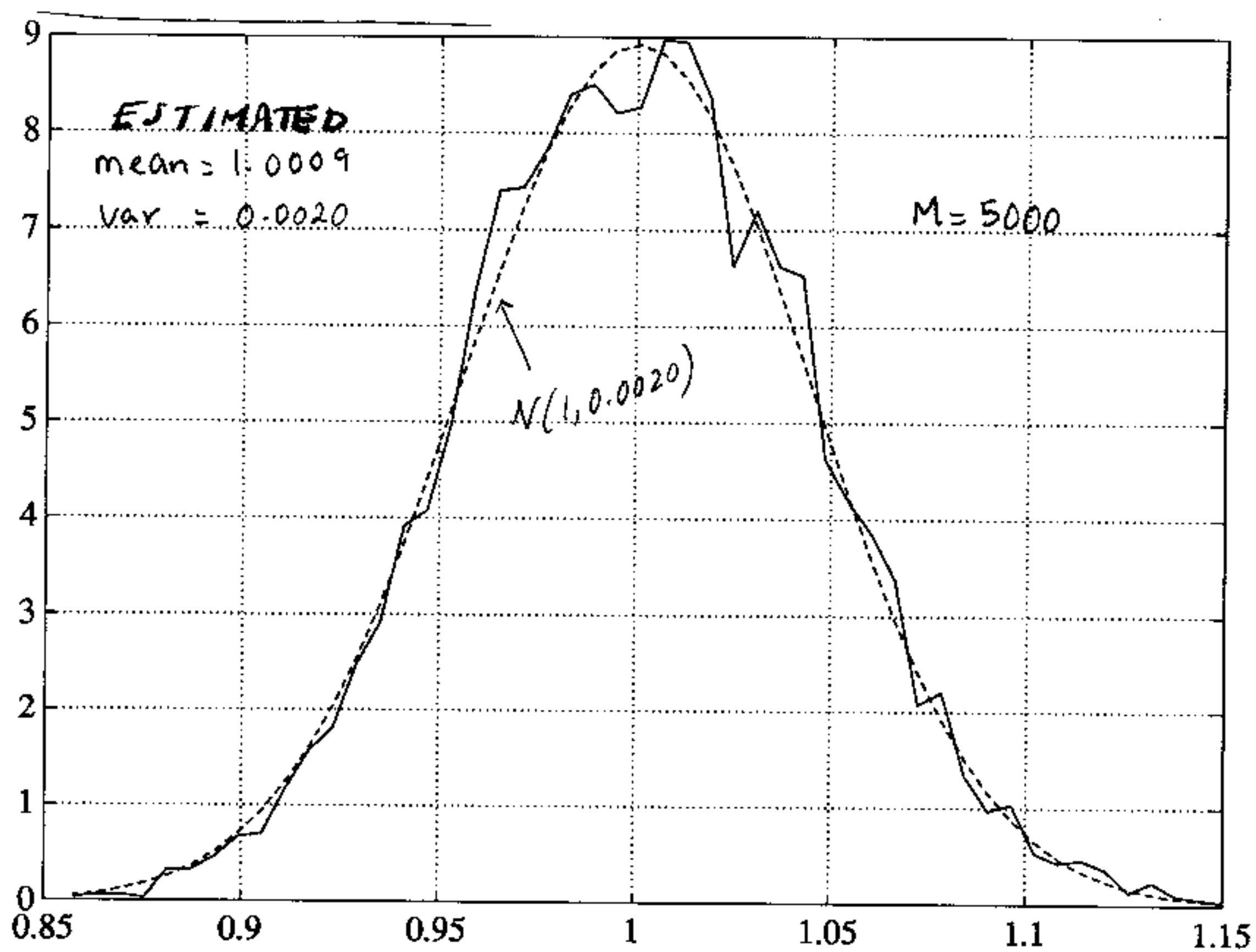
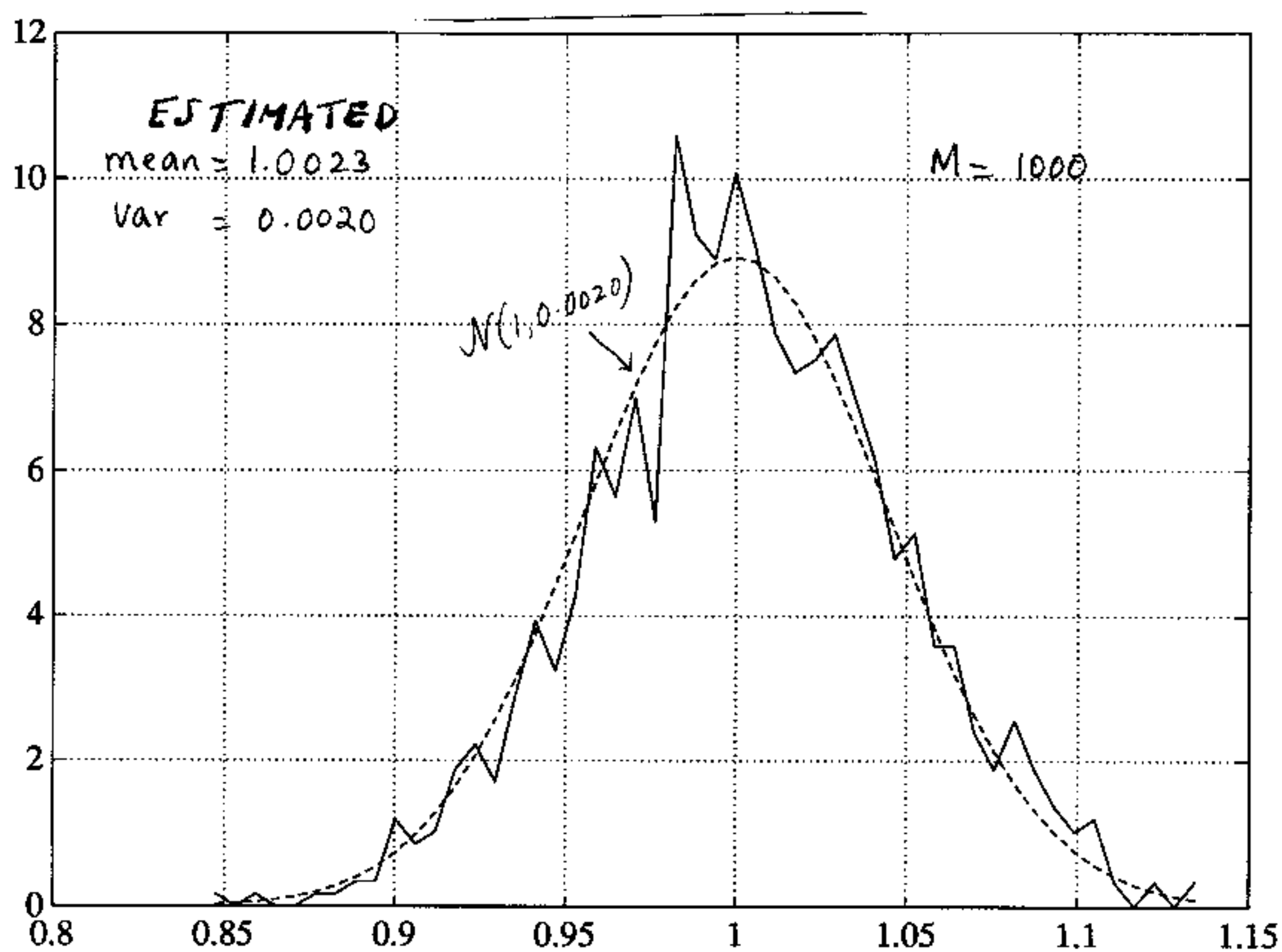
12) To find the MLE we maximize p or equivalently $\ln p$

$$\Rightarrow \frac{\partial \ln p}{\partial \theta} = 0$$

$$\text{But } \frac{\partial \ln p}{\partial \theta} = I(\theta) (\hat{\theta} - \theta)$$

Since $I(\theta) > 0$ for all θ (we always assume this - otherwise PDF does not depend on θ), the only solution is $\theta = \hat{\theta}$ or MLE is just $\hat{\theta}$, the efficient estimator.

13) See plots below.



We have a better fit as M increases.

14) See plots on next two pages.

15) E_s, E_c are zero mean since $W(n)$ is zero mean.

$$E(E_s E_c) = \frac{-4}{N^2 A^2} \sum_m \sum_n E(W(m) W(n)) \cdot \sin 2\pi f_0 m \cos 2\pi f_0 n$$

$$= \frac{-4}{N^2 A^2} \sigma^2 \sum_n \sin 2\pi f_0 n \cos 2\pi f_0 n$$

$$= \frac{-2\sigma^2}{N A^2} \sum_n \sin 4\pi f_0 n \approx 0$$

Thus, E_s, E_c are independent Gaussian random variables with zero means and variances given as follows:

$$\text{var}(E_s) = E(E_s^2) = \frac{4}{N^2 A^2} \sum_m \sum_n E(W(m) W(n))$$

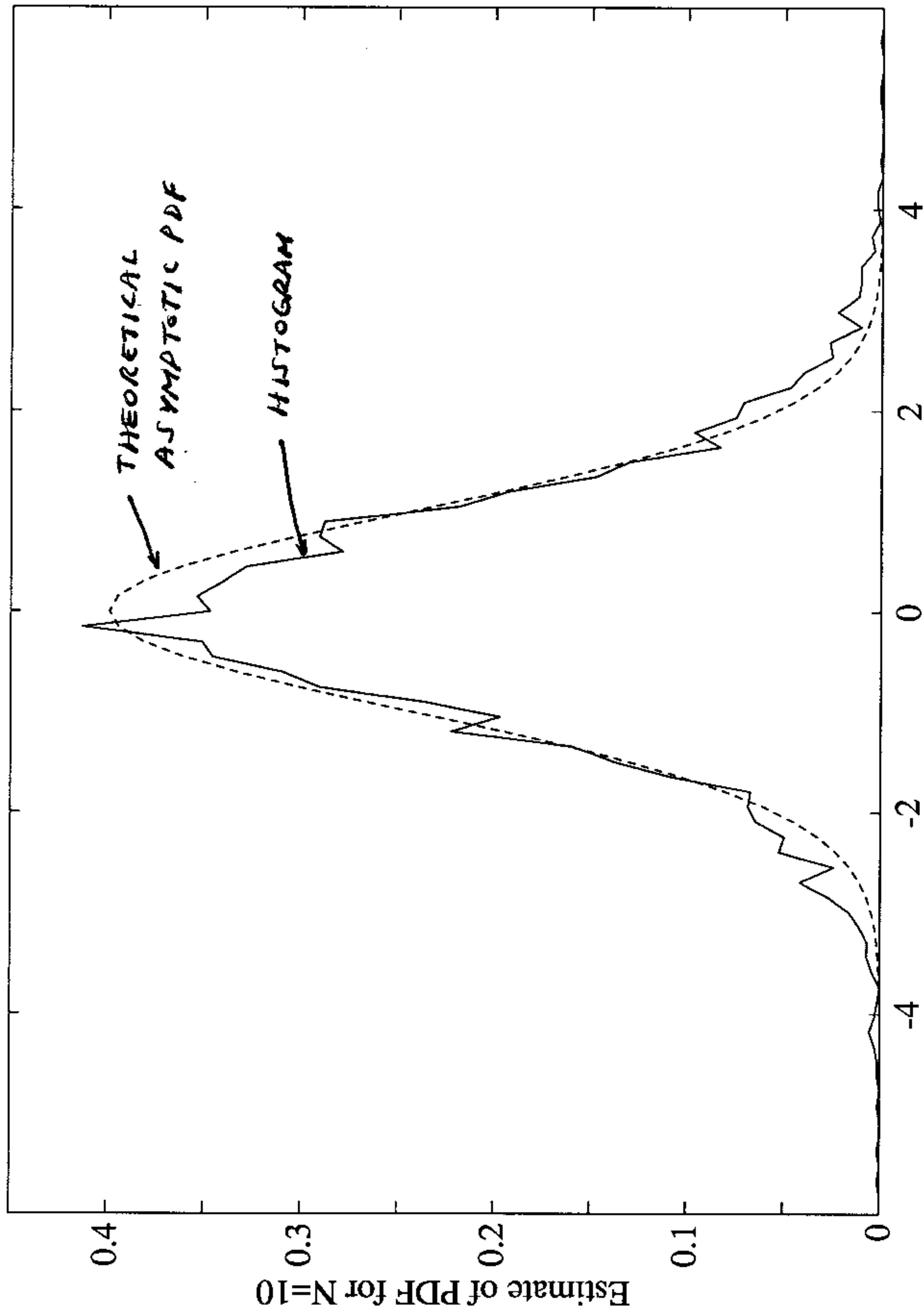
$$\cdot \sin 2\pi f_0 m \sin 2\pi f_0 n$$

$$= \frac{4\sigma^2}{N^2 A^2} \sum_n \sin^2 2\pi f_0 n$$

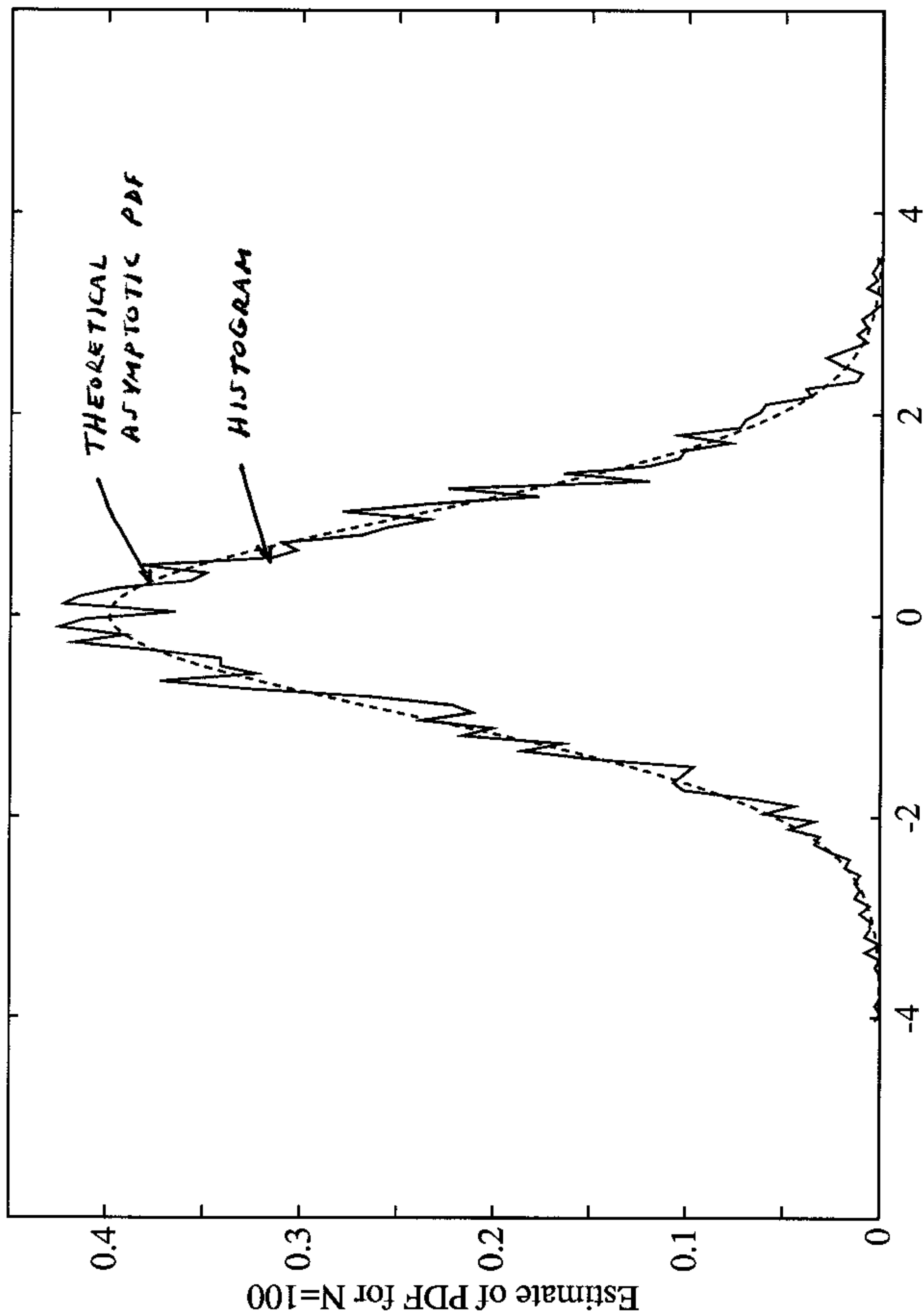
$$= \frac{4\sigma^2}{N^2 A^2} \sum_n \left(\frac{1}{2} - \frac{1}{2} \cos 4\pi f_0 n \right)$$

$$\approx \frac{4\sigma^2}{N^2 A^2} (N/2) = \frac{2\sigma^2}{N A^2}$$

Problem 7.14: Verification of Slutsky's theorem



Problem 7.14: Verification of Slutsky's theorem



and similarly, $\text{var}(\epsilon_c) \approx 2\sigma^2/NA^2$.

$$g(\epsilon_s, \epsilon_c) = \arctan \frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c}$$

$$\frac{\partial g}{\partial \epsilon_s} = \frac{1}{1 + \left(\frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c} \right)^2} \cdot \frac{1}{\cos \phi + \epsilon_c}$$

$$\left. \frac{\partial g}{\partial \epsilon_s} \right|_{\epsilon_s = \epsilon_c = 0} = \frac{\cos^2 \phi}{\cos^2 \phi + \sin^2 \phi} \cdot \frac{1}{\cos \phi} = \cos \phi$$

$$\frac{\partial g}{\partial \epsilon_c} = \frac{1}{1 + \left(\frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c} \right)^2} \cdot \frac{\sin \phi + \epsilon_s}{(\cos \phi + \epsilon_c)^2}$$

$$\left. \frac{\partial g}{\partial \epsilon_c} \right|_{\epsilon_s = \epsilon_c = 0} = \cos^2 \phi \frac{\sin \phi}{\cos^2 \phi} = \sin \phi$$

$$\hat{\phi} \approx \phi + \cos \phi \epsilon_s + \sin \phi \epsilon_c$$

$$\Rightarrow \hat{\phi} \sim N\left(\phi, \underbrace{\cos^2 \phi \text{var}(\epsilon_s) + \sin^2 \phi \text{var}(\epsilon_c)}_{\frac{2\sigma^2}{NA^2}}\right)$$

The variance is just the CRLB.

$$\begin{aligned} 16) \quad \text{var}(\hat{A}) &= \text{var}(x[0]) \\ &= \text{var}(W(0)) \\ &= \int_{-\infty}^{\infty} u^2 \frac{1}{2} e^{-|u|} du \end{aligned}$$

$$= \int_0^{\infty} u^2 e^{-u} du$$

$$= - (u^2 + 2u + 2) e^{-u} \Big|_0^{\infty} = 2$$

$$\text{var}(\hat{A}) \geq \int_{-\infty}^{\infty} \left(\frac{d p(u)}{du} \right)^2 / p(u) du$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{2} e^{-|u|} \right)^2 / \frac{1}{2} e^{-|u|} du$$

$$\text{since } dp/du = \begin{cases} -\frac{1}{2} e^{-u} & u > 0 \\ \frac{1}{2} e^u & u < 0 \end{cases}$$

$$\text{var}(\hat{A}) \geq \int_{-\infty}^{\infty} \frac{1}{2} e^{-|u|} du = 1$$

No, the MLE has variance 2 for all N .

17) Let $\alpha = 1/\theta$ so that by invariance of the MLE, $\hat{\alpha} = 1/\hat{\theta}$. But $\hat{\alpha} = 1/N \sum_{n=0}^{N-1} x^2(n)$

Since

$$p(\underline{x}; \alpha) = \frac{1}{(2\pi\alpha)^{N/2}} e^{-\frac{1}{2\alpha} \sum_{n=0}^{N-1} x^2(n)}$$

$$\frac{\partial \log}{\partial \alpha} = -\frac{N}{2} \frac{1}{\alpha} + \frac{1}{2\alpha^2} \sum x^2(n) = 0$$

$$\Rightarrow \hat{\alpha} = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n)$$

$$\text{Thus, } \hat{\theta} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} x^2(n)}$$

From Chapter 3

$$I(\alpha) = N/2\alpha^2$$

$$\text{and } I^{-1}(\alpha) = I^{-1}(\theta) \left(\frac{\partial \alpha}{\partial \theta} \right)^2$$

$$\begin{aligned} I^{-1}(\theta) &= \frac{2\alpha^2}{N} \left(\frac{\partial \theta}{\partial \alpha} \right)^2 \\ &= \frac{2\alpha^2}{N} \left(-1/\alpha^2 \right)^2 = \frac{2}{N\alpha^2} \\ &= 2\theta^2/N \end{aligned}$$

$$\Rightarrow \hat{\theta} \sim N(\theta, 2\theta^2/N)$$



$$\begin{aligned} g'(x) &= -x e^{-\frac{1}{2}x^2} - 0.1(x-10) e^{-\frac{1}{2}(x-10)^2} \\ g''(x) &= x^2 e^{-\frac{1}{2}x^2} + 0.1(x-10)^2 e^{-\frac{1}{2}(x-10)^2} \\ &\quad - 0.1 e^{-\frac{1}{2}(x-10)^2} - e^{-\frac{1}{2}x^2} \end{aligned}$$

$$x_{k+1} = x_k - \frac{dg/dx}{d^2g/dx^2} \Big|_{x=x_k}$$

Using a computer it is found that for $x_0 = 0.5$, the iteration converges to $x = 0$ and for $x_0 = 0.95$ it converges to $x = 10$. For other initial values such as $x_0 = 1$ it does not converge. To attain the maximum (global), x_0 must be close to the true value.

$$19) \quad p(\underline{x}; f_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_n (x[n] - \cos 2\pi f_0 n)^2}$$

To find MLE minimize

$$\begin{aligned} \sum_n (x[n] - \cos 2\pi f_0 n)^2 &= \\ \sum_n x^2[n] - 2 \sum_n x[n] \cos 2\pi f_0 n \\ &\quad + \sum_n \cos^2 2\pi f_0 n \end{aligned}$$

$$\approx \sum_n x^2[n] - 2 \sum_n x[n] \cos 2\pi f_0 n + N/2$$

$$\Rightarrow \text{maximize } \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n$$

over $0 < f_0 < 1/2$

Using a Newton-Raphson iteration

$$g(f_0) = \sum_n x[n] \cos 2\pi f_0 n$$

$$dg/df_0 = - \sum_n 2\pi n x[n] \sin 2\pi f_0 n$$

$$d^2g/df_0^2 = - \sum_n (2\pi n)^2 x[n] \cos 2\pi f_0 n$$

$$f_{0,k+1} = f_{0,k} - \frac{\sum_n (2\pi n) x[n] \sin 2\pi f_{0,k} n}{\sum_n (2\pi n)^2 x[n] \cos 2\pi f_{0,k} n} \bigg|_{f_0 = f_{0,k}}$$

The function to be maximized is shown on following page for 10 different realizations of $w(n)$. Next, for 500 realizations we plot the results for a grid search and a Newton-Raphson iteration with 10 iterations. In (a) the grid search produces.

$$E(\hat{f}_0) = 0.25 = f_0$$

$$\text{var}(\hat{f}_0) = 1.3 \times 10^{-6}$$

In (b), (c), (d) the Newton-Raphson produces

Initial freq. (f_0)	$E(\hat{f}_0)$	$\text{var}(\hat{f}_0)$
0.24	0.25	1.2×10^{-6}
0.26	0.25	1.2×10^{-6}
0.28	0.16	10

$$20) \quad p(\underline{x}; \underline{\xi}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \xi)^2}$$

Maximized when $\sum (x[n] - \xi)^2$ is minimized.

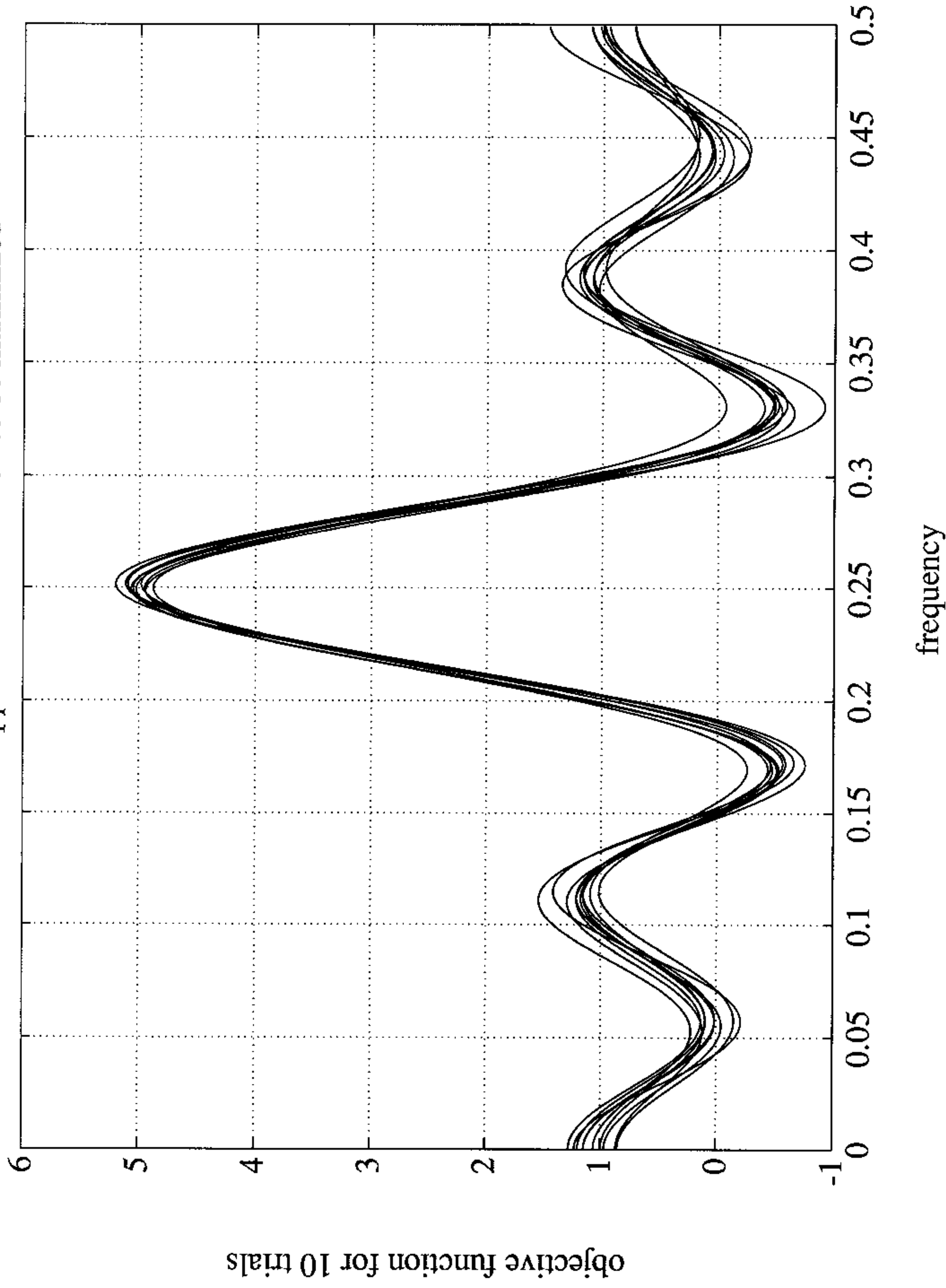
Clearly, then $\hat{\xi}[n] = x[n]$.

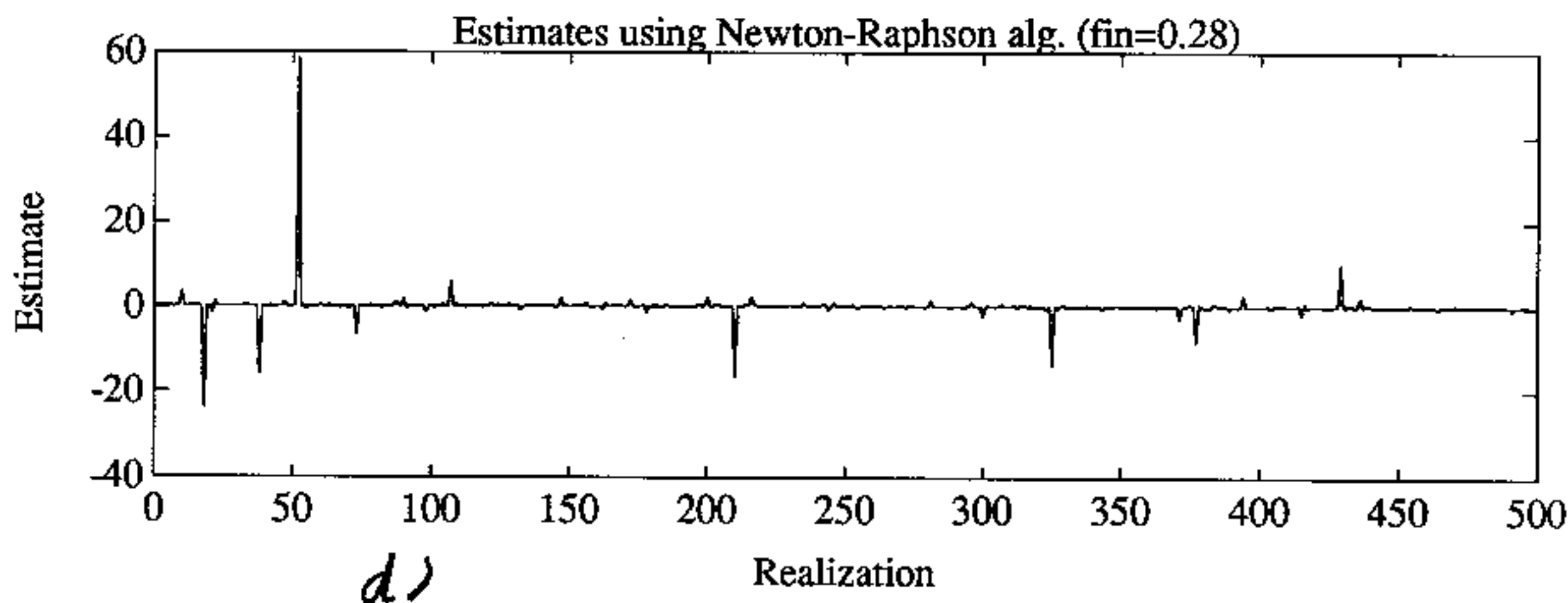
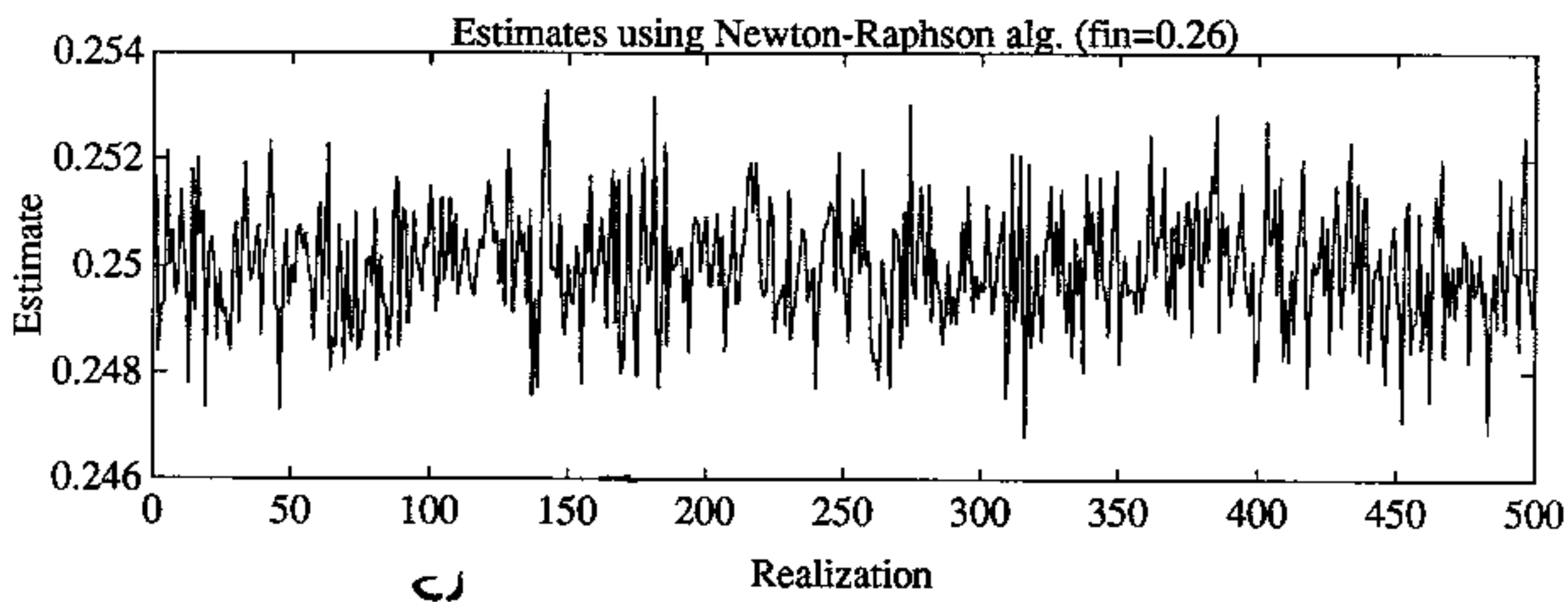
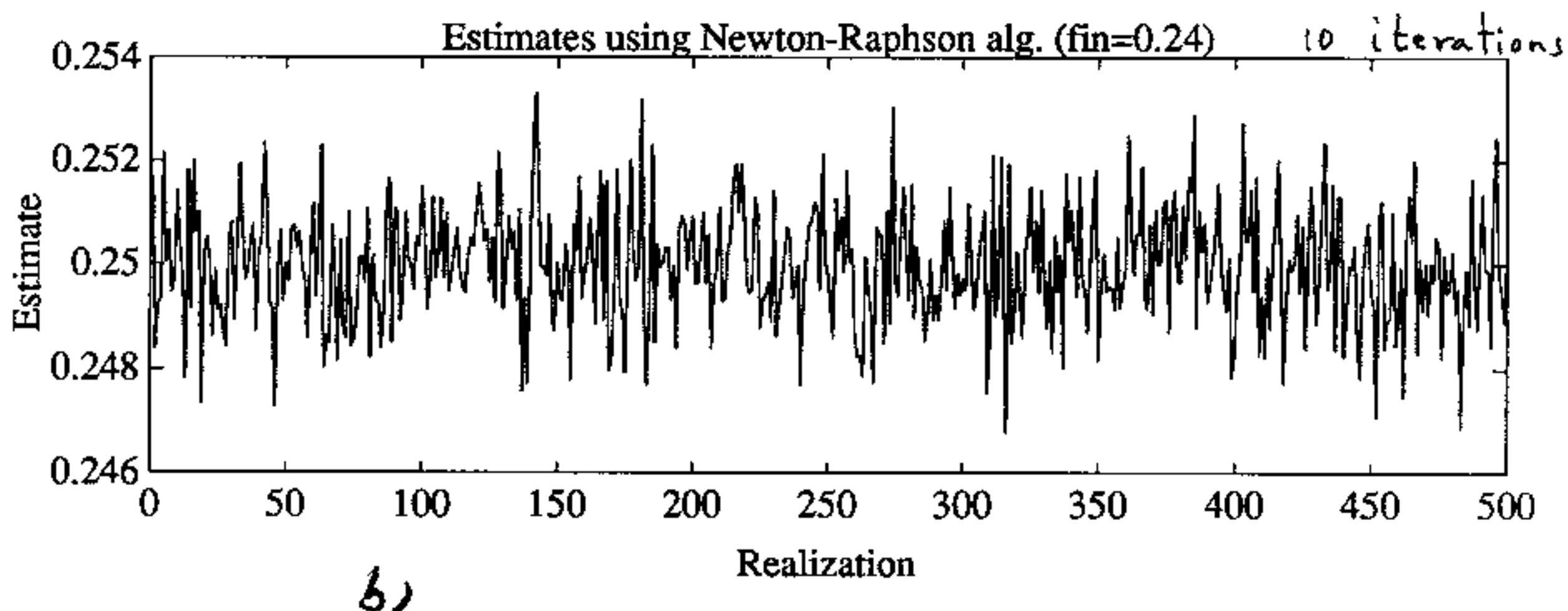
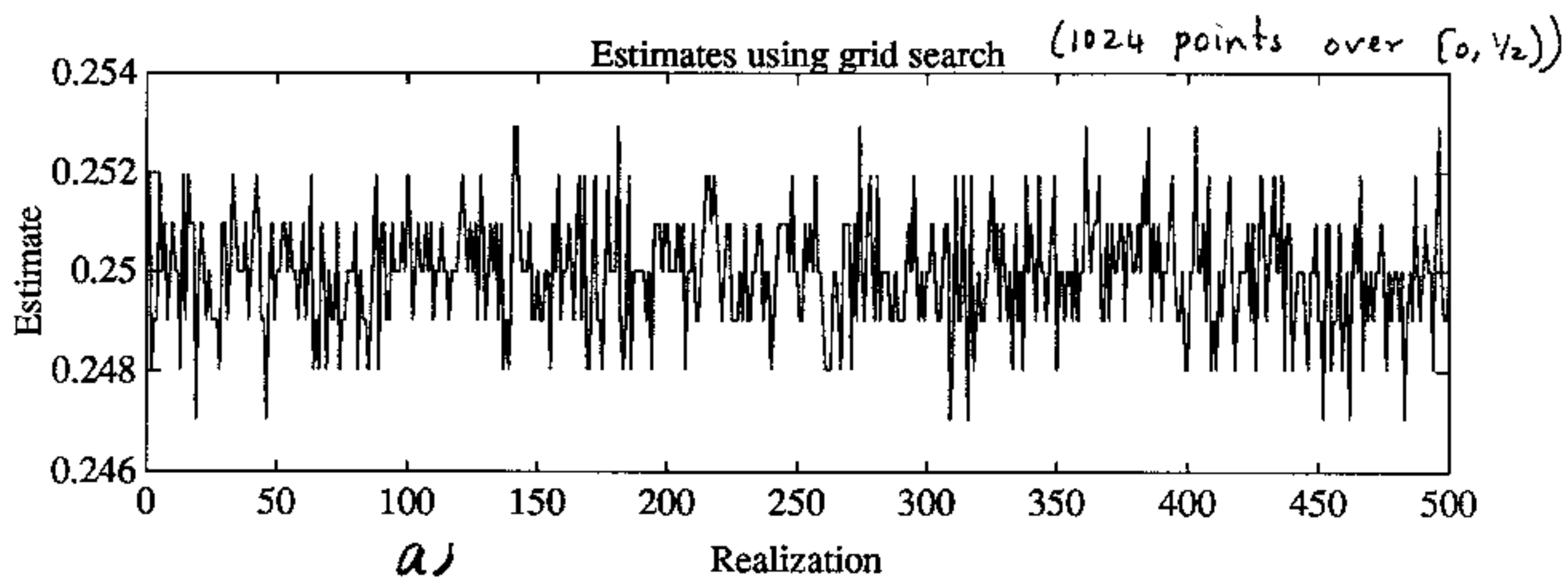
But $\hat{\xi} = \underline{x} \sim N(\underline{\xi}, \sigma^2 \underline{I})$ and thus $\hat{\xi}$ is unbiased and Gaussian. To determine if it is efficient we find the CRLB.

$$\frac{\partial \ln p}{\partial \xi[k]} = \frac{1}{\sigma^2} (x[k] - \xi[k])$$

$$\frac{\partial^2 \ln p}{\partial \xi[k] \partial \xi[l]} = -\frac{1}{\sigma^2} \delta_{kl}$$

Problem 7.19: Approx. MLE function to be maximized





$$\Rightarrow \underline{I}(\underline{\xi}) = \frac{1}{\sigma^2} \underline{I}$$

$$\text{or } \hat{\underline{\xi}} \sim N(\underline{\xi}, \underline{I}^{-1}(\underline{\xi}))$$

Hence, $\hat{\underline{\xi}}$ is efficient. However, it is not consistent since as $N \rightarrow \infty$, $\text{var}(\hat{\xi}[n]) = \sigma^2 \nrightarrow 0$.

21) From Example 7.12 $\hat{A} = \bar{X}$
 $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (X[n] - \bar{X})^2$

Hence, by the invariance

property $\hat{\xi} = \frac{\hat{A}^2}{\hat{\sigma}^2} = \frac{\bar{X}^2}{\frac{1}{N} \sum_{n=0}^{N-1} (X[n] - \bar{X})^2}$

$$22) I = \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |A(f)|^2 df = \int_{-\frac{1}{2}}^{\frac{1}{2}} [\ln A(f) + \ln A^*(f)] df$$

$$= 2 \operatorname{Re} \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln A(f) df$$

Now let

$$z = e^{j2\pi f}, \quad dz = j2\pi e^{j2\pi f} df \\ = j2\pi z df$$

$$\Rightarrow I = 2 \operatorname{Re} \oint \ln A(z) \frac{1}{j2\pi} \frac{dz}{z}$$

$$= 2 \operatorname{Re} \left\{ \frac{1}{2\pi j} \oint \ln A(z) \frac{dz}{z} \right\}$$

$$= 2 \operatorname{Re} \left\{ \mathcal{Z}^{-1} \{ \ln A(z) \} \Big|_{n=0} \right\}$$

Since $A(z)$ converges for all $z \neq 0$, it

converges for $|z| \geq 1$ and since all of its zeros are within the unit circle, $\ln A(z)$ converges on and outside the unit circle. Thus, $\ln A(z)$ has a causal inverse.

The sample at $n=0$ is found from the initial value theorem or

$$\begin{aligned} \mathcal{Z}^{-1} \{ \ln A(z) \}_{n=0} &= \lim_{z \rightarrow \infty} \ln A(z) \\ &= \lim_{z \rightarrow \infty} \ln \left[1 + \sum_{k=1}^{\infty} a[k] z^{-k} \right] = \ln 1 \\ &= 0 \end{aligned}$$

2.3) We use (7.60) which when differentiated produces

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{xx}^2(f)} \right] \frac{\partial P_{xx}(f)}{\partial P_0} df = 0$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{P_0 Q(f)} - \frac{I(f)}{P_0^2 Q^2(f)} \right] Q(f) df = 0$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(P_0 - \frac{I(f)}{Q(f)} \right) df = 0$$

$$\Rightarrow \hat{P}_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{Q(f)} df$$

If $Q(f) = 1$ for all f ,

$$\hat{P}_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f) df$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \quad \text{from Prob 7.25}$$

24) See plots on next page. For $f_0 = 0.25$ the peak of the periodogram and that of the exact function $\underline{x}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}$ are at 0.25. But for $f_0 = 0.05$ the peak of the periodogram is shifted away ($= 0.068$) from the true value. This is due to the interaction of the complex sinusoids at $f_0 = 0.05$ and $-f_0 = -0.05$, which are not adequately resolved by the periodogram.

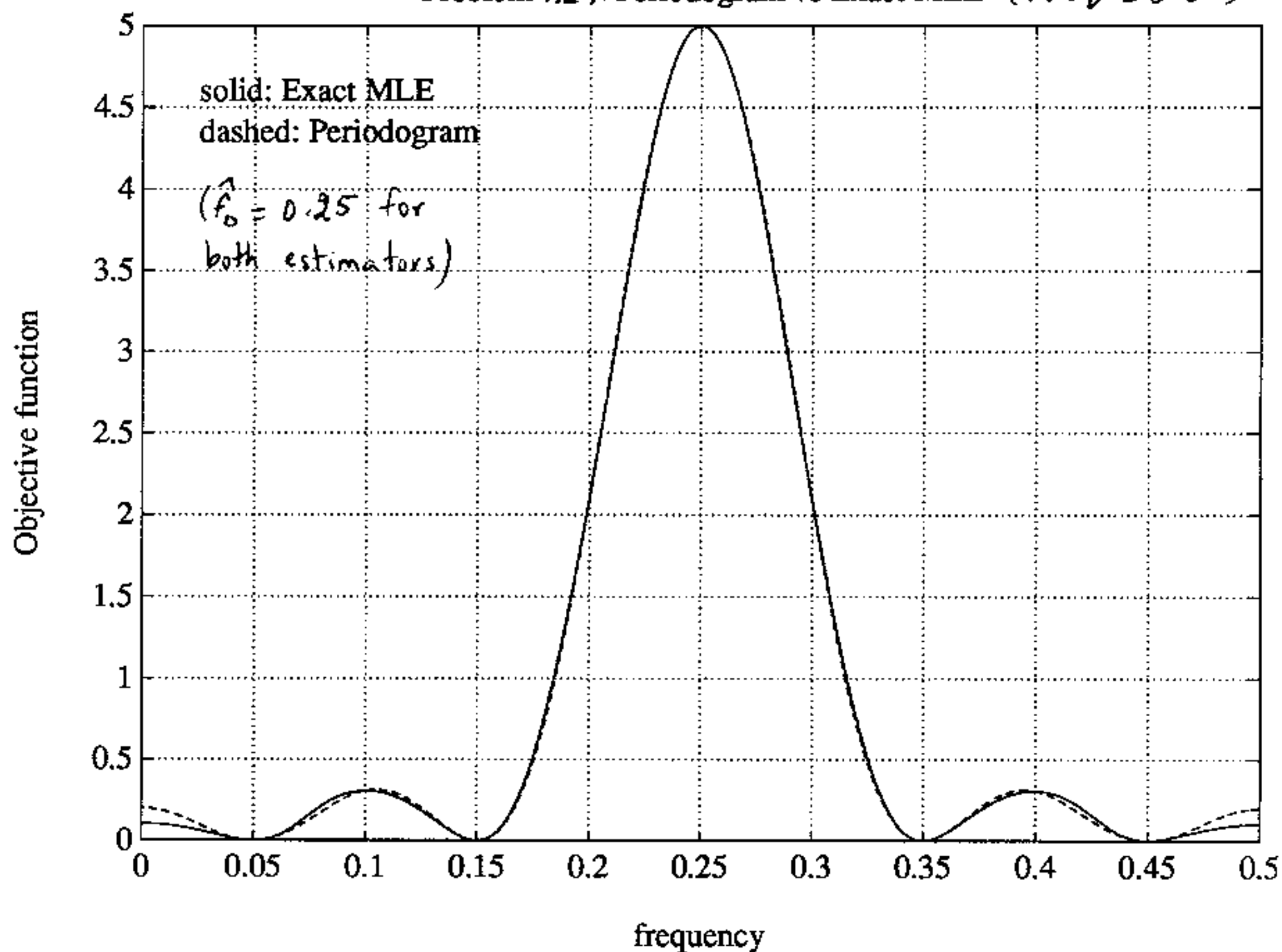
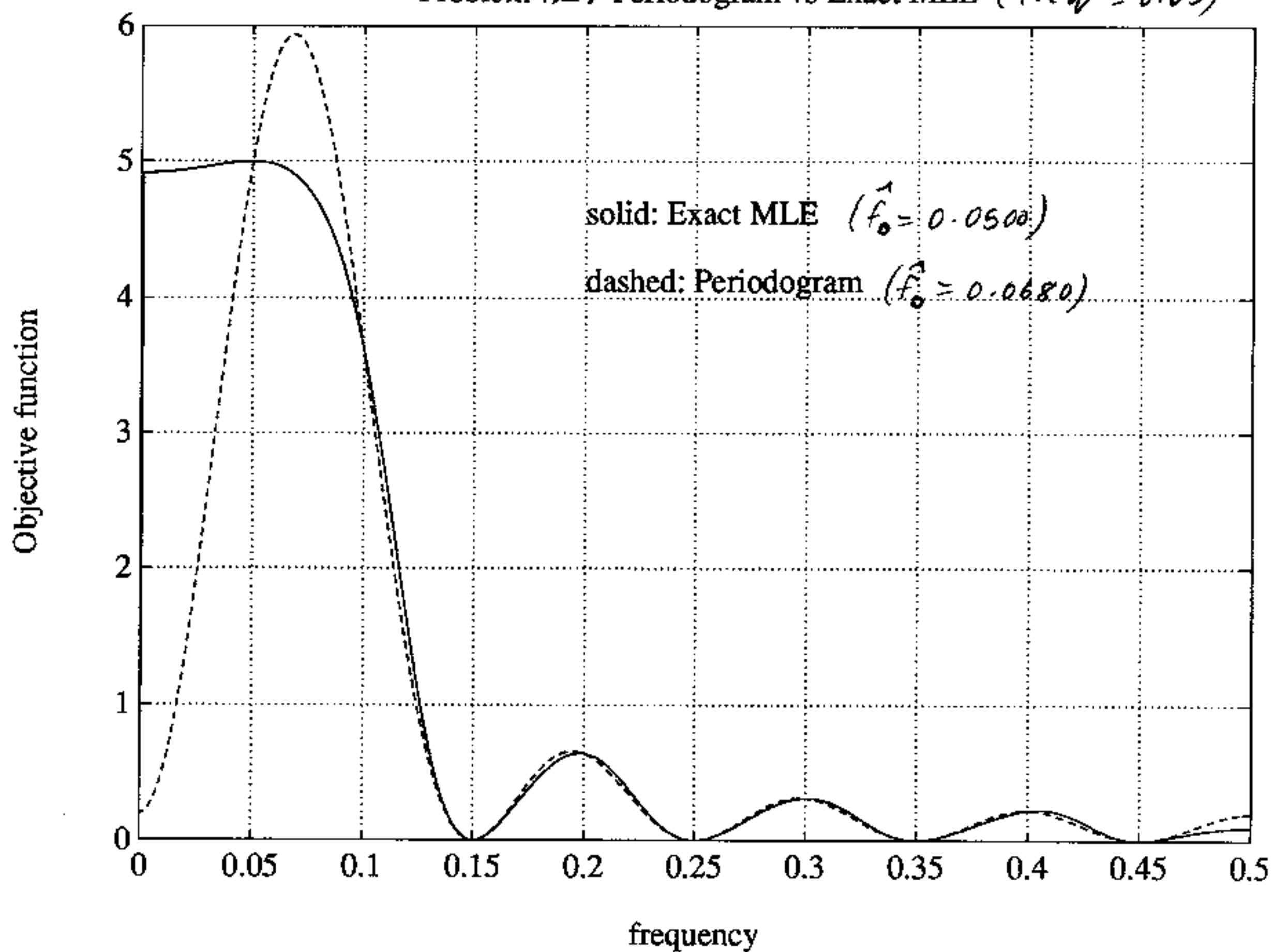
25) $\mathcal{F}^{-1} \{ I(f) \} = \frac{1}{N} \mathcal{F}^{-1} \{ x'(f) x'(f)^* \}$
 But if $x[n] \xrightarrow{\mathcal{F}} x'(f)$
 $x'[-n] \xrightarrow{\mathcal{F}} x'(f)^*$ for $x[n]$ real

$$\begin{aligned} \Rightarrow \mathcal{F}^{-1} \{ I(f) \} &= \frac{1}{N} x'[n] \star x'[-n] \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[-n] x'[k-n] \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[n] x'[n+k] \end{aligned}$$

For $k \geq 0$

$$\mathcal{F}^{-1} \{ I(f) \} = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n] x[n+k]$$

since $x'[n] = 0$ for $n < 0$ or $n > N-1$
 $= x[n]$ otherwise

Problem 7.24 Periodogram vs Exact MLE ($f_{eq} = 0.25$)Problem 7.24 Periodogram vs Exact MLE ($f_{eq} = 0.05$)

For $k \geq 0$

$$\mathcal{I}^{-1} \{ \mathcal{I}(f) \} = \frac{1}{N} \sum_{n=-k}^{N-1} x[n] x[n+k]$$

Let $l = n+k$

$$= \frac{1}{N} \sum_{l=0}^{N-1+k} x[l-k] x[l]$$

$$= \frac{1}{N} \sum_{n=0}^{N-1+k} x[n] x[n-k]$$

Combining the results we have our solution.

$$\begin{aligned} 2b) \quad \hat{a}(1) &= -\hat{r}_{xx}[1] / \hat{r}_{xx}[0] \\ \hat{\sigma}^2 &= \hat{r}_{xx}[0] + \hat{a}(1) \hat{r}_{xx}[1] \end{aligned}$$

By invariance the MLE of $P_{xx}(f_0)$ is

$$\hat{p}_{xx}(f_0) = \frac{\hat{\sigma}^2}{|1 + \hat{a}(1) e^{-j2\pi f_0}|^2}$$

But

$$\hat{\mathcal{I}}_2 = \frac{\partial \mathcal{I}}{\partial \underline{\theta}} \mathcal{I}^{-1}(\underline{\theta}) \frac{\partial \mathcal{I}}{\partial \underline{\theta}}^T \geq 0$$

$$\underline{\mathcal{I}}(\underline{\theta}) = \begin{bmatrix} \frac{N \hat{r}_{xx}[0]}{\sigma_u^2} & 0 \\ 0 & \frac{N}{2\sigma_u^4} \end{bmatrix}$$

$$\text{And } g(a[1], \sigma_u^2) = \frac{\sigma_u^2}{|1 + a[1] e^{-j2\pi f_0}|^2}$$

so that

$$\partial g / \partial a[1] = \frac{-\sigma_u^2}{|A(f)|^4} (A(f) e^{j2\pi f_0} + A^*(f) e^{-j2\pi f_0})$$

$$= \frac{-\sigma_u^2}{|A(f)|^4} 2 \operatorname{Re}(A(f) e^{j2\pi f_0})$$

$$\partial g / \partial \sigma_u^2 = \frac{1}{|A(f)|^2}$$

$$\operatorname{var}(\hat{P}_{xx}(f_0)) \approx \frac{\partial g}{\partial \theta} \mathbf{I}^{-1}(\theta) \frac{\partial g}{\partial \theta}^T$$

$$= \frac{\sigma_u^2}{N r_{xx}[0]} \frac{\sigma_u^4}{|A(f)|^8} 4 \operatorname{Re}^2(A(f) e^{j2\pi f_0})$$

$$+ \frac{2\sigma_u^4}{N} \frac{1}{|A(f)|^4}$$

$$= \frac{4 P_{xx}^4(f_0)}{N r_{xx}[0] \sigma_u^2} \operatorname{Re}^2(A(f) e^{j2\pi f_0})$$

$$+ \frac{2}{N} P_{xx}^2(f_0)$$

Chapter 8

1) Nonlinear LS due to f_0 and r .

Yes, quadratic in A, B .

Analytically, we could find the values of A and B that minimize J for given f_0 and r . Then, plug these into J , which will now be a nonquadratic function of f_0 and r . Next, use a grid search over $0 < r < 1$ and $0 \leq f_0 \leq \frac{1}{2}$.

$$2) \quad J_{\min} = \sum_{n=0}^{N-1} x^2[n] - N \bar{x}^2 \leq \sum_{n=0}^{N-1} x^2[n]$$

$$\text{Also, } J_{\min} = \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \geq 0$$

$$3) \quad J = \sum_{n=0}^{M-1} (x[n] - A)^2 + \sum_{n=M}^{N-1} (x[n] + A)^2$$

$$\frac{\partial J}{\partial A} = -2 \sum_{n=0}^{M-1} (x[n] - A) + 2 \sum_{n=M}^{N-1} (x[n] + A) = 0$$

$$\Rightarrow \sum_{n=0}^{M-1} x[n] + 2MA + 2 \sum_{n=M}^{N-1} x[n] + 2(N-M)A = 0$$

$$\hat{A} = \frac{1}{N} \left(\sum_{n=0}^{M-1} x[n] - \sum_{n=M}^{N-1} x[n] \right)$$

$$J_{\min} = \sum_{n=0}^{M-1} (x[n] - \hat{A})(x[n] - \hat{A}) + \sum_{n=M}^{N-1} (x[n] + \hat{A})(x[n] + \hat{A})$$

$$= \sum_{n=0}^{M-1} x[n](x[n] - \hat{A}) + \sum_{n=M}^{N-1} x[n](x[n] + \hat{A})$$

using the $\partial J / \partial A = 0$ equation.

$$\begin{aligned} J_{MIN} &= \sum_0^{N-1} x^2[n] - \hat{A} \left(\sum_0^{M-1} x[n] - \sum_M^{N-1} x[n] \right) \\ &= \sum_0^{N-1} x^2[n] - N \hat{A}^2 \end{aligned}$$

For $w[n]$ wgn we have

$$\begin{aligned} E(\hat{A}) &= 1/N [MA - (N-M)A] = A \\ \text{var}(\hat{A}) &= \frac{1}{N^2} \left(\text{var} \left(\sum_0^{M-1} x[n] \right) \right. \\ &\quad \left. + \text{var} \left(\sum_M^{N-1} x[n] \right) \right) \\ &= \frac{1}{N^2} \left(\sum_0^{M-1} \text{var}(x[n]) + \sum_M^{N-1} \text{var}(x[n]) \right) \\ &= \frac{1}{N^2} (M\sigma^2 + (N-M)\sigma^2) = \sigma^2/N \end{aligned}$$

$\Rightarrow \hat{A} \sim N(A, \sigma^2/N)$ since \hat{A} is a linear function of the $x[n]$'s.

4) See solution for Prob 5.19

$$5) \quad \underline{z} = \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ \cos 2\pi f_1 & \cos 2\pi f_2 & \dots & \cos 2\pi f_p \\ \vdots & & & \\ \cos 2\pi f_1(N-1) & \cos 2\pi f_2(N-1) & \dots & \cos 2\pi f_p(N-1) \end{bmatrix}}_{\underline{H}} \underbrace{\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix}}_{\underline{\theta}}$$

$$\underline{H}^T \underline{H} \hat{\underline{\theta}} = \underline{H}^T \underline{x} \quad \text{are normal equations}$$

If $f_i = i/N$, the column vectors of \underline{H} are orthogonal (see (4.13)). Thus,

$$\underline{H}^T \underline{H} = \frac{N}{2} \underline{I}$$

$$\Rightarrow \hat{\underline{\theta}} = \frac{2}{N} \underline{H}^T \underline{x} \Rightarrow \hat{A}_i = \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_i n$$

$$J_{M,N} = \underline{x}^T (\underline{I} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) \underline{x}$$

$$= \underline{x}^T (\underline{I} - \frac{2}{N} \underline{H} \underline{H}^T) \underline{x}$$

$$= \underline{x}^T \underline{x} - \frac{2}{N} \|\underline{H}^T \underline{x}\|^2$$

$$= \underline{x}^T \underline{x} - \frac{2}{N} \left(\frac{N}{2}\right)^2 \|\hat{\underline{\theta}}\|^2$$

$$= \sum_{n=0}^{N-1} x^2[n] - \frac{N}{2} \sum_{i=1}^p \hat{A}_i^2$$

For $w[n]$, $w \in \mathcal{N}$ the PDF is

$$\hat{\underline{\theta}} \sim \mathcal{N}(\underline{\theta}, \sigma^2 (\underline{H}^T \underline{H})^{-1}) \quad \text{since}$$

$$E(\hat{\underline{\theta}}) = (\underline{H}^T \underline{H})^{-1} \underline{H}^T E(\underline{x}) = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} \underline{\theta} = \underline{\theta}$$

$$C_{\hat{\underline{\theta}}} = E[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T]$$

$$= E[(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underbrace{(\underline{x} - \underline{H} \underline{\theta})}_{\underline{w}} \underbrace{(\underline{x} - \underline{H} \underline{\theta})^T}_{\underline{w}^T} \underline{H} (\underline{H}^T \underline{H})^{-1}]$$

$$= (\underline{H}^T \underline{H})^{-1} \underline{H}^T \sigma^2 \underline{I} \underline{H} (\underline{H}^T \underline{H})^{-1} = \sigma^2 (\underline{H}^T \underline{H})^{-1}$$

$$\text{or } C_{\hat{\underline{\theta}}} = 2\sigma^2/N \underline{I}$$

and since $\hat{\underline{\theta}}$ is a linear function of \underline{x} ,
we have a Gaussian PDF on

$$\hat{\underline{\theta}} \sim N(\underline{\theta}, 2\sigma^2/N \underline{I})$$

$$6) \quad \hat{\underline{\theta}} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x}$$

From Prob 8.5 we have $\hat{\underline{\theta}} \sim N(\underline{\theta}, \sigma^2 (\underline{H}^T \underline{H})^{-1})$

Yes, it is unbiased.

$$\begin{aligned} 7) \quad E(\hat{\sigma}^2) &= \frac{1}{N} E \left[\underline{x}^T (\underline{I} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) \underline{x} \right] \\ &= \frac{1}{N} \text{tr} \left[(\underline{I} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) E(\underline{x} \underline{x}^T) \right] \end{aligned}$$

$$\text{Since } E(\underline{x} \underline{x}^T) = \text{tr} (E(\underline{y} \underline{x}^T))$$

$$E(\hat{\sigma}^2) = \frac{1}{N} \text{tr} \left[(\underline{I} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) E((\underline{H}\underline{\theta} + \underline{w})(\underline{H}\underline{\theta} + \underline{w})^T) \right]$$

$$= \frac{1}{N} \text{tr} \left[(\underline{I} - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) (\underline{H} \underline{\theta} \underline{\theta}^T \underline{H}^T + \sigma^2 \underline{I}) \right]$$

$$\begin{aligned} &= \frac{1}{N} \text{tr} \left[\underline{H} \underline{\theta} \underline{\theta}^T \underline{H}^T + \sigma^2 \underline{I} \right. \\ &\quad \left. - \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} \underline{\theta} \underline{\theta}^T \underline{H}^T \right. \\ &\quad \left. - \sigma^2 \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\sigma^2}{N} \text{tr} \left\{ \underline{I} - \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \right\} \\
&= \sigma^2 - \frac{\sigma^2}{N} \text{tr} \left\{ \underline{H}(\underline{H}^T \underline{H})^{-1} \underline{H}^T \right\} \\
&= \sigma^2 - \sigma^2/N \underbrace{\text{tr} \left\{ (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} \right\}}_{\underline{I} \text{ (p \times p)}} \\
&= \sigma^2 - \sigma^2 p/N = \sigma^2 \frac{N-p}{N}
\end{aligned}$$

$\hat{\sigma}^2$ is biased. To make it unbiased use

$$\hat{\sigma}^2 = \frac{1}{N-p} J_{MIN}.$$

It is said that we lose p degrees of freedom in estimating $\underline{\theta}$.

$$\begin{aligned}
\text{Ex) } J(A) &= \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x(n) - A)^2 \\
\frac{\partial J}{\partial A} &= -2 \sum_n \frac{1}{\sigma_n^2} (x(n) - A) = 0 \\
\Rightarrow \hat{A} &= \frac{\sum_{n=0}^{N-1} x(n) / \sigma_n^2}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}
\end{aligned}$$

$$E(\hat{A}) = \frac{\sum_n E(x(n)) / \sigma_n^2}{\sum_n 1 / \sigma_n^2} = A$$

$$\begin{aligned}
 \text{var}(\hat{A}) &= \frac{1}{\left(\sum_n 1/\sigma_n^2\right)^2} \underbrace{\sum_n \text{var}(x(n)/\sigma_n^2)}_{= \sum_n 1/\sigma_n^4 \text{var}(x(n))} \\
 &= \sum_n 1/\sigma_n^4 \text{var}(x(n)) \\
 &= \sum_n 1/\sigma_n^2 \\
 &= \frac{1}{\sum_{n=0}^{N-1} 1/\sigma_n^2}
 \end{aligned}$$

$$\begin{aligned}
 9) \quad J &= (\underline{x} - \underline{H}\underline{\theta})^T \underline{W} (\underline{x} - \underline{H}\underline{\theta}) \\
 &= (\underline{x} - \underline{H}\underline{\theta})^T \underline{D}^T \underline{D} (\underline{x} - \underline{H}\underline{\theta}) \\
 &= \underbrace{(\underline{D}\underline{x} - \underline{D}\underline{H}\underline{\theta})^T}_{\underline{x}' \quad \underline{H}'} (\underline{D}\underline{x} - \underline{D}\underline{H}\underline{\theta})
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow \hat{\underline{\theta}} &= (\underline{H}'^T \underline{H}')^{-1} \underline{H}'^T \underline{x}' \\
 &= (\underline{H}^T \underline{D}^T \underline{D} \underline{H})^{-1} \underline{H}^T \underline{D}^T \underline{D} \underline{x} \\
 &= (\underline{H}^T \underline{W} \underline{H})^{-1} \underline{H}^T \underline{W} \underline{x}
 \end{aligned}$$

$$\begin{aligned}
 J_{\min} &= \underline{x}'^T (\underline{I} - \underline{H}' (\underline{H}'^T \underline{H}')^{-1} \underline{H}'^T) \underline{x}' \\
 &= \underline{x}^T \underline{D}^T (\underline{I} - \underline{D} \underline{H} (\underline{H}^T \underline{D}^T \underline{D} \underline{H})^{-1} \underline{H}^T \underline{D}^T) \underline{D} \underline{x} \\
 &= \underline{x}^T (\underline{W} - \underline{W} \underline{H} (\underline{H}^T \underline{W} \underline{H})^{-1} \underline{H}^T \underline{W}) \underline{x}
 \end{aligned}$$

$$10) \quad \hat{\underline{y}} = \underline{H} \hat{\underline{\theta}} = \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{x} = \underline{P} \underline{x}$$

$$\underline{x} - \hat{\underline{s}} = \underline{x} - \underline{P}\underline{x} = (\underline{I} - \underline{P})\underline{x}$$

$$\begin{aligned} \|\hat{\underline{s}}\|^2 + \|\underline{x} - \hat{\underline{s}}\|^2 &= \underline{x}^T \underline{P}^T \underline{P} \underline{x} \\ &\quad + \underline{x}^T (\underline{I} - \underline{P})^T (\underline{I} - \underline{P}) \underline{x} \\ &= \underline{x}^T \underline{P} \underline{x} + \underline{x}^T (\underline{I} - \underline{P}) \underline{x} = \underline{x}^T \underline{x} = \|\underline{x}\|^2 \end{aligned}$$

Since \underline{P} and $(\underline{I} - \underline{P})$ are symmetric and idempotent.

$$11) \quad \underline{x}_1 = \underline{z}_1 + \underline{z}_1^\perp \quad \underline{x}_2 = \underline{z}_2 + \underline{z}_2^\perp$$

$$\begin{aligned} l &= \underline{x}_1^T \underline{P} \underline{x}_2 - \underline{x}_2^T \underline{P} \underline{x}_1 = (\underline{z}_1 + \underline{z}_1^\perp)^T \underline{P} (\underline{z}_2 + \underline{z}_2^\perp) \\ &\quad - (\underline{z}_2 + \underline{z}_2^\perp)^T \underline{P} (\underline{z}_1 + \underline{z}_1^\perp) \\ &= (\underline{z}_1 + \underline{z}_1^\perp)^T \underline{P} \underline{z}_2 - (\underline{z}_2 + \underline{z}_2^\perp)^T \underline{P} \underline{z}_1 \end{aligned}$$

$$\text{Since } \underline{P} \underline{z}_1^\perp = 0$$

$$l = (\underline{z}_1 + \underline{z}_1^\perp)^T \underline{z}_2 - (\underline{z}_2 + \underline{z}_2^\perp)^T \underline{z}_1$$

Since $\underline{P} \underline{z}_2 = \underline{z}_2$

$$l = \underline{z}_1^T \underline{z}_2 - \underline{z}_2^T \underline{z}_1 = 0 \quad \text{Since } \underline{z}_i^T \underline{z}_j = 0$$

$$\text{Now let } \underline{x}_1 = \underline{e}_i = [0 \dots 0 \underset{\substack{\uparrow \\ i^{\text{th}} \text{ place}}}{1} 0 \dots 0]^T$$

$$\underline{x}_2 = \underline{e}_j$$

$$\underline{e}_i^T \underline{P} \underline{e}_j = \underline{e}_j^T \underline{P} \underline{e}_i \quad \text{from previous result}$$

$$\Rightarrow [\underline{P}]_{ij} = [\underline{P}]_{ji}$$

$$\begin{aligned} 12) a) \quad \underline{P}^2 &= \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T \\ &= \underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T = \underline{P} \end{aligned}$$

$$\begin{aligned} b) \quad \underline{x}^T \underline{P} \underline{x} &= \underline{x}^T \underline{P} \underline{P} \underline{x} = \underline{x}^T \underline{P}^T \underline{P} \underline{x} \\ &= (\underline{P} \underline{x})^T \underline{P} \underline{x} = \|\underline{P} \underline{x}\|^2 \geq 0 \end{aligned}$$

$$\begin{aligned} c) \quad \underline{P} \underline{x} &= \lambda \underline{x} \Rightarrow \underline{P} \underline{x} = \underline{P} \underline{P} \underline{x} = \underline{P} \lambda \underline{x} \\ &= \lambda (\underline{P} \underline{x}) = \lambda^2 \underline{x} \end{aligned}$$

\Rightarrow If λ is an eigenvalue, so is λ^2 .
By uniqueness $\lambda^2 = \lambda$ or $\lambda = 0, 1$

d) Rank = sum of nonzero eigenvalues.
To show that there are p nonzero eigenvalues, we find $\text{tr}(\underline{P}) = \sum_{i=1}^N \lambda_i$

$$\begin{aligned} \text{tr}(\underline{P}) &= \text{tr}(\underline{H} (\underline{H}^T \underline{H})^{-1} \underline{H}^T) \\ &= \text{tr}((\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{H}) \\ &= \text{tr}(\underline{I}) = p \end{aligned}$$

$$\begin{aligned} \Rightarrow \sum_{i=1}^p \lambda_i &= p \Rightarrow p \text{ eigenvalues} = 1 \\ &\Rightarrow \text{rank} = p \end{aligned}$$