ESTIMATION THEORY

- Estimation of parameters
  - General principles and methods
  - Estimation of mean and correlation function
  - Spectrum estimation (Chapter 10)

- Estimation of random variables
  - Bayes (nonlinear) estimation
  - Linear mean-square estimation

PARAMETER ESTIMATION PROBLEM

1. A set of random variables $x_1, x_2, \cdots, x_N$ is characterized by

$$f_{x_1,x_2,\cdots,x_N;\theta}(x_2, x_2, \cdots, x_N; \theta) = f_{\boldsymbol{x};\theta}(\boldsymbol{x}; \theta)$$

   where θ is a *parameter* of the density.

2. The form of the density $f_{\boldsymbol{x};\theta}(\boldsymbol{x}; \theta)$ i.e., its *dependence* on θ, is known but the *value* of θ is not known.

3. Given an observation $\mathbf{x}^o \Leftrightarrow \{x_1^o, x_2^o, \cdots, x_N^o\}$, use this to estimate θ.

MAXIMUM LIKELIHOOD ESTIMATION

- The quantity $f_{\boldsymbol{x};\theta}(\mathbf{x}^{\mathsf{O}};\theta)$ for *fixed* $\mathbf{x}^{\mathsf{O}}$ *regarded as a function of* θ is called the *likelihood function*.

- Choose θ to maximize the likelihood function

$$f_{\boldsymbol{x};\theta}(\mathbf{x}^{\mathsf{O}};\theta)$$

This makes the given observation $\mathbf{x}^{\mathsf{O}}$ the *most likely event*.

- The maximizing value of θ

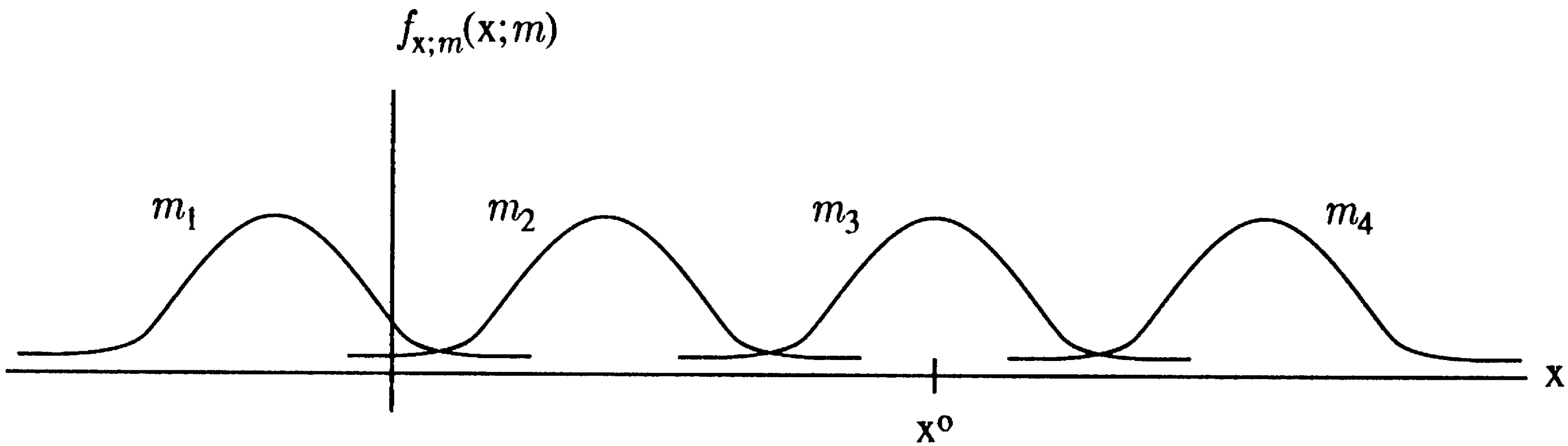
$$\widehat{\theta}_{ml}(\boldsymbol{x}) = \operatorname{argmax} \quad f_{\boldsymbol{x};\theta}(\boldsymbol{x};\theta)$$

is called the *maximum likelihood estimate*.

ILLUSTRATION OF ML ESTIMATE

$$f_{\boldsymbol{x};m}(\mathsf{x}^{\mathsf{o}};m) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mathsf{x}^{\mathsf{o}}-m)^2}$$

Gaussian density
unknown mean
$N = 1,\ \theta = m$



- Choose $\hat{m}_{ml} = m_3$: This makes x^{o} the most likely event.

ESTIMATING THE MEAN (MULTIPLE OBSERVATIONS)

Given N *independent* observations $\mathbf{x} = [\mathsf{x}_1, \mathsf{x}_2, \cdots, \mathsf{x}_N]^T$ from a Gaussian density with unknown mean, form the likelihood function

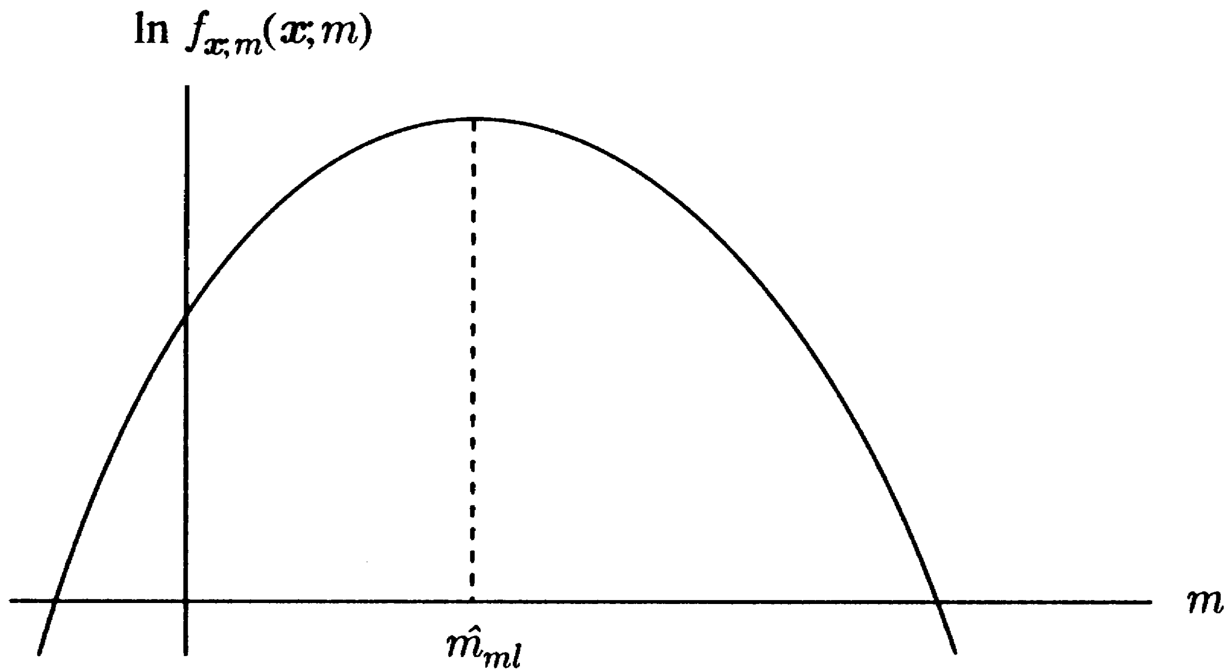
$$f_{\boldsymbol{x};m}(\boldsymbol{x};m) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{\mathsf{o}}^2}} e^{-\frac{(x_i-m)^2}{2\sigma_{\mathsf{o}}^2}}$$

Maximize: $\ln f_{\boldsymbol{x};m}(\boldsymbol{x};m) = -N \ln(\sqrt{2\pi\sigma_{\mathsf{o}}^2}) - \sum_{i=1}^{N} \frac{(x_i - m)^2}{2\sigma_{\mathsf{o}}^2}$

View this as a *function of m*:

$$\ln f_{\boldsymbol{x};m}(\boldsymbol{x};m) = \left(-\frac{N}{2\sigma_{\mathsf{o}}^2}\right) m^2 + \left(\sum_{i=1}^{N} \frac{x_i}{\sigma_{\mathsf{o}}^2}\right) m - \left(N \ln(\sqrt{2\pi\sigma_{\mathsf{o}}^2}) + \sum_{i=1}^{N} \frac{x_i^2}{2\sigma_{\mathsf{o}}^2}\right)$$

LIKELIHOOD FUNCTION FOR GAUSSIAN DENSITY WITH UNKNOWN MEAN



$$\ln f_{\boldsymbol{x};m}(\boldsymbol{x};m) = \left(-\frac{N}{2\sigma_o^2}\right) m^2 + \left(\sum_{i=1}^{N} \frac{x_i}{\sigma_o^2}\right) m - \left(N \ln(\sqrt{2\pi\sigma_o^2}) + \sum_{i=1}^{N} \frac{x_i^2}{2\sigma_o^2}\right)$$

MAXIMIZATION OF LIKELIHOOD FUNCTION

Starting with

$$\ln f_{\boldsymbol{x};m}(\boldsymbol{x};m) = \left(-\frac{N}{2\sigma_o^2}\right) m^2 + \left(\sum_{i=1}^{N} \frac{x_i}{\sigma_o^2}\right) m - \left(N \ln(\sqrt{2\pi\sigma_o^2}) + \sum_{i=1}^{N} \frac{x_i^2}{2\sigma_o^2}\right)$$

Set the derivative equal to zero:

$$\frac{\partial \ln f_{\boldsymbol{x};m}(\boldsymbol{x};m)}{\partial m} = -\frac{N}{\sigma_o^2} m + \frac{1}{\sigma_o^2} \sum_{i=1}^{N} x_i = 0$$

The result is . . .

$$\widehat{m}_{ml} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \text{the ``sample mean''}$$

Insert Example 6.1 here.

DEFINITION OF AN ESTIMATE

- An *estimate* for a parameter is a function of the observations

  $$\widehat{\theta}_N = \widehat{\theta}_N(x_1, x_2, \cdots, x_N) = \widehat{\theta}_N(\boldsymbol{x})$$

  Since the observations are random variables, $\widehat{\theta}_N$ is also a random variable.

- Any function of the samples $x_1, x_2, \cdots, x_N$ is called a *statistic*. It has an associated mean, variance, and density function. An estimate is a type of statistic.

PROPERTIES OF ESTIMATES

1. An estimate is *unbiased* if $\mathcal{E}\left\{\widehat{\theta}_N\right\} = \theta$. Otherwise it is *biased* with bias $b(\theta) = \mathcal{E}\left\{\widehat{\theta}_N\right\}$.

   It is *asymptotically unbiased* if $\lim_{N\to\infty} \mathcal{E}\left\{\widehat{\theta}_N\right\} = \theta$.

2. An estimate is *consistent* if

$$\lim_{N\to\infty} \mathrm{Pr}\left[\left|\widehat{\theta}_N - \theta\right| < \varepsilon\right] = 1$$

   for any small number ε. The sequence of estimates $\{\widehat{\theta}_N\}$ *converges in probability* to the true value θ.

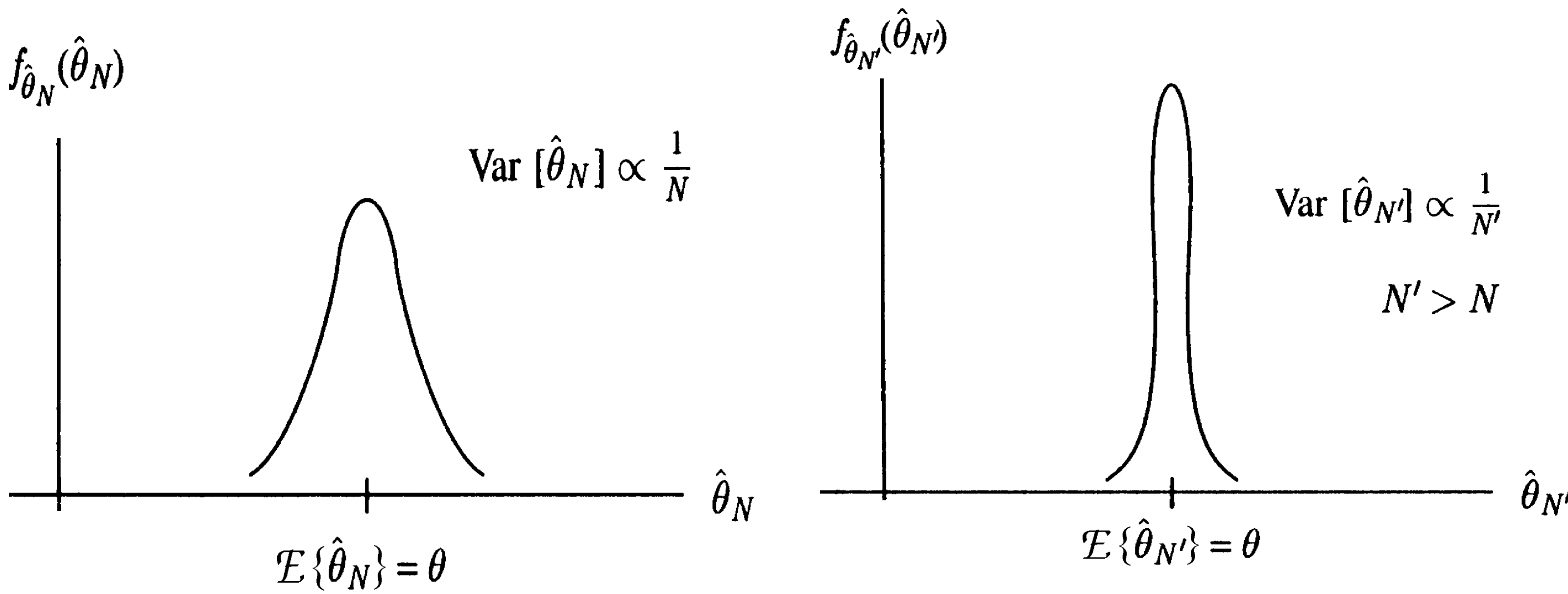
PROPERTIES OF ESTIMATES (cont'd.)

3. An estimate is *efficient* with respect to another estimate if it has a lower variance. An estimate is said to be "most efficient" if it has the lowest possible variance.

Note: Some authors use the term "efficient" to mean "most efficient."

DEMONSTRATION OF PROPERTIES

- If $\widehat{\theta}_N$ is <u>*unbiased*</u> and <u>*efficient*</u> with respect to $\widehat{\theta}_{N-1}$ then $\widehat{\theta}_N$ is a consistent estimate.



(Decrease in the variance implies convergence in probability.)

Insert Example 6.3 here.

CRAMÉR-RAO BOUND

For any unbiased estimate $\widehat{\theta}_N$

$$\text{Var}\left[\widehat{\theta}\right] \geq \frac{1}{\mathcal{E}\left\{\left(\frac{\partial \ln f_{\boldsymbol{x};\theta}(\boldsymbol{x};\theta)}{\partial\theta}\right)^2\right\}} = \frac{1}{-\mathcal{E}\left\{\frac{\partial^2 \ln f_{\boldsymbol{x};\theta}(\boldsymbol{x};\theta)}{\partial\theta^2}\right\}}$$

- If the bound is satisfied with equality, the estimate is called a *minimum-variance* estimate (or "most efficient" estimate).

  This requires that $\dfrac{\partial \ln f_{\boldsymbol{x};\theta}}{\partial\theta} = K(\theta)\cdot(\widehat{\theta}_N - \theta)$.

- If a maximum likelihood estimate exists and does not occur at a boundary, it is the minimum-variance estimate.

Insert Example 6.4 here.

(Review Appendix A slides here.)

ML ESTIMATION: VECTOR PARAMETERS

DEFINITION

$$\hat{\boldsymbol{\theta}}_{ml}(\boldsymbol{x}) = \text{argmax} \quad f_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta})$$

LIKELIHOOD EQUATIONS

$$\nabla_{\boldsymbol{\theta}^*} f_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta}) = \mathbf{0} \qquad \text{(likelihood equation)}$$

$$\nabla_{\boldsymbol{\theta}^*} \ln f_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta}) = \mathbf{0} \quad \text{(log likelihood equation)}$$

PROPERTIES OF VECTOR ESTIMATES

1. An estimate $\widehat{\boldsymbol{\theta}}_N$ is *unbiased* if $\mathcal{E}\left\{\widehat{\boldsymbol{\theta}}_N\right\} = \boldsymbol{\theta}$.
   Otherwise it is *biased* with bias $\mathbf{b}(\boldsymbol{\theta}) = \mathcal{E}\left\{\widehat{\boldsymbol{\theta}}_N\right\} - \boldsymbol{\theta}$.

   It is *asymptotically unbiased* if $\lim_{N\to\infty} \mathcal{E}\left\{\widehat{\boldsymbol{\theta}}_N\right\} = \boldsymbol{\theta}$.

2. An estimate is *consistent* if

   $$\lim_{N\to\infty} \Pr\left[\left\|\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right\| < \epsilon\right] = 1$$

   for any arbitrarily small number ϵ. The sequence of estimates $\{\widehat{\boldsymbol{\theta}}_N\}$ *converges in probability* to the parameter $\boldsymbol{\theta}$.

PROPERTIES OF VECTOR ESTIMATES (cont'd.)

3. An estimate $\widehat{\boldsymbol{\theta}}$ is <u>*efficient*</u> with respect to another estimate $\widehat{\boldsymbol{\theta}}'$ if the difference of their covariance matrices $\mathsf{C}_{\widehat{\boldsymbol{\theta}}'} - \mathsf{C}_{\widehat{\boldsymbol{\theta}}}$ is positive definite.

- Property (3) implies that the variance of every component of $\widehat{\boldsymbol{\theta}}$ is smaller than the variance of the corresponding component of $\widehat{\boldsymbol{\theta}}'$.

- If $\widehat{\boldsymbol{\theta}}_N$ is unbiased and efficient with respect to $\widehat{\boldsymbol{\theta}}_{N-1}$ for all N then $\widehat{\boldsymbol{\theta}}_N$ is a consistent estimate.

C-R BOUND FOR VECTOR PARAMETERS

Let $\boldsymbol{\theta}$ be a vector parameter and $\widehat{\boldsymbol{\theta}}$ an unbiased estimate with covariance matrix $\mathbf{C}_{\widehat{\boldsymbol{\theta}}}$. Then

$$\mathbf{C}_{\widehat{\boldsymbol{\theta}}} \geq \mathbf{J}^{-1}$$

where

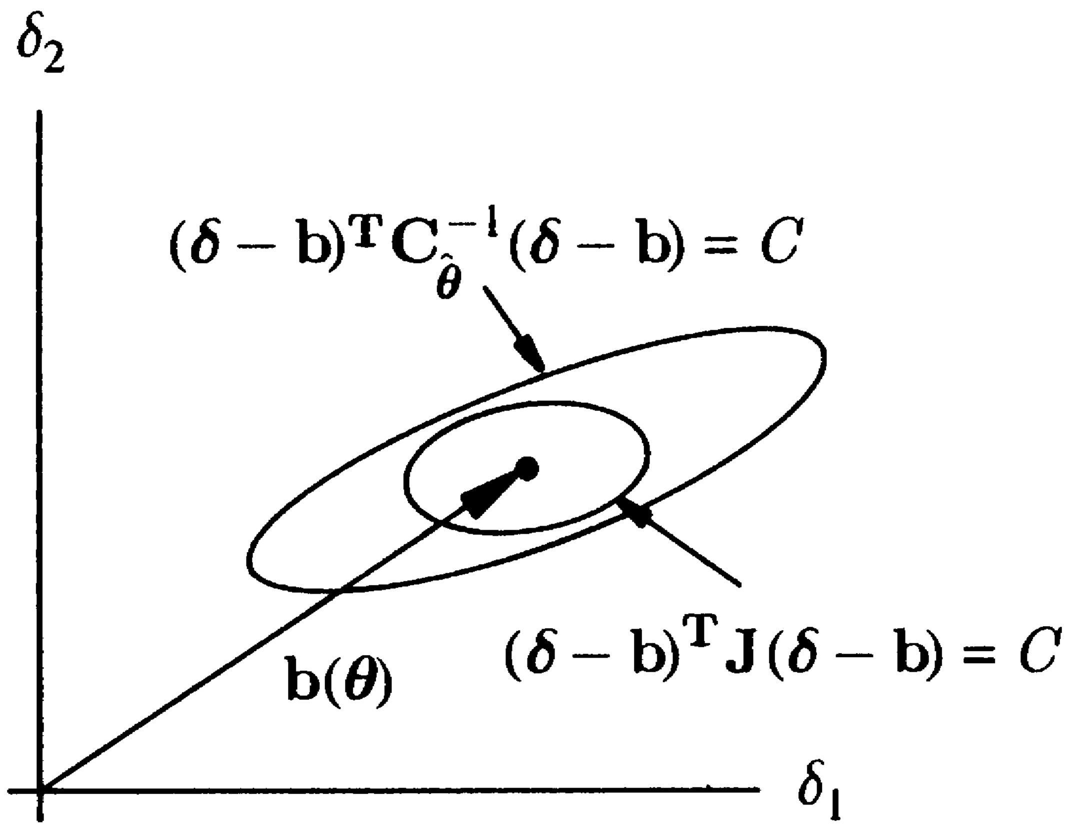
$$\mathbf{J} = \mathcal{E}\left\{\boldsymbol{s}(\boldsymbol{x};\boldsymbol{\theta})\boldsymbol{s}^T(\boldsymbol{x};\boldsymbol{\theta})\right\} \quad \text{“Fisher information matrix”}$$

and

$$\boldsymbol{s}(\boldsymbol{x};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln f_{\boldsymbol{x};\boldsymbol{\theta}}(\boldsymbol{x};\boldsymbol{\theta}) \qquad \text{“score vector”}$$

- The inequality is interpreted to mean that the matrix $\mathbf{C}_{\widehat{\boldsymbol{\theta}}} - \mathbf{J}^{-1}$ is a positive semi-definite matrix.

GEOMETRICAL INTERPRETATION OF C-R BOUND (VECTOR PARAMETERS)



$\boldsymbol{\delta} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$
"deviation"

C a constant

NOTES ON THE C-R BOUND FOR VECTOR PARAMETERS

- The stated bound applies only to real-valued parameters.

- The variances of the individual components of the vector satisfy Var $\left[\widehat{\theta}_i\right] \geq j_{ii}^{(-1)}$.

- The bound is satisfied with equality if and only if the estimate satifies an equation of the form $\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta} = \mathbf{K}(\boldsymbol{\theta}) \cdot \boldsymbol{s}(\boldsymbol{x}; \boldsymbol{\theta})$.

- Generalizations of the bound apply for biased estimates.

Insert Example 6.5 here.

ESTIMATING THE MEAN OF A RANDOM PROCESS

SAMPLE MEAN

$$\widehat{m}_x = \frac{1}{N_s} \sum_{n=0}^{N_s-1} x[n]$$

MEAN OF SAMPLE MEAN

$$\mathcal{E}\{\widehat{m}_x\} = \frac{1}{N_s} \sum_{n=0}^{N_s-1} \mathcal{E}\{x[n]\} = \frac{1}{N_s} \sum_{n=0}^{N_s-1} m_x = m_x$$

- This shows the sample mean is unbiased.

ESTIMATING THE MEAN (cont'd.)

VARIANCE OF SAMPLE MEAN

Let $\mathbf{1} = [1, 1, \ldots, 1]^T$; $\quad \boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T$ so $\widehat{m}_x = \frac{1}{N_s}\mathbf{1}^T\boldsymbol{x}$.

$$\begin{aligned}
\text{Var}\,[\widehat{m}_x] &= \mathcal{E}\left\{|\widehat{m}_x - m_x|^2\right\} = \frac{1}{N_s^2}\mathbf{1}^T\mathcal{E}\left\{(\boldsymbol{x} - \mathbf{m}_{\boldsymbol{x}})(\boldsymbol{x} - \mathbf{m}_{\boldsymbol{x}})^{*T}\right\}\mathbf{1} \\
&= \frac{1}{N_s^2}\mathbf{1}^T\mathbf{C}_{\boldsymbol{x}}\mathbf{1} = \frac{1}{N_s^2}\sum_{l=-N_s+1}^{N_s-1}[N_s - |l|]C_x[l] \\
&= \frac{1}{N_s}\sum_{l=-N_s+1}^{N_s-1}\left(1 - \frac{|l|}{N_s}\right)C_x[l]
\end{aligned}$$

- Since $\lim_{N_s\to\infty}\text{Var}\,[\widehat{m}_x] = 0$ the estimate is *consistent*.

Detail of matrix calculations

$$\mathbf{1}^T\mathbf{C}_{\boldsymbol{x}}\mathbf{1} \;=\; [1\ 1\ \cdots\ 1\ 1] \begin{bmatrix} C_x[0] & C_x[-1] & \cdots & C_x[-N_s{+}1] \\ C_x[1] & C_x[0] & \cdots & \\ \vdots & \ddots & \ddots & \ddots \\ & & & C_x[0]C_x[-1] \\ C_x[N_s{-}1] & & & C_x[1]C_x[0] \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}$$

= sum of all terms in the matrix

N_s terms $C_x[0]$ on main diagonal

$N_s - 1$ terms $C_x[1]$ on next diagonal

$N_s - 2$ terms $C_x[2]$ on next diagonal

$\vdots$

etc.

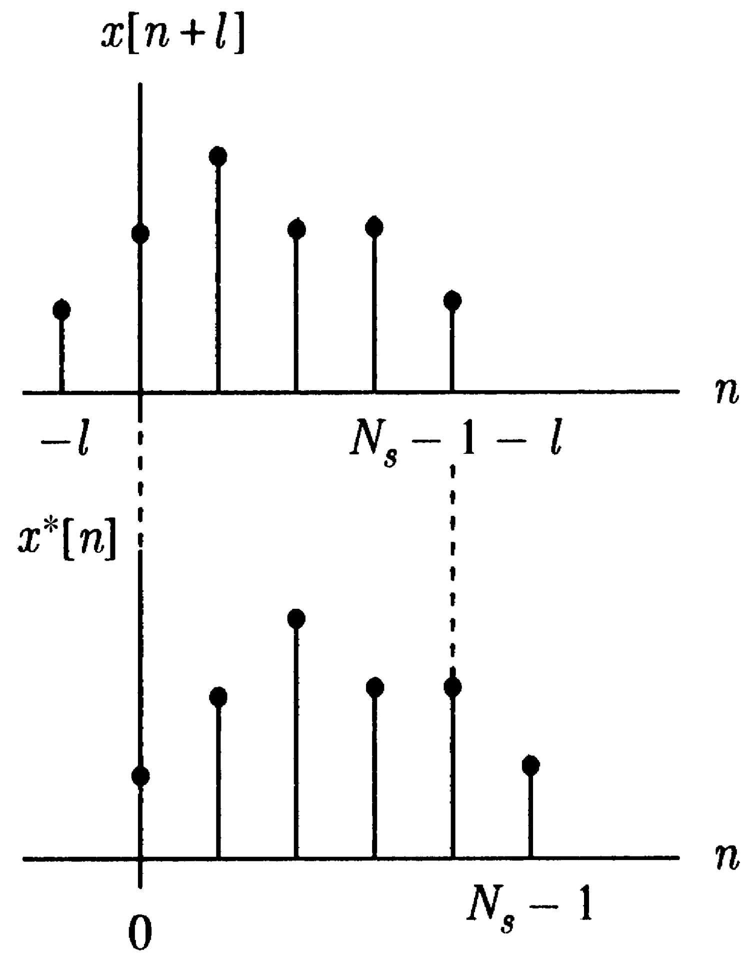
thus $$\mathbf{1}^T\mathbf{C}_{\boldsymbol{x}}\mathbf{1} = \sum_{l=-N_s+1}^{N_s-1} (N_s - |l|) \cdot C_x[l]$$

ESTIMATING THE CORRELATION FUNCTION

$$\widehat{R}'_x[l] = \frac{1}{N_s - l} \sum_{n=l}^{N_s-1} x[n]x^*[n-l]$$

$$= \frac{1}{N_s - l} \sum_{n=0}^{N_s-1-l} x[n+l]x^*[n]$$

$$0 \leq l < N_s$$



MEAN AND VARIANCE OF $\widehat{R}'_x$ ESTIMATE

MEAN

$$\mathcal{E}\left\{\widehat{R}'_x[l]\right\} = \frac{1}{N_s - l}\sum_{n=0}^{N_s-1-l}\mathcal{E}\{x[n+l]x^*[n]\} = \frac{1}{N_s - l}\sum_{n=0}^{N_s-1-l} R_x[l] = R_x[l]$$

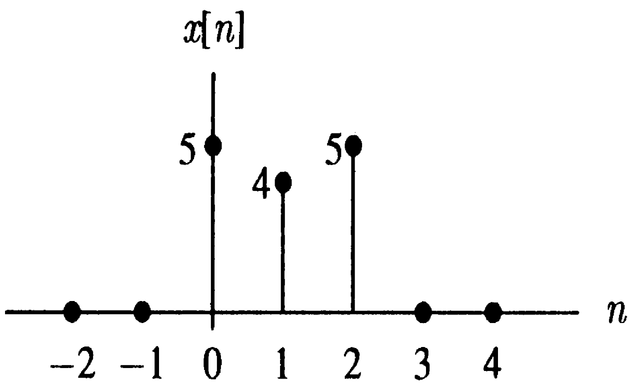
VARIANCE

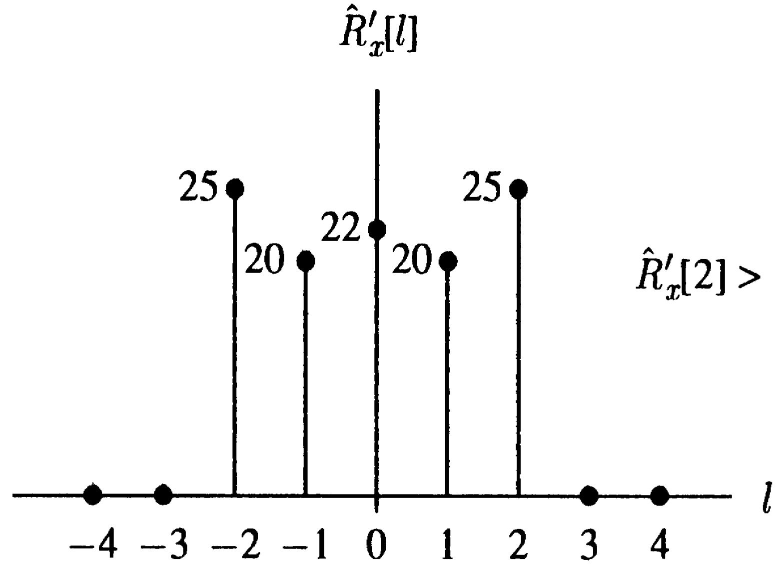
$$\mathrm{Var}\left[\widehat{R}'_x[l]\right] = \frac{1}{N_s - |l|}\sum_{k=-N_s+1+|l|}^{N_s-1-|l|}\left(1 - \frac{|k|}{N_s - |l|}\right)|R_x[k]|^2 \quad \text{(see text)}$$

- Since $\mathrm{Var}\left[\widehat{R}'_x[l]\right] \to 0$ as $N_s \to \infty$ the estimate is *consistent*.
- Estimate is *not* guarranteed to be positive semi-definite.

DEMONSTRATION BY COUNTEREXAMPLE

$\widehat{R}'_x[l]$ MAY NOT BE POSITIVE SEMI-DEFINITE





$\hat{R}'_x[2] > \hat{R}'_x[0]$

ALTERNATIVE CORRELATION FUNCTION ESTIMATE

$$\hat{R}_x[l] = \frac{1}{N_s} \sum_{n=0}^{N_s-1-l} x[n+l]x^*[n]; \qquad 0 \le l < N_s$$

MEAN

$$\mathcal{E}\left\{\hat{R}_x[l]\right\} = \frac{N_s - |l|}{N_s} R_x[l] \qquad \text{(biased)}$$

Since $\mathcal{E}\left\{\hat{R}_x[l]\right\} \to R_x[l]$ as $N_s \to \infty$ this is *asymptotically unbiased*.

VARIANCE

$$\text{Var}\left[\hat{R}_x[l]\right] = \frac{1}{N_s} \sum_{k=-N_s+1+|l|}^{N_s-1-|l|} \left(1 - \frac{|l| + |k|}{N_s}\right) |R_x[k]|^2$$

Since $\text{Var}\left[\hat{R}_x[l]\right] \to 0$ as $N_s \to \infty$, the estimate is *consistent*.

ESTIMATING THE CROSS-CORRELATION FUNCTION

BIASED ESTIMATE

$$\widehat{R}_{xy}[l] = \begin{cases} \dfrac{1}{N_s} \displaystyle\sum_{n=0}^{N_s-1-l} x[n+l]y^*[n] & 0 \leq l < N_s \\ \\ \dfrac{1}{N_s} \displaystyle\sum_{n=0}^{N_s-1-|l|} x[n]y^*[n+|l|] & -N_s < l < 0 \end{cases}$$

- Properties are similar to those of $\widehat{R}_x$.
- To obtain the unbiased estimate, replace $\dfrac{1}{N_s}$ by $\dfrac{1}{N_s - |l|}$.

ESTIMATING THE CORRELATION MATRIX

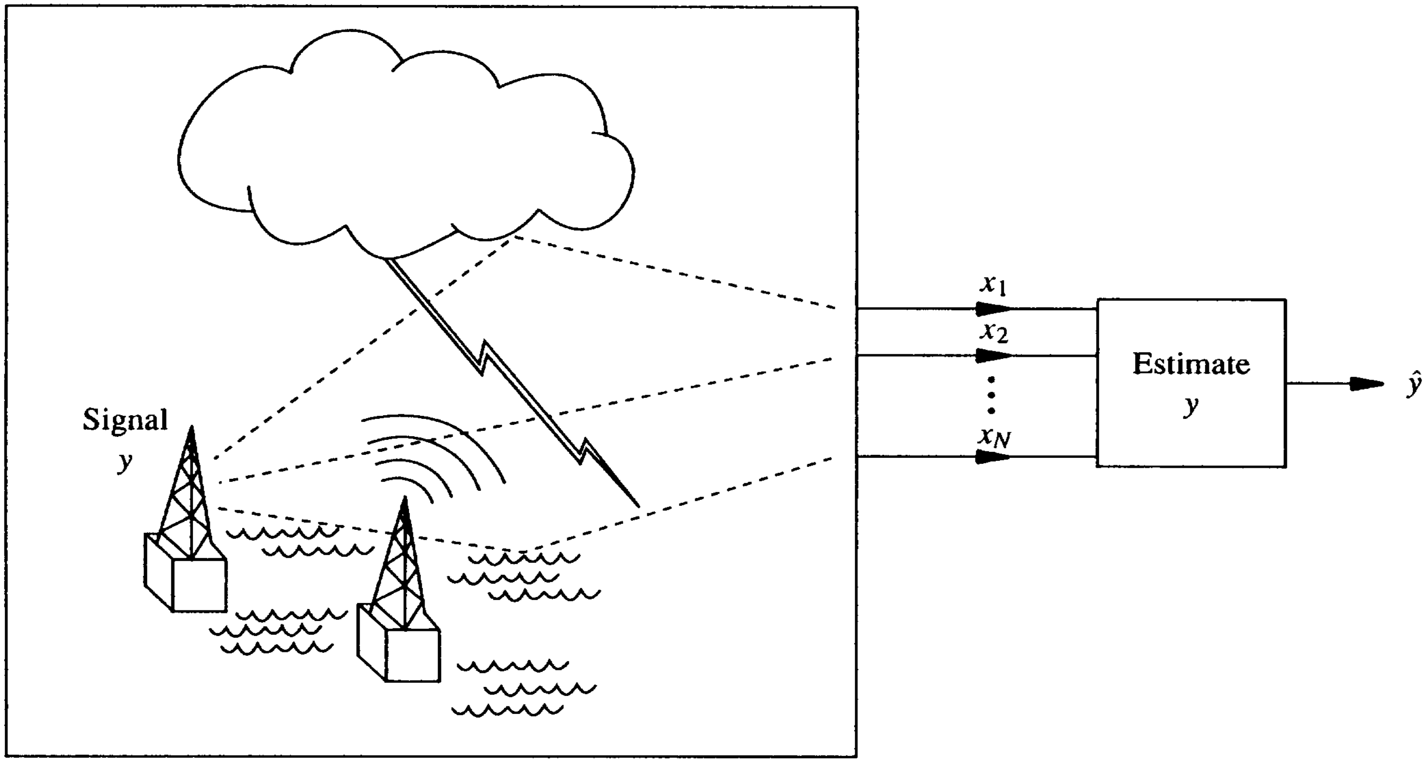
AUTOCORRELATION METHOD

(size 3×3 illustrated; $N_s = 50$)

$$\widehat{\mathbf{R}}_{\boldsymbol{x}} = \frac{1}{N_s} \cdot \underbrace{\begin{bmatrix} x^*[0] & x^*[1] & x^*[2] & \cdots & x^*[49] & 0 & 0 \\ 0 & x^*[0] & x^*[1] & \cdots & x^*[48] & x^*[49] & 0 \\ 0 & 0 & x^*[0] & \cdots & x^*[47] & x^*[48] & x^*[49] \end{bmatrix}}_{\mathbf{X}^{*T}} \underbrace{\begin{bmatrix} x[0] & 0 & 0 \\ x[1] & x[0] & 0 \\ x[2] & x[1] & x[0] \\ \vdots & \vdots & \vdots \\ x[49] & x[48] & x[47] \\ 0 & x[49] & x[48] \\ 0 & 0 & x[49] \end{bmatrix}}_{\mathbf{X}}$$

- To compute a cross-correlation matrix $\widehat{\mathbf{R}}_{\boldsymbol{xy}}$, replace the first term with $\mathbf{Y}^{*T}$.

ESTIMATION PROBLEM



y: a random variable

$x_1, \cdots, x_N$: a set of related measurements.

BAYES ESTIMATION

OBSERVATIONS: $x_1, x_2, \cdots, x_N$

ESTIMATE: $\widehat{y} = \widehat{y}(\boldsymbol{x}) = \phi(x_1, x_2, \cdots, x_N)$

BAYES APPROACH

- Assume a cost function $\mathcal{C}(y, \widehat{y}) \geq 0$

- Minimize the *risk* $\mathcal{R} = \mathcal{E}\{\mathcal{C}(y, \widehat{y})\}$

EXPRESSION FOR THE RISK

$$\mathcal{R} = \mathcal{E}\{\mathcal{C}(y,\widehat{y})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{C}(\mathsf{y},\widehat{y}(\mathbf{x}))f_{y\boldsymbol{x}}(\mathsf{y},\mathbf{x})d\mathsf{y}d\mathbf{x}$$

$$= \int_{-\infty}^{\infty} \mathcal{I}(\widehat{y})f_{\boldsymbol{x}}(\mathbf{x})d\mathbf{x} = \mathcal{E}\{\mathcal{I}(\widehat{y})\}$$

where

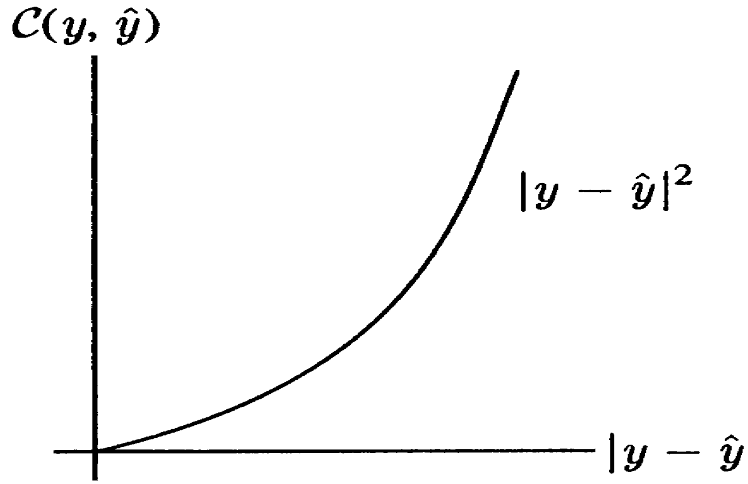
$$\mathcal{I}(\widehat{y}) = \int_{-\infty}^{\infty} \mathcal{C}(\mathsf{y},\widehat{y})f_{y|\boldsymbol{x}}(\mathsf{y}|\mathbf{x})d\mathsf{y}$$

Since $\mathcal{C}(y,\widehat{y}) \geq 0$ implies $\mathcal{I}(\widehat{y}) \geq 0$ . . .

- To minimize $\mathcal{R} = \mathcal{E}\{\mathcal{I}(\widehat{y})\}$, minimize $\mathcal{I}(\widehat{y})$ for every $\widehat{y}$.

BAYES ESTIMATE: MEAN-SQUARE COST

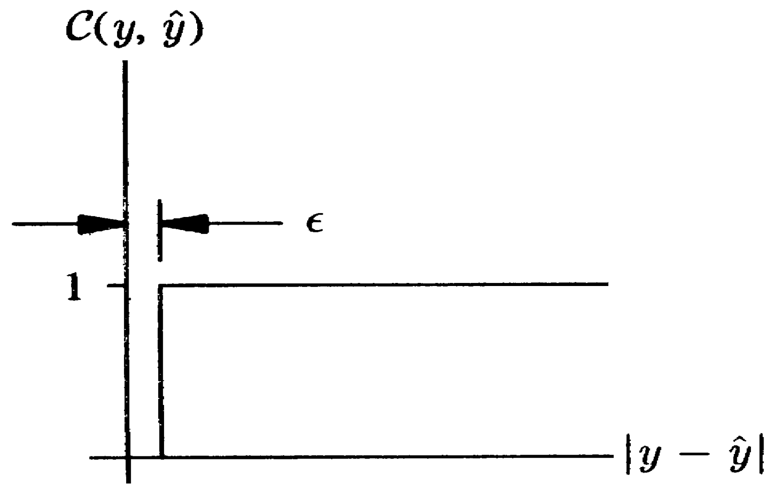
$$\mathcal{I}(\hat{y}) = \int_{-\infty}^{\infty} |\mathsf{y}-\hat{y}|^2 f_{y|\boldsymbol{x}}(\mathsf{y}|\boldsymbol{x})d\mathsf{y}$$



$$\nabla_{\hat{y}^*}\mathcal{I}(\hat{y}) = -\int_{-\infty}^{\infty}(\mathsf{y}-\hat{y})f_{y|\boldsymbol{x}}(\mathsf{y}|\boldsymbol{x})d\mathsf{y} = 0$$

$$\Rightarrow \quad \boxed{\hat{y}_{ms}(\boldsymbol{x}) = \int_{-\infty}^{\infty} \mathsf{y} f_{y|\boldsymbol{x}}(\mathsf{y}|\boldsymbol{x})d\mathsf{y}} \quad \text{“mean-square estimate”}$$

BAYES ESTIMATE: UNIFORM COST



$$\begin{aligned}\mathcal{I}(\hat{y}) &= \int_{|y-\hat{y}|\geq\epsilon} f_{y|\boldsymbol{x}}(\mathrm{y}|\boldsymbol{x})d\mathrm{y} \\ &= 1-\int_{|y-\hat{y}|<\epsilon} f_{y|\boldsymbol{x}}(\mathrm{y}|\boldsymbol{x})d\mathrm{y}\end{aligned}$$

$\mathcal{I}(\hat{y}) = 1 - f_{y|\boldsymbol{x}}(\mathrm{y}|\boldsymbol{x})\cdot 2\epsilon$ is minimized when $f_{y|\boldsymbol{x}}$ is maximized.

$$\Rightarrow \quad \boxed{\hat{y}_{MAP}(\boldsymbol{x}) = \underset{y}{\operatorname{argmax}} \quad f_{y|\boldsymbol{x}}(\mathrm{y}|\boldsymbol{x})} \quad \text{"MAP estimate"}$$

Insert Example 6.6 here.

BOUND ON MEAN-SQUARE ERROR
(ANALOGOUS TO CRAMÉR-RAO BOUND)

$$\mathcal{E}\left\{(y-\widehat{y})^2\right\} \geq \frac{1}{\mathcal{E}\left\{\left(\frac{\partial \ln f_{y\boldsymbol{x}}(y,\boldsymbol{x})}{\partial y}\right)^2\right\}} = \frac{1}{-\mathcal{E}\left\{\frac{\partial^2 \ln f_{y\boldsymbol{x}}(y,\boldsymbol{x})}{\partial y^2}\right\}}$$

Bound is met with equality when . . .

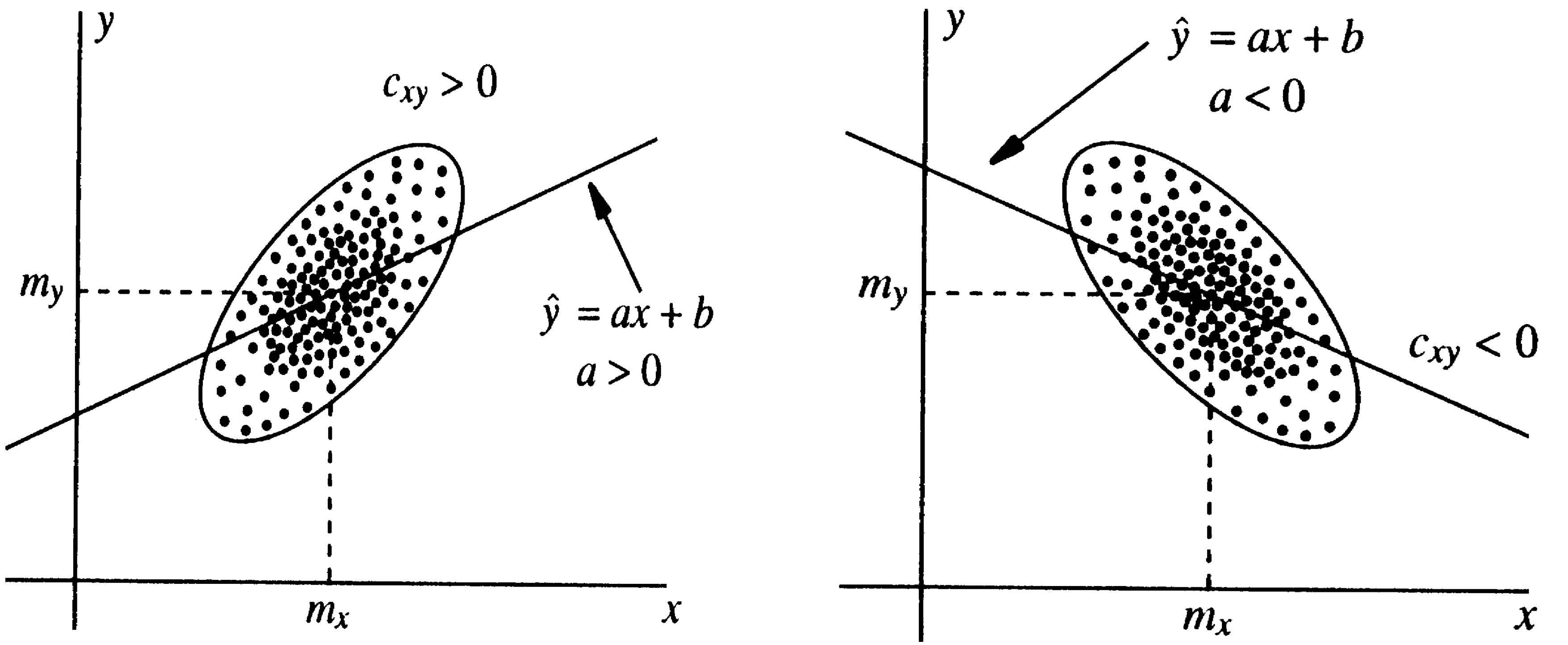
$$y - \widehat{y}(\boldsymbol{x}) = -K \cdot \frac{\partial \ln f_{y\boldsymbol{x}}(y,\boldsymbol{x})}{\partial y}$$

in which case

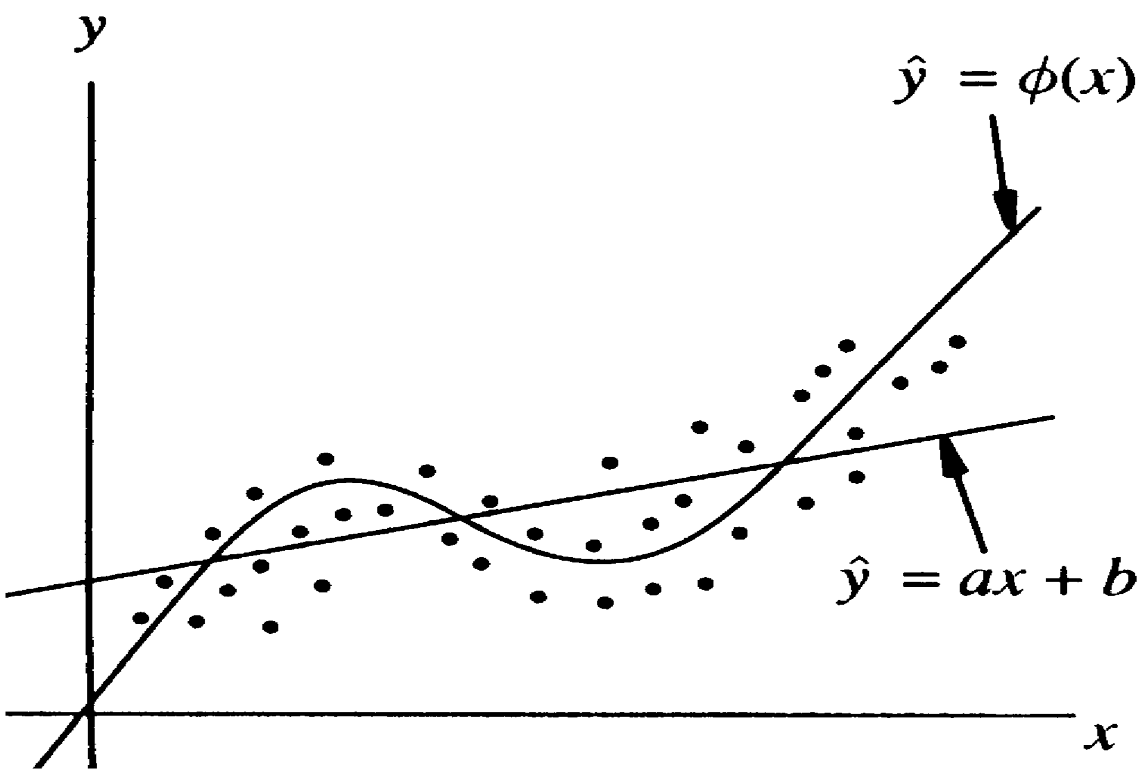
$$K = \frac{1}{\mathcal{E}\left\{\left(\frac{\partial \ln f_{y\boldsymbol{x}}(y,\boldsymbol{x})}{\partial y}\right)^2\right\}} = \frac{1}{-\mathcal{E}\left\{\frac{\partial^2 \ln f_{y\boldsymbol{x}}(y,\boldsymbol{x})}{\partial y^2}\right\}}$$

Insert Example 6.7 here.

MEAN-SQUARE ESTIMATES: GAUSSIAN DATA



LINEAR AND NONLINEAR ESTIMATES



LINEAR MEAN-SQUARE ESTIMATION

- The mean-square error $\mathcal{E}\left\{|y-\hat{y}|^2\right\}$ is minimized while requiring the estimate *to be linear*.

- The estimate $\hat{y}$ depends only on first and second moments.

- Solution involves *linear* equations.

SIMPLE LMS ESTIMATION PROBLEM

- Assume real scalar variables, x, y; estimate $\hat{y} = ax + b$.
- Choose a and b to *minimize* $\mathcal{E}_{lms} = \mathcal{E}\left\{(y - \hat{y})^2\right\}$.

POSSIBLE APPROACH

1) Substitute $\hat{y} = ax + b$ into expression for $\mathcal{E}_{lms}$.

2) Set $\dfrac{\partial \mathcal{E}_{lms}}{\partial a} = 0$ and $\dfrac{\partial \mathcal{E}_{lms}}{\partial b} = 0$
   — solve simultaneous equations for a, b.

SIMPLE LMS ESTIMATION (cont'd.)

BETTER APPROACH

1. Define $\varepsilon = y - \widehat{y}$ then $\mathcal{E}_{lms} = \mathcal{E}\left\{\varepsilon^2\right\} = \sigma_\varepsilon^2 + m_\varepsilon^2$.

   To minimize $\mathcal{E}_{lms}$, require $m_\varepsilon = 0$:

   $$m_\varepsilon = \mathcal{E}\{y - \widehat{y}\} = \mathcal{E}\{y - ax - b\} = 0$$

   thus $b = m_y - am_x$.

2. Now write $\mathcal{E}_{lms}$ as a function of a single variable:

   $$\begin{aligned} \mathcal{E}_{lms} &= \mathcal{E}\left\{(y - \widehat{y})^2\right\} = \mathcal{E}\left\{(y - ax - b)^2\right\} \\ &= \mathcal{E}\left\{[(y - m_y) - a(x - m_x)]^2\right\} \end{aligned}$$

SIMPLE LMS ESTIMATION (cont'd.)

BETTER APPROACH (cont'd.)

3. Minimize: $\mathcal{E}_{lms} = \mathcal{E}\Big\{[(y - m_y) - a(x - m_x)]^2\Big\}$

$$\frac{d\mathcal{E}_{lms}}{da} = \mathcal{E}\{-2[(y-m_y)-a(x-m_x)](x-m_x)\} = -2[c_{xy} - a\sigma_x^2] = 0$$

Thus $a = \dfrac{c_{xy}}{\sigma_x^2}$; $b = m_y - \dfrac{c_{xy}}{\sigma_x^2}m_x$ and $\widehat{y} = \left(\dfrac{c_{xy}}{\sigma_x^2}\right)x + \left(m_y - \dfrac{c_{xy}}{\sigma_x^2}m_x\right)$

ALTERNATE FORM

$$\widehat{y} = \left(\frac{\sigma_y}{\sigma_x}\rho_{xy}\right)x + \left(m_y - \frac{\sigma_y}{\sigma_x}\rho_{xy}m_x\right) \quad \text{where} \quad \rho_{xy} = \frac{c_{xy}}{\sigma_x\sigma_y}$$

(checks with Example 6.7)

SIMPLE LMS ESTIMATION (cont'd.)

MINIMUM MEAN-SQUARE ERROR

$$
\begin{aligned}
\mathcal{E}_{lms} &= \mathcal{E}\left\{(y-\hat{y})^2\right\} = \mathcal{E}\left\{[(y-m_y)-a(x-m_x)]^2\right\} \\
&= \mathcal{E}\left\{(y-m_y)^2 - 2a(x-m_x)(y-m_y) + a^2(x-m_x)^2\right\} \\
&= \sigma_y^2 - 2ac_{xy} + a^2\sigma_x^2 = \sigma_y^2 - 2\frac{c_{xy}}{\sigma_x^2}c_{xy} + \left(\frac{c_{xy}}{\sigma_x^2}\right)^2\sigma_x^2 \\
&= \sigma_y^2 - \frac{c_{xy}^2}{\sigma_x^2}
\end{aligned}
$$

Substitute $c_{xy} = \sigma_x\sigma_y\rho_{xy}$ to obtain

$$
\mathcal{E}_{lms} = \sigma_y^2(1-\rho_{xy}^2) \le \sigma_y^2
$$

Insert Example 6.9 here.

GENERAL LMS ESTIMATION PROBLEM

- Random variable y, vector of observations $\boldsymbol{x}$, both may be complex.

- Estimate of the form: $\widehat{y} = \mathbf{a}^{*T}\boldsymbol{x} + b$ where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

- Choose $\mathbf{a}$, b to minimize $\mathcal{E}_{lms} = \mathcal{E}\left\{|y - \widehat{y}|^2\right\}$

GENERAL LMS ESTIMATION: STEPS TO SOLUTION

STEP 1

Define $\varepsilon = y - \widehat{y}$ and write

$$\mathcal{E}_{lms} = \mathcal{E}\left\{|\varepsilon|^2\right\} = \sigma_\varepsilon^2 + m_\varepsilon^2$$

Require:

$$m_\varepsilon = \mathcal{E}\{y - \widehat{y}\} = \mathcal{E}\left\{y - \mathbf{a}^{*T}\boldsymbol{x} - b\right\} = 0$$

thus . . .

$$b = m_y - \mathbf{a}^{*T}\mathbf{m}_x$$

GENERAL LMS ESTIMATION (cont'd.)

STEP 2

Substitute $b = m_y - \mathbf{a}^{*T}\mathbf{m}_x$ in $\mathcal{E}_{lms}$:

$$
\begin{aligned}
\mathcal{E}_{lms} &= \mathcal{E}\left\{|y-\hat{y}|^2\right\} = \mathcal{E}\left\{|y-\mathbf{a}^{*T}\boldsymbol{x}-b|^2\right\} \\
&= \mathcal{E}\left\{[(y-m_y)-\mathbf{a}^{*T}(\boldsymbol{x}-\mathbf{m}_x)][(y-m_y)^* - (\boldsymbol{x}-\mathbf{m}_x)^{*T}\mathbf{a}]\right\} \\
&= \mathcal{E}\left\{|y-m_y|^2 - (y-m_y)(\boldsymbol{x}-\mathbf{m}_x)^{*T}\mathbf{a}\right. \\
&\qquad \left. -\mathbf{a}^{*T}(\boldsymbol{x}-\mathbf{m}_x)(y-m_y)^* + \mathbf{a}^{*T}(\boldsymbol{x}-\mathbf{m}_x)(\boldsymbol{x}-\mathbf{m}_x)^{*T}\mathbf{a}\right\}
\end{aligned}
$$

or . . . $\quad \mathcal{E}_{lms} = \sigma_y^2 - \mathbf{c}_{xy}^{*T}\mathbf{a} - \mathbf{a}^{*T}\mathbf{c}_{xy} + \mathbf{a}^{*T}\mathbf{C}_x\mathbf{a}$

GENERAL LMS ESTIMATION (cont'd.)

STEP 3

Find stationary point and solve for $\mathbf{a}$:

$$\begin{aligned}
\mathcal{E}_{lms} &= \sigma_y^2 - \mathbf{c}_{xy}^{*T}\mathbf{a} - \mathbf{a}^{*T}\mathbf{c}_{xy} + \mathbf{a}^{*T}\mathbf{C}_x\mathbf{a} \\
\nabla_{\mathbf{a}^*}\mathcal{E}_{lms} &= \mathbf{0} - \mathbf{0} - \mathbf{c}_{xy} + \mathbf{C}_x\mathbf{a} = \mathbf{0}
\end{aligned}$$

Therefore . . . $\boxed{\mathbf{C}_x\mathbf{a} = \mathbf{c}_{xy} \quad \text{or} \quad \mathbf{a} = \mathbf{C}_x^{-1}\mathbf{c}_{xy}}$

GENERAL LMS ESTIMATION (cont'd.)

MINIMUM MEAN-SQUARE ERROR

$$\begin{aligned}\mathcal{E}_{lms} &= \sigma_y^2 - \mathbf{c}_{xy}^{*T}\mathbf{a} - \mathbf{a}^{*T}\mathbf{c}_{xy} + \mathbf{a}^{*T}\mathbf{C}_x\mathbf{a} \\ &= \sigma_y^2 - \mathbf{c}_{xy}^{*T}\mathbf{a} - \mathbf{a}^{*T}(\underbrace{\mathbf{c}_{xy} - \mathbf{C}_x\mathbf{a}}_{\mathbf{0}})\end{aligned}$$

$$\boxed{\mathcal{E}_{lms} = \sigma_y^2 - \mathbf{c}_{xy}^{*T}\mathbf{a} = \sigma_y^2 - \mathbf{c}_{xy}\mathbf{C}_x^{-1}\mathbf{c}_{xy}}$$

Insert Example 6.10 here.