

MSc and EEE/ISE PART IV: MEng and ACGI

Wednesday, 11 May 10:00 am

Time allowed: 3:00 hours

There are FIVE questions on this paper.

Answer FOUR questions.

All questions carry equal marks

Examiners responsible

First Marker(s) :	R.B. Vinter
Second Marker(s) :	S. Evangelou

Information for candidates:

The Multivariate Normal Density:

The probability density $N(m, Q)$ of an n -vector, normal random variable with mean m and covariance matrix Q ($Q > 0$) is

$$N(m, Q)(x) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det Q)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - m)^T Q^{-1} (x - m) \right\} .$$

In the case that $n = 1$, m is a scalar and $Q = \sigma^2$ ($\sigma^2 > 0$),

$$N(m, \sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x - m)^2}{2\sigma^2} \right)$$

and, if X is a scalar random variable with probability density $N(m, \sigma^2)$,

$$\text{Prob}\{m - 2\sigma \leq X \leq m + 2\sigma\} \approx 0.95 .$$

The Cramer-Rao Lower Bound:

Take a family of probability densities $\{p(\mathbf{y}; \theta)\}$ parameterised by the k -vector θ . Let $\hat{\theta}(\mathbf{y})$ be any unbiased estimate of θ given \mathbf{y} . Then the covariance of $\hat{\theta}(\mathbf{y})$ satisfies

$$\text{cov}\{\hat{\theta}(\mathbf{y})\} \geq M^{-1}(\theta)$$

where $M(\theta)$ is the $k \times k$ Fisher Information Matrix, with components $\{m_{ij}\}$ defined by:

$$m_{ij} = -E_{\theta} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log_e f(Y, \theta) \right\} .$$

Spectral Density Relations for Linear Systems :

Consider the stationary n -vector discrete time process y_t satisfying

$$y_t = G(z)e_t ,$$

in which $G(z)$ is the $n \times k$ matrix transfer function of a stable linear system and e_t is a k -vector white noise sequence with covariance Σ . Then y_t has spectral density

$$\Phi_y(\omega) = G(e^{j\omega}) \Sigma G^T(e^{-j\omega}) .$$

1. A zero mean stationary scalar process x_t satisfies the equation

$$x_t + ax_{t-1} = e_t,$$

in which e_t is white noise with variance σ_1^2 . a is a given parameter, $|a| < 1$. Noisy observations y_t are taken of x_t

$$y_t = x_t + v_t$$

in which v_t is a white noise process with variance σ_2^2 , independent of e_t .

- (i): Develop a difference equation for y_t involving the vector noise process

$$\mathbf{w}_t = \begin{bmatrix} e_t \\ v_t \end{bmatrix}$$

of the form

$$A(z)y_t = \mathbf{b}^T(z)\mathbf{w}_t. \quad [2]$$

- (ii): Compute the covariance function of y_t :

$$R_y(l), \quad \text{for } l = \dots, -1, 0, +1, \dots \quad \text{--- page 1} \quad [10]$$

- (iii): Using (i) and the spectral density formulae on page ~~2~~ or otherwise, show that the spectral density of y_t is

$$\Phi_y(\omega) = \frac{(1 + a^2 + 2a \cos \omega)\sigma_2^2 + \sigma_1^2}{(1 + a^2 + 2a \cos \omega)}. \quad [6]$$

- (iv): Show that y_t can be realized as an ARMA process

$$y_t + ay_{t-1} = w_t + cw_{t-1}$$

involving a *scalar* white noise w_t , in which c is a root of the function

$$\gamma(z) := \cancel{\sigma_2^2} + \sigma_2^2(1 + az^{-1})(1 + az). \quad [2]$$

σ_1^2

2. The output y_t of a stochastic linear discrete time system is related to the applied control u_t by the equation

$$y_t = u_{t-1} + e_t + 2e_{t-1} ,$$

in which e_t is white noise with unit variance. Notice the one-step time delay in the control term.

The control system is operated in steady state with the feedback control

$$u_t = -ky_t$$

for some constant gain parameter k , constrained to satisfy $|k| < 1$.

- (i): Derive the output variance for the system under closed loop operation

$$\sigma_y^2(k) := E[y_t^2] ,$$

for a general value of k .

[14]

- (ii): Determine the value k^* of k which minimizes the output variance $\sigma_y(k)$ and also the minimum value of the variance, $\sigma_y(k^*)$.

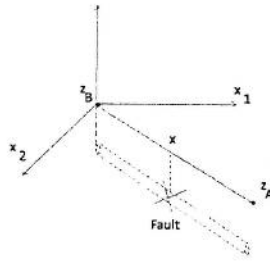
[4]

- (iii): Explain why it is not practical to choose the feedback gain

$$k = 2 ,$$

even though this value of the gain appears to give a lower output variance.

[2]



3. A fault has occurred in a straight water-pipe lying at constant depth under a flat field. A detector is used to measure the position \mathbf{x} , in 2D space, of a point in the field vertically above the fault. (See Figure.)

N independent measurements, $\mathbf{y}_1, \dots, \mathbf{y}_N$, of the position are made. The measurements are 2-vectors, modelled as

$$\mathbf{y}_t = \mathbf{x} + \mathbf{e}_t \quad \text{for } t = 1, \dots, N.$$

Here the \mathbf{e}_t 's are independent, Gaussian, zero-mean 2-vector random variables, each with covariance matrix $\sigma^2 I_{2 \times 2}$.

Assume that the position \mathbf{x} is modelled as

$$\mathbf{x} = \mathbf{z}_B + \frac{\theta}{\|\mathbf{z}_A - \mathbf{z}_B\|} (\mathbf{z}_A - \mathbf{z}_B)$$

where the 2-vectors \mathbf{z}_A and \mathbf{z}_B are two known locations in 2D space under which the pipe passes, and θ is an unknown parameter, giving the distance of the fault along the pipe from the point under \mathbf{z}_B .

- (i): Show that the linear least squares estimate $\hat{\mathbf{x}}$ is

$$\hat{\mathbf{x}} = \mathbf{z}_B + \left(\frac{\frac{1}{N} \sum_{t=1}^N (\mathbf{y}_t - \mathbf{z}_B)^T (\mathbf{z}_A - \mathbf{z}_B)}{\|\mathbf{z}_A - \mathbf{z}_B\|^2} \right) (\mathbf{z}_A - \mathbf{z}_B).$$

[14]

- (ii): Show that the estimate $\hat{\mathbf{x}}$ of \mathbf{x} is unbiased.

[2]

- (iii): Now suppose that the variance $\sigma^2 = 100m^2$. How many measurements need to be taken, in order that the fault is determined within $1m$ with 0.95 probability, i.e such that

$$P \{ \|\mathbf{x} - \hat{\mathbf{x}}\| \leq 1 \} \geq 0.95 ?$$

[4]

4. Consider a stationary, ergodic process y_t satisfying

$$y_t + ay_{t-1} = n_t \quad (1)$$

in which n_t is a 'coloured noise' process satisfying

$$n_t + dn_{t-1} = e_t. \quad (2)$$

Here, e_t is white noise with unit variance. In these equations, a and d are unknown parameters, satisfying $|a| < 1$, $|d| < 1$.

(i): Show that the covariance function of y_t satisfies

$$\frac{R_y(1)}{R_y(0)} = -\frac{(a+d)}{(1+ad)}. \quad [2]$$

Hint: This formula can be simply derived from (1) and (2), without the need to calculate $R_y(0)$ and $R_y(1)$ separately.

(ii): An estimate \hat{a} of the parameter a in (1) is obtained by the linear least squares method, without reference to the fact that the noise is coloured.

Show that the asymptotic bias in the estimate as $N \rightarrow \infty$ is

$$\hat{a} - a = \frac{d(1-a^2)}{1+ad}. \quad [10]$$

(iii): Describe in detail an algorithm for obtaining consistent estimates of both a and d . [8]

5. Measurements y_1, \dots, y_N are taken of a stationary, ergodic process described by the equation

$$y_t + ay_{t-1} = e_t ,$$

in which a is a constant ($|a| < 1$) and e_t is Gaussian white noise with variance σ^2 . a and σ^2 are to be regarded as unknown modelling parameters.

Derive the log likelihood function for a and σ^2 given the measurements y_1, \dots, y_N :

$$LLF(a, \sigma^2) = p(y_1, \dots, y_N | a, \sigma^2) .$$

(You may assume as starting value $y_0 = 0$.) [4]

Show that the Maximum Likelihood estimates \hat{a} and $\hat{\sigma}^2$ of a and σ^2 given (y_1, \dots, y_N) are

$$\hat{a} = -\hat{R}_y(1) / \hat{R}_y(0)$$

and

$$\hat{\sigma}^2 = \hat{R}_y(0) - \hat{R}_y^2(1) / \hat{R}_y(0) ,$$

where $\hat{R}_y(0)$ and $\hat{R}_y(1)$ are the sample covariances

$$\hat{R}_y(0) = \frac{1}{N} \sum_{t=1}^N y_t^2, \quad \hat{R}_y(1) = \frac{1}{N} \sum_{t=1}^N y_t y_{t-1} .$$

[10]

Show that the estimates \hat{a} and $\hat{\sigma}^2$ are consistent, i.e., with probability one,

$$\hat{a} \rightarrow a \quad \text{and} \quad \hat{\sigma}^2 \rightarrow \sigma^2 .$$

[6]

Identification Exam 2011. Model Answers

1. (i) Eliminating x_t yields: $y_t - v_t + a(y_{t-1} - v_{t-1}) = e_t$
 giving $y_t + a y_{t-1} = e_t + v_t + a v_{t-1}$ — (A)

(ii) $E\{(y_t + a y_{t-1})^2\} = E\{(e_t + v_t + a v_{t-1})^2\}$. So
 $(1+a^2)R_y(0) + 2a R_y(1) = \sigma_1^2 + (1+a^2)\sigma_2^2$ — (1)
 $E\{(y_t + a y_{t-1})y_{t-1}\} = E\{(e_t + v_t + a v_{t-1})y_{t-1}\}$
 gives $R_y(1) + a R_y(0) = 0 + 0 + a R_{vy}(0)$
 Also $E\{(y_t + a y_{t-1})v_{t-1}\} = E\{(e_t + v_t + a v_{t-1})v_{t-1}\}$
 gives $a R_{vy}(0) = a \sigma_2^2$. Hence
 $a R_y(0) + R_y(1) = a \sigma_2^2$ — (2)

(1) and (2) can be solved for $R_y(0)$ and $R_y(1)$:
 $R_y(0) = \frac{\sigma_1^2 + (1-a^2)\sigma_2^2}{(1-a^2)} = (1-a^2)^{-1} \sigma_1^2 + \sigma_2^2$, $R_y(1) = -\frac{a}{(1-a^2)} \sigma_1^2$

From (A), $E\{(y_t + a y_{t-1})y_{t-1}\} = R_y(1) + a R_y(0)$, $t \geq 2$
 Hence, $R_y(l) = \frac{(-a)^l}{(1-a^2)} \sigma_1^2$ for $l = \pm 2, \pm 3, \dots$

(iii) y_t can be described in terms of the vector white process $\begin{bmatrix} e_t \\ v_t \end{bmatrix}$:
 $(1 + a z^{-1}) y_t = \begin{bmatrix} 1 & 1 + a z^{-1} \end{bmatrix} \begin{bmatrix} e_t \\ v_t \end{bmatrix}$
 and so $y_t = G(z) \begin{bmatrix} e_t \\ v_t \end{bmatrix}$, where $G(z) = (1 + a z^{-1})^{-1} \begin{bmatrix} 1 & 1 + a z^{-1} \end{bmatrix}$

Then $\Phi_y(\omega) = \frac{1}{(1 + a z^{-1})(1 + a z)} \begin{bmatrix} 1 & 1 + a z^{-1} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 + a z \end{bmatrix} \Big|_{z=e^{j\omega}}$
 $\Phi_y(\omega) = \frac{\sigma_1^2 + \sigma_2^2 (1 + a z^{-1})(1 + a z)}{(1 + a z^{-1})(1 + a z)} \Big|_{z=e^{j\omega}}$ — (3)
 $= \frac{(1 + a^2 + 2a \cos \omega) \sigma_2^2 + \sigma_1^2}{1 + a^2 + 2a \cos \omega}$

(iv) (3) factorizes as constant $\times \frac{(1 + c z^{-1})(1 + c z)}{(1 + a z^{-1})(1 + a z)}$
 where c is a root of $\gamma(z) = \sigma_1^2 + \sigma_2^2 (1 + a z^{-1})(1 + a z)$.
 $(1 + a z^{-1}) y_t = (1 + c z^{-1}) w_t$ (w_t white) provides a realization by
 standard theory

2. Inserting $u_t = -ky_t$ into the system equations gives

$$y_t + ky_{t-1} = e_t - ze_{t-1}$$

Squaring both sides and taking expectations gives

$$(1+k^2)R_y(0) + 2kR_y(1) = 5 \quad \text{--- (1)}$$

$E\{\dots \times y_{t-1}\}$ gives

$$R_y(1) + kR_y(0) = 0 + 2R_{ye}(0)$$

But $E\{\dots \times e_t\}$ gives

$$R_{ye}(0) + 0 = 1 + 0$$

So $R_y(1) + kR_y(0) = 2$

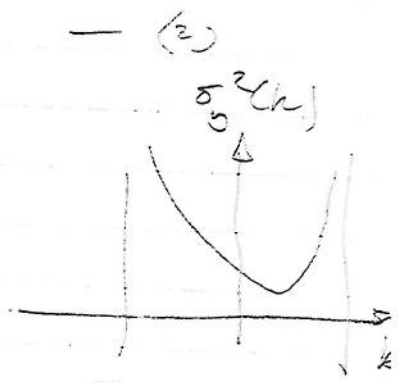
From (1) and (2)

$$(1+k^2)R_y(0) + 2k(2 - kR_y(0)) = 5$$

so $(1-k^2)R_y(0) = 5 - 4k$

and

$$\sigma_y^2(k) = E[y_t^2] = \frac{5-4k}{(1-k^2)}$$



$$\frac{\partial}{\partial k} = 0 \Rightarrow \frac{-4}{(1-k^2)^2} + \frac{(5-4k) \times 2k}{(1-k^2)^2} = 0$$

or $(5-4k) \times 2k = 4(1-k^2)$

or $4k^2 - 10k + 4 = 0$ or $k^2 - 2\frac{5}{4}k + 1 = 0$

The roots are $\frac{5}{4} \pm \sqrt{\frac{25}{16} - 1} = \frac{5}{4} \pm \frac{3}{4} = 2$ or $\frac{1}{2}$.

The values of $\sigma_k^2(k)$ are

$\sigma_k^2(k=2) = 1$ and $\sigma_k^2(k=\frac{1}{2}) = \frac{5-2}{3/4} = 4$

We cannot choose $k=2$, because this value violates the ^{"stability"} constraint on k . It is clear from the graph of $\sigma_y^2(k)$ however that $k=\frac{1}{2}$ is minimizing in the range $-1 < k < +1$.

If you attempted to implement $u_t = 2y_t$, the closed loop system would be

$$(1+2z^{-1})y_t = (1+z^{-1})e_t \Rightarrow y_t = z_t$$

which is stable. But even very small modelling errors for the transfer function $G(z) = (1+2z^{-1})$ would mean that the unstable pole $z = -2$ would not be cancelled, and the time τ to stabilize would be too long.

3. (i) The measurements are related to x according to:

$$\begin{bmatrix} y_1 - z_B \\ \vdots \\ y_N - z_B \end{bmatrix} = \frac{1}{\|z_A - z_B\|} \begin{bmatrix} z_A - z_B \\ \vdots \\ z_A - z_B \end{bmatrix} \theta + \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix} \quad \checkmark$$

or $y' = X\theta + e$ where $y' = \alpha^1 \{y_t - z_B\}$, $X = \frac{1}{\|z_A - z_B\|} \alpha^1 \{z_A - z_B\}$

The LLS estimate $\hat{\theta}$ of θ , given y is

$$\hat{\theta} = (X^T X)^{-1} X^T y'$$

in which

$$(X^T X) = \left[(z_A - z_B)^T \dots (z_A - z_B)^T \right] \begin{bmatrix} z_A - z_B \\ \vdots \\ z_A - z_B \end{bmatrix} \times \frac{1}{\|z_A - z_B\|^2} = 1$$

and

$$X^T y' = \left[(z_A - z_B)^T \dots (z_A - z_B)^T \right] \begin{bmatrix} y_1 - z_B \\ \vdots \\ y_N - z_B \end{bmatrix} \times \frac{1}{\|z_A - z_B\|} = \frac{1}{\|z_A - z_B\|} \sum_{t=1}^N (y_t - z_B) \frac{(z_A - z_B)^T}{\|z_A - z_B\|}$$

So

$$\hat{\theta} = \|z_A - z_B\|^2 \times \frac{1}{N} \sum_{t=1}^N (y_t - z_B)^T (z_A - z_B)$$

The estimated position is then

$$\hat{x} = z_B + \left(\frac{\frac{1}{N} \sum_{t=1}^N (y_t - z_B)^T (z_A - z_B)}{\|z_A - z_B\|^2} \right) (z_A - z_B)$$

$$(ii) \quad \hat{\theta} = (X^T X)^{-1} X^T y' = (X^T X)^{-1} X^T (X\theta + e) = \theta + (X^T X)^{-1} X^T e$$

$$\text{So } E\{\hat{\theta}\} = \theta + 0$$

$$\begin{aligned} \text{Then } E\{\hat{x}\} &= z_B + \frac{E[\hat{\theta}]}{\|z_A - z_B\|} (z_A - z_B) \\ &= z_B + \frac{\theta}{\|z_A - z_B\|} (z_A - z_B) = x \quad (\text{unbiased}) \end{aligned}$$

$$(iii) \quad \text{The error variance for } \hat{\theta} \text{ is } E\{\hat{\theta} - \theta\}^2 = (X^T X)^{-1} \sigma^2 = \frac{\sigma^2}{N}$$

$$\text{But } \sigma^2 = 100, \text{ So } \hat{\theta} - \theta \sim N(0, 100/N)$$

Since θ measures unit length along pipe, we require

$$\hat{\theta} - \theta = \pm 1 \text{ w.p. } 0.95, \text{ Satisfied if } \hat{\theta} = \frac{1}{2} \text{ or } \sqrt{\frac{100}{N}} = \frac{1}{2} \Rightarrow \underline{N = 400}$$

4(i) From $y_t + a y_{t-1} = n_t$ and $n_t + d n_{t-1} = e_t$
we deduce

$$(1 + a z^{-1})(1 + d z^{-1}) y_t = e_t \quad \text{or} \quad y_t + (a+d) y_{t-1} + a d y_{t-2} = e_t.$$

$$E\{\dots y_{t-1}\} \Rightarrow E\{y_t y_{t-1} + (a+d) y_{t-1} y_{t-1} + a d y_{t-1} y_{t-2}\} = E\{e_t y_{t-1}\}$$

$$\text{or } (1+ad) R_y(1) + (a+d) R_y(0) = 0$$

Hence

$$R_y(1) / R_y(0) = - \frac{a+d}{1+ad}$$

(ii) The LLSE of a based on $y_t + a y_{t-1} = \text{'noise'}$ however is

$$\hat{a} = - \frac{\sum_{t=1}^N y_t y_{t-1}}{\sum_{t=1}^N y_t^2} = - \frac{R_y(1)}{R_y(0)}$$

By ergodicity, as $N \rightarrow \infty$,

$$\hat{a} \rightarrow - \frac{R_y(1)}{R_y(0)} = \frac{a+d}{1+ad} = a + \frac{d(1-a^2)}{(1+ad)}$$

So the asymptotic bias is $\frac{d(1-a^2)}{1+ad}$

(iii) The Generalized Least Squares algorithm can give consistent estimates of a and d . This generates a sequence of "improved" estimates $\hat{a}_0, \hat{d}_1, \hat{a}_1, \hat{d}_2, \hat{a}_2, \dots$, given a starting value \hat{d}_0 .

Step 1 Filter data: $y_t^{(0)} = (1 + \hat{d}_0 z^{-1}) y_t$.

Obtain LLS estimate \hat{a}_0 from $(1 + \hat{a}_0 z^{-1}) y_t^{(0)} = \text{'error'}$

Step 2 Calculate residuals $e_t^{(0)} = (1 + \hat{a}_0 z^{-1}) y_t^{(0)}$

and obtain LLS estimate \hat{d}_1 from

$$(1 + \hat{d}_1 z^{-1}) e_t^{(0)} = \text{'error'}$$

Go back to step 1, replacing \hat{d}_0 by \hat{d}_1 and so on. Continue until estimates converge and $e_t^{(i)}$ satisfies a whiteness test. Algorithm is not guaranteed to converge.

5. If y_t satisfies $y_t + a y_{t-1} = e_t$, $t = 0, 1, \dots, N$ then

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ a & 1 & \dots & 0 \\ 0 & a & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$
 Write $A = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a & 1 & \dots & 0 \\ 0 & a & \dots & 1 \end{bmatrix}$ & $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$

So the log likelihood function of a and σ^2 given y_t is

$$\begin{aligned} LLF(a, \sigma^2) &= \log P(y_1, y_2, \dots, y_N | a, \sigma^2) \\ &= \log \left(\frac{1}{(2\pi)^{N/2}} \right) - \log(\sigma^N) - \frac{1}{2\sigma^2} y^T (A^T A)^{-1} y = \end{aligned}$$

$$= \text{constant} - \log(\sigma^N) - \frac{1}{2\sigma^2} y^T A^T A y$$

But $Ay = [y_1 + ay_0, y_2 + ay_1, \dots, y_N + ay_{N-1}]^T$ (where $y_0 = 0$)
 So $y^T A^T A y = \sum_{i=1}^N (y_i + a y_{i-1})^2 = N(R_y(0) + 2a \hat{R}_y(1) + a^2 \hat{R}_y(2))$

The maximum likelihood estimate \hat{a} of a is given by setting

$$\partial/\partial a \text{ LLF}(a, \sigma^2) = 0$$

$$\text{This gives } \hat{a} = -\hat{R}_y(1) / \hat{R}_y(0)$$

Setting $\partial/\partial \sigma \text{ LLF}(a, \sigma^2) = 0$ yields

$$-N/\hat{\sigma} + N/\hat{\sigma}^3 \|\hat{A}y\|^2 = 0 \text{ where } \hat{A} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \hat{a} & 1 & \dots & 0 \\ 0 & \hat{a} & \dots & 1 \end{bmatrix}$$

$$\text{or } \hat{\sigma}^2 = \frac{1}{N} \|\hat{A}y\|^2 = \frac{1}{N} \sum (y_t + \hat{a} y_{t-1})^2$$

$$= \hat{R}_y(0) + 2\hat{a} \hat{R}_y(1) + \hat{a}^2 \hat{R}_y(2)$$

$$= \hat{R}_y(0) - \frac{\hat{R}_y(1)^2}{\hat{R}_y(0)} \quad \checkmark$$

But, $E[(y_t + a y_{t-1})^2] = E[e_t^2] \Rightarrow \sigma^2 = (1 + a^2) R_y(0) + 2a R_y(1)$

and $E[(y_t + a y_{t-1}) y_{t-1}] = E[e_t y_{t-1}] = 0 \Rightarrow R_y(1) + a R_y(0) = 0$

Hence $a = -R_y(1)/R_y(0)$ and $\sigma^2 = \hat{R}_y(0) - (R_y(1)/R_y(0))^2 R_y(0)$

By ergodicity $\hat{R}_y(0) \rightarrow R_y(0)$, $\hat{R}_y(1) \rightarrow R_y(1)$ as $N \rightarrow \infty$

So $\hat{a} \rightarrow a$, $\hat{\sigma}^2 \rightarrow \sigma^2$ as $N \rightarrow \infty$ (consistency).