Covariance estimation for asynchronously observed diffusions

Second-order asymptotic expansion

Arnak Dalalyan and Nakahiro Yoshida

dalalyan@ccr.jussieu.fr

Université Paris 6, France

Outline of my talk

- Model description
- Quick review
- A lower bound
- Second-order expansion: *diffusions without drift*
- The case of Poisson sampling

Model description

Diffusions

Consider $\{X_t = (X_{1t}, X_{2t}), t \in [0,T]\}$ given by

$$\begin{cases} dX_{1t} = \beta_{1t}\, dt + \sigma_{1t}\, dB_{1t}, & t \in [0,T] \\ dX_{2t} = \beta_{2t}\, dt + \sigma_{2t}\, dB_{2t}, & t \in [0,T], \end{cases}$$

where $B = ((B_{1t}, B_{2t})^\top,\ t \geq 0)$ is a 2-dim. Gaussian process with independent increments, zero mean and

$$E[B_t \cdot B_t^\top] = \begin{pmatrix} t & \int_0^t \rho_s\, ds \\ \int_0^t \rho_s\, ds & t \end{pmatrix},\ \forall t \geq 0.$$

Model description

Coefficients

- σ_1, σ_2 and ρ are either deterministic or independent of B. They are unknown.
- ρ takes values in the interval $[-1, 1]$.
- β_{it}, ($i = 1, 2$) are unknown progressively measurable processes.

We are interested in the estimation of

$$\boxed{\theta = \mathrm{Cov}(X_{1,T}, X_{2,T}) = \int_0^T \sigma_{1t}\sigma_{2t}\rho_t \, dt.}$$

Model description

Observations

We have at our disposal the data

$$(X_{1T_{11}}, \ldots, X_{1T_{1N_1}}) \text{ and } (X_{2T_{21}}, \ldots, X_{2T_{2N_2}}).$$

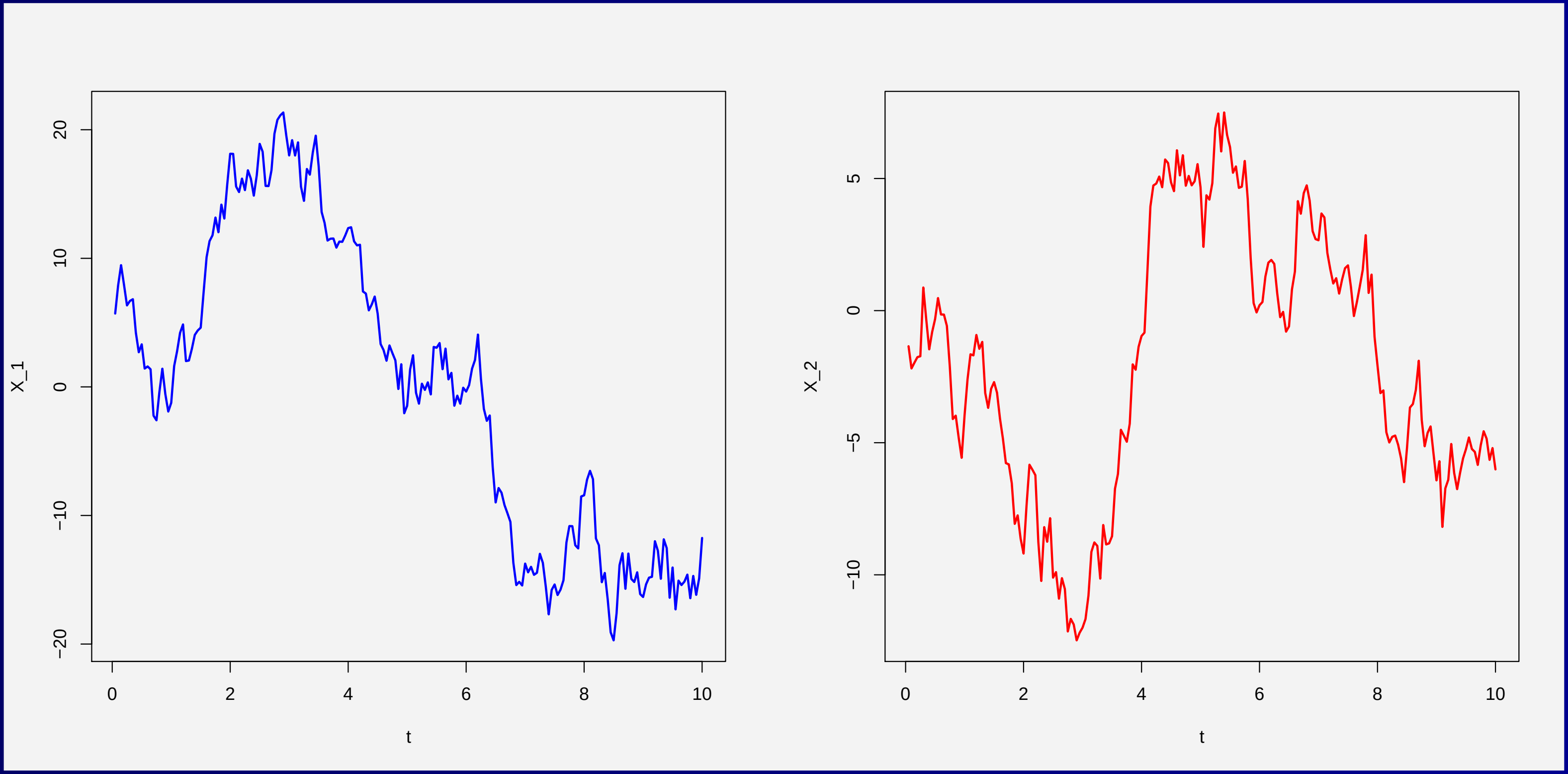
The time instants are random but $(T_{1i}, i \leq N_1)$ is independent of $(T_{2i}, i \leq N_2)$.

- Case 1: deterministic sampling (T_{ik} are not random),
- Case 2: Poisson sampling ($\{T_{1i}\}_i$ and $\{T_{2i}\}_i$ are independent Poisson point processes).

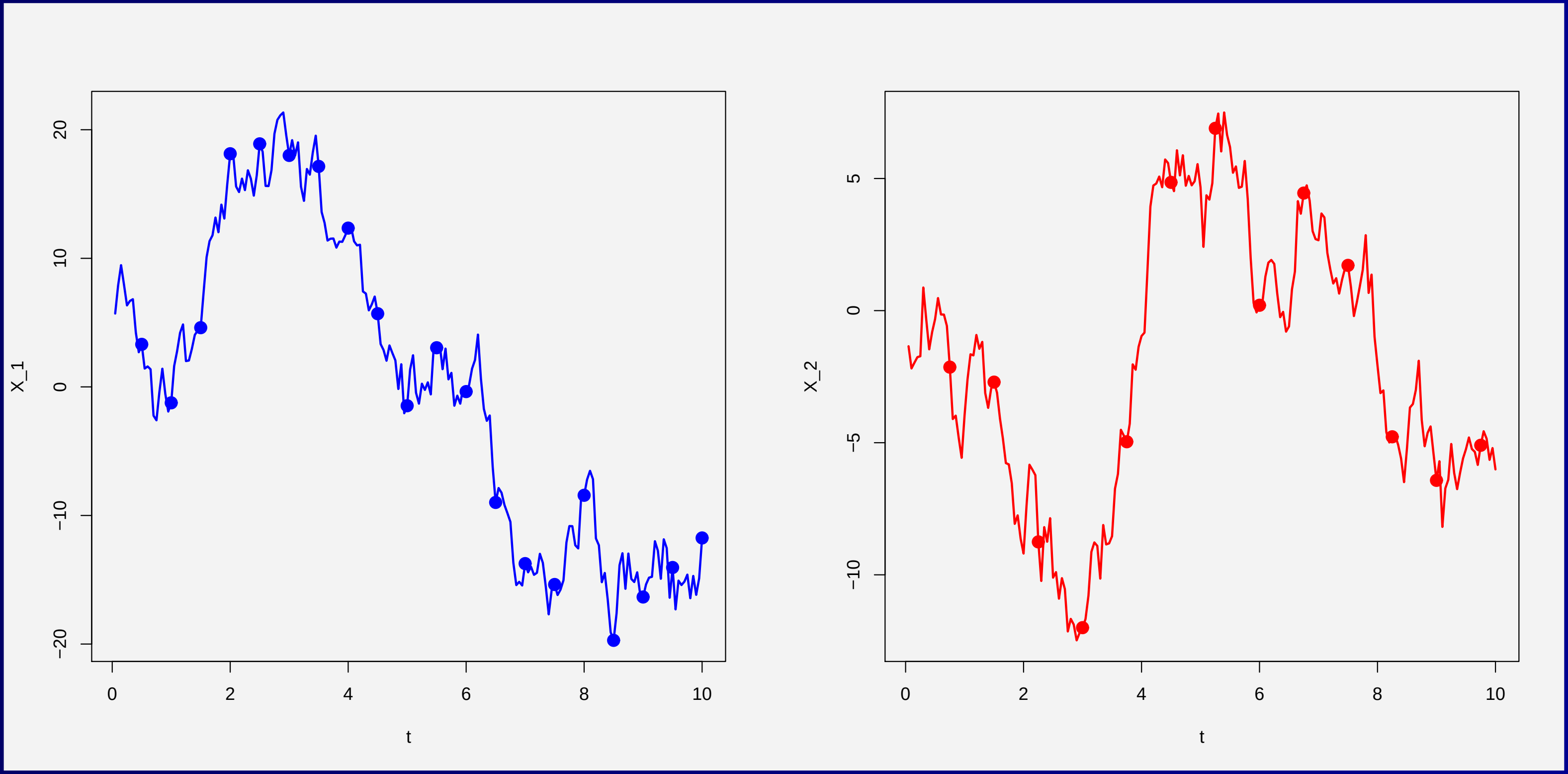
Model description

Our goal is to study the asymptotic properties of the estimators when the step of discretization tends to zero. To this end, we assume that the sampling processes $\mathcal{T}_i = \{T_{ik}\}_k$, $i = 1, 2$, depend on some parameter n and $r_n = \max_{k,k'} |T_{1(k+1)} - T_{1k}| \vee |T_{2(k'+1)} - T_{2k'}|$ tends to zero in probability as $n \to \infty$.

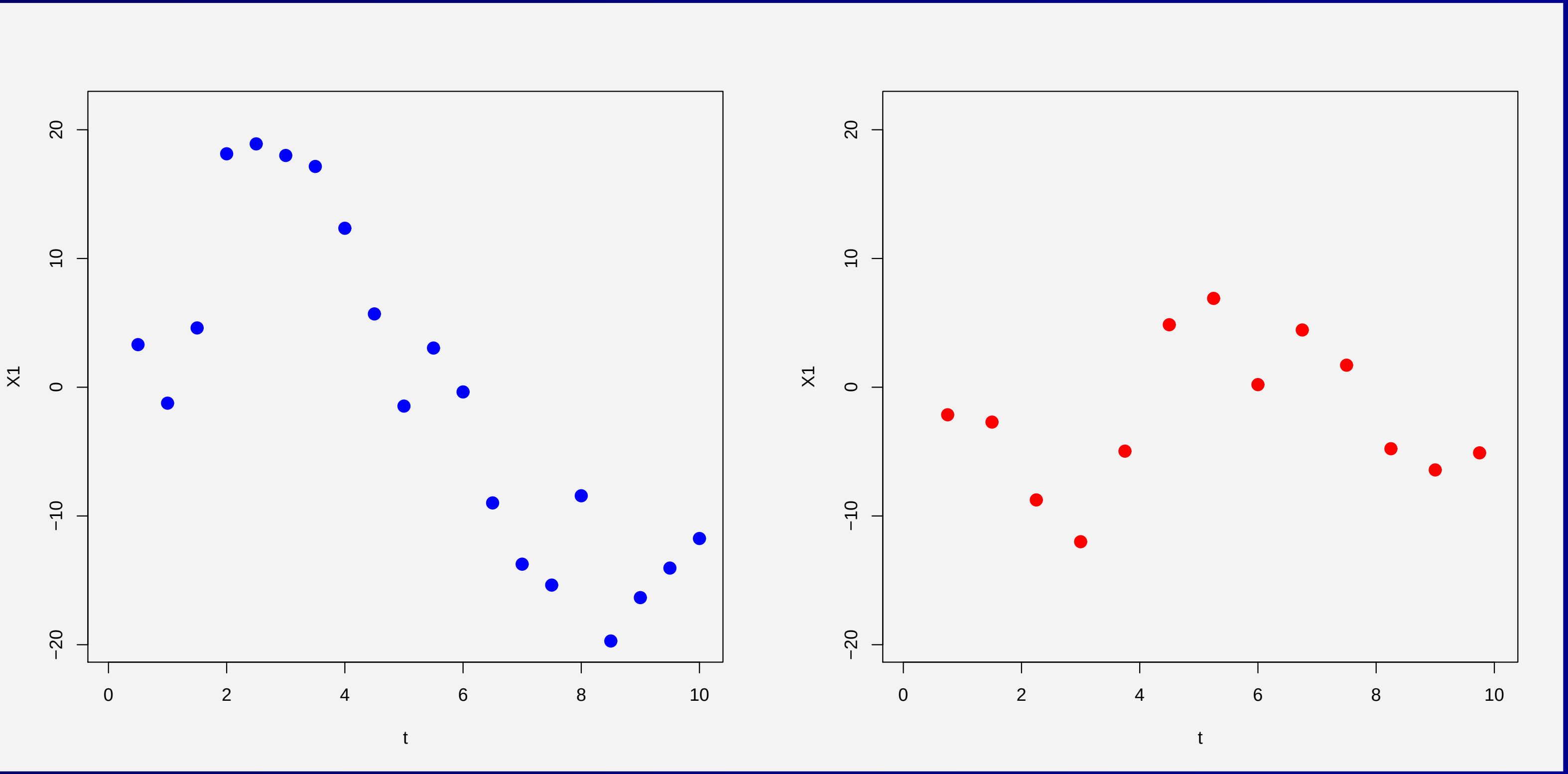
Two diffusions



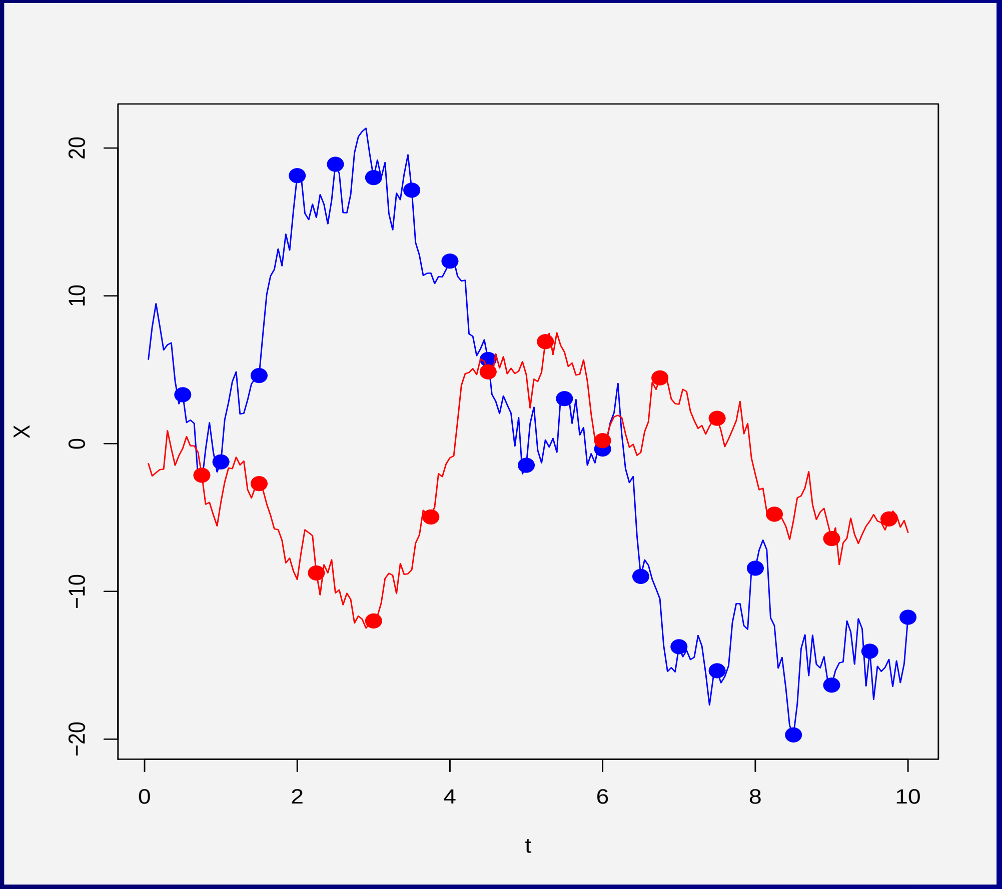
Two diffusions and the observed values



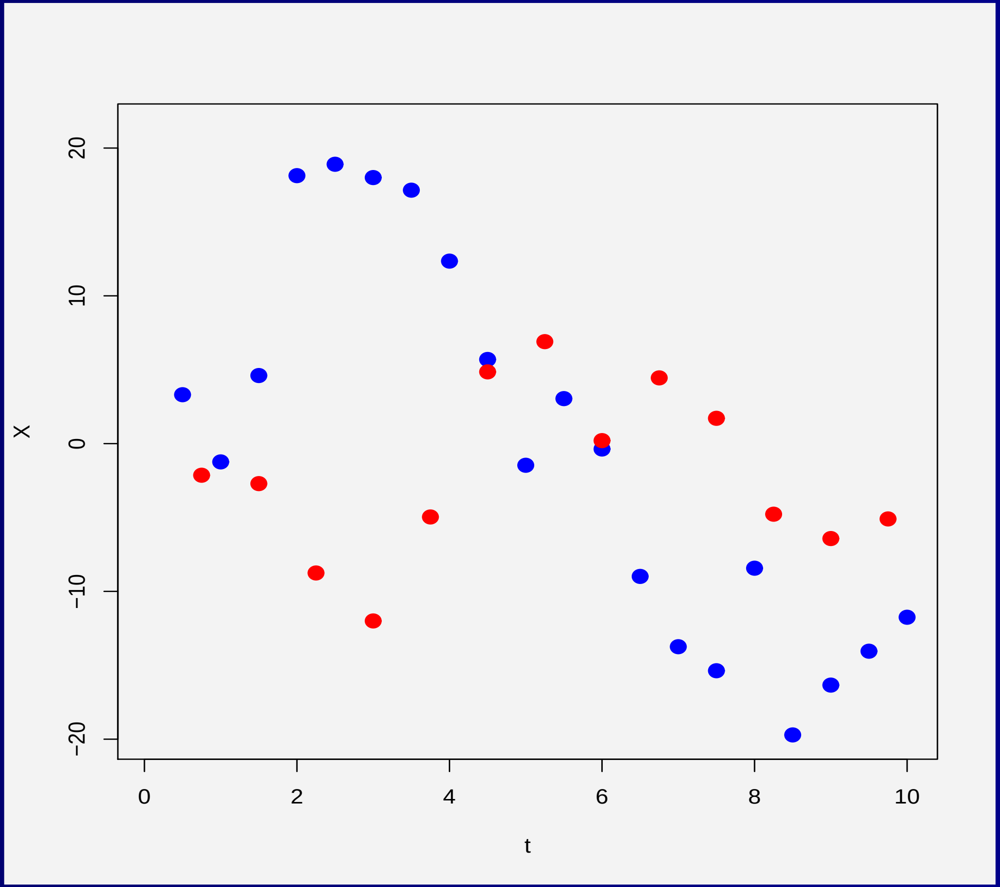
Observed data



Observed data



Observed data



Short review

- The synchronous case is well studied, e.g. *Genon-Catalot and Jacod ('93)*
- A consistent estimator is proposed by *Hayashi and Yoshida ('05)*
- Asymptotic normality of that procedure is proved by *Hayashi and Yoshida ('07)*
- Estimation of the asymptotic variance, *Mykland ('06)*
- ...

Estimator and a Cramér-Rao bound

Assume that there is no drift, then

- the data $\boldsymbol{\mathcal{X}}$ is a Gaussian vector with zero mean and covariance $\boldsymbol{\Sigma}$,
- $\boldsymbol{\theta}$ is a linear functional of $\boldsymbol{\Sigma}$: $\boldsymbol{\theta} = \boldsymbol{A}(\boldsymbol{\Sigma})$,
- the MLE of $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\mathcal{X}} \cdot \boldsymbol{\mathcal{X}}^\top$,
- the MLE of $\boldsymbol{\theta}$ is then $\hat{\boldsymbol{\theta}} = \boldsymbol{A}(\widehat{\boldsymbol{\Sigma}})$,
- the Cramér-Rao inequality holds: for every unbiased estimator $\bar{\boldsymbol{\theta}}$,

$$\boxed{\boldsymbol{E}[(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})^2] \geq \boldsymbol{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2].}$$

Estimator and a Cramér-Rao bound

Assume that there is no drift, then

- the data $\mathcal{X}$ is $\mathcal{N}_{N_1+N_2}(0, \Sigma)$,
- θ is a linear functional of Σ: $\theta = A(\Sigma)$,
- the MLE of θ is then

$$\boxed{\hat{\theta} = \sum_{i,j} (\Delta_i X_1)(\Delta_j X_2) K_{ij},}$$

  where $K_{i,j} = 1_{\{[T_{1i}, T_{1(i+1)}] \cap [T_{2j}, T_{2(j+1)}] \neq \emptyset\}}$,

- $\inf_{\bar{\theta} \text{ unbiased}} E[(\bar{\theta} - \theta)^2] = E[(\hat{\theta} - \theta)^2]$.

Second-order expansion: diffusion without drift

Set $r_n = \max_i |T_{1i} - T_{1(i+1)}| \vee \max_j |T_{2j} - T_{2(j+1)}|$

Theorem 1 (Hayashi and Yoshida).

(a) *If* σ_1, σ_2 *are bounded and* $r_n \xrightarrow{P} 0$, *then* $\hat{\theta}_n$ *is consistent.*

(b) *If, moreover, for some deterministic sequence* b_n, $b_n^{-1}\mathrm{Var}(\hat{\theta}|\mathcal{T}) \xrightarrow{P} \mathfrak{s}^2$ *and* $r_n^2/b_n \xrightarrow{P} 0$, *then*

$$b_n^{-1/2}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0, \mathfrak{s}^2).$$

Second-order expansion: diffusion without drift

Theorem 2. *If, in addition to the assumptions of Theorem 1, we have*

$$E[r_n^2] + |\mathrm{Var}[\hat{\theta}] - b_n \mathfrak{s}^2| = o(b_n^{1+\alpha})$$

with some $\alpha > 0$*, then for every bounded measurable function* f

$$E\Big[f\Big(\frac{\{\hat{\theta}_n - \theta\}}{\sqrt{b_n}}\Big) - f(\zeta)\Big] \sim \frac{\mu_n}{\sqrt{b_n}} \int_{\mathrm{R}} f\, \phi^{(3)}_{\mathfrak{s}^2}.$$

Poisson sampling

Consider the case when $\mathcal{T}_1$ and $\mathcal{T}_2$ are induced by two independent Poisson point processes with intensities np_1 and np_2.

Proposition 1. *If the functions* $\sigma_1,\ \sigma_2$ *and* ρ *are Lipschitz continuous, then Theorem 2 holds true with* $b_n = n^{-1}$*,*

$$\mathfrak{s}^2 = \left(\tfrac{2}{p_1} + \tfrac{2}{p_2}\right)\|\sigma_1\sigma_2\sqrt{1+\rho^2}\|_2^2 - \tfrac{2}{p_1+p_2}\|h_2\|_2^2.$$

$$\mu_n \sim \left(\tfrac{2}{p_1^2} + \tfrac{2}{p_2^2}\right)\|h_3\|_2^2 + \tfrac{(6p_1^2+4p_1p_2+6p_2^2)}{p_1^2p_2^2}\|h_3/\rho\|_2^2,$$

where $h_k = (\sigma_1\sigma_2\rho)^{k/2}$*.*

Concluding remarks

- The boundedness of f can be relaxed.
- In the case where the drifts are not zero, similar results are proved under the assumption that the Malliavin derivatives of the drifts exist and satisfy suitable boundedness assumptions
- These results, combined with the estimator of the variance proposed by Mykland, may be applied to assess the accuracy of a confidence interval for θ