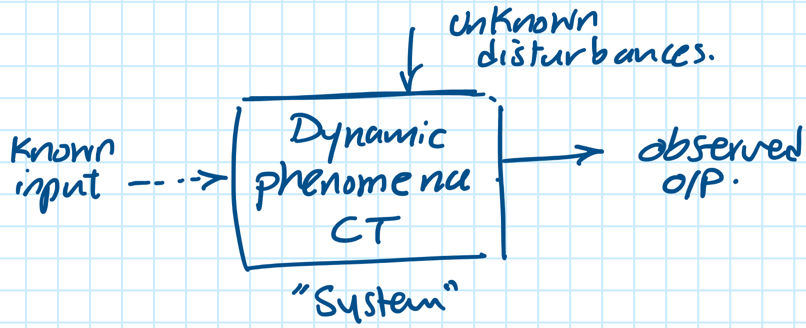
# Parameter estimation for stochastic CT dynamical models

Mohamed Abdalmoaty




One framework of modelling continuous-time dynamical systems adapts SDEs.
To facilitate mathematical analysis, Itô calculus is usually used.

Itô SDEs:

$$dX(t) = f(X(t), u(t); \theta)\,dt + \sigma(X(t);\theta)\,dW(t)$$
$$dY(t) = g(X(t), u(t); \theta)\,dt + \lambda(X(t);\theta)\,dV(t)$$

- $\theta \in \mathbb{R}^d$ unknown parameter. $\subset \Theta$ compact
- $W, V$ are Wiener processes.
- $f, g, \sigma$ and $\lambda$ are functions with appropriate dimensions that satisfy the required regularity condition for existence and uniqueness of solutions.
- $X$ latent state, $u$ known signal
  $Y$ observed (measured) signal.

## Possible setups

- $g =$ identity function, $\lambda = 0$.

then $Y = X$
and we observe the state directly.

- $g \neq$ identity $\Rightarrow$ partial observation.

- in many applications:
  the measurement process is modelled in discrete-time; i.e.

  $Y$ is a discrete-time process

$$y(t_K) = g(x(t_K), u(t_K); \theta) + v_K$$

$v_K$ is discrete-time white noise.

Alternatively:

$$y(t_K) \sim p(y(t_K) \mid x(t_K); \theta)$$

for some pdf $p$.

## Experiment design

- An integral part of the estimation procedure.
- involves designing $u(t)$ (if any), and choosing sampling times (if possible) among other things.

Not discussed in this lecture!

## Problem.

Given a parameterized SDE, and a data set, the goal is to construct an estimator $\hat{\theta}$ of $\theta$.

$$\text{Data set} \longmapsto \hat{\theta} \in \Theta$$

thus, we assume $f$, $\sigma$ etc are fixed to some known form, and parameterized by $\theta$.

## ML based on continuous record likelihood

in the discrete-time case, the ML estimator is straightforward to define. The likelihood function based on a sequence $y_1, \dots, y_N$ is interpreted as follows.

Given a value $\theta$, the RVs $y_1, \dots, y_N$ induces a measure $\mathbb{P}_Y$ on $\mathbb{R}^N$, which is absolutely continuous w.r.t the Lebesgue measure $dY$ on $\mathbb{R}^N$, and the likelihood function is defined to be the corresponding Radon-Nikodym derivative. $\frac{d\mathbb{P}_Y}{dY}(y;\theta)$ seen as a function of $\theta$.

### Extension to CT

Possible for models on the form.

$$dX(t) = f(X(t);\theta)\,dt + \sigma(X(t))\,dW(t)$$

$$Y = X.$$

For simplicity, assume a scalar case.

$\mathbb{C} = C[0,T]$ space of continuous functions $[0,T] \to \mathbb{R}$

Let $\mathbb{P}_Y$ be the measure induced by the observation process $X(t,\omega)$ on $0 \leq t \leq T$

A likelihood functional can be defined if we find a fixed measure on $\mathbb{C}$ such that $\mathbb{P}_Y$ is absolutely continuous w.r.t it.

This can always be done when $\sigma$ is known.

$$L_T(\theta) = \exp\left( \int_0^T \frac{f(X(s);\theta)}{\sigma^2(X(s))} dX(s) - \frac{1}{2}\int_0^T \frac{f^2(X(s);\theta)}{\sigma^2(X(s))} ds \right)$$

Motivation (Ornstein-Uhlenbeck process).

$$dX(t) = -\alpha X(t)dt + \sigma dW(t),$$

$\alpha > 0$ unknown, $\sigma > 0$ known.

Euler-Maruyama: $\Delta$ small

Denote $X_t = X(t_k)$, $X_{t+1} = X(t_{k+1})$, $t_k = k\Delta$

$$X_{t+1} = (1-\alpha\Delta)\ X_t + \sigma\sqrt{\Delta}\ Z_k, \quad Z_k \sim N(0,1)$$

Take a reference process:

$$dX(t) = \sigma dW(t)$$

then $X_{t+1} = X_t + \sigma\sqrt{\Delta}\ Z_k$

$$P_1^\theta(X_{[1:T]}) = (2\pi\sigma^2\Delta)^{-T/2} \exp\left[\frac{-1}{2\sigma^2\Delta}\sum(X_{t+1} - (1-\alpha\Delta)X_t)^2\right]$$

$$P_2^\theta(X_{[1:T]}) = (2\pi\sigma^2\Delta)^{-T/2} \exp\left[\frac{-1}{2\sigma^2\Delta}\sum(X_{t+1} - X_t)^2\right]$$

Define.

$$L_T(\alpha) = \frac{P_1^\theta}{P_2^\theta} = \exp\left(\frac{-1}{2\sigma^2\Delta}\sum (X_{t+1}-(1-\alpha\Delta)\ X_t)^2 - (X_{t+1}-X_t)^2\right)$$

$\vdots$

$\vdots$

$$= \exp\Bigg(\underbrace{\frac{-\alpha}{\sigma^2}\sum_t x_t(x_{t+1}-x_t)}\Bigg)\exp\Bigg(\underbrace{\frac{-\alpha^2}{2\sigma^2}\sum x_t^2}\Bigg)$$

$$\to \quad \frac{-\alpha}{\sigma^2}\int_0^T X(s)\,dX(s) \qquad \to -\frac{\alpha^2}{2\sigma^2}\int_0^T X^2(s)\,ds$$

as $\Delta \to 0$.

ML estimator of $\alpha$.

It's integral.

$$\max_\alpha \; L_T(\alpha).$$

$$\Rightarrow \quad \hat{\alpha}_T = \frac{\frac{1}{T}\int_0^T X(s)\,dX(s)}{\frac{1}{T}\int_0^T X^2(s)\,ds}$$

easy to show $\quad \hat{\alpha}_T \longrightarrow$ true $\alpha$ as $T\to\infty$

In practice, a continuous record of $X(t)$ is not available & $\hat{\alpha}_T$ is not feasible. Yet, with sufficiently small $\Delta$, we may use <u>discrete-data</u> to approximate $\hat{\alpha}_T$ by approximating the likelihood function or directly its closed form, if available.

## Remark on asymptotics.

- One is usually interested in consistent estimators.

Fixed $\Delta$ scheme: $\quad \Delta_n = \Delta > 0$

fixed sampling time.

process is observed on $[0, T = N\Delta]$

- asymptotics in $N \to \infty$.

Variable $\Delta$ scheme $\Delta_N \to 0$ as $N \to \infty$

process is observed on $[0, T = N\Delta]$

Two alternatives

↳ $T = N\Delta$ is fixed

↳ $T = N\Delta \to \infty$

---

what to do if the continuous-record likelihood function isn't available?

Likelihood for discrete-data

Due to the Markov properties of the solution of Itō SDEs, we can write the likelihood function for sampled data as.

$$P(X(t_1), \ldots, X(t_N)) = \prod_{n=0}^{N-1} P(X(t_{n+1}) | X(t_n); \theta)$$

- we assumed time-homogeneous case to simplify notations.

- $P(X(t_{n+1}) | X(t_n); \theta)$ is the transition density of the SDE.

↳ Solutions of FPK equations.
usually not available analytically!

$\Rightarrow$ unknown form for the likelihood and approximations have to be used.

Approximate.

The model itself | The transition density. | The likelihood function.

## The Euler approximation and the like

rewrite the model as:

$$X_{t+1} = X_t + f(X_t;\theta)\Delta + \sigma(X(t);\theta)\sqrt{\Delta}\,\varepsilon_t$$
$$\varepsilon_t \overset{iid}{\sim} N(0,1)$$

then $X_{t+1} \mid X_t \sim N\left(X_t + f(X_t;\theta)\Delta,\ \sigma^2(X_t;\theta)\Delta\right)$

$\equiv$ transition density.

when $\Delta$ is small, this should give a good accuracy.

Generally, we will get a bias; called discretization bias.

- instead of Euler-Maruyama, one can use Milstein's scheme or any other discretization of the model.

## Simulated likelihood methods.

Suppose $M-1$ auxiliary points are introduced between $(n-1)\Delta$ and $n\Delta$

$$(n-1)\Delta \equiv \tau_0, \tau_1, \ldots, \tau_{M-1}, \tau_M \equiv n\Delta$$

By Chapman-Kolmogorov equation, and the Markov property

$$P(X_t | X_{t-1}; \theta) = \int \cdots \int \prod_{m=1}^{M} P(X_{\tau_m} | X_{\tau_{m-1}}; \theta)\, dX_{\tau_1} \cdots dX_{\tau_{m-1}}$$

The idea is to approximate $P(X_{\tau_m} | X_{\tau_{m-1}}; \theta)$ then evaluate the multidimensional integral using importance sampling.

when $M$ is large enough, $P(X_{\tau_m} | X_{\tau_{m-1}}; \theta)$ can be approximated well using Euler discretization of the model.

So, a simple option is to draw $X_{\tau_1} \ldots X_{\tau_{M-1}}$ via simulations from the Euler scheme.

(This ignores the end point $X_{\tau_M}$ during sampling $\rightarrow$ only forward simulation).

Same as before, one may use other model discretizations.

Many other methods are available for approximating transition densities / likelihood function.

For example, closed-form analytic approx.

## Partially observed SDEs.

$$dX(t) = f(X(t), u(t); \theta)\,dt + \sigma(X(t), \theta)\,dW(t)$$

$$y_n \sim P(y_n \mid X(t_n); \theta).$$

· Here, the problem is closely related to parameter estimation in discrete-time state-space models.

The common approach there is to work with an appropriate discretization of the state equation.

Then apply the Expectation-Maximization algorithm + Approximation of the posterior density of X (Particle Smoother, or analytical approximation).

## Input/output approaches.

When the state $X(t)$ is observed indirectly via $y(t)$, the parameter estimation problem becomes entangled with state smoothing if a likelihood based method is to be used.

Another alternative is to look at the problem using an input output approach.

In what follows, we will only consider the linear case. To simplify further, we limit the discussion to the LTI scenario.

$$\begin{cases} dX(t) = A(\theta)X(t) + B_u(\theta)u(t) + B_w(\theta)dW(t) \\ \bar{y}(t) = C(\theta)X(t) \end{cases}$$

with measurement noise in discrete-time.

i.e. the measured signal

$$y(t_k) = \bar{y}(t_k) + v_k, \qquad v_k \text{ iid zero mean var. } \sigma_v^2$$

we will consider two frameworks

1- Noise is modelled in discrete-time.
requires uniform (equidistance) sampling $\Delta_n = \Delta \;\; \forall n$

2- Noise in CT
can deal with irregular sampling.

## Why an input-output approach is possible?

Regardless of the used model the output process in continuous time (after removing the mean) has a rational spectrum.

$$\Phi_c(\omega) = \lambda \left| \frac{B(i\omega)}{A(i\omega)} \right|^2$$

$A(s)$, $B(s)$, $s \in \mathbb{C}$ are stable polynomials of orders $n$ & $m$ respectively.

$n \geq m$.

$$A(s) = \prod_{i=1}^{n} (s - p_i) \qquad p_i, z_i \in LHP(\mathbb{C})$$

$$B(s) = \prod_{i=1}^{m} (s - z_i)$$

Thus, we model the spectrum!

if $u(t) = 0$ and $B(s) = 1$, we get

CAR process, continuous-time auto-regressive.

Correlation: $r(\tau) = \mathbb{E}[y(t+\tau)y(t)]$

$$\Phi_c(\omega) = \int_{-\infty}^{\infty} r(\tau) e^{-s\tau} d\tau$$

$$r(\tau) = \frac{1}{2\pi i} \int_{s=i\omega} \phi_c(s) e^{s\tau} ds$$

Then.

$$A(p)\, y(t) = B_e(p)\, e_c(t)$$

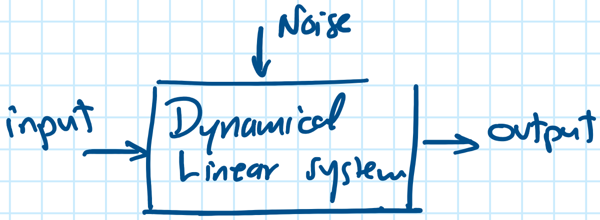
$$\mathbb{E}[e_c(t) e_c(s)] = \sigma_e^2\, \delta(t-s)$$

$p = d/dt$

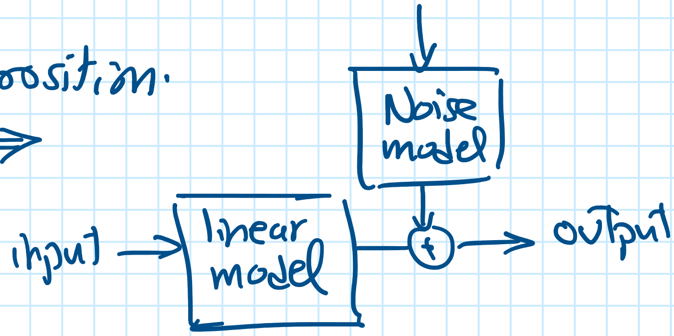
State-space form.

$$dx(t) = Ax(t)dt + B\,dw(t)$$
$$y(t) = Cx(t).$$

## Noise model




superposition. $\Longrightarrow$




we can discretize all parts of the model or only the noise model.

A very common setup in engineering:

$$z(t) = G(p)\,u(t) \qquad \text{CT dynamic model}$$
$$v_k = H(q^{-1})\,e_k \qquad e_k \text{ DT noise}$$
$$y(t_k) = x(t_k) + v_k \qquad \begin{cases} q \text{ shift operator} \\ q^{-1} e_k = e_{k-1} \end{cases}$$

So, we get

$$y(t_k) = G(p)\,u(t_k) + H(q^{-1})\,v_k$$

where we informly mixed the operators here.

Note $\quad dx(t) = A(\theta)x(t)\,dt + B_u(\theta)\,u(t)\,dt + B_w(\theta)\,dw(t)$

$$x(t) = e^{A(t-t_0)}x(t_0) + \underbrace{\int_{t_0}^{t} e^{A(t-s)} B_u\, u(s)\,ds}_{\text{due to } u} + \underbrace{\int_{t_0}^{t} e^{A(t-s)} B_w\, dW(s)}_{\tilde{w}(t) :=}$$

$$\dot{\tilde{x}}(t) = A(\theta)\tilde{x}(t) + B_u(\theta)\,u(t)$$
$$x(t) = \tilde{x}(t) + \tilde{w}(t)$$

so we can always move the noise to the o/p.

Issues in CT identification.

$$A(p)\,y(t) = B(p)\,u(t) + e_c(t).$$

Assume equidistance/uniform sampling.

At any time $t_k$, one can assume a model

$$y^{(n)}(t_k) = \varphi^T(t_k)\,\theta + v(t_k)$$

$$\varphi^T(t_k) = [-y^{(n-1)}(t_k) \;\dots\; -y(t_k) \;\; u^{(m)}(t_k) \dots u(t_k)]$$

$$\theta^T = [a_1 \cdots a_n \;\; b_0 \cdots b_m]$$

we used notation $y^{(k)}(t) = p^k y(t)$ , $p = d/dt$

Similarly for $u^{(k)}(t)$.

Regressors contain input and output time derivatives. Not available in data!
what is more, they are not well defined.

## The State-variable filter SVF

This issue is solved by the so-called State-variable filter (SVF)

$$F(p) = \frac{1}{(p+\eta)^n}$$

$\eta$ is a parameter controlling the B.W. of F.

Then,

$$A(p)\,F(p)\,y(t) = B(p)\,F(p)\,u(t) + F(p)\,v(t)$$

$$\left(\frac{p^n}{(p+\eta)^n} + a_1\frac{p^{n-1}}{(p+\eta)^n} + \cdots + a_n\frac{1}{(p+\eta)^n}\right) y(t)$$

$$= \left(b_0\frac{p^m}{(p+\eta)^n} + \cdots + b_m\frac{1}{(p+\eta)^n}\right) u(t)$$

$$+ F(p)\,v(t).$$

Define $F_i(p) = \frac{p^i}{(p+\eta)^n}$, $i = 0, \dots, n$

Then define $y_f^{(i)}(t) = F_i(p)\, y(t)$.

similarly for $u_f^{(i)}(t)$

then,

$$y_f^{(n)}(t_k) = \varphi_f^T(t_k)\,\theta + \gamma(t_k)$$

$$\varphi_f^T(t_k) = \left[-y_f^{(n-1)}(t_k) \;\dots\; -y_f^{(0)}(t_k)\;\; u_f^{(m)}(t_k) \cdots u_f^{(0)}(t)\right]$$

Then. $$\hat{\theta}^n_{LS-SVF} = \left[\frac{1}{N}\sum \varphi_f(t_k)\,\varphi_f^T(t_k)\right]^{-1} \frac{1}{N}\sum \varphi_f^T(t_k)\, y_f^{(n)}(t_k)$$

But then $\hat{\theta}_{LS-SVF}$ is <u>not consistent</u>,

Due to filtering, $\gamma(t_k)$ is not white and correlated with the regressor $\varphi_f(t_k)$.

$$\hat{\theta}_{LS-SVF} - \theta^* = \left[\frac{1}{N}\sum \varphi_f(t_k)\,\varphi_f(t_k)\right]^{-1}\left[\frac{1}{N}\sum \varphi_f(t_k)\,\gamma(t_k)\right]$$

$$\longrightarrow \beta \neq 0$$

## Simplified refined instrumental variable method

To solve the consistency problem, we may use an instrumental variable. This means that we change the definition.

of the LS-SVF to

$$\hat{\theta} = \left[\frac{1}{N}\sum \hat{\varphi}_f(t_k)\,\varphi_f^T(t_k)\right]^{-1}\frac{1}{N}\sum \hat{\varphi}_f^T(t_k)\, y_f^{(n)}(t_k)$$

IV regressors

$$\hat{\varphi}_f(t_k) = \left[-\hat{y}_f^{(n-1)}(t_k)\ \cdots\ -\hat{y}_f^{(0)}(t_k)\ \ u_f^{(m)}(t_k)\ \cdots\ u_f^{(0)}(t_k)\right]$$

$$\hat{y}_f^{(i)}(t_k) = F_i(P)\,\frac{B(P, \hat{\theta}_{LS\text{-}SVF})}{A(P, \hat{\theta}_{LS\text{-}SVF})}\,u(t_k).$$

- This is called an instrumental variable!

## Better choice for F(p)

if the true noise model $H(q;\theta) = 1$.
Then the optimal prediction errors.

$$\varepsilon(t_k) = y(t_k) - \frac{B(P)}{A(P)}u(t_k)$$

$$= A(P)\underbrace{\frac{1}{A(P)}}_{F(P)}y(t_k) - B(P)\underbrace{\frac{1}{A(P)}}_{F(P)}u(t_k).$$

But this "optimal" filter is unknown.
Also, $\varepsilon(t_k)$ may not be white.

The following SRIVC algorithm is used:

- instead of $A(P)$ use $A(P;\theta_j)$
  where $\theta_j$ is the estimate at hand.

- instead of $A(p)$ use $A(p;\theta_j)$ where $\theta_j$ is the estimate at hand.
- iterate until max no. of iterations reach or convergence criterion met.
- initialize with LS-SVF.

## Estimating a Noise model in discrete-time

Assume, at hand, $\theta_j$ with the associated model

$$G_j(p) = \frac{B_j(p)}{A_j(p)}$$

- First, an estimate of the noise model

$H_{j+1}(q) = \dfrac{C_{j+1}(q)}{D_{j+1}(q)}$ is obtained by fitting ARMA model to $y(t_k) - G_j(p)u(t_k)$ using discrete-time methods.

After that, use

$$F_{j+1} = H^{-1}_{j+1}(q)\underbrace{\frac{1}{A_j(p)}}$$

where we used informal notation mixing $p$ & $q$.

discretized using zero-order hold assumption!

What we mean is that $1/A(p)$ will be discretized first!

$$\frac{z-1}{z}\, \mathcal{Z}\left\{ \mathcal{L}^{-1}\left(\frac{1}{sA_j(s)}\right)_{t=k\Delta} \right\}$$

## Identification of stochastic LTI models with CT noise model.

- Use exact discretization:

$$\begin{cases} dx(t) = A(\theta)\,x(t)\,dt + B(\theta)\,u(t)\,dt + \underset{\sim}{B}(\theta)\,dw(t) \\ y(t) = C(\theta)\,x(t) \end{cases}$$

, $w \sim$ Wiener process with incremental cov. $\Sigma\, dt$

$\downarrow$ discretization

$$\begin{cases} x(t_{k+1}) = F_\theta(\Delta_k)\,x(t_k) + \mu_\theta(t_k) + v_\theta(t_k) \\ y(t_k) = C(\theta)\,x(t_k). \\ \Delta_k = t_{k+1} - t_k. \\ F_\theta(\Delta_k) = \exp(A(\theta)\Delta_k). \\ \mu_\theta(\Delta_k) = \int_{t_k}^{t_{k+1}} e^{A(\theta)s} B(\theta)\,u(s)\,ds. \end{cases}$$

and $v_\theta(t_k)$ is Gaussian discrete-time white noise with covariance matrix.

$$\int_0^{\Delta_k} e^{A(\theta)s}\,\Sigma\, e^{A^T(\theta)s}\,ds.$$

Then, likelihood function can be computed using a KF. But this requires running a KF for each candidate $\theta$ within a numerical optimization algorithm.

Also, if $\Delta_k$ is small, we get numerical problems!

To see this, assume

$\Delta_k = \Delta$. Then $F_\theta = \exp(A(\theta)\Delta)$.

Then when $\Delta \to 0$, all eigenvalues of $F_\theta$ converge to $0$!.

## Alternatives?

input/output approaches with CT noise model!

Replace differentials by differences in a way that gaurantee consistency.

we clarify the problems with examples:

Example

$$dy(t) = -a\,y(t)\,dt + b\,u(t) + \sigma\,dw(t), \quad a>0$$

Sol. $$y(t) = e^{-a(t-t_0)} y(t_0) + b\int_{t_0}^{t} e^{-a(t-s)} u(s)\,ds + \sigma\int_{t_0}^{t} e^{-a(t-s)}\,dw(s)$$ ← Itô integral.

ARX

Assume zero initial conditions.

$$y(t) = \frac{b}{p+a}u(t) + \underbrace{\frac{\sigma}{p+a}e_c(t)}_{\text{Stationary with a spectral density}}$$ ↓

spectral density

$$\Phi_{y_e}(\omega) = \frac{\sigma^2}{|i\omega + a|^2}$$

Approximations of $p$:

$$\delta_f = \frac{q-1}{\Delta}, \quad \delta_b = \frac{1-q^{-1}}{\Delta}$$

both are $O(\Delta)$ approximations.

Heuristic models. Assume $b=0$, estimate $a$

$$(p+a)\, y(t) = e_c(t)$$

$$\delta_b y(t) = -y(t)\, a_b + \varepsilon_1(t)$$

LS? $\hat{a}_b = -\frac{(y(t) - y(t-\Delta))\, y(t)}{\Delta\, y^2(t)}$

asymptotically.

$$\hat{a}_b \to -\frac{\frac{1}{\Delta}\mathbb{E}\left[(y(t) - y(t-\Delta))\, y(t)\right]}{\mathbb{E}\left[y^2(t)\right]}$$

$$= \frac{1}{\Delta}\left(-1 + \frac{r(\Delta)}{r(0)}\right)$$

$r$ is the correlation function.

$$r(\tau) = \mathbb{E}[y(t)\, y(t+\tau)]$$

satisfies

$$r(\tau) = \frac{\sigma^2}{2a} e^{-a|\tau|}$$

thus, $\hat{a}_b \to \frac{1}{\Delta}(-1 + e^{-a\Delta}) = -a + O(\Delta)$

which is far from the truth even when $\Delta \to 0$ and $N \to \infty$.

Try $\delta_f$:

$$\delta_f y(t) = -y(t)\, a_f + \varepsilon_2(t).$$

$$\hat{a}_f \to \frac{-\frac{1}{\Delta}\mathbb{E}[(y(t+h)-y(t))\,y(t)]}{\mathbb{E}[y^2(t)]}$$

$$= \frac{1}{h}\left(1-\frac{r(\Delta)}{r(0)}\right) = \frac{1}{\Delta}\left(1-e^{-a\Delta}\right)$$

$$= a + O(\Delta).$$

and we get accurate estimates for $N\to\infty$ & $\Delta\to 0$.

Try <u>central difference</u>:

$$\delta_c y(t) = \frac{1}{2h}\left(y(t+h)-y(t-h)\right)$$

$$\hat{a}_c \to \frac{r(\Delta)-r(\Delta)}{2\Delta\, r(0)} = 0$$

useless estimator!

Does $\delta_f$ always work? No!

Consider the process

$$p^2 y(t) + (a+b)\,p\,y(t) + ab\,y(t) = ab\,e_c(t)$$

$$\delta_f^2 y(t) + \hat{a}_1\,\delta_f y(t) + \hat{a}_2\,y(t) = \varepsilon(t).$$

then use LS.

$$\lim_{\substack{\Delta\to 0\\ N\to\infty}} \hat{a}_1 = \tfrac{2}{3}(a+b) \qquad \text{wrong!}$$

$$\lim_{\substack{\Delta\to 0\\ N\to\infty}} \hat{a}_2 = ab \qquad \text{correct!}$$

There are ways to correct this.
we give here one of these ways without going into details:

Let $D^K$ be an approximation of $p^K$ such that for a smooth function $f(t)$.

$$\begin{cases} D^K f(t) = p^K f(t) + O(\Delta), & K = 0, \dots, n \\ \quad = \frac{1}{\Delta^K} \sum_j \beta_{K,j} f(t + j\Delta). & \end{cases}$$

This holds if.

$$\sum_j \beta_{K,j} j^\nu = \begin{cases} 0 & \nu = 0, \dots, K-1. \\ K!, & \nu = K \end{cases}$$

then, The LS estimator is consistent if

$$\sum_j \beta_{n-1,j} \sum_\ell \beta_{n,\ell} \left( |\ell - j|^{2n-1} - (\ell - j)^{2n-1} \right) = 0$$

this holds when $\ell \geq j$

i.e. All measurements used when forming $D^n y(t)$ are at least as recent as those used when forming $D^{n-1} y(t)$.

<u>Thus</u>. a simple modification of $\delta_f^2$ above gives.

all as recent

$$D^2 y(t) = \frac{1}{\Delta^2} \left[ y(t+3\Delta) - 2y(t+2\Delta) + \underline{\underline{y(t+\Delta)}} \right]$$

(here, $\ell = \{1, 2, 3\}$ and $j = \{0, 1\}$).

because

$$D' y(t) = \delta_f = \frac{y(t+\Delta) - y(t)}{\Delta}$$