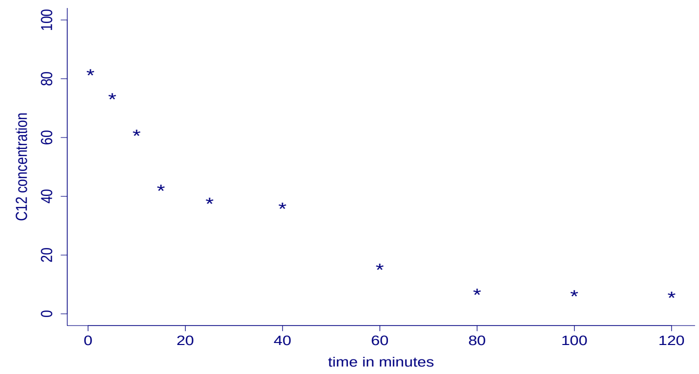
Parameter estimation for discretely observed diffusions

Susanne Ditlevsen
Department of Mathematical Sciences
University of Copenhagen, Denmark
Email: `susanne@math.ku.dk`

Summer school 4–12 August 2008, Middelfart, Denmark

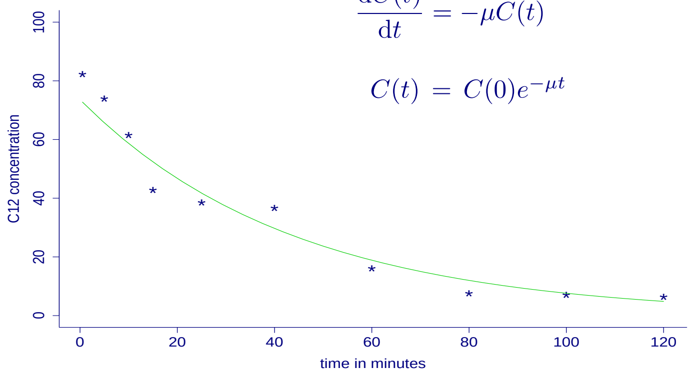
The concentration of a drug in blood



Exponential decay

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = -\mu C(t)$$

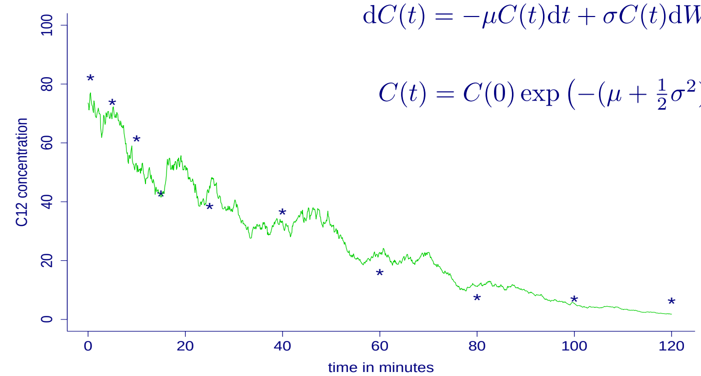
$$C(t) = C(0)e^{-\mu t}$$



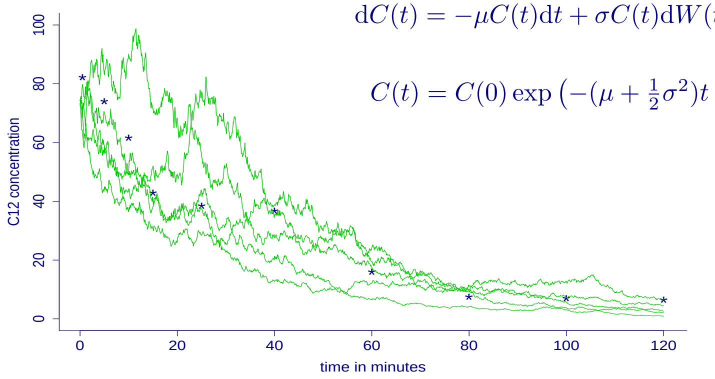
Exponential decay with noise

$$\mathrm{d}C(t) = -\mu C(t)\mathrm{d}t + \sigma C(t)\mathrm{d}W(t)$$

$$C(t) = C(0)\exp\left(-(\mu + \tfrac{1}{2}\sigma^2)t + \sigma W(t)\right)$$



Different realizations



$$\mathrm{d}C(t) = -\mu C(t)\mathrm{d}t + \sigma C(t)\mathrm{d}W(t)$$

$$C(t) = C(0)\exp\left(-(\mu + \tfrac{1}{2}\sigma^2)t + \sigma W(t)\right)$$

Estimation for discretely observed diffusions

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

X, b and W d-dimensional, σ $d \times d$-matrix

State space: $D \subseteq \mathbb{R}^d$

For $d = 1$, $D = (\ell, r)$, $-\infty \le \ell < r \le \infty$

Data: $X_{t_0}, \cdots, X_{t_n}$, $t_i = \Delta i$

The likelihood function

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

Data: $X_{t_1}, \cdots, X_{t_n}$, $t_1 < \cdots < t_n$

Likelihood-function:

$$L_n(\theta) = p(X_{t_1}, \cdots, X_{t_n};\theta)$$

IF the observations were iid (independent and identically distributed) - which they are NOT - then we could write the likelihood-function:

$$\begin{aligned} L_n(\theta) &= p_1(X_{t_1};\theta) \times \cdots \times p_n(X_{t_n};\theta) \\ &= p_1(X_{t_1};\theta) \times \cdots \times p_1(X_{t_n};\theta) \end{aligned}$$

If e.g. the process is stationary, this is an approximation to the true likelihood, ignoring dependence between observations.

Example: the Ornstein-Uhlenbeck process

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t$$

where $\beta > 0, \alpha \in \mathbb{R}, \sigma > 0$ and $X_0 = x_0$.

Solution:

$$X_t = \alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s$$

Note that this is a sum of deterministic terms and an integral of a deterministic function with respect to a Wiener process with normally distributed increments. The distribution is thus normal.

The conditional expectation is

$$\begin{aligned} E[X_t|X_0 = x_0] &= E\left[\alpha + (x_0 - \alpha)e^{-\beta t} + \sigma \int_0^t e^{-\beta(t-s)} dW_s\right] \\ &= \alpha + (x_0 - \alpha)e^{-\beta t} \end{aligned}$$

The conditional variance is

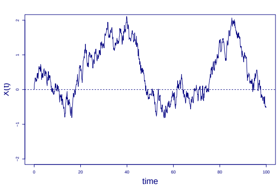
$$\text{Var}[X_t|X_0 = x_0] = E\left[(\sigma \int_0^t e^{-\beta(t-s)} dW_s)^2\right]$$

Use Ito's isometry to obtain

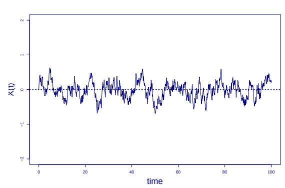
$$\text{Var}[X_t|X_0 = x_0] = \sigma^2 E\left[\int_0^t e^{-2\beta(t-s)} ds\right] = \frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta t}\right)$$

Thus $(X_t|X_0 = x_0) \sim N(\alpha + (x_0 - \alpha)e^{-\beta t}, \frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta t}\right))$.
Asymptotically $X_t \sim N(\alpha, \frac{\sigma^2}{2\beta})$ (or always if $X_0 \sim N(\alpha, \frac{\sigma^2}{2\beta})$).

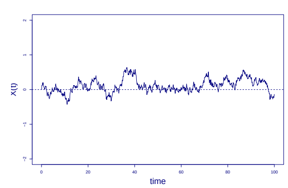
Parameter interpretation in the OU-process



$\beta = 0.01, \sigma = 1$



$\beta = 0.1, \sigma = 1$



$\beta = 0.01, \sigma = 0.5$

β	:	how "strongly" the system reacts to perturbations (the "decay-rate" or "growth-rate")
σ^2	:	the variation or the size of the noise.
α	:	the asymptotic mean

Back to maximum likelihood estimation

Consider the OU-process (for simplicity with only one parameter):

$$dX_t = -\theta X_t \, dt + dW_t$$

If $X_0 \sim N(0, \frac{1}{2\theta})$ then $X_t \sim N(0, \frac{1}{2\theta})$ for all t.

Assuming (wrongly) independence between observations, we write the likelihood

$$\begin{aligned} L_n(\theta) &= p_1(X_{t_1}; \theta) \times \cdots \times p_1(X_{t_n}; \theta) \\ &= \frac{1}{\sqrt{2\pi/2\theta}} \exp\left\{-\frac{X_{t_1}^2}{2/2\theta}\right\} \times \cdots \times \frac{1}{\sqrt{2\pi/2\theta}} \exp\left\{-\frac{X_{t_n}^2}{2/2\theta}\right\} \\ &= \left(\frac{\theta}{\pi}\right)^{\frac{n}{2}} \exp\left\{-\theta \sum_{i=1}^n X_{t_i}^2\right\} \end{aligned}$$

Maximizing the likelihood yields the maximum likelihood estimator.
It is easier to maximize the log-likelihood (it has the same maximum):

$$\log L_n(\theta) \;=\; l_n(\theta) \quad = \quad \frac{n}{2}\log(\theta/\pi) - \theta \sum_{i=1}^{n} X_{t_i}^2$$

We differentiate to find the maximum:

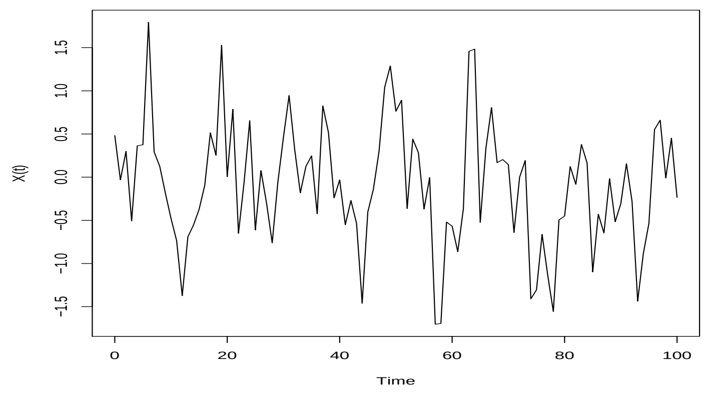
$$\frac{\mathrm{d}\, l_n(\theta)}{\mathrm{d}\,\theta} \;=\; \partial_\theta l_n(\theta) \quad = \quad \frac{n}{2\theta} - \sum_{i=1}^{n} X_{t_i}^2$$

This is called the *score function*. An estimator $\hat{\theta}$ is found by equating the score function to 0:

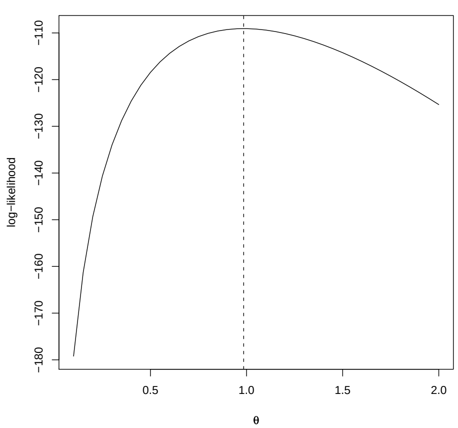
$$\hat{\theta} \quad = \quad \frac{n}{2\sum_{i=1}^{n} X_{t_i}^2}$$

Note that if $\theta < \hat{\theta}$ then $\partial_\theta l_n(\theta) > 0$ and if $\theta > \hat{\theta}$ then $\partial_\theta l_n(\theta) < 0$.
Thus, $\hat{\theta}$ is a (unique) maximum.

Assume observations from this process (simulated with $\theta = 1$):



Log-likelihood function: $\partial_\theta l_n(\theta) = \frac{n}{2\theta} - \sum_{i=1}^{n} X_{t_i}^2$. For this data set is $\hat{\theta} = 0.984972$.



The score function is an example of an *estimating function*:

$$G_n(\theta; X_{t_1}, \cdots, X_{t_n}) \quad : \quad \Theta \times D \mapsto \mathbb{R}^p$$

which is a p-dimensional function of the parameter θ and the data. Usually we simply write $G_n(\theta)$. An estimator is obtained by solving the equation:

$$G_n(\theta) \quad = \quad 0$$

In the previous example:

$$G_n(\theta) \quad = \quad \frac{n}{2\theta} - \sum_{i=1}^{n} X_{t_i}^2$$

The likelihood function

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

Data: $X_{t_1}, \cdots, X_{t_n}, \quad t_1 < \cdots < t_n$

Remember that we WRONGLY assumed independence. We have the ("correct") likelihood-function:

$$L_n(\theta) = p(X_{t_0}, X_{t_1}, \cdots, X_{t_n};\theta)$$

which by Bayes' theorem can be expressed as

$$L_n(\theta) = p(X_{t_n}|X_{t_0}, X_{t_1}, \cdots, X_{t_{n-1}};\theta) \times p(X_{t_0}, X_{t_1}, \cdots, X_{t_{n-1}};\theta)$$

Continuing this way we obtain

$$\begin{aligned} L_n(\theta) &= p(X_{t_n}|X_{t_0}, X_{t_1}, \cdots, X_{t_{n-1}};\theta) \times \\ & \quad p(X_{t_{n-1}}|X_{t_0}, X_{t_1}, \cdots, X_{t_{n-2}};\theta) \times \cdots \\ & \quad \cdots \times p(X_{t_2}|X_{t_1};\theta) \times p(X_{t_0};\theta) \end{aligned}$$

A very nice feature of our observations which they inherit from the diffusion process: they are a Markov process. Thus

$$p(X_{t_n}|X_{t_0}, X_{t_1}, \cdots, X_{t_{n-1}};\theta) = p(X_{t_n}|X_{t_{n-1}};\theta)$$

and therefore

$$L_n(\theta) = p(X_{t_n}|X_{t_{n-1}};\theta)\, p(X_{t_{n-1}}|X_{t_{n-2}};\theta) \cdots p(X_{t_2}|X_{t_1};\theta)\, p(X_{t_0};\theta)$$

Transition densities:

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p$$

Data: $X_{t_0}, X_{t_1}, \cdots, X_{t_n}, \quad 0 = t_0 < t_1 < \cdots < t_n$

Likelihood-function:

$$L_n(\theta) = \prod_{i=1}^{n} p(\Delta_i, X_{t_{i-1}}, X_{t_i};\theta),$$

where $\Delta_i = t_i - t_{i-1}$

$y \mapsto p(\Delta, x, y;\theta)$ is the transition density, i.e. probability density function of the conditional distribution of $X_{t+\Delta}$ given that $X_t = x$. Also conditional density of $X_{t+s+\Delta}$ given $X_{t+s} = x$.

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t.$$

$$y \mapsto p(t, x, y)$$

Conditional density of X_t given $X_0 = x$;

Data: $X_{t_1}, \cdots, X_{t_n}, \quad t_1 < \cdots < t_n$.

Likelihood function:

$$L(\theta) = \prod_{i=1}^{n} p(t_i - t_{i-1}, X_{t_{i-1}}, X_{t_i};\theta)$$

• Cox-Ingersoll-Ross

$$dX_t = -\theta(X_t - \alpha)dt + \sigma\sqrt{X_t}dW_t$$

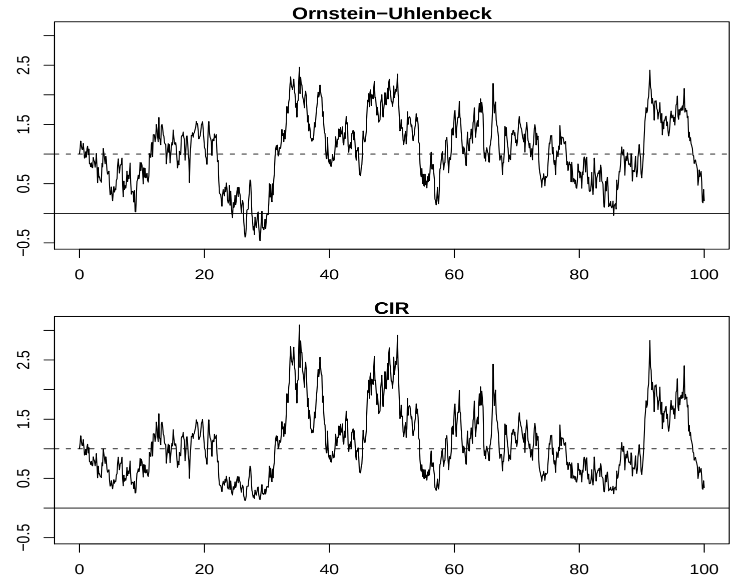
$\theta > 0$, $\alpha > 0$, $\sigma > 0$.

$$\begin{aligned} p(t,x,y) &= \frac{\beta(y/x)^{\frac{1}{2}\nu}\exp(\frac{1}{2}\theta\nu t - \beta y)}{\Gamma(\beta\alpha)(1-\exp(-\theta t))} \\ &\quad \times \exp\left[\frac{-\beta(x+y)}{\exp(\theta t)-1}\right] \mathrm{I}_\nu\left(\frac{\beta\sqrt{xy}}{\sinh(\frac{1}{2}\theta t)}\right), \end{aligned}$$

where $\beta = 2\theta\sigma^{-2}$ and $\nu = \beta\alpha - 1$.

I_ν is a modified Bessel function with index ν.

The transition density is a non-central χ^2-distribution.



The score function

$$U_n(\theta) = \partial_\theta \log L_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)$$

$$\partial_\theta f(\theta) = \left(\frac{\partial f}{\partial \theta_1}(\theta), \ldots, \frac{\partial f}{\partial \theta_p}(\theta)\right)^T$$

Under weak regularity conditions, the score function is a P_θ-martingale with respect to $\{\mathcal{F}_n\}$,
$\mathcal{F}_n = \sigma(X_{t_1}, \ldots, X_{t_n})$, $n = 1, 2, \ldots$.

$$\begin{aligned} & E_\theta\left(\partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \,\middle|\, \mathcal{F}_{i-1}\right) \\ &= E_\theta\left(\frac{\partial_\theta p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)}{p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta)} \,\middle|\, X_{t_{i-1}}\right) \\ &= \int_E \frac{\partial_\theta p(\Delta_i, X_{t_{i-1}}, y; \theta)}{p(\Delta_i, X_{t_{i-1}}, y; \theta)} p(\Delta_i, X_{t_{i-1}}, y; \theta) dy \\ &= \int_E \partial_\theta p(\Delta_i, X_{t_{i-1}}, y; \theta) dy \\ &= \partial_\theta \underbrace{\int_E p(\Delta_i, X_{t_{i-1}}, y; \theta) dy}_{=1} = 0 \end{aligned}$$

Local dominated integrability

Lemma. *Consider a real function* $f(x;\theta), (x,\theta) \in D \times \Theta$, *where* $\Theta \subseteq \mathbb{R}$. *Suppose* $\frac{\partial}{\partial\theta} f(x;\theta)$ *is locally dominated integrable w.r.t. a measure* μ *on* D. *Then*

$$\frac{\partial}{\partial\theta}\int_D f(x;\theta)\mu(dx) = \int_D \frac{\partial}{\partial\theta} f(x;\theta)\mu(dx).$$

A real function $h(x;\theta)$ $(x \in D \subseteq \mathbb{R}^d)$ is called locally dominated integrable w.r.t. the measure μ on D if, for each $\theta_0 \in \Theta$, there exists a neighbourhood U_{θ_0} of θ_0 and a non-negative μ-integrable function $g_{\theta_0}(x)$ such that

$$|\, h(x;\theta)\,| \leq g_{\theta_0}(x)$$

for all $(x,\theta) \in D \times U_{\theta_0}$.

Example

$$dX_t = \theta X_t dt + \sigma dW_t,\ X_0 = x_0, \quad \theta \in \mathbb{R}.$$

Ornstein-Uhlenbeck process.

Data: $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ for some $\Delta > 0$.

$$X_t = X_s e^{\theta(t-s)} + \sigma \int_s^t e^{\theta(t-u)} dW_u$$

for $0 \leq s < t$.

$p(\Delta, x, y; \theta)$ is Gaussian with

$$E_\theta(X_{i\Delta} \,|\, X_{(i-1)\Delta} = x) = x e^{\theta\Delta}$$

and

$$\mathrm{Var}_\theta(X_{i\Delta} \,|\, X_{(i-1)\Delta}) = \frac{e^{2\theta\Delta} - 1}{2\theta}\sigma^2$$

Find an estimator for θ by minimizing

$$K_n(\theta) = \sum_{i=1}^n (X_{i\Delta} - e^{\theta\Delta} X_{(i-1)\Delta})^2$$

Least squares estimation or minimum contrast estimation.

Solve $\frac{d}{d\theta} K_n(\theta) = 0$ or $G_n(\theta) = 0$, where

$$G_n(\theta) = \sum_{i=1}^n X_{(i-1)\Delta}(X_{i\Delta} - e^{\theta\Delta} X_{(i-1)\Delta})$$

$$\widehat{\theta}_n = \frac{1}{\Delta} \log\left(\frac{\sum_{i=1}^n X_{(i-1)\Delta} X_{i\Delta}}{\sum_{i=1}^n X_{(i-1)\Delta}^2}\right),$$

provided that $\sum_{i=1}^n X_{(i-1)\Delta} X_{i\Delta} > 0$.

$G_n(\theta)$ is a martingale estimating function:

$$\begin{aligned} & E_\theta(X_{(i-1)\Delta}(X_{i\Delta} - e^{\theta\Delta} X_{(i-1)\Delta}) \,|\, \mathcal{F}_{i-1}) \\ = \ & X_{(i-1)\Delta}\{\underbrace{E_\theta(X_\Delta | X_{(i-1)\Delta})}_{= e^{\theta\Delta} X_{(i-1)\Delta}} - e^{\theta\Delta} X_{(i-1)\Delta}\} = 0 \end{aligned}$$

Approximate likelihood inference

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \qquad \theta \in \Theta \subseteq \mathbb{R}^p \qquad d = 1$$

Approximate transition density

$$p(\Delta, x, y;\theta) \approx q(\Delta, x, y;\theta) = \frac{1}{\sqrt{2\pi\Phi(\Delta, x;\theta)}} \exp\left[\frac{(y - F(\Delta, x;\theta))^2}{2\Phi(\Delta, x;\theta)}\right]$$

$$F(x;\theta) = E_\theta(X_\Delta | X_0 = x) \quad \text{and} \quad \Phi(x;\theta) = \text{Var}_\theta(X_\Delta | X_0 = x)$$

Approximate likelihood function

$$L_n(\theta) \approx \tilde{L}_n(\theta) = \prod_{i=1}^{n} q(\Delta_i, X_{t_{i-1}}, X_{t_i};\theta)$$

Approximate likelihood inference

Approximate log-likelihood function

$$\begin{aligned}\log \tilde{L}_n(\theta) &= \sum_{i=1}^{n} \log q(\Delta_i, X_{t_{i-1}}, X_{t_i};\theta) \\ &= \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\Phi - \frac{(y-F)^2}{2\Phi}\right]\end{aligned}$$

Approximate score function:

$$\partial_\theta \log \tilde{L}_n(\theta) = \sum_{i=1}^{n} \left[-\frac{1}{2}\frac{\partial_\theta \Phi}{\Phi} + \frac{y-F}{\Phi}\partial_\theta F + \frac{(y-F)^2}{2\Phi^2}\partial_\theta \Phi\right]$$

Approximate score function

$$\begin{aligned}\partial_\theta \log \tilde{L}_n(\theta) = \sum_{i=1}^{n} \Bigg\{ & \frac{\partial_\theta F(\Delta_i, X_{t_{i-1}};\theta)}{\Phi(\Delta_i, X_{t_{i-1}};\theta)}[X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)] \\ & + \frac{\partial_\theta \Phi(\Delta_i, X_{t_{i-1}};\theta)}{2\Phi(\Delta_i, X_{t_{i-1}};\theta)^2}\left[(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta))^2 - \Phi(\Delta_i, X_{t_{i-1}};\theta)\right]\Bigg\}\end{aligned}$$

Quadratic martingale estimating function

$$\begin{aligned}G_n(\theta) = \sum_{i=1}^{n} \big\{ & a_1(X_{t_{i-1}}, \Delta_i;\theta)(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)) \\ & + a_2(X_{t_{i-1}}, \Delta_i;\theta)\left[(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta))^2 - \Phi(\Delta_i, X_{t_{i-1}};\theta)\right]\big\}\end{aligned}$$

Bibby and Sørensen (1995,1996)

Martingale estimating functions

$$\begin{aligned}G_n(\theta) &= \sum_{i=1}^{n} g(\Delta_i, X_{t_i}, X_{t_{i-1}};\theta) \\ g(\Delta, y, x;\theta) &= \sum_{j=1}^{N} a_j(x, \Delta;\theta)[f_j(y) - \pi_\theta^\Delta f_j(x)]\end{aligned}$$

↑ p-dimensional ↑ real valued

Transition operator: $\pi_\theta^\Delta f(x) = E_\theta(f(X_\Delta) \,|\, X_0 = x)$

$G_n(\theta)$ is a P_θ-martingale:

$$E_\theta(a_j(X_{t_{i-1}},\Delta_i;\theta)[f_j(X_{t_i}) - \pi_\theta^{\Delta_i} f_j(X_{t_{i-1}})] \,|\, X_{t_1},\cdots,X_{t_{i-1}}) \quad = \quad$$
$$a_j(X_{t_{i-1}},\Delta_i;\theta)E_\theta([f_j(X_{t_i}) - \pi_\theta^{\Delta_i} f_j(X_{t_{i-1}})] \,|\, X_{t_1},\cdots,X_{t_{i-1}}) \quad = \quad 0$$

G_n–estimator(s): $\quad G_n(\hat{\theta}_n) = 0$

Martingale estimating functions

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_i}, X_{t_{i-1}};\theta),$$

$$g(\Delta, y, x;\theta) = \sum_{j=1}^{N} a_j(x,\Delta;\theta)[f_j(y) - \pi_\theta^\Delta f_j(x)]$$

- Easy asymptotics
- Simple expression for optimal estimating function
- The score function is a P_θ-martingale

Simulation

$\pi_\theta^\Delta f(x) = E_\theta(f(X_\Delta)|X_0 = x)$ is usually not explicitly known

Fix θ

Simulate numerically M independent trajectories of $\{X_t : t \in [0,\Delta]\}$ with $X_0 = x$

$$\pi_\theta^\Delta f(x) \approx \frac{1}{M}\sum_{i=1}^{M} f(X_\Delta^{(i)})$$

Taylor expansions

Review of deterministic expansions:

Consider

$$\frac{d}{dt}x_t \quad = \quad a(x_t)$$

with initial value x_{t_0} for $t \in [t_0, T]$, and $a(\cdot)$ is sufficiently smooth. We can write

$$x_t \quad = \quad x_{t_0} + \int_{t_0}^{T} a(x_s)ds$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function. By the chain rule

$$\frac{d}{dt} f(x_t) = \frac{d}{dt} x_t \, f'(x_t) = a(x_t) f'(x_t)$$

Define the operator

$$Lf = af'$$

where ' denotes differentiation with respect to x. Express the above equation for $f(x)$ in integral form

$$f(x_t) = f(x_{t_0}) + \int_{t_0}^{t} Lf(x_s) ds$$

Note that if $f(x) = x$ then $Lf = a, L^2 f = La$ and

$$x_t = x_{t_0} + \int_{t_0}^{t} a(x_s) ds$$

If $f(x) = a(x)$ then $La = aa'$ and

$$a(x_s) = a(x_{t_0}) + \int_{t_0}^{s} La(x_z) dz$$

Apply this to the equation for x_t

$$\begin{aligned} x_t &= x_{t_0} + \int_{t_0}^{t} \left(a(x_{t_0}) + \int_{t_0}^{s} La(x_z) dz \right) ds \\ &= x_{t_0} + a(x_{t_0}) \int_{t_0}^{t} ds + \int_{t_0}^{t} \int_{t_0}^{s} La(x_z) dz \, ds \\ &= x_{t_0} + a(x_{t_0})(t - t_0) + R_1 \end{aligned}$$

which is the simplest non-trivial expansion for x_t.

Apply again to the function $f = La$ to obtain

$$\begin{aligned} x_t &= x_{t_0} + a(x_{t_0}) \int_{t_0}^{t} ds + \int_{t_0}^{t} \int_{t_0}^{s} La(x_z) dz ds \\ &= x_{t_0} + a(x_{t_0}) \int_{t_0}^{t} ds + La(x_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s} dz \, ds + R_2 \\ &= x_{t_0} + a(x_{t_0})(t - t_0) + La(x_{t_0}) \frac{1}{2} (t - t_0)^2 + R_2 \end{aligned}$$

where

$$R_2 = \int_{t_0}^{t} \int_{t_0}^{s} \int_{t_0}^{z} L^2 a(x_u) du \, dz \, ds$$

For a general $r + 1$ times continuously differentiable function f we obtain the classical Taylor formula in integral form

$$f(x_t) = f(x_{t_0}) + \sum_{l=1}^{r} \frac{(t - t_0)^l}{l!} L^l f(x_{t_0}) + \int_{t_0}^{t} \cdots \int_{t_0}^{s_r} L^{r+1} f(x_{s_1}) ds_1 \ldots ds_{r+1}$$

The Ito-Taylor expansion

Iterated application of Ito's formula!

Consider

$$X_t = X_{t_0} + \int_{t_0}^{t} b(X_s)ds + \int_{t_0}^{t} \sigma(X_s)dWs$$

We introduce the operators

$$L^0 f = bf' + \frac{1}{2}\sigma^2 f''$$
$$L^1 f = \sigma f'$$

For f twice continuously differentiable, Ito's formula yields

$$\begin{aligned} f(X_t) &= f(X_{t_0}) + \int_{t_0}^{t} \left(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s) \right) ds \\ &\quad + \int_{t_0}^{t} \sigma(X_s)f'(X_s)dWs \\ &= f(X_{t_0}) + \int_{t_0}^{t} L^0 f(X_s)ds + \int_{t_0}^{t} L^1 f'(X_s)dWs \end{aligned}$$

Note that for $f(x) = x$ we have $L^0 f = b$ and $L^1 f = \sigma$, and the original equation for X_t is obtained

$$X_t = X_{t_0} + \int_{t_0}^{t} b(X_s)ds + \int_{t_0}^{t} \sigma(X_s)dWs$$

Like in the deterministic expansions, we apply Ito's formula to the functions $f = b$ and $f = \sigma$ and obtain

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} \left(b(X_{t_0}) + \int_{t_0}^{s} L^0 b(X_z)dz + \int_{t_0}^{s} L^1 b'(X_z)dWz \right) ds \\ &\quad + \int_{t_0}^{t} \left(\sigma(X_{t_0}) + \int_{t_0}^{s} L^0 \sigma(X_z)dz + \int_{t_0}^{s} L^1 \sigma'(X_z)dWz \right) dWs \\ &= X_{t_0} + b(X_{t_0}) \int_{t_0}^{t} ds + \sigma(X_{t_0}) \int_{t_0}^{t} dWs + R \\ &= X_{t_0} + b(X_{t_0})(t - t_0) + \sigma(X_{t_0})(W_t - W_{t_0}) + R \end{aligned}$$

This is the simplest non-trivial Ito-Taylor expansion of X_t involving single integrals with respect to both time and the Wiener process. The remainder contains multiple integrals with respect to both.

In the previous expansion we had

$$\begin{aligned} R &= \int_{t_0}^{t} \int_{t_0}^{s} L^0 b(X_z)dz\, ds + \int_{t_0}^{t} \int_{t_0}^{s} L^1 b(X_z)dW_z\, ds \\ &\quad + \int_{t_0}^{t} \int_{t_0}^{s} L^0 \sigma(X_z)dz\, dW_s + \int_{t_0}^{t} \int_{t_0}^{s} L^1 \sigma(X_z)dW_z\, dW_s \end{aligned}$$

Note that $dz\, ds, dW_z\, ds$ and $dz\, dW_s$ "scales like 0", whereas $dW_z\, dW_s$ scales like dt, comparable to the terms in the simplest expansion with two single integrals.

We therefore continue the expansion by applying the Ito formula to $f = L^1\sigma$.

The next Ito-Taylor expansion becomes

$$X_t = X_{t_0} + b(X_{t_0}) \int_{t_0}^{t} ds + \sigma(X_{t_0}) \int_{t_0}^{t} dWs + L^1\sigma(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s} dW_z \, dW_s + \bar{R}$$

$$= X_{t_0} + b(X_{t_0})\Delta t + \sigma(X_{t_0})\Delta W_t + \sigma(X_{t_0})\sigma'(X_{t_0})\frac{1}{2}(\Delta W_t^2 - \Delta t) + \bar{R}$$

with remainder

$$\bar{R} = \int_{t_0}^{t}\int_{t_0}^{s} L^0 b(X_z)dz\,ds + \int_{t_0}^{t}\int_{t_0}^{s} L^1 b(X_z)dW_z\,ds$$
$$+ \int_{t_0}^{t}\int_{t_0}^{s} L^0\sigma(X_z)dz\,dW_s + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{z} L^0L^1\sigma(X_u)du\,dW_z\,dW_s$$
$$+ \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{z} L^1L^1\sigma(X_u)dW_u\,dW_z\,dW_s$$

Numeric solutions

When no explicit solution is available we can approximate different characteristics of the process by simulation. (Realizations, moments, qualitative behavior etc). We use the approximations from the Ito-Taylor expansions.

- Different schemes (Euler, Milstein, higher order schemes...)
- Rate of convergence (Weak and strong)

Consider the Itô stochastic differential equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t$$

and a time discretization

$$0 = t_0 < t_1 < \cdots < t_j < \cdots < t_N = T$$

Put

$$\Delta_j = t_{j+1} - t_j$$
$$\Delta W_j = W_{t_{j+1}} - W_{t_j}$$

Then

$$\Delta W_j \sim N(0, \Delta_j)$$

The Euler-Maruyama scheme

We approximate the process X_t given by

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t \;;\; X(0) = x_0$$

at the discrete time-points $t_j, 1 \leq j \leq N$ by

$$Y_{t_{j+1}} = Y_{t_j} + b(Y_{t_j})\Delta_j + \sigma(Y_{t_j})\Delta W_j \;;\; Y_{t_0} = x_0$$

where $\Delta W_j = \sqrt{\Delta_j} \cdot Z_j$, with $Z_j \sim N(0,1)$ for all j.

The Euler-Maruyama scheme

Let us consider the expectation of the absolute error at the final time instant T:

There exist constants $K > 0$ and $\delta_0 > 0$ such that

$$E(|X_T - Y_{t_N}|) \quad \leq \quad K\delta^{0.5}$$

for any time discretization with maximum step size $\delta \in (0, \delta_0)$.

We say that the approximating process Y *converges in the strong sense* with order 0.5.

(Compare with the Euler scheme for an ODE which has order 1).

The Euler-Maruyama scheme

Sometimes we do not need a close *pathwise* approximation, but only some function of the value at a given final time T (e.g. $E(X_T)$, $E(X_T^2)$ or generally $E(g(X_T))$):

There exist constants $K > 0$ and $\delta_0 > 0$ such that for any polynomial g

$$|E(g(X_T) - E(g(Y_{t_N}))| \quad \leq \quad K\delta$$

for any time discretization with maximum step size $\delta \in (0, \delta_0)$.

We say that the approximating process Y *converges in the weak sense* with order 1.

The Milstein scheme

We can even do better!

We approximate X_t by

$$\begin{aligned} Y_{t_{j+1}} \quad &= \quad Y_{t_j} + b(Y_{t_j})\Delta_j + \sigma(Y_{t_j})\Delta W_j \\ &\quad + \frac{1}{2}\sigma(Y_{t_j})\sigma'(Y_{t_j})\{(\Delta W_j)^2 - \Delta_j\} \quad \text{(now Milstein...)} \end{aligned}$$

where the prime $'$ denotes the derivative.

The Milstein scheme

The Milstein scheme converges in the strong sense with order 1:

$$E(|X_T - Y_{t_N}|) \quad \leq \quad K\delta$$

We could regard the Milstein scheme as the proper generalization of the deterministic Euler-scheme.

If $b(X_t)$ does not depend on X_t the Euler-Maruyama and the Milstein scheme coincide.

Multi-dimensional diffusions:

Euler scheme: Similar.

Milstein scheme: Involves multiple Wiener integrals.

$$\int_{n\delta}^{(n+1)\delta} \int_{n\delta}^{s} dW_u^{(1)} dW_s^{(2)}$$

Simulation schemes are based on stochastic Ito-Taylor expansions that are formally obtained by iterated use of Ito's formula.

Kloeden and Platen (1992)