A Tutorial Introduction to Stochastic Differential Equations: Continuous-time Gaussian Markov Processes

Chris Williams

Institute for Adaptive and Neural Computation
School of Informatics, University of Edinburgh, UK

Presented: 9 December, minor revisions 13 December 2006

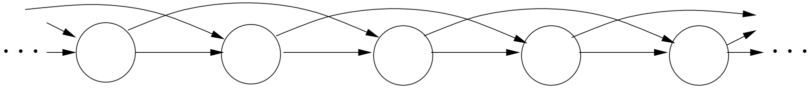
AR Processes: Discrete-time Gaussian Markov Processes

A discrete-time autoregressive (AR) process of order p:

$$X_t = \sum_{k=1}^{p} a_k X_{t-k} + b_0 Z_t,$$

where $Z_t \sim \mathcal{N}(0, 1)$ and all Z_t's are iid.

AR(2) example:



Linear combinations of Gaussians are Gaussian

From discrete to continuous time

- In continuous time, have not only the function value but also p of its derivatives at time t

$$a_p X^{(p)}(t) + a_{p-1} X^{(p-1)}(t) + \ldots + a_0 X(t) \;=\; b_0 Z(t),$$

  where $Z(t)$ is a white Gaussian noise process with covariance $\delta(t - t')$, and $a_p = 1$.
- This is a *stochastic differential equation* (SDE)
- Applications in many fields, e.g. chemistry, epidemiology, finance, neural modelling
- We will consider only SDEs driven by Gaussian white noise; this can be relaxed

Vector processes

- An AR(p) process can be written as a vector AR(1) process if one stores X_t and the previous $p-1$ values in $\mathbf{X}_t$
- Similarly for the pth order SDE

$$X^{(p)}(t) + a_{p-1}X^{(p-1)}(t) + \ldots + a_0 X(t) \;=\; b_0 Z(t),$$

$$\begin{aligned} X_1(t) &= X(t) \\ X_2(t) &= \dot{X}_1(t) = \dot{X}(t) \\ &\vdots \\ X_p(t) &= \dot{X}_{p-1}(t) = X^{(p-1)}(t) \end{aligned}$$

$$\dot{X}_p(t) + a_{p-1}X_p(t) + \ldots + a_1 X_2(t) + a_0 X_1(t) = b_0 Z(t)$$

or, in matrix form

$$\dot{\mathbf{X}}(t) = F\mathbf{X}(t) + B\mathbf{Z}(t)$$

for $\mathbf{Z}(t)$ being a p-dimensional white noise process, with

$$F = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{p-2} & -a_{p-1} \end{pmatrix}$$

and

$$B = \mathrm{diag}(0,\ 0, \dots, 0,\ 1)$$

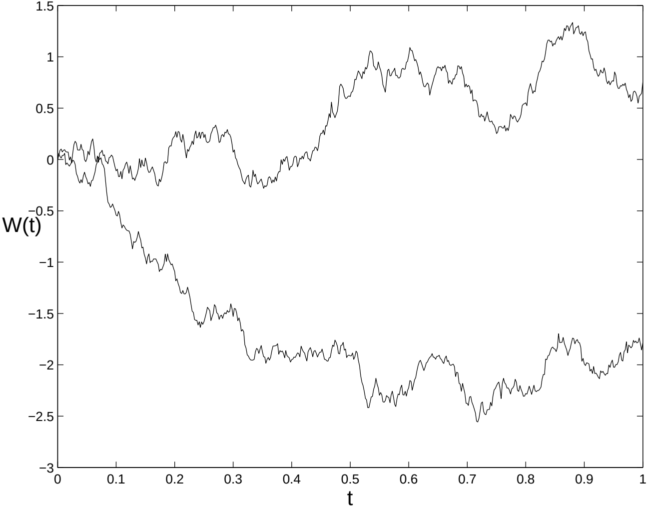
Overview

- Wiener process
- SDEs and simulation
- Stationary processes and covariance functions
- Inference (Gaussian process prediction)
- Fokker-Planck equations
- 3 views: SDE vs covariance function vs Fokker-Planck

The Wiener Process

- $W(t)$ is continuous and $W(0) = 0$
- $W(t) \sim N(0, t)$
- Independent increments: $W(t+s) - W(s) \sim N(0, t)$ and is independent of the history of the process up to time t
- $\mathrm{cov}(W(s), W(t)) = \min(s, t)$
- Interpret $dW(t) = W(t + dt) - W(t)$

Discretized Wiener Process



Gaussian Processes

- For a stochastic process $X(t)$, mean function is

$$\mu(t) = \mathbb{E}[X(t)]$$

- Covariance function

$$k(t, t') = \mathbb{E}[(X(t) - \mu(t))(X(t') - \mu(t'))]$$

- Gaussian processes are stochastic processes defined by their mean and covariance functions

SDEs

- Consider the SDE

$$\dot{\mathbf{X}}(t) = F\mathbf{X}(t) + B\mathbf{Z}(t)$$

- This is a *Langevin* equation
- A problem is that we want to think of $\mathbf{Z}(t)$ as being the derivative of a Wiener process, but the Wiener process is (with probability one) nowhere differentiable ...
- The "kosher" way of writing this SDE is

$$d\mathbf{X}(t) = F\mathbf{X}(t)dt + B\, d\mathbf{W}(t)$$

where $\mathbf{W}(t)$ is a vector of Wiener processes

Simulation of an SDE

Times $t_0 < t_1 < t_2 < \ldots < t_n$, $\Delta t_i = t_{i+1} - t_i$

$$\mathbf{X}_{i+1} = \mathbf{X}_i + F\mathbf{X}_i \Delta t_i + B\mathbf{Z}_i \sqrt{\Delta t_i}$$

where $\mathbf{Z}_i \sim N(0, I)$

This is the Euler-Maruyama method; higher-order methods are also possible (Milstein)

Stochastic Integration

- Riemann sum

$$\int_0^T h(t)dt = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(t_{j+1} - t_j)$$

for $t_j = jT/N$

- Itô stochastic integral

$$\int_0^T h(t)dW(t) = \mathrm{m.s.}\ \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(W(t_{j+1}) - W(t_j))$$

- Example

$$\int_0^T W(t)dW(t) = \frac{1}{2}W(T)^2 - \frac{1}{2}T$$

- Mnemonically we have $dW(t)^2 = dt$

General form of a Diffusion process

$$d\mathbf{X}(t) = \mathbf{a}(\mathbf{X}, t)dt + B(\mathbf{X}, t)\, d\mathbf{W}(t)$$

where the functions $\mathbf{a}(\mathbf{X}, t)$ and $B(\mathbf{X}, t)$ must be non-anticipating, corresponding to the integral form

$$\mathbf{X}(t) - \mathbf{X}(0) = \int_0^t \mathbf{a}(\mathbf{X}(t'), t')dt' + \int_0^t B(\mathbf{X}(t'), t')\, d\mathbf{W}(t')$$

- $\mathbf{a}(\mathbf{X}, t)$ is the drift vector, $B(\mathbf{X}, t)$ is the diffusion matrix
- Sample paths of a diffusion process are continuous

$$d\mathbf{X}(t) = F(t)\mathbf{X}(t)dt + B(t)\, d\mathbf{W}(t)$$

is the most general form that is a *Gaussian* process

Simple Examples

- Wiener process

$$dX = dW \qquad X(t) = X(0) + W(t)$$

- Wiener process with scaling and drift

$$dX = adt + \sigma dW \qquad X(t) = X(0) + at + \sigma W(t)$$

- Ornstein-Uhlenbeck process

$$dX = -aXdt + \sigma dW \qquad X(t) = X(0)e^{-at} + \sigma \int_0^t e^{-a(t-t')} dW(t')$$

Infinitesimal moments

$$\Delta\mathbf{X}(t) = \mathbf{X}(t+\Delta t) - \mathbf{X}(t) = \mathbf{a}(\mathbf{X}, t)\Delta t + B(\mathbf{X}, t)\mathbf{Z}_t\sqrt{\Delta t}$$

- First moment: drift

$$\lim_{\Delta t\to 0} \frac{\mathbb{E}[\Delta\mathbf{X}(t)]}{\Delta t} = \mathbf{a}(X, t)$$

- Second moment: diffusion

$$\lim_{\Delta t\to 0} \frac{\mathbb{E}[(\Delta\mathbf{X}(t))(\Delta\mathbf{X}(t))^T]}{\Delta t} = B(X, t)B^T(X, t)$$

Notice that $\mathbb{E}[(\sqrt{\Delta t}\mathbf{Z}_t)(\sqrt{\Delta t}\mathbf{Z}_t)^T] = (\Delta t)I$

Stationary Processes

- Assume time-invariant coefficients of univariate SDE of order p
- If the coefficients are such that eigenvalues of F are in the left half plane (negative real parts) then the SDE will have a *stationary distribution*, such that $\mathbb{E}[X(t)X(t')] = k(t - t')$
- Can generalize this to vector-valued processes, when k is a matrix-valued function

Fourier Analysis

Sinusoids are eigenfunctions of LTI systems

$$\tilde{X}(s) \;=\; \int_{-\infty}^{\infty} X(t)e^{-2\pi ist}\,dt, \qquad X(t) \;=\; \int_{-\infty}^{\infty} \tilde{X}(s)e^{2\pi ist}\,ds,$$

$$X^{(k)}(t) \;=\; \int_{-\infty}^{\infty} (2\pi is)^k \tilde{X}(s)e^{2\pi ist}\,ds.$$

$$\sum_{k=0}^{p} a_k X^{(k)}(t) \;=\; b_0 Z(t), \qquad \sum_{k=0}^{p} a_k (2\pi is)^k \tilde{X}(s) \;=\; b_0 \tilde{Z}(s)$$

Power spectrum of SDE

Wiener-Khintchine Theorem

$$k(\boldsymbol{\tau}) = \int S(\mathbf{s}) e^{2\pi i \mathbf{s}\cdot\boldsymbol{\tau}}\, d\mathbf{s}, \qquad S(\mathbf{s}) = \int k(\boldsymbol{\tau}) e^{-2\pi i \mathbf{s}\cdot\boldsymbol{\tau}}\, d\boldsymbol{\tau}.$$

so

$$\langle \tilde{X}(s_1)\tilde{X}^*(s_2)\rangle = S(s_1)\delta(s_1 - s_2)$$

and thus

$$S(s) = \frac{b_0^2}{|A(2\pi i s)|^2}$$

where $A(z) = \sum_{k=0}^{p} a_k z^k$. Require that roots of $A(z)$ lie in left half plane for stationarity

Examples

- First order SDE

$$\dot{X}(t) + a_0 X(t) = b_0 Z(t), \quad S(s) = \frac{b_0^2}{(2\pi s)^2 + a_0^2}, \quad k(t) = \frac{b_0^2}{2a_0} e^{-a_0|t|}$$

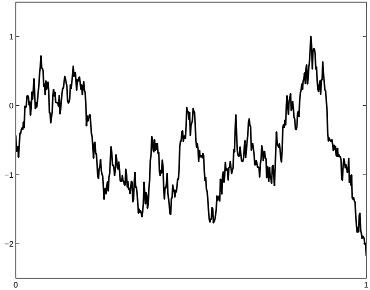
- Damped simple harmonic oscillator (second order SDE)

$$\ddot{X}(t) + a_1 \dot{X}(t) + a_0 X(t) = b_0 Z(t), \quad S(s) = \frac{b_0^2}{(a_0 - (2\pi s)^2)^2 + a_1^2 (2\pi s)^2}$$

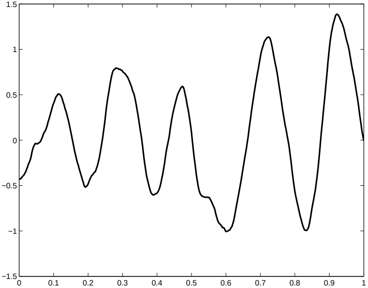
if $a_1^2 < 4a_0$ (weak damping) then

$$k(t) = \frac{b_0^2}{2a_0 a_1} e^{-\alpha|t|}(\cos(\beta t) + \frac{\alpha}{\beta}\sin(\beta|t|))$$

where $\alpha = a_1/2$, and $\alpha^2 + \beta^2 = a_0$.



OU process
$a_0 = 5$



Damped harmonic oscillator
$a_0 = 500,\ a_1 = 5$

Vector OU process

$$dX(t) = -AX(t) + BdW(t)$$

solution is

$$X(t) = \exp(-At)X(0) + \int_0^t \exp(-A(t-t'))B\, dW(t')$$

For stationary solution remove $\mathbf{X}(0)$ dependence

$$\begin{aligned}\langle \mathbf{X}(t)\mathbf{X}^T(s)\rangle &\overset{def}{=} \Sigma(t-s) \\ &= \int_{-\infty}^{\min(t,s)} \exp(-A(t-t'))BB^T \exp(-A^T(s-t'))\, dt'\end{aligned}$$

- Can show that

$$A\Sigma(0) + \Sigma(0)A^T = BB^T$$

and

$$\Sigma(t-s) = \exp(-A(t-s))\Sigma(0) \qquad \text{for } t > s$$

and $\Sigma(t-s) = \Sigma^T(s-t)$
- Can also do spectral analysis of vector OU process
- See Gardiner (1985, §4.4.6) for more details

Mean square differentiability

$$a_p X^{(p)}(t) + a_{p-1} X^{(p-1)}(t) + \ldots + a_0 X(t) \ = \ b_0 Z(t),$$

- SDEs of order p are $p - 1$ times mean square differentiable
- This is easy to see intuitively from the above equation, as $X^{(p)}(t)$ is like white noise
- Note that a process gets rougher the more times it is differentiated

Relating Discrete-time and Sampled Continuous-time GMPs

- Discrete time ARMA(p, q) process

$$X_t = \sum_{i=1}^{p} X_{t-i} + \sum_{j=0}^{q} b_j Z_{t-j}$$

- A continuous-time ARMA process has spectral density

$$S(s) \ = \ \frac{|B(2\pi is)|^2}{|A(2\pi is)|^2}$$

- Theorem (e.g. Ihara, 1993): Let X be a continuous-time stationary Gaussian process and X_h be the discretization of this process. If X is an ARMA process then X_h is also an ARMA process. However, if X is an AR process then X_h is not necessarily an AR process
- A discretized continuous-time AR(1) process is a discrete-time AR(1) process
- However, a discretized continuous-time AR(2) process is not, in general, a discrete-time AR(2) process.

Inference

- Given observations of X at times $t_1, t_2, \ldots, t_n$, compute posterior distribution at t_*
- Note that for OU process, the Markov property means that we need only condition on t_P and t_F, the nearest times to the past and future of t_*
- Caveat: observations must be noise free, otherwise all observations will count
- This is just Gaussian process prediction:

$$X(t_*)|X(t_1), \dots X(t_n) \sim \mathcal{N}(\mu(t_*), \sigma^2(t_*))$$

with

$$\mu(t_*) = (k_{*P}, k_{*F}) \begin{pmatrix} k_{PP} & k_{PF} \\ k_{PF} & k_{FF} \end{pmatrix}^{-1} \begin{pmatrix} X_P \\ X_F \end{pmatrix}$$

$$\sigma^2(t_*) = k_{**} - (k_{*P}, k_{*F}) \begin{pmatrix} k_{PP} & k_{PF} \\ k_{PF} & k_{FF} \end{pmatrix}^{-1} \begin{pmatrix} k_{*P} \\ k_{*F} \end{pmatrix}$$

where $k_{*P} = k(t_*, t_P)$ etc

Vector process works similarly

Fokker-Planck Equations

- Consider the transition pdf $p \stackrel{def}{=} p(\mathbf{x}, t | \mathbf{x}_0, t_0)$. This evolves according to the (forward) Fokker-Planck equation

$$\partial_t p = -\sum_i \partial_i (a_i(\mathbf{x}, t) p) + \frac{1}{2} \partial_i \partial_j [B(\mathbf{x}, t) B^T(\mathbf{x}, t)]_{ij} p]$$

  corresponding to the SDE

$$d\mathbf{X}(t) = \mathbf{a}(\mathbf{X}, t) dt + B(\mathbf{X}, t)\, d\mathbf{W}(t)$$

- This is just the differential form of the Chapman-Kolmogorov equation
- There is also a "backward" equation

Simple example: Wiener process with drift

- Wiener process with scaling and drift

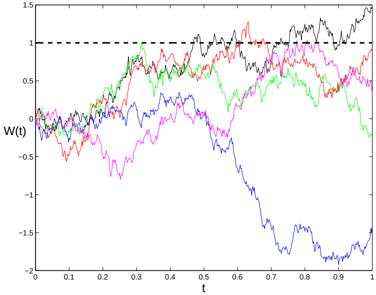
$$dX = a dt + \sigma dW \qquad X(t) = X(0) + at + \sigma W(t)$$

$$p(x, t|x_0, 0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(x - x_0 - at)^2}{2\sigma^2 t}\right)$$

Fokker-Planck Boundary Conditions

Feller, 1952

- Regular
    - Absorbing
    - Reflecting
    - ...
- Exit
- Entrance
- Natural



Parameter Estimation

- If we have observations $\mathbf{X} = (X(t_1), \ldots, X(t_n))^T$ of a Gaussian process at some set of finite times $t_1, \ldots, t_n$, then

$$\log p(\mathbf{X}|\theta) = -\frac{1}{2}\log|K_\theta| - \frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_\theta)^T K_\theta^{-1}(\mathbf{X} - \boldsymbol{\mu}_\theta) - \frac{n}{2}\log(2\pi)$$

- Can use e.g. numerical methods to optimize parameters θ
- For continuous observations, see e.g. Feigin (1976)

Summary

- Relationship of SDEs driven by Gaussian white noise to Gaussian Markov processes
- Formal mathematical framework of stochastic integration
- As Gaussian processes we can compute their mean and covariance functions, and do inference
- Markov properties are to the fore for Fokker-Planck equations
- Extend to allow observation noise: continuous-time Kalman filter (Kalman and Bucy, 1961)
- Challenges of the workshop: nonlinear dynamics, nonlinear observation

References

- C. W. Gardiner, *Handbook of Stochastic Methods* (second edition) Springer-Verlag, 1985 [3rd ed, 2004]
- B. Øksendal, *Stochastic Differential Equations*, Springer-Verlag, 1985
- D. J. Higham, An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, *SIAM Review* 43(3) 525-546, 2001
- P. E. Kloeden and E. Platen, *The Numerical Solution of Stochastic Differential Equations* Springer-Verlag, 1999
- P. D. Feigin, Maximum likelihood estimation for continuous-time stochastic processes, *Adv. Appl. Prob.* 55, 468-519, 1976
- C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* [see Appendix B], MIT Press, 2006

Stratonovich stochastic integral

- Itô stochastic integral

$$\int_0^T h(t)dW(t) = \text{m.s. } \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(W(t_{j+1}) - W(t_j))$$

- Stratonovich integral

$$\int_0^T h(t)dW(t) = \text{m.s. } \lim_{N\to\infty} \sum_{j=0}^{N-1} h(\frac{t_j + t_{j+1}}{2})(W(t_{j+1}) - W(t_j))$$

- Some authors use $\frac{1}{2}(h(t_j) + h(t_{j+1}))(W(t_{j+1}) - W(t_j))$ instead

Itô's formula

- Let the stochastic process X satisfy

$$dX = a(X,t)dt + b(X,t)dW$$

- Then $Y = f(X,t)$ satisfies

$$dY = \left(a(X,t)f_x(X,t) + \frac{1}{2}b^2(X,t)f_{xx}(X,t) + f_t(X,t) \right) dt \\ + (b(X,t)f_x(X,t))dW$$

- Example: $Y(t) = X(t)^2$, $dX = dW$ (Wiener process)

$$dY = dt + 2\sqrt{Y}\, dW$$

$\int_0^T W(t)dW(t)$

$$
\begin{aligned}
&\sum_{j=0}^{N-1} W(t_j)(W(t_{j+1}) - W(t_j)) \\
&= \frac{1}{2}\sum_{j=0}^{N-1} \left(W(t_{j+1})^2 - W(t_j)^2 - (W(t_{j+1}) - W(t_j))^2\right) \\
&= \frac{1}{2}\left(W(T)^2 - W(0)^2 - \sum_{j=0}^{N-1}(W(t_{j+1}) - W(t_j))^2\right)
\end{aligned}
$$

Last term has expected value T and variance $O(\delta t)$, Thus

$$\int_0^T W(t)dW(t) = \frac{1}{2}W(T)^2 - \frac{1}{2}T$$

for the Itô integral

Geometric Wiener Process

$$dX = X(\mu dt + \sigma dW)$$

$$X(t) = \exp(\sigma W(t) + (\mu - \frac{1}{2}\sigma^2)t)$$

An essential part of the Black-Scholes model for option pricing