Stochastic Differential Equations

Continuous Evolving Variables

I. Holmes

Department of Bioengineering
University of California, Berkeley

Spring semester

Outline

- Review of salient facts about Gaussian distributions (Gardiner p36-37)
  - Multivariate Gaussian: if $\mathbf{x}$ is a vector of n Gaussian r.v.s,

$$P(\mathbf{x}) = [2\pi \det(\sigma)]^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T \sigma^{-1} (\mathbf{x}-\bar{\mathbf{x}})\right)$$

    where $\bar{\mathbf{x}}$ is mean and σ is (symmetric) covariance matrix.
  - Characteristic function

$$\phi(\mathbf{s}) = \langle \exp(i\mathbf{s}^T\mathbf{x})\rangle = \exp(i\mathbf{s}^T\bar{\mathbf{x}} - \frac{1}{2}\mathbf{s}^T\sigma\mathbf{s})$$

- General formulae for moments when $\bar{\mathbf{x}} = 0$: odd moments are zero, higher moments satisfy

$$\langle x_i x_j x_k \ldots \rangle = \frac{2N!}{N!2^N} \{\sigma_{ij}\sigma_{kl}\sigma_{mn}\ldots\}\text{sym}$$

where "sym" means the symmetrized form of the product of σ's, and $2N$ is the order of the moment, e.g.

$$\begin{aligned}
\langle x_i x_j \rangle &= \sigma_{ij} \\
\langle x_1 x_2 x_3 x_4 \rangle &= \frac{4!}{2!2^2} \left\{ \frac{1}{3} [\sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} \right\} \\
&= \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} \\
\langle x_i^4 \rangle &= 3\sigma_{ii}^2
\end{aligned}$$

Central limit theorem (van Kampen p26): consider arbitrary $P_X(x)$ with $\langle x \rangle = 0$, $\langle x^2 \rangle = \sigma$ and let $z = n^{-1/2} \sum_n x_n$
Characteristic function for P_X is

$$G_X(k) = \int \exp(ikx) P_X(x) dx = 1 - \frac{1}{2} k^2 \sigma + O(k^4)$$

Thus characteristic function for P_Z is

$$G_Z(k) = \left[G_X \left(\frac{k}{\sqrt{n}} \right) \right]^n = \left[1 - \frac{\sigma k^2}{2r} + O\left(\frac{k^4}{r^{3/2}} \right) \right]^n \rightarrow \exp(-\frac{1}{2} \sigma k^2)$$

(using the limit $\lim_{n \to \infty} (1 + y/n)^{-n} = \exp(-y)$).
Therefore, in the limit $n \to \infty$, z is Gaussian-distributed.

- Definition of a Gaussian process (van Kampen p63-64)
  - "Hierarchy of Distribution Functions" (van Kampen p61+). Consider timepoints $t_1 < t_2 < t_3 \ldots t_n$. Define

    $$P_n(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) \equiv P(x(t_1) = x_1, x(t_2) = x_2, \ldots, x(t_n) = x_n$$

  - If P_n is an n-dimensional Gaussian $\forall n, \{t_1 \ldots t_n\}$, then $x(t)$ is a *Gaussian process*
  - The covariance matrix is $\sigma_{ij} = \langle x(t_i) x(t_j) \rangle$
  - Marginals of a multivariate Gaussian are themselves multivariate Gaussians. The full distribution $P(x(t))$ can be thought of as an infinite-dimensional Gaussian, P_∞
  - A Gaussian process is effectively a prior over functions, that can be fully specified by the covariance function

- The *characteristic functional*, $G([k])$, plays a role analogous to the characteristic function for discrete processes. Define an arbitray auxiliary test function, $k(t)$. Then $G([k])$ is the following functional of $k(t)$

$$G([k]) = \langle \exp\left[i\int_{-\infty}^{\infty} k(t)x(t)dt\right]\rangle = \exp\left[i\int k(t_1)\langle x(t_1)\rangle dt_1 - \frac{1}{2}\int$$

- Inference, prediction, clustering with GPs (MacKay chapter 45, p535-548; MacKay 1998, "Introduction to Gaussian Processes")
  - Suppose we have N datapoints, $\{\mathbf{x}^{(n)}, t_n\}_{n=1}^{N}$. The input variables $\mathbf{x}^{(n)}$ are I-dimensional vectors. The target variables t_n will be assumed real scalars (corresponding to interpolation or regression problems).
  - Goal: fit some (nonlinear) function $y(\mathbf{x})$. Posterior probability of $y(\mathbf{x})$ is

$$P(y(\mathbf{x})|\mathbf{t}_N, \mathbf{X}_N) = \frac{P(\mathbf{t}_N|y(\mathbf{x}), \mathbf{X}_N)P(y(\mathbf{x}))}{P(\mathbf{t}_N|\mathbf{X}_N)}$$

Typically $t_k = y(x_k) +$ separable Gaussian noise.

$$P(y(\mathbf{x})|\mathbf{t}_N, \mathbf{X}_N) = \frac{P(\mathbf{t}_N|y(\mathbf{x}), \mathbf{X}_N)P(y(\mathbf{x}))}{P(\mathbf{t}_N|\mathbf{X}_N)}$$

- In parametric approaches, $y(\mathbf{x}) \equiv y(\mathbf{x}; \mathbf{w})$ where $\mathbf{w}$ is a set of parameters over which we place some prior. In nonparametric approaches (e.g. Gaussian processes), we place a prior directly on $P(y(\mathbf{x}))$.
- A Gaussian process can be defined as a probability distribution over functions, $P(y(\mathbf{x}))$, of the form

$$P(y(\mathbf{x})|\mu(\mathbf{x}), \mathbf{A}) = \frac{1}{Z} \exp\left[-\frac{1}{2}(y(\mathbf{x}) - \mu(\mathbf{x}))^T \mathbf{A}(y(\mathbf{x}) - \mu(\mathbf{x}))\right]$$

where $\mathbf{A}$ is a linear operator and the inner product of two functions is

- Parametric approaches; fixed, adaptive basis functions; neural nets (MacKay p536-537)
  - Consider a set of basis functions, $\{\phi_h(\mathbf{x})\}_{h=1}^{H}$.
  - Case #1: fixed basis functions (parameters independent of $\mathbf{w}$)

$$y(\mathbf{x}; \mathbf{w}) = \sum_{h=1}^{H} w_h \phi_h(\mathbf{x})$$

    e.g. radial basis functions

$$\phi_h(\mathbf{x}) = \exp\left[-\frac{(\mathbf{x} - \mathbf{c}_h)^2}{2r^2}\right]$$

    In this model, y is a linear function of $\mathbf{w}$.
    - Let $R_{nh} = \phi_h(\mathbf{x}^{(n)})$. Then $y^{(n)} = \sum_h R_{nh} w_h$. Let $\mathbf{y} = (y^{(1)}, y^{(2)} \dots y^{(N)})$ be the vector of y-values and let $\mathbf{w} = (w^{(1)}, w^{(2)} \dots w^{(N)})$ be the vector of corresponding

- Kramers-Moyal expansion (treatment follows van Kampen p197-198; see also Gillespie p74+)
  - The most general form of the *master equation* for a continuous-time stochastic process can be written

    $$\frac{\partial}{\partial t}p(x,t) = \int W(x-r;r)p(x-r,t)dr - p(x,t)\int W(x;r)dr$$

    where $W(x;r)$ is the rate from x to $x+r$. In the notation we used for discrete state spaces, $W(x;r) \equiv R_{x,x+r}$
  - Assuming that $W(x;r)$ varies smoothly in x and is sharply peaked in r, we can write the term $W(x-r;r)p(x-r,t)$ in the first integral as a Taylor expansion in x:

    $$\frac{\partial}{\partial t}p(x,t) = \sum_{n=0}^{\infty}\int \frac{(-r)^n}{n!}\frac{\partial^n}{\partial x^n}\left\{W(x;r)p(x,t)\right\}dr - p(x,t)\int W(x;r$$

    (Note that we're only allowed to expand $W(x;r)$ in x, not in

The Wiener process (undamped Brownian motion, diffusive drift, limit of random walk...)

- Derivation of Fick's equations for one-dimensional diffusion (Berg, "Random Walks in Biology", p18-20)
  - Discrete random walk: $x(n) = \sum_{i=1}^{n} d_i$ where $P(d_i = +\delta) = P(d_i = -\delta) = 1/2$
    - Implies that $\langle x(n) \rangle = 0$ and $\langle x(n)^2 \rangle = n\delta^2$
    - If each step takes time τ then $n = t/\tau$, so $\langle x(n)^2 \rangle = \frac{\delta^2}{\tau} t = 2Dt$ where $D = \delta^2/2\tau$ is the diffusion constant
    - Let $r(x,t) = P(x(t) = x)$. In time τ, a particle at x has probability $1/2$ of drifting to $x + \delta$, and a particle at $x + \delta$ has probability $1/2$ of drifting to x. The net flux of probability mass from x to $x + \delta$ is

$$J(x) = \frac{1}{\tau}\left(\frac{r(x,t)}{2} - \frac{r(x+\delta,t)}{2}\right) = D\frac{1}{\delta}\left(\frac{r(x,t)}{\delta} - \frac{r(x+\delta,t)}{\delta}\right)$$

The Ornstein-Uhlenbeck process: Brownian motion with exponential decay (van Kampen p83-85)

- Originally constructed to describe the *velocity* of a Brownian particle (van Kampen p84)
- Fokker-Planck equation (Gardiner p74-77)

$$\frac{\partial}{\partial t}p(x,t) = \frac{\partial}{\partial x}(kxp(x,t)) + \frac{1}{2}D\frac{\partial^2}{\partial x^2}p(x,t)$$

  Boundary condition is $p(x,0) = \delta(x - x_0)$.
- Characteristic equation for $\phi(s,t) = \langle \exp(\imath sx) \rangle$

$$\frac{\partial}{\partial t}\phi(s,t) + ks\frac{\partial}{\partial s}\phi(s,t) = -\frac{1}{2}Ds^2\phi(s,t) \tag{2}$$

  Boundary condition is $\phi(s,0) = \exp(\imath sx_0)$.
  (Here we have used $\int \exp(\imath sx)\frac{\partial}{\partial x}(xp)dx =$

- Case study of an Ornstein-Uhlenbeck process in stochastic systems biology: the enzyme futile cycle
    - Samoilov M, Plyasunov S, Arkin AP. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc Natl Acad Sci U S A. 2005 Feb 15;102(7):2310-5.
- Multivariate Ornstein-Uhlenbeck process (Gardiner p109-112)
- Case study of inference using a multivariate OU process: relationship between CD4 and beta-2-microglobulin in AIDS patients
    - Sy JP, Taylor JM, Cumberland WG. A stochastic model for the analysis of bivariate longitudinal AIDS data. Biometrics. 1997 Jun;53(2):542-55.

- Felsenstein, chapter 23 (p391-414)
- Consider tree (.x_1. (.x_2. .x_4. .x_5.) (.x_3. .x_6. .x_7.)) where x_n are Brownian variables.
- For a (parent,child) pair (p, c) let t_c be distance from p to c and let $d_c = x_c - x_p$. We have $\langle d_c \rangle = 0$ and $\langle d_c^2 \rangle = Dt_c$.
- Covariance matrix
- Pruning algorithm

Summary

- SCFGs