Stochastic Differential Equations

Continuous Evolving Variables

I. Holmes

Department of Bioengineering
University of California, Berkeley

Spring semester

Outline

Texts:

- Stochastic Processes in Physics and Chemistry.
  N.G. Van Kampen
- Stochastic Methods: A Handbook for the Natural and Social Sciences.
  C. Gardiner
- Information Theory, Inference, and Learning Algorithms.
  D. MacKay
- Gillespie
- Berg

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

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

(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 $k(t)$

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

$$= \exp\left[i \int k(t_1)\langle x(t_1)\rangle dt_1 - \frac{1}{2} \int \int k(t_1)k(t_2)\langle\langle x(t_1)x(t_2)\rangle\rangle dt_1 dt_2 \right]$$

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

$$y(\mathbf{x})^T z(\mathbf{x}) = \int y(\mathbf{x}) z(\mathbf{x}) d\mathbf{x}$$

The operator $\mathbf{A}$ must be *positive definite*, i.e. $y(\mathbf{x})^T \mathbf{A} y(\mathbf{x}) > 0$ for all functions except $y(\mathbf{x}) = 0$.

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 indep. 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 w-values. Thus $\mathbf{y} = \mathbf{R}\mathbf{w}$. If $\mathbf{w}$ is Gaussian-distributed

$$P(\mathbf{w}) = \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I})$$

then $\mathbf{y}$ is also Gaussian with covariance matrix

$$\langle \mathbf{y}\mathbf{y}^T \rangle = \langle \mathbf{R}\mathbf{w}\mathbf{w}^T\mathbf{R}^T \rangle = \mathbf{R}\langle \mathbf{w}\mathbf{w}^T \rangle \mathbf{R}^T = \sigma_w^2 \langle \mathbf{R}\mathbf{R}^T \rangle$$

Additive noise: if $\mathbf{t} = \mathbf{y} + \mathbf{v}$ where $v_k \sim \mathcal{N}(0, \sigma_v^2)$ then

$$P(\mathbf{w}) = \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{R}\mathbf{R}^T + \sigma_v^2 \mathbf{I})$$

Case #2: adaptive basis functions (parameters dependent on $\mathbf{w}$)

$$y(\mathbf{x};\mathbf{w}) = \sum_{h=1}^{H} w_h^{(2)} \tanh\left(\sum_{i=1}^{I} w_{hi}^{(1)} x_i + w_{h0}^{(1)}\right) + w_0^{(2)}$$

This is equivalent to a two-layer feedforward neural network with nonlinear hidden units and a linear output. The input weights are $\{w_{hi}^{(1)}\}$, the hidden unit biases $\{w_{h0}^{(1)}\}$, the output weights $\{w_h^{(2)}\}$ and the output bias $w_0^{(2)}$.
In this model, y is a nonlinear function of $\mathbf{w}$.

Nonparametric approaches: the spline smoothing method (MacKay p538-541) attempts to minimize the functional

$$M(y(x)) = \frac{1}{2}\beta \sum_{n=1}^{N} (y(x^{(n)}) - t_n)^2 + \frac{1}{2}\alpha \int \left[\frac{d^k y}{dx^k}\right]^2 dx$$

(If $k = 2$ then $y = \text{argmin} M$ is a *cubic spline* with discontinuities in $\frac{d^2 y}{dx^2}$ at the $x^{(n)}$.)

$$M(y(x)) = \frac{1}{2}\beta \sum_{n=1}^{N} (y(x^{(n)}) - t_n)^2 + \frac{1}{2}\alpha \int \left[\frac{d^k y}{dx^k}\right]^2 dx$$

The term involving α is equivalent to the following prior over $y(x)$

$$P(y(x)|\alpha) = \text{const.} \times \exp\left(-\frac{1}{2}\alpha \int \left[\frac{d^k y}{dx^k}\right]^2 dx\right)$$

which is a Gaussian process prior with $\mathbf{A} = [D^k]^T D^k$. Combined with linearly independent Gaussian noise on each measurement, this gives a Gaussian process model with MAP estimates identical to those produced by splines.

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, $W(x;r) \equiv R_{x,x+r}$

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

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:

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

(Note that we're only allowed to expand $W(x;r)$ in x, not in r, since it varies smoothly in x but rapidly in r.)

We then rewrite the terms in the expansion using the *jump moments*

$$a_n(x) = \int_{-\infty}^{\infty} r^n W(x; r) dr$$

so that the master equation becomes the Kramers-Moyal equation

$$\begin{aligned} \frac{\partial}{\partial t} p(x,t) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \{a_n(x) p(x,t)\} - p(x,t) \int W(x; r) dr \\ &= \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \{a_n(x) p(x,t)\} \end{aligned}$$

Truncating the Taylor expansion to second order gives

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left\{a_1(x)p(x,t)\right\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left\{a_2(x)p(x,t)\right\}$$

which is a form of the Fokker-Planck equation; see below.

Consider the discrete-time process x_n where $t = n\tau$. We have

$$x_{n+1} = x_n + \Xi_n$$

where the Ξ_n are random variables distributed $\sim W(x_n; \Xi)\tau$. Whatever the precise form of $W(x; r)$, we're effectively assuming that we can characterize it (and hence Ξ_n) by its first two moments, a_1 and a_2. Since $x_n = \sum \Xi_n$, the process x_n and hence $x(t)$ tends towards a Gaussian, by the central limit theorem.

Fokker-Planck equation (Gillespie p121; van Kampen p193+)
Fokker-Planck describes the time evolution of the probability density for a continuous stochastic process

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}A(x,t)p(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x,t)p(x,t)$$

By comparison with the Kramers-Moyal expansion we see that $A = a_1$ and $B = a_2$, so A and B are the mean and variance of the drift (i.e. the jump rate $W(x;r)$).
When $B = 0$, we have a (deterministic) Liouville process. When $A = 0$ and B is constant, we have Brownian motion, aka the Wiener process. When $A = -kx$ and B is constant, we have Brownian motion with exponential decay, aka the Ornstein-Uhlenbeck process.

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}A(x,t)p(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x,t)p(x,t)$$

Note that the terms $A(x,t)$ and $B(x,t)$ are time-dependent, unlike our earlier treatment of the Kramers-Moyal expansion. This is just because we didn't allow the jump rate $W(x;r)$ to be a function of t. It's straightforward to repeat the Kramers-Moyal expansion using time-dependent jump rates and moments, $W(x;r;t)$ and $a_n(x;t)$.

Using the (time-dependent) propagator we have $x(t+dt) = x(t) + \Xi(dt; x(t), t)$ and hence

$$\begin{aligned} x^n(t+dt) &= [x(t) + \Xi(dt; x(t), t)]^n \\ &= x^n(t) + \sum_{k=1}^{n} \binom{n}{k} \times x^{n-k}(t)\Xi^k(dt; x(t), t) \end{aligned}$$

To find the expectation of this we use the following result

$$\langle x^j(t)\Xi^k(dt; x(t), t)\rangle = \langle x^j(t)b_k(t)\rangle dt + O(dt) \qquad (1)$$

(where $b_k(t) \equiv a_k(x(t))$ is just an alternate notation for the k'th jump moment).

Hence we arrive at the *moment evolution equations*

$$\frac{d}{dt}\langle x^n(t)\rangle = \sum_{k=1}^{n} \binom{n}{k} \langle x^{n-k}(t) b_k(t)\rangle$$

with the initial conditions $\langle x^n(0)\rangle = x_0^n$. A further application of (1) gives the evolution of the autocorrelation function

$$\begin{aligned}
\langle x(t_1)x(t_2+dt_2)\rangle &= \langle x(t_1)x(t_2) + x(t_1)\Xi(dt; x(t_2), t_2)\rangle \\
&= \langle x(t_1)x(t_2)\rangle + \langle x(t_1)b_1(t_2)\rangle \\
\frac{d}{dt_2}\langle x(t_1)x(t_2)\rangle &= \langle x(t_1)b_1(t_2)\rangle
\end{aligned}$$

Putting these together, we obtain the following equations for the evolution of the mean, variance and covariance

$$\begin{aligned}
\frac{d}{dt}\langle x(t)\rangle &= \langle b_1(t)\rangle \\
\frac{d}{dt}\mathrm{var}\{x(t)\} &= \frac{d}{dt}\left(\langle x^2(t)\rangle - \langle x(t)\rangle^2\right) \\
&= 2\left(\langle x(t)b_1(t)\rangle - \langle x(t)\rangle\langle b_1(t)\rangle\right) + \langle b_2(t)\rangle \\
\frac{d}{dt_2}\mathrm{cov}\{x(t_1)x(t_2)\} &= \frac{d}{dt_2}\left(\langle x(t_1)x(t_2)\rangle - \langle x(t_1)\rangle\langle x(t_2)\rangle\right) \\
&= \langle x(t_1)b_1(t_2)\rangle - \langle x(t_1)\rangle\langle b_1(t_2)\rangle
\end{aligned}$$

with the initial conditions $\langle x(t_0)\rangle = x_0$, $\mathrm{var}\{x(t_0)\} = 0$, $\mathrm{cov}\{x(t_1)x(t_2 = t_1)\} = \mathrm{var}\{x(t_1)\}$.
These equations are closed iff (for $n = 1, 2$) $b_n(t) = a_n(x(t))$ is a polynomial in x of degree $\leq n$ (Gillespie p86).

The covariance is all we need to do inference with Gaussian processes, but it's useful to look deeper.
Liouville processes (Gillespie p126)
When $B(x,t) = 0$, the jump rate $W(x;r)$ has no variance. It must therefore be a delta function in r, and we can write

$$x(t+dt) - x(t) = A(x(t),t)dt$$

that is, $\frac{d}{dt} = A(x,t)$. This is a completely deterministic process, or *Liouville process*. The Liouville process guides intuition as to the role of $A(x,t)$ in the Fokker-Planck equation.

Discrete random walk
$x(n) = \sum_{i=1}^{n} d_i$ where $P(d_i = +\delta) = P(d_i = -\delta) = 1/2$
Implies that

$$\begin{aligned} \langle x(n) \rangle &= 0 \\ \langle x(n)^2 \rangle &= n\delta^2 \end{aligned}$$

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 continuous limit of $r(x,t)/\delta$ is the probability density $p(x,t)$ and so

$$J(x) = -D\frac{\partial p}{\partial x}$$

This is a version of *Fick's first equation* (Berg p18).

Consider the interval from x to $x + \delta$. Flux from the left is $J(x)$ and from the right $-J(x + \delta)$. Since probability mass is conserved, we have

$$[p(x, t + \tau) - p(x, t)] \times \delta = [J(x) - J(x + \delta)] \times \tau$$

Dividing through by $\delta\tau$, taking the continuous limit and substituting the expression for $J(x)$, we have

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}$$

This is *Fick's second equation* (Berg p20), aka the master equation (or Fokker-Planck equation) for Brownian motion.

In the continuous limit ($\tau \to 0$), $x(t)$ is the sum of a large number ($n = t/\tau$) of IID rvs (the d_i's). By the central limit theorem, the finite-time increments $x(t_2) - x(t_1)$ must be Gaussian-distributed, so $x(t)$ is a Gaussian process. Furthermore, $x(t_2) - x(t_1)$ is independent of $x(t_3) - x(t_2)$. This is called *independence of increments*.

Fokker-Planck equation for Wiener process (Gardiner p66-70). Choose units such that $\delta^2 = \tau, D = \frac{1}{2}$. Then

$$\frac{\partial}{\partial t}p(x,t) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p(x,t) \qquad p(x,0) = \delta(x)$$

Fourier-transforming the Fokker-Planck equation gives for the characteristic function $\phi(s,t) = \langle \exp(\imath s x(t)) \rangle_{x(t)}$

$$\frac{\partial}{\partial t}\phi = -\frac{1}{2}s^2\phi \qquad \phi(s,0) = 1$$

which has the solution $\phi(s,t) = \exp(-\frac{1}{2}s^2 t)$. The Fourier inversion gives

$$p(x,t) = (2\pi t)^{-1/2} \exp\left(-\frac{x^2}{2t}\right)$$

Autocorrelation function:

$$\langle x(t_1)x(t_2)\rangle = \langle x(t_1)^2\rangle + \langle x(t_1)(x(t_2) - x(t_1))\rangle = t_1$$

since the second expectation vanishes due to independence of increments.

Non-differentiable/fractal nature of sample trajectories:

$$\begin{aligned} P(|x(t+\epsilon)-x(t)|>k\epsilon) &= 2\int_{k\epsilon}^{\infty} p(x,\epsilon)dx \\ &= 2\int_{k\epsilon}^{\infty} (2\pi\epsilon)^{-1/2}\exp\left(-\frac{x^2}{2\epsilon}\right)dx \end{aligned}$$

In the limit $\epsilon \to 0$, this is one (intuitively, $|x(\epsilon)|_{\text{rms}} \sim \sqrt{\epsilon}$, so ϵ approaches zero faster than x does). This means that the derivative of $x(t)$ is infinite "almost surely" (in probabilistic terminology).

The Ornstein-Uhlenbeck process: random walk with damping.
Describes *velocity* of Brownian particle.

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

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

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

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

$$\frac{\partial}{\partial t}\phi(s,t) + ks\frac{\partial}{\partial s}\phi(s,t) = -\frac{1}{2}Ds^2\phi(s,t) \tag{3}$$

Method of characteristics: let $t = t(r)$ and $s = s(r)$

$$\frac{d\phi}{dr} = \frac{\partial\phi}{\partial t}\frac{dt}{dr} + \frac{\partial\phi}{\partial s}\frac{ds}{dr}$$

which is the same as (3) if

$$dr = \frac{dt}{1} = \frac{ds}{ks} = -\frac{d\phi}{\frac{1}{2}Ds^2\phi}$$

(which Gardiner calls the “subsidary equation”).

$$dr = \frac{dt}{1} = \frac{ds}{ks} = -\frac{d\phi}{\frac{1}{2}Ds^2\phi}$$

Integrating the equation involving ds and dt, and the equation involving ds and $d\phi$, gives

$$s = a\exp(kt) \qquad \phi = b\exp(-Ds^2/4k)$$

where a, b are arbitrary constants. This can be rearranged to give the "characteristic directions"

$$\begin{aligned} u(s,t,\phi) &= s\exp(-kt) &&= a \\ v(s,t,\phi) &= \phi\exp(Ds^2/4k) &&= b \end{aligned}$$

$$\begin{array}{rcccc} dr & = & \frac{dt}{1} & = & \frac{ds}{ks} & = & -\frac{d\phi}{\frac{1}{2}Ds^2\phi} \\ u(s,t,\phi) & = & s\exp(-kt) & = & a \\ v(s,t,\phi) & = & \phi\exp(Ds^2/4k) & = & b \end{array}$$

By the subsidary equation, $du = dv = 0$. Thus a general solution is $f(u, v) = 0$ with f an arbitrary function, or (equivalently) $v = g(u)$ with g an arbitrary function. Therefore

$$\phi(s,t) = \exp(-Ds^2/4k)g\left[s\exp(-kt)\right]$$

The boundary condition $\phi(s,0) = \exp(\imath s x_0)$ requires that $g(s) = \exp(Ds^2/4k + \imath s x_0)$. Hence

$$\phi(s,t) = \exp\left(-\frac{Ds^2}{4k}(1-\exp(-2kt)) + \imath s x_0 \exp(-kt)\right)$$

which is the characteristic function of a Gaussian with

$$\begin{aligned} \langle x(t)\rangle &= x_0 \exp(-kt) \\ \text{var}\{x(t)\} &= \frac{D}{2k}(1-\exp(-2kt)) \end{aligned}$$

Thus

$$p(x,t) = \left(\frac{\pi D}{k}(1-\exp(-2kt))\right)^{-1/2} \exp\left[-\frac{k}{D}\frac{(x-x_0\exp(-kt))^2}{1-\exp(-2kt)}\right]$$

The canonical form has $D = 2k$ (van Kampen p83).

The Langevin equation is a stochastic differential equation

$$dx = A(x,t)dt + B(x,t)dW(t)$$

where $dW(t)$ is a "white noise" term derived from the Wiener process $W(t)$, for which $\langle W(t)W(t+u)\rangle = t$.
This is equivalent to the Fokker-Planck equation

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}A(x,t)p(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x,t)p(x,t)$$

The key to the equivalence is the property of the Wiener process that

$$\langle dW(t)\rangle^2 = dt$$

See e.g. Gardiner pp95-96.