

# MINI-COURSE ON STOCHASTIC PROCESSES

## Contents

<b>1 Stochastic Process</b>	<b>2</b>
<b>2 Wiener process</b>	<b>3</b>
<b>3 Stochastic differential equations</b>	<b>7</b>
<b>4 Stochastic calculus</b>	<b>8</b>
<b>5 White noise</b>	<b>13</b>
<b>6 Scaling properties</b>	<b>14</b>
<b>7 Estimators of a process</b>	<b>14</b>
<b>8 Processes with independent increments</b>	<b>15</b>

<b>9 Gaussian processes</b>	<b>17</b>
<b>10 Stationary processes</b>	<b>18</b>
<b>11 Markov processes</b>	<b>20</b>
<b>12 Markov chains</b>	<b>21</b>
<b>13 Markov diffusions</b>	<b>24</b>
<b>14 The Ornstein-Uhlenbeck process</b>	<b>25</b>

## 1 Stochastic Process

A **random function** is a family  $\{Y_t\}_{t \in \mathbb{T}}$  of random variables indexed by  $t \in \mathbb{T}$  and taking values in some space  $\Sigma$ , the **state space** of the process. If  $\mathbb{T} = \mathbb{R}$  or some bounded interval  $[0, T]$ , for  $T \in (0, +\infty)$  the random function is called **stochastic process** and  $\mathbb{T}$  is called the **time set** of the process. For some processes, time is assumed to be

the set of discrete time points  $t_0, t_1, t_2, \dots$ , where  $t_0 < t_1 < t_2 < \dots$ . If  $\mathbb{T}$  is  $\mathbb{R}^d$  or a subset thereof, the random function is called **random field**.  $\Sigma$  is called the state space of the random function.  $\Sigma$  could be fairly general, but one is mostly interested in the case in which it is a set with finitely many elements or the  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ ,  $d = 1, 2, \dots$ , or a subset thereof. A random function is specified by the **joint probabilities**

$$\rho(x_1, t_1; \dots; x_n, t_n) \quad (1)$$

of the random variables  $Y_{t_1}, \dots, Y_{t_n}$  for any set  $\{t_1, t_2, \dots, t_n\}$ .

## 2 Wiener process

In mathematics, the limiting processes  $W(t)$  defined by

$$X_\tau(t) \equiv \sum_{k=0}^{t/\tau} \Delta X_k \rightarrow W(t) \quad \text{as } \tau \rightarrow 0 \quad (2)$$

where  $\Delta X_k$  are i.i.d. random variable with mean 0 and unit variance is called the **Brownian motion** or **Wiener process**. Here is the formal definition:

### Proposition 1 (Wiener process)

A family of random variables  $\{W(t)\}$  indexed by the continuous variable  $t$  ranging over  $[0, \infty)$  is called Wiener process (or Brownian motion) if and only if it satisfies the following conditions:

- (i)  $W(0) = 0$ ;
- (ii) the increments  $W(t_i + \Delta t_i) - W(t_i)$  over an arbitrary finite set of disjoint intervals  $(t_i, t_i + \Delta t_i)$ , are independent random variables;
- (iii) for each  $s \geq 0$ ,  $t \geq 0$ ,  $W(t + \Delta t) - W(t)$  has normal distribution  $N(0, \Delta t)$ .

For each constant  $y$ , the process  $W_y(t) = W(t) + y$ , is called the Wiener process starting at  $y$ .

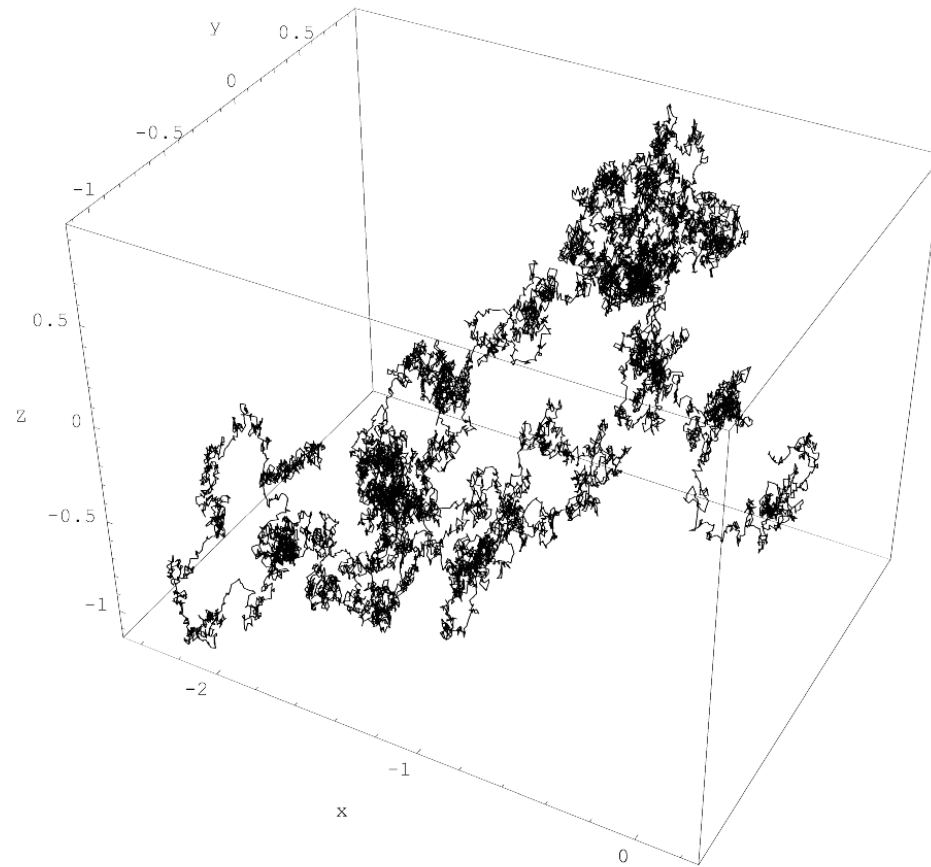
We have seen that the process constructed above by a limiting passage from symmetric random walks has the property (iii). Property (ii) comes from the fact that increments over disjoint intervals are obtained by summing the displacements  $\Delta X_k$  in disjoint blocks; hence the sums are independent.

The probability distribution of  $W(t)$  satisfies the equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \quad (3)$$

and thus the diffusion constant is  $1/2$ . Hence (exercise) a Brownian motion  $X(t)$  with diffusion constant  $D$  can be written as

$$X(t) = \sqrt{2D}W(t)$$



(Wiener process in 3 dimensions)

**Exercise** Consider the asymmetric random walk considered in the previous lecture. Show that the limit process is

$$X(t) = vt + \sqrt{2D}W(t). \quad (4)$$

### 3 Stochastic differential equations

When eq. (4) is written in differential form,

$$dX_t = vdt + \sqrt{2D}dW_t, \quad (5)$$

is a particular instance of **stochastic differential equation**, that is an equation of the type

$$dX_t = b(X_t, t)dt + a(X_t, t)dW_t, \quad (6)$$

While to the differential  $dW$  can be a mathematical meaning in terms of the so-called Ito integrals

$$\int f(W_t)dW_t,$$

(suitable limits of Riemann sums), the derivative

$$\dot{W}(t) = \frac{dW_t}{dt}$$

does not exist. This is due to the fact that the sample paths of the Wiener process are **nowhere differentiable** with probability 1. This is easily seen heuristically:

$$\dot{W} \sim \frac{\Delta W}{\Delta t} \sim \frac{\sqrt{\Delta t}}{\Delta t} \sim \frac{1}{\sqrt{\Delta t}}$$

which is infinite when  $\Delta t \rightarrow 0$

## 4 Stochastic calculus

To solve differential equations we need calculus. Stochastic calculus differs from standard calculus because  $(dW_t)^2$  is of order  $dt$ . So in computing the differential of a function of the process, say,  $f(X_t, t)$ , one need



to go to second order terms in the following way:

$$df = \frac{\partial f}{\partial X}dX + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(dX)^2$$

Now, at the lowest order,  $(dX)^2 = a^2dt$ , whence

$$\begin{aligned}df &= \frac{\partial f}{\partial X}dX + \frac{\partial f}{\partial t}dt + \frac{1}{2}a^2\frac{\partial^2 f}{\partial X^2}dt \\&= \frac{\partial f}{\partial X} [bdt + adW] + \frac{1}{2}a^2\frac{\partial^2 f}{\partial X^2}dt \\&= \left[ b\frac{\partial f}{\partial X} + \frac{1}{2}a^2\frac{\partial^2 f}{\partial X^2} \right] dt + adW\end{aligned}$$

This is the fundamental rule of the so-called ***Ito calculus***.

**Example** Solve the equation for ***geometric Brownian motion***

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \tag{7}$$

where  $\mu$  and  $\sigma$  are constants, which is the equation for the dynamics of the price of a stock in the Black Scholes options pricing model of financial mathematics.

Let's apply the technique of separation of variables, then the equation becomes

$$\frac{dX}{X} = \mu dt + \sigma dW$$

Consider the function  $f(X) = \log X$ . According to Ito calculus

$$\begin{aligned} d(\log X) &= \frac{dX}{X} - \frac{1}{2} \frac{1}{X^2} (dX)^2 \\ &= \mu dt + \sigma dW - \frac{1}{2} (\mu dt + \sigma dW)^2 \\ &= \mu dt + \sigma dW - \frac{1}{2} \sigma^2 dt \\ &= \left( \mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dW \end{aligned}$$

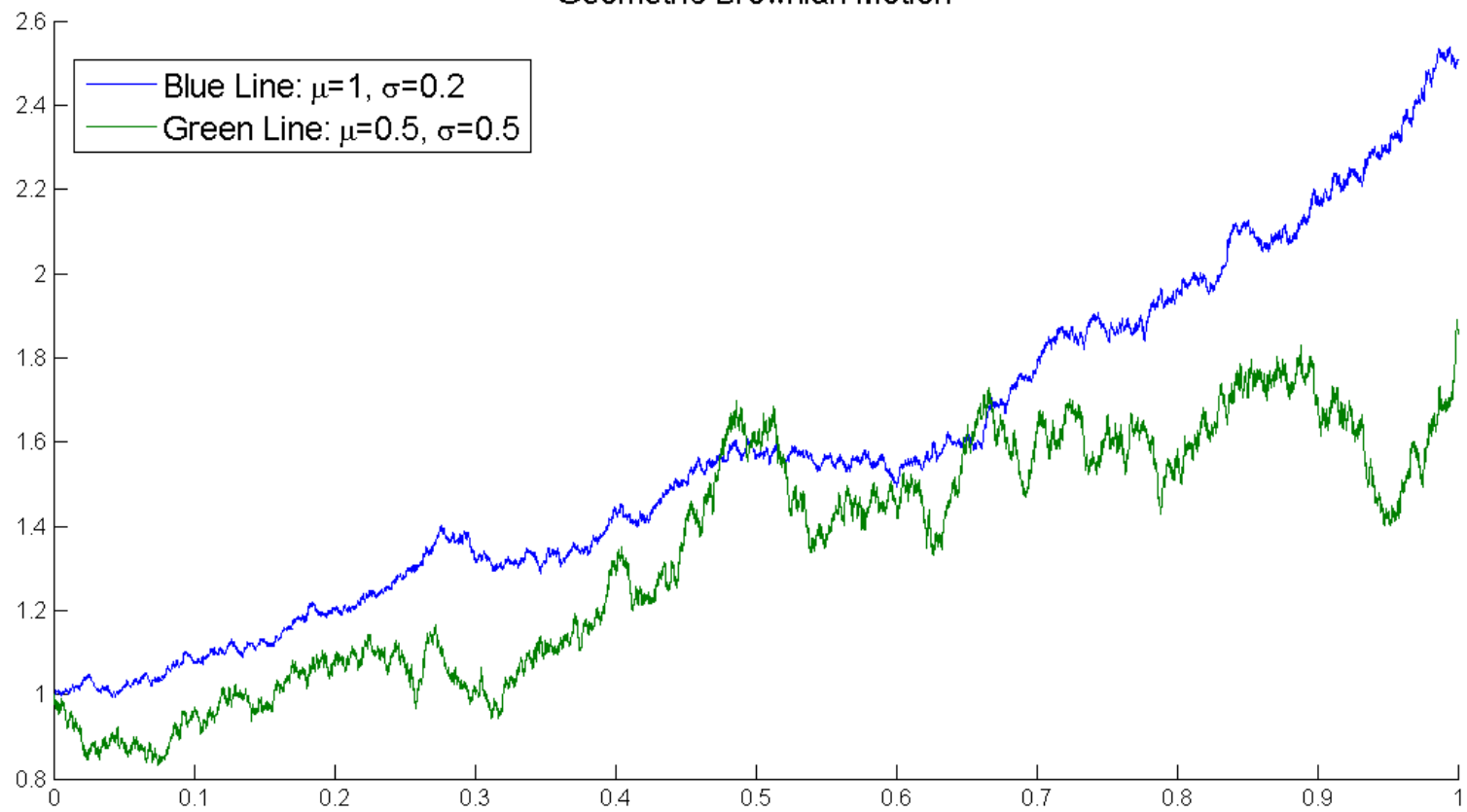
Now we can integrate:

$$\log X_t - \log X_0 = \left( \mu - \frac{1}{2}\sigma^2 \right) t + \sigma W_t$$

whence

$$X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

### Geometric Brownian Motion



## 5 White noise

It is useful to treat formally  $\dot{W}(t)$  as a process. From the definition of Wiener process we see that

(1) no correlation exists between the values of  $\dot{W}(t)$  at different time instants, i.e.,

$$\mathbb{E}\{\dot{W}(t)\dot{W}(s)\} = \delta(t - s); \quad (8)$$

(2) it is a Gaussian process;

(3) it has mean zero.

In brief,  $\dot{W}(t)$  is a **white noise**, that is a mean zero Gaussian process with covariance

$$C(t, s) = \delta(t - s). \quad (9)$$

## 6 Scaling properties

From its very definition, the Wiener process has the scaling property

$$W(\alpha t) = \sqrt{\alpha}W(t). \quad (10)$$

for any  $\alpha > 0$ . By (formal) differentiation of both sides

$$\alpha\dot{W}(\alpha t) = \sqrt{\alpha}\dot{W}(t),$$

whence the scaling property of white noise

$$\dot{W}(\alpha t) = \frac{1}{\sqrt{\alpha}}\dot{W}(t) \quad (11)$$

## 7 Estimators of a process

For both continuous and discrete time, it is useful to distinguish various classes of stochastic processes according to their specific temporal

relationships. Assuming that the expressions exist, the **expectations** and **variances**

$$\mathbb{E}\{Y_t\}, \quad \text{Var}(Y_t) \quad (12)$$

at each instant  $t \in \mathbb{T}$  and the **covariances**

$$C(t, s) = \mathbb{E}\{(Y_t - \mathbb{E}\{Y_t\})(Y_s - \mathbb{E}\{Y_s\})\} = \mathbb{E}\{Y_t Y_s\} - \mathbb{E}\{Y_t\} \mathbb{E}\{Y_s\} \quad (13)$$

at distinct instants  $s, t \in \mathbb{T}$  provide information about the time variability of a stochastic process. The quantity

$$R(t, s) = \mathbb{E}\{Y_t Y_s\}. \quad (14)$$

is usually called the **correlation function** of the process. It is not an independent estimator since

$$C(t, s) = R(t, s) - \mathbb{E}\{Y_t\} \mathbb{E}\{Y_s\}. \quad (15)$$

## 8 Processes with independent increments

An important class of stochastic processes are those with independent increments, that is for which the random variables  $Y_{t_{j+1}} - Y_{t_j}$ ,  $j = 0, 1, 2, \dots, n-$

1 are independent for any finite combination of time instants  $t_0 < t_1 < \dots < t_n$  in  $\mathbb{T}$ . If  $t_0$  is the smallest time instant in  $\mathbb{T}$  then the random variables  $Y_{t_0}$  and  $Y_{t_j} - Y_{t_0}$  for any other  $t_j \in \mathbb{T}$  are also required to be independent.

**Poisson process** A Poisson process with intensity  $\lambda$  is a process  $N_t, t \geq 0$ , with state space the positive integers, such that

- (i)  $N_0 = 0$ ,
- (ii)  $N_t - N_s$  is a Poisson distributed random variable with parameter  $\lambda(t-s)$  for all  $0 \leq s < t$ ,
- (iii) the increments  $N_{t_2} - N_{t_1}$  and  $N_{t_4} - N_{t_3}$  are independent for all  $0 \leq t_1 < t_2 \leq t_3 < t_4$ .

Its means, variances and covariances are

$$\mathbb{E}\{N(t)\} = \lambda t, \quad \text{Var}(N(t)) = \lambda t, \quad C(s, t) = \lambda \min\{s, t\}, \quad (16)$$

respectively, for all  $s, t > 0$ .



A Poisson process is an example of a continuous time stochastic process with independent increments. The process is named after the French mathematician Poisson and is a good model of radioactive decay and many other phenomena which involve the counting of the number of events in a given time interval.

## 9 Gaussian processes

When the the joint probability distributions (1) are all Gaussian the process is a *Gaussian process*.

**Wiener process** The standard Wiener processes  $W_t, t \geq 0$  is another example of process with independent increments. It can be equivalently defined as a Gaussian process with independent increments for which

$$W_0 = 0 \text{ (with probability 1), } \mathbb{E}\{W_t\} = 0, \quad \text{Var}(W_t - W_s) = t - s, \quad (17)$$

for all  $0 < s < t$ . We leave as an exercise to show that the covariance of the Wiener process is  $C(s, t) = \min\{s, t\}$ . Thus, the Wiener process is not

wide-sense stationary.

## 10 Stationary processes

A stochastic process whose joint probability distribution does not change when shifted in time is called a (strictly ) *stationary processes*. Consequently, parameters such as the mean and variance, if they are present, also do not change over time. For example, a sequence of i.i.d. random variables is strictly stationary. On the other hand, if there is a constant  $m$  and a function  $c(t)$  such that the means, variances and covariances of a stochastic process satisfy

$$\mathbb{E}\{Y(t)\} = m, \quad \text{Var}(Y(t)) = c(0) \quad \text{and} \quad C(t, s) = c(t - s) \quad (18)$$

for all  $s, t \in \mathbb{T}$  the process is said *wide-sense stationary*. This means the process is only stationary with respect to its first and second moments. Clearly, also the correlation function of such a process depends only on the time difference,

$$R(t, s) = \mathbb{E}\{Y_t Y_s\} = r(t - s) \quad (19)$$

with  $c(\tau) = r(\tau) - m^2$ . It is straightforward to show that a strictly stationary process is wide-sense stationary if its means, variances and covariances are all finite, but a wide-sense stationary process need not be strictly stationary.

Note that neither the Poisson process nor the Wiener process are wide-sense stationary.

It is important to pay attention to the fact that "stationarity" here is used in a probabilistic sense; it does not, of course, imply that the individual sample functions do not vary in time. For example, if one considers the fluctuations of a thermodynamic quantity, such as the pressure exerted by a gas on the walls of its container, this quantity varies in time even when the system is in thermodynamic equilibrium. The single-time probability distribution of the quantity is independent of time, but the correlation at different times depends on the time difference.

# 11 Markov processes

Conditional probabilities provide a more refined tool than mean values and covariances for analyzing the relationships between the random variables of a stochastic process at different time instants. In particular, they allow to characterize the processes that have the following property:

**Markov property:** *The future is independent from the past, given the present.*  
(20)

A process that has this property is called a *Markov process* or a *Markovian processes*.

Markov introduced this property and defined what are now called Markov chains (see below) to weaken the assumption of strict statistical independence between variables to mere conditional independence. Markov wanted to show that independence was not a necessary condition for the law of large numbers to hold. It turns out that all the classical limit theorems of probability—the laws of large numbers and the central limit theorem—work perfectly well for Markov processes, as

well as for i.i.d. variables.

Processes with independent increments are Markovian. Thus, the Poisson process and the Wiener processes are Markovian.

## 12 Markov chains

Consider a discrete time process  $Y_1, Y_2, \dots$  taking values in a finite set  $\Sigma = \{y_1, y_2, \dots, y_N\}$ . Let  $t = n$  be the present, so that  $Y_n$  is the state of the process in the present and  $Y_{n+1}$  its state in the immediate future; the past of the process is  $Y_1, Y_2, \dots, Y_{n-1}$ . Then (20) becomes

$$\mathbb{P} \{Y_{n+1} | Y_1, Y_2, \dots, Y_n\} = \mathbb{P} \{Y_{n+1} | Y_n\} . \quad (21)$$

and the the probability law of the process simplifies considerably:

$$\begin{aligned}
\mathbb{P} \{Y_{n+1}, Y_n, \dots, Y_2, Y_1\} &= \mathbb{P} \{Y_{n+1} | Y_1, Y_2, \dots, Y_n\} \mathbb{P} \{Y_1, Y_2, \dots, Y_n\} \\
&= \mathbb{P} \{Y_{n+1} | Y_n\} \mathbb{P} \{Y_n | Y_1, Y_2, \dots, Y_{n-1}\} \mathbb{P} \{Y_1, Y_2, \dots, Y_{n-1}\} \\
&= \mathbb{P} \{Y_{n+1} | Y_n\} \mathbb{P} \{Y_n | Y_{n-1}\} \mathbb{P} \{Y_1, Y_2, \dots, Y_{n-1}\} \\
&= \dots \\
&= \mathbb{P} \{Y_{n+1} | Y_n\} \mathbb{P} \{Y_n | Y_{n-1}\} \cdots \mathbb{P} \{Y_2 | Y_1\} \mathbb{P} \{Y_1\} .
\end{aligned}$$

Thus we see that given the family of  $N \times N$  matrices  $P(n) = [p^{ij}(n)]$ ,  $n = 0, 1, 2, \dots$ , with components

$$p^{ij}(n) = \mathbb{P} \{Y_{n+1} = y_j | Y_n = y_i\} \quad \text{for } i, j = 1, 2, \dots, N, \quad (22)$$

and the initial distribution of  $Y_1$ , all the distributions of the process are determined. The process so defined is called a *Markov chain* and the matrix  $P(n)$  is called its *transition matrix*.

Obviously,  $0 \leq p^{ij}(n) \leq 1$  and, since  $Y_{n+1}$  can only attain states in  $\Sigma$ ,

$$\sum_{j=1}^N p^{ij}(n) = 1$$

for each  $i = 1, 2, \dots, N$  and  $n = 1, 2, 3, \dots$ . A probability vector on  $\Sigma$  is a vector  $p = (p_1, p_2, \dots, p_N) \in \mathbb{R}^N$ , with  $0 \leq p_i \leq 1$ , for  $i = 1, 2, \dots, N$ , and  $\sum_i p_i = 1$ . Thus if  $p(n)$  is the probability vector corresponding to the distribution of the random variable  $Y_n$ , that is  $p^{ij}(n) = \mathbb{P}\{Y_{n+1} = Y_j | Y_n = Y_i\}$  for  $i = 1, 2, \dots, N$ , then the probability vector  $p(n+1)$  corresponding  $\mathbb{P}\{Y_{n+1}\}$  is related to it through the vector equation

$$p(n+1) = p(n)P(n). \quad (23)$$

Hence, if we know the initial distribution probability vector  $p(1)$ , then we have

$$p(n) = p(1)P(1)P(2) \cdots P(n-1) \quad (24)$$

for  $n = 2, 3, \dots$  by applying eq. (23) recursively. In the case that the transition matrices are all the same, that is  $P(n) = P$  for  $n = 1, 2, 3, \dots$ , the Markov chain is said *homogeneous*. For a homogeneous Markov chain eq. (24) takes the form

$$p(n) = p(1)P^{n-1}. \quad (25)$$

## 13 Markov diffusions

A Markov diffusion is the continuous version of a Markov chain and is defined by a conditional probability density or **transition density**

$$p(x, t|y, s)$$

in terms of which,  $\rho(x, t)$ , the probability law of the process at time  $t$  is obtained from the probability law at time  $s < t$ :

$$\rho(x, t) = \int p(x, t|y, s)\rho(y, s)dy \quad (26)$$

A Markov diffusion is characterized by two functions:

(1) the **drift coefficient**  $b = b(x, t)$

(2) the (square of the) **diffusion coefficient**  $D = D(x, t)$

that we shall assume to be regular functions. Then the process is defined by the stochastic differential equation

$$dX_t = b dt + a dW_t, \quad D = \frac{1}{2}a^2 \quad (27)$$



and the transition density of the process satisfy the Fokker -Planck equation

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial x}bp = \frac{\partial^2}{\partial x^2}Dp \quad (28)$$

with initial condition

$$\lim_{t \downarrow s} p(x, t|y, s) = \delta(x - y) \quad (29)$$

## 14 The Ornstein-Uhlenbeck process

It is a Gaussian process  $V_t$ ,  $t \geq 0$ , with state space  $\mathbb{R}$ , parameters  $\gamma > 0$  and  $\sigma$ , and

$$\mathbb{E}\{V_t\} = 0, \quad C(s, t) = \mathbb{E}\{V_s V_t\} = \frac{\sigma^2}{2\gamma} e^{-\gamma|t-s|}, \quad (30)$$

for all  $0 < s < t$ . Hence it is a wide-sense stationary process. It is also strictly stationary. The process is the Markovian diffusion

$$dV = -\gamma V dt + \sigma dW$$

and is the only nontrivial process that is stationary, Gaussian, and Markovian, up to allowing linear transformations of the velocity and time variables. If initially  $V_0$  is not gaussian distributed with mean 0, the process is not anymore stationary. However, over time, it will tend to drift towards the stationary process.