# A Short Introduction to Diffusion Processes and Ito Calculus

Cédric Archambeau

University College, London
Center for Computational Statistics and Machine Learning
`c.archambeau@cs.ucl.ac.uk`

January 24, 2007

Notes for the Reading Group on Stochastic Differential Equations (SDEs). The text is largely based on the book *Numerical Solution of Stochastic Differential Equations* by E. Kloeden & E. Platen (1992) [3]. Other good references are [4] and [2], as well as [1]

## 1   Elements of Measure and Probability Theory

In this section, we review some important concepts and definitions, which will be extensively used in the next sections.

**Definition 1.** A collection $\mathcal{A}$ of subsets of $\Omega$ is a *σ-algebra* if

$$\Omega \in \mathcal{A}, \tag{1}$$

$$A^c \in \mathcal{A} \text{ if } A \in \mathcal{A}, \tag{2}$$

$$\bigcup_n A_n \text{ if } A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A}. \tag{3}$$

This means that $\mathcal{A}$ is a collection of subsets of $\Omega$ containing $\Omega$ and which is closed under the set of operations of complementation and countable unions. Note that this implies that $\mathcal{A}$ is also closed under countable intersections.

**Definition 2.** Let $(\Omega, \mathcal{A})$ be a measurable space, i.e. $(\Omega, \mathcal{A})$ is an ordered pair consisting of a non-empty set $\Omega$ and a $\sigma$-algebra $\mathcal{A}$ of subsets of $\Omega$. A *measure* $\mu$ on $(\Omega, \mathcal{A})$ is a nonnegative valued set function on $\mathcal{A}$ satisfying

$$\mu(\emptyset) = 0, \tag{4}$$

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n), \tag{5}$$

for any sequence $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$.

From (5) it follows that $\mu(A) \leq \mu(B)$ for all $A \subseteq B$ in $\mathcal{A}$. The measure $\mu$ is finite if $0 \leq \mu(\Omega) \leq \infty$. Hence, it can be normalized to obtain a *probability measure* $P$ with $P(A) = \mu(A)/\mu(\Omega) \in [0,1]$ for all $A \in \mathcal{A}$.

An important measure is the *Borel measure* $\mu_B$ on the $\sigma$-algebra $\mathcal{B}$ of Borel subsets[1] of $\mathbb{R}$, which assigns to each finite interval its length. However, the measure space $(\mathbb{R}, \mathcal{B}, \mu_B)$ is not complete in the sense that there exist subsets $B^*$ of $\mathbb{R}$ with $B^* \notin \mathcal{B}$, but $B^* \subset B$ for some $B \in \mathcal{B}$ with $\mu_B(B) = 0$. Therefore, we enlarge the $\sigma$-algebra $\mathcal{B}$ to a $\sigma$-algebra $\mathcal{L}$ and extend the measure $\mu_B$ uniquely to the measure $\mu_L$ on $\mathcal{L}$ so that $(\mathbb{R}, \mathcal{L}, \mu_L)$ is complete, that is $L^* \in \mathcal{L}$ with $\mu_L(L^*) = 0$ whenever $L^* \subset L$ for some $L \in \mathcal{L}$ with $\mu_L = 0$. We call $\mathcal{L}$ the *Lebesgue subsets* of $\mathbb{R}$ and $\mu_L$ the *Lebesgue measure*.

**Definition 3.** Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be two measurable spaces. The function $f : \Omega_1 \to \Omega_2$ is $\mathcal{A}_1 : \mathcal{A}_2$-*measurable* if

$$f^{-1}(A_2) = \{\omega_1 \in \Omega_1 : f(\omega_1) \in A_2\} \in \mathcal{A}_1, \tag{6}$$

for all $A_2 \in \mathcal{A}_2$.

This means that the pre-image of any $A_2 \in \mathcal{A}_2$ is in $\mathcal{A}_1$.

**Definition 4.** Let $\Omega$ be the sample space, the $\sigma$-algebra $\mathcal{A}$ a collection of events and $P$ the associated probability measure. We call a triplet $(\Omega, \mathcal{A}, P)$ a *probability space* if $\mathcal{A}$ and $P$ satisfy the following properties:

$$A^c = \Omega \setminus A, \; A \cup B, \; A \cap B \in \mathcal{A} \text{ if } A, B \in \mathcal{A} \tag{7}$$

$$0 \leq P(A) \leq 1, \; P(A^c) = 1 - P(A), \; P(\emptyset) = 0, \; P(\Omega) = 1, \tag{8}$$

$$\textstyle\bigcup_n A_n, \; \bigcap_n A_n \in \mathcal{A} \text{ if } \{A_n, A_n \in \mathcal{A}\}, \tag{9}$$

$$P\left(\textstyle\bigcup_n A_n\right) = \sum_n P(A_n) \text{ if } \{A_n, A_n \in \mathcal{A}\} \text{ and } A_i \cap A_j = \emptyset \text{ for all } i \neq j. \tag{10}$$

The last two properties hold for a countably (finite or infinite) number of outcomes. For uncountably many basic outcomes we restrict attention to countable combinations of natural events, which are subintervals of $\Omega$ and to which non-zero probabilities are (possibly) assigned.

When $(\Omega_1, \mathcal{A}_1, P)$ is a probability space and $(\Omega_2, \mathcal{A}_2)$ is either $(\mathbb{R}, \mathcal{B})$ or $(\mathbb{R}, \mathcal{L})$, we call the measurable function $f : \Omega_1 \to \mathbb{R}$ a *random variable* and denote it usually by $X$.

## 2 Stochastic Processes

In this section, we review the general properties of standard stochastic processes and discuss Markov chains and diffusion processes.

**Definition 5.** Let $T$ denote the time set under consideration and let $(\Omega, \mathcal{A}, P)$ be a common underlying probability space. A *stochastic process* $X = \{X_t, t \in T\}$ is a function of two variables $X : T \times \Omega \to \mathbb{R}$, where

---

[1]Any subset of $\mathbb{R}$ generated from countable unions, intersections or complements of the semi-infinite intervals $\{x \in \mathbb{R} : -\infty < x \leq a\}$ is a *Borel subset* of $\mathbb{R}$.

- $X_t = X(t, \cdot) : \Omega \to \mathbb{R}$ is a random variable for each $t \in T$,

- $X(\cdot, \omega) : T \to \mathbb{R}$ is a realization or *sample path* for each $\omega \in \Omega$.

Depending on $T$ being a discrete or a continuous time set, we call the stochastic process a *discrete* or a *continuous* time process.

**Example.** A Gaussian process *is a stochastic process for which the probability law is Gaussian, that is any joint distribution* $F_{t_{i_1}, t_{i_2}, \ldots, t_{i_n}}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_n})$ *is Gaussian for all* $t_{i_j} \in T$.

The time variability of a stochastic process is described by all its conditional probabilities (see Section 2.1 and 2.2). However, substantial information can already be gained from by the following quantities:

- The *means*: $\mu_t = E\{X_t\}$ for each $t \in T$.

- The *variances*: $\sigma_t^2 = E\{(X_t - \mu_t)^2\}$ for each $t \in T$.

- The (two-time) *covariances*: $C_{s,t} = E\{(X_s - \mu_s)(X_t - \mu_t)\}$ for distinct time instants $s, t \in T$.

**Definition 6.** A stochastic process $X = \{X_t, t \in T\}$ for which the random variables $X_{t_{j+1}} - X_{t_j}$ with $j = 1, \ldots, n-1$ are independent for any finite combination of time instants $t_1 < \ldots < t_n$ in $T$ is a stochastic process with *independent increments*.

**Example.** A Poisson process *is a continuous time stochastic process* $X = \{X_t, t \in \mathbb{R}^+\}$ *with (non-overlapping) independent increments for which*

$$X_0 = 0 \ \text{ w.p. 1},$$
$$E\{X_t\} = 0,$$
$$X_t - X_s \sim \mathcal{P}(\lambda(t - s)),$$

*for all* $0 \le s \le t$ *and where* $\lambda$ *is the intensity parameter.*

As a consequence, the means, the variances and the covariances of a Poisson process are respectively given by $\mu_t = \lambda t$, $\sigma_t^2 = \lambda t$ and $C_{s,t} = \lambda \min\{s, t\}$.

**Property 6.1.** A stochastic process is *strictly stationary* if all its joint distributions are invariant under time displacement, that is $F_{t_{i_1}+h, t_{i_2}+h, \ldots, t_{i_n}+h} = F_{t_{i_1}, t_{i_2}, \ldots, t_{i_n}}$ for all $t_{i_j}, t_{i_{j+1}} \in T$ with $h \ge 0$.

This constraint can be relaxed to stationarity with respect to the first and the second moments only.

**Property 6.2.** A stochastic process $X = \{X_t, t \in \mathbb{R}^+\}$ is *wide-sense stationary* if there exists a constant $\mu \in \mathbb{R}$ and a function $c : \mathbb{R}^+ \to \mathbb{R}$, such that

$$\mu_t = \mu, \quad \sigma_t^2 = c(0) \quad \text{and} \quad C_{s,t} = c(t - s), \tag{11}$$

for all $s, t \in T$.

**Example.** *The* Ornstein-Uhlenbeck process *with parameter* $\gamma > 0$ *is a wide-sense stationary Gaussian process* $X = \{X_t, t \in \mathbb{R}^+\}$ *for which*

$$X_0 \sim \mathcal{N}(0, 1),$$
$$E\{X_t\} = 0,$$
$$C_{s,t} = e^{-\gamma|t-s|},$$

*for all* $s, t \in \mathbb{R}^+$.

**Definition 7.** Let $(\Omega, \mathcal{A}, P)$ be the probability space and $\{\mathcal{A}_t, t \geq 0\}$ an increasing family of sub-$\sigma$-algebras[2] of $\mathcal{A}$. The stochastic process $X = \{X_t, t \in \mathbb{R}^+\}$, with $X_t$ being $\mathcal{A}_t$-measurable for each $t \geq 0$, is a *martingale* if

$$E\{X_t|\mathcal{A}_s\} = X_s, \quad \text{w.p. 1}, \tag{12}$$

for all $0 \leq s < t$.

Equivalently, we can write $E\{X_t - X_s|\mathcal{A}_s\} = 0$. Thus, a stochastic process is a martingale if the expectation of some future event given the past and the present is always the same as if given only the present.

When the process $X_t$ satisfies the Markov property (see Sections 2.1 and 2.2), we have $E\{X_t|\mathcal{A}_s\} = E\{X_t|X_s\}$.

## 2.1 Markov chains

We first describe discrete time Markov chains and then generalize to their continuous time counterpart.

**Definition 8.** Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be the set of a finite number of discrete states. The discrete time stochastic process $X = \{X_t, t \in T\}$ is a *discrete time Markov chain* if it satisfies the Markov property, that is

$$P(X_{n+1} = x_j|X_n = x_{i_n}) = P(X_{n+1} = x_j|X_1 = x_{i_1}, \ldots, X_n = x_{i_n}) \tag{13}$$

for all possible $x_j, x_{i_1}, \ldots, x_{i_n} \in \mathcal{X}$ with $n = 1, 2, \ldots$.

This means that only the present value of $X_n$ is needed to determine the future value of $X_{n+1}$.

The entries of the *transition matrix* $\mathbf{P}_n \in \mathbb{R}^{N \times N}$ of the Markov chain are given by

$$p_n^{(i,j)} = P(X_{n+1} = x_j|X_n = x_i) \tag{14}$$

for $i, j = 1, \ldots, N$. We call them the *transition probabilities*. They satisfy $\sum_j^N p_n^{(i,j)} = 1$ for each $i$, as $X_{n+1}$ can only attain states in $\mathcal{X}$.

Let $\mathbf{p}_n$ be the column vector of the marginal probabilities $P(X_n = x_i)$ for $i = 1, \ldots, N$. The probability vector $\mathbf{p}_{n+1}$ is then given by $\mathbf{p}_{n+1} = \mathbf{P}_n^{\mathrm{T}} \mathbf{p}_n$.

---

[2]The sequence $\{\mathcal{A}_t, t \geq 0\}$ is called an *increasing family of sub-$\sigma$-algebras of* $\mathcal{A}$ if $\mathcal{A}_s \subseteq \mathcal{A}_t$ for any $0 \leq s \leq t$. This means that more information becomes available with increasing time.

**Property 8.1.** A discrete time Markov chain is *homogeneous* if $\mathbf{P}_n = \bar{\mathbf{P}}$ for all $n = 1, 2, \ldots$.

As a consequence, the probability vector of a homogenous Markov chain satisfies $\mathbf{p}_{n+k} = \left(\bar{\mathbf{P}}^k\right)^{\mathrm{T}} \mathbf{p}_n$ for any $k = \mathbb{N} \setminus \{0\}$. The probability distributions depend only on the time that has elapsed. However, this does not mean that the Markov chain is strictly stationary. In order to be so, it is also required that $\mathbf{p}_n = \bar{\mathbf{p}}$ for each $n = 1, 2, \ldots$, which implies that the probability distributions are equal for all times such that $\bar{\mathbf{p}} = \bar{\mathbf{P}}^{\mathrm{T}} \bar{\mathbf{p}}$.

It can be shown that a homogenous Markov chain has at least one stationary probability vector solution. Therefore, it is sufficient that the initial random variable $X_1$ is distributed according to one of its stationary probability vectors for the Markov chain to be strictly stationary.

**Property 8.2.** Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be the set of discrete states and $f : \mathcal{X} \to \mathbb{R}$. The discrete time homogeneous Markov chain $X = \{X_n, n = 1, 2, \ldots\}$ is *ergodic* if

$$\lim_{T \to \infty} \frac{1}{T} \sum_{n=1}^{T} f(X_n) = \sum_{i=1}^{N} f(x_i) \bar{p}^{(i)} \tag{15}$$

where $\bar{\mathbf{p}}$ is a stationary probability vector and all $x_i \in \mathcal{X}$.

This property is an extension of the Law of Large Numbers to stochastic processes. A sufficient condition for a Markov chain to be ergodic is that all the components of some $k^{\text{th}}$ power $\bar{\mathbf{P}}^k$ are nonzero. The same result holds if all the entries of the unique $\bar{\mathbf{p}}$ are nonzero.

**Definition 9.** Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be the set of a finite number of discrete states. The stochastic process $X = \{X_t, t \in \mathbb{R}^+\}$ is a *continuous time Markov chain* if it satisfies the following Markov property:

$$P(X_t = x_j | X_s = x_i) = P(X_t = x_j | X_{r_1} = x_{i_1}, \ldots, X_{r_n} = x_{i_n}, X_s = x_i) \tag{16}$$

for $0 \leq r_1 \leq \ldots \leq r_n < s < t$ and all $x_{i_1}, \ldots, x_{i_n}, x_i, x_j \in \mathcal{X}$.

The entries of the *transition matrix* $\mathbf{P}_{s,t} \in \mathbb{R}^{N \times N}$ and the probability vectors are now respectively given by $p_{s,t}^{(i,j)} = P(X_t = x_j | X_s = x_i)$ and $\mathbf{p}_t = \mathbf{P}_{s,t}^{\mathrm{T}} \mathbf{p}_s$ for any $0 \leq s \leq t$. Furthermore, the transition matrices satisfy the relationship $\mathbf{P}_{r,t} = \mathbf{P}_{r,s} \mathbf{P}_{s,t}$ for any $0 \leq r \leq s \leq t$.

**Property 9.1.** If all its transition matrices depend only on the time differences, then the continuous time Markov chain is *homogeneous* and we write $\mathbf{P}_{s,t} = \mathbf{P}_{0,t-s} \equiv \mathbf{P}_{t-s}$ for any $0 \leq s \leq t$.

Hence, we have $\mathbf{P}_{t+s} = \mathbf{P}_t \mathbf{P}_s = \mathbf{P}_s \mathbf{P}_t$ for all $s, t \geq 0$.

**Example.** *The Poisson process is a continuous time homogenous Markov chain on the countably infinite state space $\mathbb{N}$. Its transition matrix is given by*

$$\mathbf{P}_t = \frac{(\lambda t)^m}{m!} \, e^{-\lambda t}, \tag{17}$$

*where $m \in \mathbb{N}$.*

Indeed, invoking the independent increments of the Poisson process we get

$$
\begin{aligned}
P(X_s = m_s, X_t = m_t) &= P(X_s = m_s, X_t - X_s = m_t - m_s) \\
&= P(X_s = m_s)P(X_t - X_s = m_t - m_s) \\
&= \frac{\lambda^{m_s} s^{m_s}}{m_s!} e^{-\lambda s} \, \frac{\lambda^{m_t - m_s}(t-s)^{m_t - m_s}}{(m_t - m_s)!} e^{-\lambda(t-s)}.
\end{aligned}
$$

The second factor on the right hand side corresponds to the transition probability $P(X_t = m_t | X_s = m_s)$ for $m_s \leq m_t$. Hence, the Poisson process is homogeneous since

$$P(X_{t+h} = m_t + m | X_t = m_t) = \frac{(\lambda h)^m}{m!} \, e^{-\lambda h},$$

where $h \geq 0$.

**Property 9.2.** Let $f : \mathcal{X} \to \mathbb{R}$. The continuous time homogeneous Markov chain $X = \{X_t, t \in \mathbb{R}^+\}$ is *ergodic* if

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(X_t) \, dt = \sum_{i=1}^N f(x_i) \bar{p}^{(i)} \tag{18}$$

where $\bar{\mathbf{p}}$ is a stationary probability vector and all $x_i \in \mathcal{X}$.

**Definition 10.** The *intensity matrix* $\mathbf{A} \in \mathbb{R}^{N \times N}$ of a homogeneous continuous time Markov chain is defined as follows:

$$a^{(i,j)} = \begin{cases} \lim_{t \to 0} \frac{p_t^{(i,j)}}{t} & \text{if } i \neq j, \\ \lim_{t \to 0} \frac{p_t^{(i,j)} - 1}{t} & \text{if } i = j. \end{cases} \tag{19}$$

**Theorem 1.** The homogeneous continuous time Markov chain $X = \{X_t, t \in \mathbb{R}^+\}$ is completely characterized by the initial probability vector $\mathbf{p}_0$ and its intensity matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. Furthermore, if all the diagonal elements of $\mathbf{A}$ are finite, then the transition probabilities satisfy the *Kolmogorov forward equation*

$$\frac{d\mathbf{P}_t}{dt} - \mathbf{P}_t \mathbf{A} = \mathbf{0}, \tag{20}$$

and the *Kolmogorov backward equation*

$$\frac{d\mathbf{P}_t}{dt} - \mathbf{A}^{\mathrm{T}} \mathbf{P}_t = \mathbf{0}. \tag{21}$$

## 2.2 Diffusion processes

Diffusion processes are an important class of continuous time continuous state Markov processes, that is $\mathcal{X} \subseteq \mathbb{R}$.

**Definition 11.** The stochastic process $X = \{X_t, t \in \mathbb{R}^+\}$ is a *(continuous time continuous state) Markov process* if it satisfies the following Markov property:

$$P(X_t \in B | X_s = x) = P(X_t \in B | X_{r_1} = x_1, \ldots, X_{r_n} = x_n, X_s = x) \qquad (22)$$

for all Borel subsets $B \subseteq \mathbb{R}$, time instants $0 \leq r_1 \leq \ldots \leq r_n \leq s \leq t$ and all $x_1, \ldots, x_n, x \in \mathbb{R}$ for which the conditional probabilities are defined.

For fixed $s, x$ and $t$ the transition probability $P(X_t \in B | X_s = x)$ is a probability measure on the $\sigma$-algebra $\mathcal{B}$ of Borel subsets of $\mathbb{R}$ such that

$$P(X_t \in B | X_s = x) = \int_B p(s, x; t, y) \, dy \qquad (23)$$

for all $B \in \mathcal{B}$. The quantity $p(s, x; t, \cdot)$ is the *transition density*. It plays a similar role as the transition matrix in Markov chains.

From the Markov property, it follows that

$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, x; \tau, z) p(\tau, z; t, y) \, dz \qquad (24)$$

for all $s \leq \tau \leq t$ and $x, y \in \mathbb{R}$. This equation is known as the *Chapman-Kolmogorov equation*.

**Property 11.1.** If all its transition densities depend only on the time differences, then the Markov process is *homogeneous* and we write $p(s, x; t, y) = p(0, x; t - s, y) \equiv p(x; t - s, y)$ for any $0 \leq s \leq t$.

**Example.** *The transition probability of the Ornstein-Uhlenbeck process with parameter $\gamma > 0$ is given by*

$$p(s, x; t, y) = \frac{1}{\sqrt{2\pi(1 - e^{-2\gamma(t-s)})}} \exp\left\{ -\frac{\left(y - xe^{-\gamma(t-s)}\right)^2}{2\left(1 - e^{-2\gamma(t-s)}\right)} \right\} \qquad (25)$$

*for all $0 \leq s \leq t$. Hence, it is a homogeneous Markov process.*

**Property 11.2.** Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function. The Markov process $X = \{X_t, t \in \mathbb{R}^+\}$ is *ergodic* if

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(X_t) \, dt = \int_{-\infty}^{\infty} f(y) \bar{p}(y) \, dy \qquad (26)$$

where $\bar{p}(y) = \int_{-\infty}^{\infty} p(s, x; t, y) \bar{p}(x) dx$ is a stationary probability density.

This means that the time average limit coincide with the spatial average. In practice, this is often difficult to verify.

**Definition 12.** A Markov process $X = \{X_t, t \in \mathbb{R}^+\}$ is a *diffusion process* if the following limits exist for all $\epsilon > 0$, $s \geq 0$ and $x \in \mathbb{R}$:

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|>\epsilon} p(s,x;t,y) \, dy = 0, \tag{27}$$

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} (y-x)p(s,x;t,y) \, dy = \alpha(s,x), \tag{28}$$

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|<\epsilon} (y-x)^2 p(s,x;t,y) \, dy = \beta^2(s,x), \tag{29}$$

where $\alpha(s,x)$ is the *drift* and $\beta(s,x)$ is the *diffusion coefficient* at time $s$ and position $x$.

In other words, condition (27) prevents the diffusion process from having instantaneous jumps. From (28) and (29) one can see that $\alpha(s,x)$ and $\beta^2(s,x)$ are respectively the instantaneous rate of change of the mean and the instantaneous rate of change of the squared fluctuations of the process, given that $X_s = x$.

**Example.** *The Ornstein-Uhlenbeck process is a diffusion process with drift* $\alpha(s,x) = -\gamma x$ *and diffusion coefficient* $\beta(s,x) = \sqrt{2\gamma}$.

To proof this we first note that $p(s,x;t,y) = \mathcal{N}\left(y|xe^{-\gamma(t-s)}, (1 - e^{-2\gamma(t-s)})\right)$ for any $t \geq s$. Therefore, we have

$$\alpha(x,s) = \lim_{t \downarrow s} \frac{E\{y\} - x}{t-s} = -x \lim_{t \downarrow s} \frac{1 - e^{-\gamma(t-s)}}{t-s} = -x \cdot \gamma,$$

$$\beta^2(x,s) = \lim_{t \downarrow s} \frac{E\{y^2\} - 2E\{y\}x + x^2}{t-s}$$

$$= \lim_{t \downarrow s} \left\{ \frac{1 - e^{-2\gamma(t-s)}}{t-s} + x^2 \frac{(1 - e^{-\gamma(t-s)})^2}{t-s} \right\}$$

$$= 2\gamma + x^2 \cdot 0.$$

Diffusion processes are *almost surely continuous* functions of time, but they need not to be differentiable. Without going into the mathematical details, the continuity of a stochastic process can be defined in terms of continuity with probability one, mean square continuity and continuity in probability or distribution (see for example [3]).

Another interesting criterion is *Kolmogorov's continuity criterion*, which states that a continuous time stochastic process $X = \{X_t, t \in T\}$ has continuous sample paths if there exists $a, b, c, h > 0$ such that

$$E\{|X_t - X_s|^a\} \leq c|t-s|^{1+b} \tag{30}$$

for all $s, t \in T$ and $|t - s| \leq h$.

**Theorem 2.** Let the stochastic process $X = \{X_t, t \in \mathbb{R}^+\}$ be a diffusion process for which $\alpha$ and $\beta$ are moderately smooth. The forward evolution of its

transition density $p(s, x; t, y)$ is given by the *Kolmogorov forward equation* (also known as the *Fokker-Planck equation*)

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}\{\alpha(t, y)p\} - \frac{1}{2}\frac{\partial^2}{\partial y^2}\{\beta^2(t, y)p\} = 0, \tag{31}$$

for a fixed initial state $(s, x)$. The backward evolution of the transition density $p(s, x; t, y)$ is given by the *Kolmogorov backward equation*

$$\frac{\partial p}{\partial s} + \alpha(s, x)\frac{\partial p}{\partial x} + \frac{1}{2}\beta^2(s, x)\frac{\partial^2 p}{\partial x^2} = 0, \tag{32}$$

for a fixed final state $(t, y)$.

A rough proof of (32) is the following. Consider the approximate time discrete continuous state process with two equally probable jumps from $(s, x)$ to $(s + \Delta s, x + \alpha \Delta s \pm \beta\sqrt{\Delta s})$, which is consistent with (28) and (29). The approximate transition probability is then given by

$$\hat{p}(s, x; t, y) = \frac{1}{2}\hat{p}(s + \Delta s \, , \, x + \alpha \Delta s + \beta\sqrt{\Delta s} \, ; \, t \, , \, y)$$
$$+ \frac{1}{2}\hat{p}(s + \Delta s \, , \, x + \alpha \Delta s - \beta\sqrt{\Delta s} \, ; \, t \, , \, y).$$

Taking Taylor expansions up to the first order in $\Delta s$ about $(s, x; t, y)$ leads to

$$0 = \frac{\partial \hat{p}}{\partial s}\Delta s + \alpha\frac{\partial \hat{p}}{\partial x}\Delta s + \frac{1}{2}\beta^2\frac{\partial^2 \hat{p}}{\partial x^2}\Delta s + \mathcal{O}\left((\Delta s)^{3/2}\right).$$

Since the discrete time process converges in distribution to the diffusion process, we obtain the backward Kolmogorov equation when $\Delta s \to 0$.

**Example.** *The Kolmogorov foward and the backward equations for the Ornstein-Uhlenbeck process with parameter $\gamma > 0$ are respectively given by*

$$\frac{\partial p}{\partial t} - \gamma\frac{\partial}{\partial y}\{yp\} - \gamma\frac{\partial^2 p}{\partial y^2} = 0, \tag{33}$$

$$\frac{\partial p}{\partial s} - \gamma x\frac{\partial p}{\partial x} + \gamma\frac{\partial^2 p}{\partial x^2} = 0. \tag{34}$$

## 2.3  Wiener processes

The Wiener process was proposed by Wiener as mathematical description of *Brownian motion*. This physical process characterizes the erratic motion (i.e. diffusion) of a grain pollen on a water surface due to the fact that is continually bombarded by water molecules. The resulting motion can be viewed as a scaled random walk on any finite time interval and is almost surely continuous, w.p. 1.

**Definition 13.** A standard *Wiener process* is a continuous time Gaussian Markov process $W = \{W_t, t \in \mathbb{R}^+\}$ with (non-overlapping) independent increments for which

$$W_0 = 0 \quad \text{w.p. 1,}$$
$$E\{W_t\} = 0,$$
$$W_t - W_s \sim \mathcal{N}(0, t - s),$$

for all $0 \leq s \leq t$.

Its covariance is given by $C_{s,t} = \min\{s, t\}$. Indeed, if $0 \leq s < t$, then

$$\begin{aligned}
C_{s,t} &= E\{(W_t - \mu_t)(W_s - \mu_s)\} \\
&= E\{W_t W_s\} \\
&= E\{(W_t - W_s + W_s)W_s\} \\
&= E\{W_t - W_s\}E\{W_s\} + E\{W_s^2\} \\
&= 0 \cdot 0 + s.
\end{aligned}$$

Hence, it is not a wide-sense stationary process. However, it is a homogeneous Markov process since its transition probability is given by

$$p(s, x; t, y) = \frac{1}{\sqrt{2\pi(t - s)}} \exp\left\{-\frac{(y - x)^2}{2(t - s)}\right\} \tag{35}$$

Although the sample paths of Wiener processes are almost surely continuous functions of time (the Kolmogorov continuity criterion (30) is satisfied for $a = 4$, $b = 1$ and $c = 3$), they are *almost surely nowhere differentiable*. Consider the partition of a bounded time interval $[s, t]$ into subintervals $[\tau_k^{(n)}, \tau_{k+1}^{(n)}]$ of equal length $(t - s)/2^n$, where $\tau_k^{(n)} = s + k(t - s)/2^n$ for $k = 0, 1, \ldots, 2^n - 1$. It can be shown [3, p. 72] that

$$\lim_{n \to \infty} \sum_{k=0}^{2^n - 1} \left(W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega)\right)^2 = t - s, \quad \text{w.p. 1,}$$

where $W_\tau(\omega)$ is a sample path of the standard Wiener process $W = \{W_\tau, \tau \in [s, t]\}$ for any $\omega \in \Omega$. Hence,

$$\begin{aligned}
t - s \leq{} &\limsup_{n \to \infty} \max_{0 \leq k \leq 2^n - 1} \left|W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega)\right| \\
&\times \sum_{k=0}^{2^n - 1} \left|W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega)\right|.
\end{aligned}$$

(Note that $\limsup$ is the limit superior or supremum limit, that is the supremum[3] of all the limit points.) From the sample path continuity, we have

$$\max_{0 \leq k \leq 2^n - 1} \left|W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega)\right| \to 0, \quad \text{w.p. 1 as } n \to \infty$$

---

[3] For $S \subseteq T$, the supremum of $S$ is the least element of $T$, which is greater or equal to all elements of $S$.

and therefore

$$\sum_{k=0}^{2^n-1} \left| W_{\tau_{k+1}^{(n)}}(\omega) - W_{\tau_k^{(n)}}(\omega) \right| \to \infty, \ \text{ w.p. 1 as } n \to \infty.$$

As a consequence, the sample paths do, almost surely, not have bounded variation on $[s,t]$ and cannot be differentiated.

The standard Wiener process is also a diffusion process with drift $\alpha(s,x) = 0$ and diffusion coefficient $\beta(s,x) = 1$. Indeed, we have

$$\alpha(x,s) = \lim_{t\downarrow s} \frac{E\{y\} - x}{t - s} = 0,$$

$$\beta^2(x,s) = \lim_{t\downarrow s} \frac{E\{y^2\} - 2E\{y\}x + x^2}{t - s} = \lim_{t\downarrow s} \left\{ \frac{t - s}{t - s} + 0 \right\} = 1.$$

Hence, the Kolmogorov forward and backward equations are given by

$$\frac{\partial p}{\partial t} - \frac{1}{2}\frac{\partial^2 p}{\partial y^2} = 0, \tag{36}$$

$$\frac{\partial p}{\partial s} + \frac{1}{2}\frac{\partial^2 p}{\partial x^2} = 0. \tag{37}$$

Directly evaluating the partial derivatives of the transition density leads to the same results.

Note finally that the Wiener process $W = \{W_t, t \in \mathbb{R}^+\}$ is a martingale. Since $E\{W_t - W_s | W_s\} = 0$, w.p. 1, and $E\{W_s | W_s\} = W_s$, w.p. 1, we have $E\{W_t | W_s\} = W_s$, w.p. 1.

## 2.4  The Brownian Bridge

**Definition 14.** Let $W : \mathbb{R}^+ \times \Omega \to \mathbb{R}$ be a standard Wiener process. The *Brownian bridge* $B_t^{(0,x;T,y)}$ is a stochastic process defined sample pathwise such that

$$B_t^{(0,x;T,y)}(\omega) = x + W_t(\omega) - \frac{t}{T}(W_T(\omega) - y + x) \tag{38}$$

for $0 \le t \le T$.

This means that the Brownian bridge is a Wiener process for which the sample paths all start at $B_t^{(0,x;T,y)}(\omega) = x \in \mathbb{R}$ and pass through a given point $B_T^{(0,x;T,y)}(\omega) = y \in \mathbb{R}$ at a later time $T$ for all $\omega \in \Omega$.

**Property 14.1.** A Brownian bridge $B_t^{(0,x;T,y)}$ is a Gaussian process with means and covariances respectively given by

$$\mu_t = x - \frac{t}{T}(x - y), \tag{39}$$

$$C_{s,t} = \min\{s,t\} - \frac{st}{T}, \tag{40}$$

for $0 \le s, t \le T$.

## 2.5 White Noise

**Definition 15.** The *power spectral density* of a (wide-sense) stationary process $X = \{X_t, t \in \mathbb{R}\}$ is defined as the Fourier transform of its covariance:

$$S(\omega) = \int_{-\infty}^{\infty} C_t e^{-j\omega t}\, dt, \tag{41}$$

where $\omega = 2\pi f$ and $C_{0,t} \equiv C_t$.

The spectral density measures the power per unit frequency at frequency $f$ and the variance of the process can be interpreted as the average power (or energy):

$$\mathrm{Var}\{X_t\} = C_0 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\, d\omega. \tag{42}$$

Note that the covariance $C_t = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} S(\omega)\, d\omega$ is the inverse Fourier transform of the spectral density.

**Definition 16.** *Gaussian white noise* is a zero-mean wide-sense stationary process with constant nonzero spectral density $S(\omega) = S_0$ for all $\omega \in \mathbb{R}$.

Hence, its covariances satisfy $C_t = S_0 \delta(t)$ for all $t \in \mathbb{R}^+$. Without loss of generality, if we assume that $S_0 = 1$, it can be shown that Gaussian white noise corresponds to the following limit process:

$$\lim_{h \to 0} X_t^h = \frac{W_{t+h} - W_t}{h}, \tag{43}$$

where $W = \{W_t, t \in \mathbb{R}^+\}$. Hence, it is the derivative of the Wiener process. However, the sample paths of a Wiener process are not differentiable anywhere. Therefore, Gaussian white noise cannot be realized physically, but can be approached by allowing a broad banded spectrum (that is let $h$ grow).

A similar result holds for the Ornstein-Uhlenbeck process with $\gamma \to \infty$.

## 2.6 Multivariate Diffusion Processes

See for example [3, p. 68].

# 3 Ito Stochastic Calculus

The *ordinary differential equation* $dx/dt = \alpha(t, x)$ can be viewed as a degenerate form a stochastic differential equation as no randomness is involved. It can be written in symbolic differential form

$$dx = \alpha(t, x)\, dt \tag{44}$$

or as an integral equation

$$x(t) = x_0 + \int_{t_0}^{t} \alpha(s, x(s)) \, ds \tag{45}$$

where $x(t; t_0, x_0)$ is a solution satisfying the initial condition $x_0 = x(t_0)$. For some regularity conditions on $\alpha$, this solution is unique, which means that the future is completely defined by the present given the initial condition.

The symbolic differential form of a *stochastic differential equation* is written as follows:

$$dX_t = \alpha(t, X_t) \, dt + \beta(t, X_t)\xi_t \, dt \tag{46}$$

where $X = \{X_t, t \in \mathbb{R}^+\}$ is a diffusion process and $\xi_t \sim \mathcal{N}(0, 1)$ for each $t$, i.e. it is a Gaussian process. In particular, (46) is called the *Langevin equation* if $\alpha(t, X_t) = -\bar{\alpha}X_t$ and $\beta(t, X_t) = \bar{\beta}$ for constant $\bar{\alpha}$ and $\bar{\beta}$.

This symbolic differential can be interpreted as the integral equation along a sample path

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^{t} \alpha(s, X_s(\omega)) \, ds + \int_{t_0}^{t} \beta(s, X_s(\omega))\xi_s(\omega) \, ds \tag{47}$$

for each $\omega \in \Omega$. Now, for the case where $a \equiv 0$ and $\beta \equiv 1$, we see that $\xi_t$ should be the derivative of Wiener process $W_t = X_t$, i.e it is Gaussian white noise. This suggests that (47) can be written as follows:

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^{t} \alpha(s, X_s(\omega)) \, ds + \int_{t_0}^{t} \beta(s, X_s(\omega)) \, dW_s(\omega). \tag{48}$$

The problem with this formulation is that the Wiener process $W_t$ is (almost surely) nowhere differentiable such that the white noise process $\xi_t$ does not exist as a conventional function of $t$. As a result, the second integral in (48) cannot be understood as an ordinary (Riemann or Lebesgue) integral. Worse, it is not a Riemann-Stieltjes integral since the continuous sample paths of a Wiener process are not of bounded variation for each sample path. Hence, it is at this point that Ito's stochastic integral comes into play!

## 3.1 Ito Stochastic Integral

In this section, we consider a probability space $(\Omega, \mathcal{A}, P)$, a Wiener process $W = \{W_t, t \in \mathbb{R}^+\}$ and an increasing family $\{\mathcal{A}_t, t \geq 0\}$ of sub-$\sigma$-algebras of $\mathcal{A}$ such that $W_t$ is $\mathcal{A}_t$-measurable for each $t \geq 0$ and with

$$E\{W_t | \mathcal{A}_0\} = 0 \quad \text{and} \quad E\{W_t - W_s | \mathcal{A}_s\} = 0, \quad \text{w.p. 1,}$$

for $0 \leq s \leq t$.

Ito's starting point is the following. For constant $\beta(t, x) \equiv \bar{\beta}$ the second integral in (48) is expected to be equal to $\bar{\beta}\{W_t(\omega) - W_{t_0}(\omega)\}$.

We consider the integral of the random function $f : T \times \Omega \to \mathbb{R}$ on the unit time interval:

$$I[f](\omega) = \int_0^1 f(s, \omega) \, dW_s(\omega). \tag{49}$$

First, if the function $f$ is a nonrandom step function, that is $f(t, \omega) = f_j$ on $t_j \leq t < t_{j+1}$ for $j = 1, 2, \ldots, n-1$ with $0 = t_1 < t_2 < \ldots < t_n = 1$, then we should obviously have

$$I[f](\omega) = \sum_{j=1}^{n-1} f_j \{ W_{t_{j+1}}(\omega) - W_{t_j}(\omega) \}, \quad \text{w.p. 1.} \tag{50}$$

Note that this integral is a random variable with zero mean as it is a sum of random variables with zero mean. Furthermore, we have the following result

$$E\{ I[f](\omega) \} = \sum_{j=1}^{n-1} f_j^2 (t_{j+1} - t_j). \tag{51}$$

Second, if the function $f$ is a random step function, that is $f(t, \omega) = f_j(\omega)$ on $t_j \leq t < t_{j+1}$ for $j = 1, 2, \ldots, n-1$ with $t_1 < t_2 < \ldots < t_n$ is $\mathcal{A}_{t_j}$-measurable and mean square integrable over $\Omega$, that is $E\{f_j^2\} < \infty$ for $j = 1, 2, \ldots, n$. The stochastic integral $I[f](\omega)$ is defined as follows:

$$I[f](\omega) = \sum_{j=1}^{n-1} f_j(\omega) \{ W_{t_{j+1}}(\omega) - W_{t_j}(\omega) \}, \quad \text{w.p. 1.} \tag{52}$$

**Lemma.** For any $a, b \in \mathbb{R}$ and any random step function $f, g$ such that $f_j, g_j$ on $t_j \leq t < t_{j+1}$ for $j = 1, 2, \ldots, n-1$ with $0 = t_1 < t_2 < \ldots < t_n = 1$ is $\mathcal{A}_{t_j}$-measurable and mean square integrable, the stochastic integral (52) satisfies the following properties:

$$I[f] \text{ is } \mathcal{A}_1 - \text{measurable,} \tag{53}$$
$$E\{I[f]\} = 0, \tag{54}$$
$$E\{I^2[f]\} = \sum_j E\{f_j^2\}(t_{j+1} - t_j), \tag{55}$$
$$I[af + bg] = aI[f] + bI[g], \quad \text{w.p. 1.} \tag{56}$$

Since $f_j$ is $\mathcal{A}_{t_j}$-measurable and $\{W_{t_{j+1}} - W_{t_j}\}$ is $\mathcal{A}_{t_{j+1}}$-measurable, each term $f_j\{W_{t_{j+1}} - W_{t_j}\}$ is $\mathcal{A}_{t_{j+1}}$-measurable and thus $\mathcal{A}_1$-measurable. Hence, $I[f]$ is $\mathcal{A}_1$-measurable.

From the Cauchy-Schwarz inequality[4] and the fact that each term in (52) is mean-squre integrable, it follows that $I[f]$ is integrable. Hence, $I[f](\omega)$ is again

---

[4]The *Cauchy-Schwarz inequality* states that $\left| \int_a^b fg \, dx \right|^2 \leq \left| \int_a^b f^2 \, dx \right| \left| \int_a^b g^2 \, dx \right|$.

a zero mean random variable:

$$E\{I[f]\} = \sum_{j=1}^{n-1} E\{f_j(W_{t_{j+1}} - W_{t_j})\} = \sum_{j=1}^{n-1} E\{f_j E\{W_{t_{j+1}} - W_{t_j} | \mathcal{A}_{t_j}\}\} = 0.$$

Furthermore, $I[f]$ is mean square integrable:

$$E\{I^2[f]\} = \sum_{j=1}^{n-1} E\{f_j^2\} E\{(W_{t_{j+1}} - W_{t_j})^2 | \mathcal{A}_{t_j}\} = \sum_{j=1}^{n-1} E\{f_j^2\}(t_{j+1} - t_j).$$

Finally, $af + bg$ is a step random step function for any $a, b \in \mathbb{R}$. Therefore, we obtain (56), w.p. 1, after algebraic rearrangement.

Third, if the (continuous) function $f$ is a general integrand such that it $f(t, \cdot)$ is $\mathcal{A}_t$-measurable and mean square integrable, then we define the stochastic integral $I[f]$ as the limit of integrals $I[f^{(n)}]$ of random step functions $f^{(n)}$ converging to $f$. The problem is thus to characterize the limit of the following finite sums:

$$I[f^{(n)}](\omega) = \sum_{j=1}^{n-1} f(t_j^{(n)}, \omega)\{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\}, \quad \text{w.p. 1}. \tag{57}$$

where $f^{(n)}(t, \omega) = f(t_j^{(n)}, \omega)$ on $t_j \leq t \leq t_{j+1}$ for $j = 1, 2, \ldots, n - 1$ with $t_1 < t_2 < \ldots < t_n$. From (55), we get

$$E\{I^2[f^{(n)}]\} = \sum_{j=1}^{n-1} E\{f^2(t_j^{(n)}, \cdot)\}(t_{j+1} - t_j).$$

This converges to the Riemann integral $\int_0^1 E\{f^2(s, \cdot)\} \, ds$ for $n \to \infty$. This result, along with the well-behaved mean square property of the Wiener process, i.e. $E\{(W_t - W_s)^2\} = t - s$, suggests defining the stochastic integral in terms of mean square convergence.

**Theorem 3.** The *Ito (stochastic) integral $I[f]$* of a function $f : T \times \Omega \to \mathbb{R}$ is the (unique) mean square limit of sequences $I[f^{(n)}]$ for any sequence of random step functions $f^{(n)}$ converging[5] to $f$:

$$I[f](\omega) = \underset{n \to \infty}{\text{m.s. lim}} \sum_{j=1}^{n-1} f(t_j^{(n)}, \omega)\{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\}, \quad \text{w.p. 1}. \tag{59}$$

---

[5]In particular, we call the sequence $f^{(n)}$ mean square convergent to $f$ if

$$E\left\{\int_s^t (f^{(n)}(\tau, \omega) - f(\tau, \omega))^2 d\tau\right\} \to 0, \quad \text{for } n \to \infty. \tag{58}$$

15

The properties (53–56) still apply, but we write $E\{I^2[f]\} = \int_0^1 E\{f^2(t,\cdot)\}\,dt$ for (56) and call it the *Ito isometry* (on the unit time interval).

Similarly, the time-dependent Ito integral is a random variable defined on any interval $[t_0, t]$:

$$X_t(\omega) = \int_{t_0}^t f(s,\omega)\,dW_s(\omega), \tag{60}$$

which is $\mathcal{A}_t$-measurable and mean square integrable. From the independence of non-overlapping increments of a Wiener process, we have $E\{X_t - X_s | \mathcal{A}_s\} = 0$, w.p. 1, for any $t_0 \leq s \leq t$. Hence, the process $X_t$ is a martingale.

As the Riemann and the Riemann-Stieltjes integrals, (60) satisfies conventional properties such as the linearity property and the additivity property. However, it has also the unusual property that

$$\int_0^t W_s(\omega)\,dW_s(\omega) = \frac{1}{2}W_t^2(\omega) - \frac{1}{2}t, \quad \text{w.p. 1}, \tag{61}$$

where $W_0 = 0$, w.p. 1. Note that this expression follows from the fact that

$$\textstyle\sum_j W_{t_j}(W_{t_{j+1}} - W_{t_j}) = \frac{1}{2}W_t^2 - \frac{1}{2}\sum_j(W_{t_{j+1}} - W_{t_j})^2, \tag{62}$$

where the second term tends to $t$ in mean square sense.

By contrast, standard non-stochastic calculus would give $\int_0^t w(s)\,dw(s) = \frac{1}{2}w^2(t)$ if $w(0) = 0$.

## 3.2 The Ito Formula

The main advantage of the Ito stochastic integral is the martingale property. However, a consequence is that stochastic differentials, which are interpreted as stochastic integrals, do not follow the chain rule of classical calculus! Roughly speaking, an additional term is appearing due to the fact that $dW_t^2$ is equal to $dt$ in the mean square sense.

Consider the stochastic process $Y = \{Y_t = U(t, X_t), t \geq 0\}$ with $U(t, x)$ having continuous second order partial derivatives.

If $X_t$ were continuously differentiable, the chain rule of classical calculus would give the following expression:

$$dY_t = \frac{\partial U}{\partial t}\,dt + \frac{\partial U}{\partial x}\,dX_t. \tag{63}$$

This follows from a Taylor expansion of $U$ in $\Delta Y_t$ and discarding the second and higher order terms in $\Delta t$.

When $X_t$ is a process of the form (60), we get (with equality interpreted in the mean square sense)

$$dY_t = \left\{\frac{\partial U}{\partial t} + \frac{1}{2}f^2\frac{\partial^2 U}{\partial x^2}\right\}\,dt + \frac{\partial U}{\partial x}\,dX_t, \tag{64}$$

16

where $dX_t = f\ dW_t$ is the symbolic differential form of (60). The additional term is due to the fact that $E\{dX_t^2\} = E\{f^2\}dt$ gives rise to an additional term of the first order in $\Delta t$ of the Taylor expansion for $U$:

$$\Delta Y_t = \left\{ \frac{\partial U}{\partial t}\Delta t + \frac{\partial U}{\partial x}\ \Delta x \right\} + \frac{1}{2}\left\{ \frac{\partial^2 U}{\partial t^2}\Delta t^2 + 2\frac{\partial^2 U}{\partial t \partial x}\Delta t \Delta x + \frac{\partial^2 U}{\partial x^2}\ \Delta x^2 \right\} + \dots.$$

**Theorem 4.** Consider the following general stochastic differential:

$$X_t(\omega) - X_s(\omega) = \int_s^t e(u,\omega)\ du + \int_s^t f(u,\omega)\ dW_u(\omega). \tag{65}$$

Let $Y_t = U(t, X_t)$ with $U$ having continuous partial derivatives $\frac{\partial U}{\partial t}$, $\frac{\partial U}{\partial x}$ and $\frac{\partial^2 U}{\partial x^2}$. The *Ito formula* is the following stochastic chain rule:

$$Y_t - Y_s = \int_s^t \left\{ \frac{\partial U}{\partial t} + e_u \frac{\partial U}{\partial x} + \frac{1}{2}f_u^2\frac{\partial^2 U}{\partial x^2} \right\}\ du + \int_s^t \frac{\partial U}{\partial x}\ dX_u, \quad \text{w.p. 1.} \tag{66}$$

The partial derivatives of $U$ are evaluated at $(u, X_u)$.

From the Ito formula, one can recover (61). For $X_t = W_t$ and $u(x) = x^m$, we have

$$d(W_t^m) = mW_t^{(m-1)}\ dW_t + \frac{m(m-1)}{2}W_t^{(m-2)}\ dt. \tag{67}$$

In the special case $m = 2$, this reads $d(W_t^2) = 2W_t\ dW_t + dt$, which leads to

$$W_t^2 - W_s^2 = 2\int_s^t W_t\ dW_t + (t-s). \tag{68}$$

Hence, we recover $\int_0^t W_t\ dW_t = \frac{1}{2}W_t^2 - \frac{1}{2}t$ for $s = 0$.

## 3.3 Multivariate case

See [3, p. 97].

# A  Probability Distributions

**Definition 17.** The probability density function of the *Poisson distribution* with parameter $\lambda > 0$ is defined as follows:

$$\mathcal{P}(n|\lambda) = \frac{\lambda^n}{n!}\ e^{-\lambda}, \tag{69}$$

for $n \in \mathbb{N}$. Note that $E\{n\} = \lambda$ and $E\{(n - E\{n\})^2\} = \lambda$.

**Definition 18.** The probability density function of the multivariate *Gaussian distribution* with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-D/2}|\boldsymbol{\Sigma}|^{-1/2}\ e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{70}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$ is symmetric and positive definite.

# References

[1] Lawrence E. Evans. *An Introduction to Stochastic Differential Equations.* Lecture notes (Department of Mathematics, UC Berkeley), available from http://math.berkeley.edu/∼evans/.

[2] Crispin W. Gardiner. *Handbook of Stochastic Methods.* Springer-Verlag, Berlin, 2004 (3rd edition).

[3] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, Berlin, 1992.

[4] Bernt Øksendael. *Stochastic Differential Equations.* Springer-Verlag, Berlin, 1985.