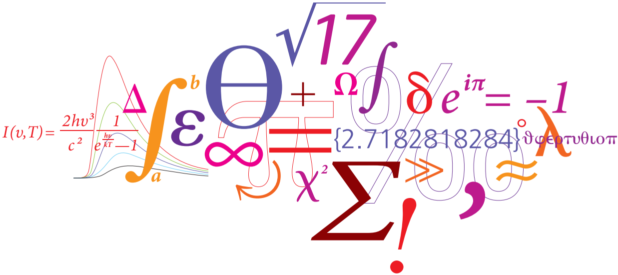
An Introduction to Stochastic Differential Equations

Pages 1–19

Lawrence C. Evans

presentation by: Arun Suresh

Georgia State University



Chapter 1 - Introduction

Chapter 1 - Motivation

- We start by considering $x_0 \in \mathbb{R}^n$ and the ODE

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{b}(\mathbf{x}(t)) \quad (t > 0) \\ \mathbf{x}(0) = x_0 \end{cases}$$

  where $\mathbf{b}(x(t))$ is a vector field

- Let us motivate our study with an example: The Lotka Volterra model
  - Let $x(t)$ denote the population count of rabbits (prey) at time t and $y(t)$ denote the population count of foxes (predators) at time t.
  - The predator prey population dynamics can be theoretically modelled by the following ODE system

$$\frac{dx}{dt} = \alpha x - \beta xy \quad \text{and} \quad \frac{dy}{dt} = \delta xy - \gamma y$$

  - where $\alpha, \beta, \gamma, \delta$ are constants determined based on the environmental conditions.

Chapter 1 - Motivation

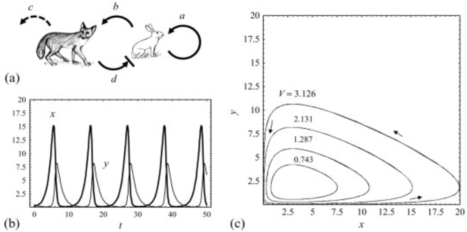
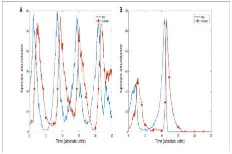
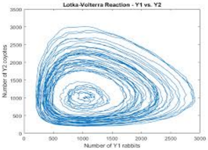


Figure: Theoretical LV



(a) Time Series



(b) Phase Space

Figure: Observed LV

Chapter 1 - Motivation

- So, there is a need to account for this "randomness" in the data we observe.
- A way to do this is simply by modifying the ODE system to introduce a "white noise" factor

$$\begin{cases} \dot{\mathbf{X}}(t) = \mathbf{b}(\mathbf{X}(t)) + \mathbf{B}(\mathbf{X}(t))\boldsymbol{\xi}(t) \quad (t > 0) \\ \mathbf{X}(0) = x_0 \end{cases} \tag{1}$$

  where $\mathbf{B} : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ and $\boldsymbol{\xi}(t) := m$ dimensional "white noise"

- However, we are mathematicians who thrive in rigor.
- Quotation marks, hand waving and intuition simply doesn't cut it.
- So, we have the following ambitious list of goals for this semester
  - Define $\boldsymbol{\xi}(t)$ rigorously
  - Understand what it even means to solve (1)
  - Ponder the uniqueness of solution to (1) and discuss other behavior of the (hopefully unique) solution that would (also, hopefully) interest us.

Chapter 1 - Background and Notations

- Consider (1) under the constraints $m = n,\ x_0 = 0,\ \mathbf{b} \equiv 0$ and $\mathbf{B} \equiv I$
  - The solution is the so called *Wiener process* denoted by $\mathbf{W}(t)$.
  - Simplifying (1) gives $\dot{\mathbf{W}}(t) = \boldsymbol{\xi}(t)$
- Going back to the general form in (1) we obtain (after some simplification)

$$\begin{cases} d\mathbf{X}(t) = \mathbf{b}(\mathbf{X}(t))dt + \mathbf{B}(\mathbf{X}(t))d\mathbf{W}(t) \\ \mathbf{X}(0) = x_0 \end{cases} \tag{2}$$

- Eq (2) is what we will fondly call *Stochastic Differential Equation* from now on.
- Ambitious as we are, we are going to try and solve (2)
- As an attempt to (falsely) pride ourselves in making progress, we rewrite the solution to the above SDE in the integral form

$$\mathbf{X}(t) = x_0 + \int_0^t \mathbf{b}(\mathbf{X}(s))ds + \int_0^t \mathbf{B}(\mathbf{X}(s))d\mathbf{W} \tag{3}$$

Chapter 1 - Background and Notations

$$\mathbf{X}(t) = x_0 + \int_0^t \mathbf{b}(\mathbf{X}(s))ds + \int_0^t \mathbf{B}(\mathbf{X}(s))d\mathbf{W}$$

- Of-course writing it this way raises more questions than it answers.
- However, these are important, more concrete questions
  - How do we construct $\mathbf{W}$?
  - How do we integrate with respect to $d\mathbf{W}$?
  - How do we even know that solution to (3) actually exists?
- Upon answering the above questions we have, *you guessed it*, more questions!
  - How can we be sure this SDE models reality?
  - Is $\boldsymbol{\xi}(t)$ really just pure randomness or does it have a "Fourier flavor" to it?

Chapter 1 - Itô's formula

- A very **very** important tool to deal with SDEs is the so called Itô's formula.
  - Think of it as a "Integration by parts" formula for SDE's
  - Helps us deal with $\int_0^t [...] d\mathbf{W}$ with relative ease
- Motivation for the formula (we will make this rigorous when we are ready)
  - Suppose $X(t)$ is the solution to the following SDE

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

    and we have $u : \mathbb{R}^2 \to \mathbb{R}$ a given smooth function.
  - We want to understand what SDE satisfies $\mathbf{Y}(t) = u(\mathbf{X}(t))$
  - The whole idea being, given such an SDE
    - We will have $d[u(\mathbf{X}(t))]$ on the left hand side...
    - There are hopes for integrating both sides to obtain an expression for $\mathbf{X}(t)$
    - Of course, the choice of u is crucial here to make progress.

Theorem (Itô's formula)

Suppose $\boldsymbol{X}(t)$ is the solution of the following SDE

$$d\boldsymbol{X}(t) = \boldsymbol{b}(\boldsymbol{X}(t))dt + \boldsymbol{B}(\boldsymbol{X}(t))d\boldsymbol{W}(t)$$

and $u(x,t)$ is a real valued function defined for $x \in \mathbb{R}$ with continuous partial derivatives $\partial u/\partial t$, $\partial u/\partial x$ and $\partial^2 u/\partial x^2$ then

$$du(\boldsymbol{X}(t)) = f(\boldsymbol{X}(t))dt + g(\boldsymbol{X}(t))d\boldsymbol{W}(t)$$

where

$$\begin{aligned} f(x,t) &= \frac{\partial u}{\partial t} + \boldsymbol{b}\frac{\partial u}{\partial x} + \frac{1}{2}\boldsymbol{B}^2\frac{\partial^2 u}{\partial x^2} \\ g(x,t) &= \boldsymbol{B}\frac{\partial u}{\partial x} \end{aligned}$$

Chapter 1 - Itô's formula

- Chapter 4 derives the formula. So, for now will simply illustrate its sheer power
- Assume (for the time being) we know $\int_a^b d\mathbf{W} = \mathbf{W}(b) - \mathbf{W}(a)$ and $\mathbf{W}(0) = 0$

Example

Suppose $\mathbf{X}(s)$ is the solution to

$$d\mathbf{X}(s) = \mathbf{X}(s)d\mathbf{W}(s) \quad \text{with } X(0) = 1$$

let $u(x,t) = \ln(x)$. Applying Itô's formula

$$d[\ln(X(s))] = -\frac{1}{2}dt + 1d\mathbf{W}(s)$$

Now integrating both sides from 0 to t simply yields

$$\ln(\mathbf{X}(t)) = -\frac{1}{2}t + \mathbf{W}(t)$$
$$\implies \mathbf{X}(t) = \exp\left(-\frac{1}{2}t + \mathbf{W}(t)\right)$$

Chapter 2 - Basic probability theory

Definition

Given a set Ω, a σ-algebra is a collection $\mathcal{U}$ of subsets of Ω. We call $P : \mathcal{U} \to [0,1]$ a probability measure provided

1. $P(\varnothing) = 0$ and $P(\Omega) = 1$
2. If $\{A_i\}_i \in \mathcal{U}$ then
$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k)$$
   with equality holding if and only if A_k are pairwise disjoint.

The triple $(\Omega, \mathcal{U}, P)$ is called a probability space.

Definition

1. $A \in \mathcal{U}$ is called an *event*, points $\omega \in \Omega$ are *sample points*
2. $P(A)$ is the *probability of event A*
3. A property that is true except for an event of probability zero is said to hold *almost surely*.

Chapter 2 - Probability Spaces

Example

The smallest σ algebra containing all the open subsets of $\mathbb{R}^n$ is called the Borel σ-algebra, denote by $\mathcal{B}$. Assume f is non-negative, integrable real valued function with $\int_{\mathbb{R}^n} f(x)dx = 1$. Define for any $B \in \mathcal{B}$

$$P(B) = \int_B f(x)dx$$

$(\mathbb{R}^n, \mathcal{B}, P)$ is a probability space. f is the density of the probability measure P.

Side note: Think of $\mathcal{B}$, informally, as the set of all "nice" subsets of $\mathbb{R}^n$.

Chapter 2 - Random Variables

- Probability spaces are simply necessary constructs for rigor
    - they do not usually represent the "observable"
- We need a way to translate objects in Ω to $\mathbb{R}^n$, the values of which we may observe...

Definition

Let $(\Omega, \mathcal{U}, P)$ be a probability space. A mapping $\mathbf{X} : \Omega \to \mathbb{R}^n$ is called an n-dimensional random variable if for each $B \in \mathcal{B}$ we have $\mathbf{X}^{-1}(B) \in \mathcal{U}$.

- For the initiated (i.e, completed graduate analysis 1), the random variable $\mathbf{X}$ is simply a $\mathcal{U}-$measurable map.
    - Unpacking $\mathbf{X}$ we see that, $\mathbf{X} : (\Omega, \mathcal{U}, P) \to (\mathbb{R}^n, \mathcal{B}, P)$.
    - It should be straight forward now why $\mathbf{X}$ is $\mathcal{U}-$measurable.
- **Notations and Clarification:**
    - Disregarding sample dependence, we simply write $\mathbf{X}(w)$ as $\mathbf{X}$
    - We denote $P(\mathbf{X}^{-1}(B))$ as $P(\mathbf{X} \in B)$
    - Capital letters = Random Variables, Boldface = Vector valued map.

Chapter 2 - Random Variables

Definition

Let $\mathbf{X} : \Omega \to \mathbb{R}^n$ be a random variable. Then $\mathcal{U}(\mathbf{X}) := \{\mathbf{X}^{-1}(B) | B \in \mathcal{B}\}$ is the σ-algebra generated by $\mathbf{X}$. It is also the smallest σ-algebra wrt which $\mathbf{X}$ is measurable.

Theorem

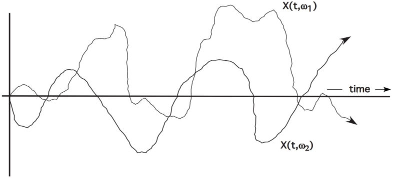
A random variable $Y : \Omega \to \mathbb{R}$ is $\mathcal{U}(\boldsymbol{X})-$measurable if and only if $Y = \Phi(\boldsymbol{X})$ for some Φ.

Chapter 2 - Stochastic Processes

- There is a need for random variables to depend on time...

Definition

1. A collection $\{\mathbf{X}(t)|t \geq 0\}$ of random variables is called a *stochastic process*
2. For each point $\omega \in \Omega$, the mapping $t \rightarrow \mathbf{X}(t, \omega)$ is the corresponding *sample path*



Two sample paths of a stochastic process

- **Remark:** Informally speaking, as t changes, **X** changes even if ω is fixed.
  - To make this explicit – $X(0)$ and $X(1)$, for example, are completely different mappings from $\Omega \rightarrow \mathbb{R}^n$. Essentially, **X** evolves with time.

- We start by recalling certain notions from measure theory

Definition

If $(\Omega, \mathcal{U}, P)$ is a probability space and $X = \sum_{i=1}^{k} a_i \chi_{A_i}$ is a real valued random variable then

$$\int_\Omega X dP = \sum_{i=1}^{k} a_i P(A_i)$$

If X is a non-negative random variable, then

$$\int_\Omega X dP := \sup_{\substack{Y \leq X \\ Y \ simple}} \int_\Omega Y dP$$

Finally, $X : \Omega \to \mathbb{R}$ is a random variable, then

$$\int_\Omega X dP = \int_\Omega X^+ dP - \int_\Omega X^- dP$$

For vector $\mathbf{X} = (X_i)_{i=1}^n$, we simply define $\int_\Omega \mathbf{X} dP = (\int_\Omega X_i dP)_{i=1}^n$

Chapter 2 - Expected value and Variance

Definition

Expected value of **X**:

$$E(\mathbf{X}) := \int_{\Omega} \mathbf{X} dP$$

Variance of **X**:

$$V(\mathbf{X}) := \int_{\Omega} |\mathbf{X} - E(\mathbf{X})|^2 dP$$

- It is easy to show $E(E(\mathbf{X})) = E(\mathbf{X})$
- It is also easy to show $V(\mathbf{X}) = E(|\mathbf{X}|^2) - |E(\mathbf{X})|^2$
- Chebyshev's inequality: Given $1 \leq p < \infty$ and $\lambda > 0$, $P(|\mathbf{X}| \geq \lambda) \leq \frac{1}{\lambda^p} E(|\mathbf{X}|^p)$
    - *Proof:*

$$\int_{\Omega} |\mathbf{X}|^p \geq \int_{\{|\mathbf{X}| \geq \lambda\}} |\mathbf{X}|^p dP \geq \int_{\{|\mathbf{X}| \geq \lambda\}} \lambda^p dP = \lambda^p P(|\mathbf{X}| \geq \lambda)$$

Chapter 2 - Distribution Functions

- **Notation:** For $x = (x_i)_{i=1}^n, y = (y_i)_{i=1}^n \in \mathbb{R}^n$, we say $x \leq y$ if $x_i \leq y_i$ for all i.

Definition

1. The *distribution function* of $\mathbf{X}$ is $F_{\mathbf{X}} : \mathbb{R}^n \to [0,1]$ where

$$F_{\mathbf{X}}(x) := P(\mathbf{X} \leq x)$$

2. If $\mathbf{X}_1, \ldots, \mathbf{X}_m : \Omega \to \mathbb{R}^n$ are random variables, then their *joint distribution function* $F_{\mathbf{X}_1,\ldots,\mathbf{X}_m} : (\mathbb{R}^n)^m \to [0,1]$ given by

$$F_{\mathbf{X}_1,\ldots,\mathbf{X}_m}(x_1,\ldots,x_m) := P(\mathbf{X}_1 \leq x_1, \ldots, \mathbf{X}_m \leq x_m) \quad \text{for all } x_i \in \mathbb{R}^n$$

3. Given $\mathbf{X}$ and its associated $F_{\mathbf{X}}$, if there exists non-negative integrable $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$F_{\mathbf{X}}(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1,\ldots,y_n) dy_n \ldots dy_1$$

then f is called the *density function* of $\mathbf{X}$.

Chapter 2 - Distribution Functions

- An important observation is that we now have

$$P(\mathbf{X} \in B) = \int_B f(x)dx \quad \text{for all } B \in \mathcal{B}$$

  - The nice fact being, we have the capacity to compute integrals
    - Be it analytically or numerically...

Example

If $\mathbf{X} : \Omega \to \mathbb{R}^n$ has density

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp\left(-\frac{1}{2}(x-m)^T C^{-1}(x-m)\right)$$

for $x, m \in \mathbb{R}^n$ and C symmetric, positive semi-definite, then we say $\mathbf{X}$ is a *normal* distribution with mean m and covarience matrix C.

We notate this by $\mathbf{X} = N(m, C)$ or equivalently $\mathbf{X} \sim N(m, C)$

Chapter 2 - Distribution Functions

Lemma

Let $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ be a random variable with density f. Suppose $g : \mathbb{R}^n \to \mathbb{R}$ and that $Y = g(\boldsymbol{X})$ is integrable. Then

$$E(Y) = \int_{\mathbb{R}}^{n} g(x) f(x) dx$$

- In particular, we have

$$E(\mathbf{X}) = \int_{\mathbb{R}^n} x f(x) dx \quad \text{and} \quad V(\mathbf{X}) = \int_{\mathbb{R}^n} |x - E(\mathbf{X})|^2 f(x) dx$$

- Computing E and V of $\mathbf{X}$ as integrals let us translate information from the "unobservable" probability space $(\Omega, \mathcal{U}, P)$ to the observable $\mathbb{R}^n$.

Chapter 2 - Independence

- Let $(\Omega, \mathcal{U}, P)$ be a probability space, with $A, B \in \mathcal{U}$ with $P(B) > 0$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- A and B are said to be *independent events* if $P(A) = P(A|B)$.
    - Rewriting the main equation gives us an equivalent condition for independence, namely

$$P(A \cap B) = P(A)P(B)$$

    - If (A, B) is an independent pair, then so are (A^c, B), (A, B^c) and (A^c, B^c).
    - This notion is easily extended for more than two events

$$P\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i)$$

Chapter 2 - Independence

- Independence can also be extended to σ-algebras and random variables

Definition

Let $\mathcal{U}_i \subseteq \mathcal{U}$ be σ-algebras. We say that $\{\mathcal{U}_i\}_{i=1}^{\infty}$ are independent if for all choices of $1 \leq k_1 \leq k_2 \leq \cdots \leq k_m$ and events $A_{k_i} \in \mathcal{U}_{k_i}$ we have

$$P\left(\bigcap_{i=1}^{m} A_{k_i}\right) = \prod_{i=1}^{m} P(A_{k_i})$$

Definition

Let $\mathbf{X}_i : \Omega \to \mathbb{R}^n$ be random variables. We say that they are *independent* if for all integers $k \geq 2$ and all borel sets $B_1, \ldots, B_k \subseteq \mathbb{R}^n$, we have

$$P(\mathbf{X}_1 \in B_1, \ldots, \mathbf{X}_k \in B_k) = \prod_{i=1}^{k} P(\mathbf{X}_i \in B_i)$$

Chapter 2 - Independence

Theorem

The random variables $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m : \Omega \to \mathbb{R}^n$ *are independent if and only if*

$$F_{\boldsymbol{X}_1, \dots, \boldsymbol{X}_m}(x_1, \dots, x_m) = \prod_{i=1}^{m} F_{\boldsymbol{X}_i}(x_i) \tag{4}$$

if the random variables have densities, the above equation is equivalent to

$$f_{\boldsymbol{X}_1, \dots, \boldsymbol{X}_m}(x_1, \dots, x_m) = \prod_{i=1}^{m} f_{\boldsymbol{X}_i}(x_i) \tag{5}$$

Proof.

If $\{\mathbf{X}_i\}_{i=1}^m$ are independent then

$$\begin{aligned} F_{\mathbf{X}_1,\ldots,\mathbf{X}_m}(x_1,\ldots,x_m) &= P(\mathbf{X}_1 \le x_1,\ldots,\mathbf{X}_m \le x_m) \\ &= P(\mathbf{X}_1 \le x_1)P(\mathbf{X}_2 \le x_2)\ldots P(\mathbf{X}_m \le x_m) \\ &= \prod_{i=1}^m F_{\mathbf{X}_i}(x_i) \end{aligned} \tag{6}$$

Now, notice that from the definitions of F and f we have the following relation

$$f_{\mathbf{X}_1,\ldots,\mathbf{X}_m}(x_1,\ldots,x_m) = \frac{\partial^m}{\partial x_1 \ldots \partial x_m} F_{\mathbf{X}_1,\ldots,\mathbf{X}_m}(x_1,\ldots,x_m)$$

Now using, the above relation on equation (4) gives equation (5), which shows the equivalence*. Now, we show the converse using (5).

Note: *This assumes $\{\mathbf{X}_i\}$ has a joint density function. It is unclear what happens when it doesn't... □

Proof.

In this case, pick $A_i \in \mathcal{U}(\mathbf{X}_i)$, and this gives us that $A_i = \mathbf{X}_i^{-1}(B_i)$ for some $B_i \in \mathcal{B}$. Now we have that

$$\begin{aligned}
P(A_1 \cap A_2 \cap \cdots \cap A_m) &= P(\mathbf{X}_1 \in B_1, \ldots, \mathbf{X}_m \in B_m) \\
&= \int_{B_1 \times \cdots \times B_m} f_{\mathbf{X}_1, \ldots, \mathbf{X}_m}(x_1, \ldots, x_m) dx_1 \ldots dx_m \\
&= \left(\int_{B_1} f_{\mathbf{X}_1}(x_1) \right) \ldots \left(\int_{B_m} f_{\mathbf{X}_m}(x_m) \right) \\
&= P(\mathbf{X}_1 \in B_1) \ldots P(\mathbf{X}_m \in B_m) \\
&= P(A_1) \ldots P(A_m)
\end{aligned}$$

which completes the proof as desired. $\square$

Chapter 2 - Independence

- A very exciting consequence of the above theorem is the following result

Corollary

If $X_1, \dots, X_m$ are independent, real valued random variables with $E(|X_i|) < \infty$ and $V(X_i) < \infty$ then

$$E(X_1 \dots X_m) = E(X_1) \dots E(X_m)$$

and

$$V(X_1 + \dots + X_m) = V(X_1) + \dots + V(X_m)$$

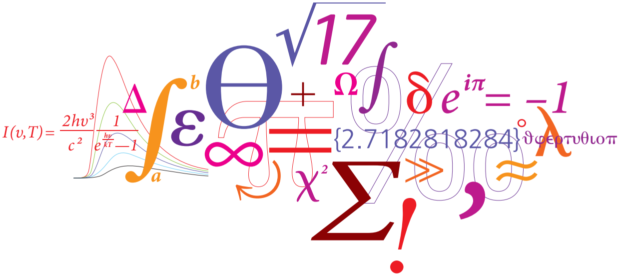
Proof.

Proof of the first statement is identical to the converse of the main theorem. For the second statement, we proceed by induction after letting $E(X_1) = m_1$ and $E(X_2) = m_2$. The base case is trivial, now notice

$$\begin{aligned} V(X_1 + X_2) &= \int_\Omega (X_1 + X_2 - (m_1 + m_2))^2 f dP \\ &= \int_\Omega (X_1 - m_1)^2 f dP + \int_\Omega (X_2 - m_2)^2 f dP \\ &\quad - 2\int_\Omega (X_1 - m_1)(X_2 - m_2) f dP \\ &= V(X_1) + V(X_2) - 2E(X_1 - m_1)E(X_2 - m_2) \\ &= V(X_1) + V(X_2) \end{aligned} \tag{7}$$

□

Thank you!


$I(\nu,T)=\frac{2h\nu^3}{c^2}\frac{1}{e^{\frac{h\nu}{kT}}-1}$
$\sqrt{17}$
$e^{i\pi}=-1$
{2.7182818284}
ϑφερτυθιοπ