# Generalization of stochastic calculus and its applications in large deviations theory

Sudip Sinha

2019-04-05

Advisors
Prof Hui-Hsiung Kuo
Prof Padmanabhan Sundar

# Outline

# § 1
# Introduction and motivation

# Quick revision and notations

- Let $T \in (0, \infty)$, and denote $\mathbb{T} = [0, T]$ as the index set for $t$.
- Let $(\Omega, \mathcal{F}, \mathcal{F}_{\cdot}, \mathbb{P})$ be a filtered probability space.
- $B_{\cdot}$ is a Brownian motion on $(\Omega, \mathcal{F}, \mathcal{F}_{\cdot}, \mathbb{P})$.
- Properties of $B_{\cdot}$
  - starts at 0
  - has independent increments
  - $B_t - B_s \sim \mathcal{N}(0, t - s)$
  - continuous paths
  - has unbounded linear variation ☹
  - has bounded quadratic variation ☺
  - $\mathbb{E}(B_t B_s) = s \wedge t$
  - martingale
- Naive stochastic integration w.r.t. $B_t$: not possible.
- A stochastic process $X_t$ is called $(\mathcal{F}_t)$-adapted if $\forall t$, $X_t$ is measurable w.r.t. $\mathcal{F}_t$.

# Wiener integral ($f \in L^2[0,T]$)

- Definition

  1. Step functions $f = \sum_{j=0}^{n-1} c_j \mathbb{1}_{[t_j,t_{j+1})}(t)$: Define $\int_0^T f(t)\,\mathrm{d}B_t = \sum_{j=0}^{n-1} c_j \Delta B_j$, where $\Delta B_j = B_{t_{j+1}} - B_{t_j}$.
  2. $f \in L^2[0,T]$: Use step functions approximating $f$ to extend the integral a.s.

- Properties

  ⋆ Linear
  ⋆ Gaussian distribution with mean 0 and variance $\|f\|^2_{L^2[0,T]}$ (Itô isometry)
  ⋆ Corresponds to the Riemann–Stieltjes integral for $f \in C[0,T]$

- The associated process: $I_t = \int_0^t X_t\,\mathrm{d}B_t$

  ⋆ continuous
  ⋆ martingale

- Problem: Cannot integrate stochastic processes.

# Itô integral ($X \in L^2_{\text{ad}}([0,T]\times\Omega)$)

- Definition

  1. Adapted step processes $X_t(\omega) = \sum_{j=0}^{n-1} \xi_j(\omega)\mathbb{1}_{[t_j,t_{j+1})}(t)$: define $\int_0^T X_t \,\mathrm{d}B_t = \sum_{j=0}^{n-1} \xi_j \Delta B_j$.
  2. $X \in L^2_{\text{ad}}([0,T]\times\Omega)$: use step processes approximating $X$ to extend the integral in $L^2(\Omega)$.

- Properties

  - ⋆ Linear
  - ⋆ Mean 0 and variance $\|f\|^2_{L^2[0,T]}$ (Itô isometry)
  - ⋆ For $X_\cdot$ continuous, $\int_0^T X_t \,\mathrm{d}B_t = \lim \int_0^T X_{\lfloor \frac{tn}{n} \rfloor} \,\mathrm{d}B_t$, for example $\int_0^t B_s \,\mathrm{d}B_s = \frac{1}{2}\left(B_t^2 - t\right)$

- The associated process: $I_t = \int_0^t X_t \,\mathrm{d}B_t$

  - ⋆ continuous
  - ⋆ martingale

- Example: $\int_0^T B_t \,\mathrm{d}B_t = \frac{1}{2}(B_T - T)$.

# Itô integral ($\int_0^T X_t^2 \, \mathrm{d}t < \infty$ a.s.)

- Definition: Use sequences of processes in $L^2_{\mathrm{ad}}([0,T] \times \Omega)$ approximating $X$ in probability to extend the integral in probability.

- Properties

  ⋆ Linear
  ⋆ Mean 0, but variance? 😐

- The associated process: $I_t = \int_0^t X_t \, \mathrm{d}B_t$

  ⋆ continuous
  ⋆ local martingale

- Example: $\int_0^T e^{B_t^2} \, \mathrm{d}B_t = \int_0^{B_1} e^{t^2} \, \mathrm{d}t - \int_0^T B_t e^{B_t^2} \, \mathrm{d}t.$

# Itô formula

- An Itô process is a process of the form $X_t = X_0 + \int_0^t m(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s$, equivalently expressed as $\mathrm{d}X_t = m(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t$.
  [Only makes sense when $\int_0^T \left(|m(s, X_s)| + |\sigma(s, X_s)|^2\right)\mathrm{d}s < \infty$ a.s.]

**Theorem ([Itô44])** Let $X_t$ be a $d$-dimensional Itô process, and let $Y_t = f(X_t)$, where $f \in C^2(\mathbb{R})$. Then $f(X_t)$ is also a $d$-dimensional Itô process, and

$$\mathrm{d}f(X_t) = \langle (\mathrm{D}f)(X_t), \mathrm{d}X_t\rangle + \frac{1}{2}\left\langle \mathrm{d}X_t, (D^2 f)(X_t)\,\mathrm{d}X_t\right\rangle,$$

where we use the rule $\mathrm{d}B_t \otimes \mathrm{d}B_t = I_d\,\mathrm{d}t$.

- Example: For $\sigma$ constant, $\mathcal{E}_t = \exp\left(\sigma B_t - \frac{1}{2}\sigma^2 t\right)$, $\mathrm{d}\mathcal{E}_t = -\frac{1}{2}\sigma^2\mathcal{E}_t\,\mathrm{d}t + \sigma\mathcal{E}_t\,\mathrm{d}B_t + \frac{1}{2}\sigma^2\mathcal{E}_t(\mathrm{d}B_t)^2$.

# Exponential processes and Girsanov theorem

TODO

Let $h \in L^2[0, T]$. Then the translated stochastic process $W_t = B_t - \int_0^t h(s)\,\mathrm{d}s$ is a Brownian motion under the probability measure $\tilde{\mathbb{P}}$ defined by the Radon-Nikodym derivative $\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} = \exp\left(\sigma B_T - \frac{1}{2}\sigma^2 T\right) =: \mathcal{E}_T^h$.

Then $\tilde{\mathbb{P}} \sim \mathbb{P}$ and the process $Z_t := \mathbb{E}\left(\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} \mid \mathcal{F}_t\right)$ is a martingale.

# Stochastic differential equations

- Let $\xi \in L^2(\Omega)$ be independent of $B_\cdot$, and $m, \sigma : [0,T] \times \mathbb{R} \times \Omega \to \mathbb{R}$ be $\mathcal{B}[0,T] \times \mathcal{B}(\mathbb{R}) \times \mathcal{F}$ measurable such that $m(t,\cdot,\cdot)$ and $\sigma(t,\cdot,\cdot)$ are $\mathcal{B}(\mathbb{R}) \times \mathcal{F}_t$ measurable $\forall t$.
  Then a $\mathcal{F}_t$-adapted stochastic process $X_t$ is called a solution of the stochastic *integral* equation $X_t = \xi + \int_0^t m(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s$ if for each $t$, the $X_t$ satisfies the integral equation a.s.

- The stochastic *differential* equation $\mathrm{d}X_t = m(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t,\ X_0 = \xi$ is a *symbolic representation* of the stochastic integral equation.

**Theorem (Existence and uniqueness, Markov property)** The stochastic differential equation above has a unique solution if there exists an $M > 0$ such that the following two conditions are satisfied:

- ⋆ (Lipschitz condition) $|m(t,x) - m(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \leq M|x-y|^2$ a.s.
- ⋆ (growth condition) $|m(t,x)|^2 + |\sigma(t,y)|^2 \leq M\left(1 + |x|^2\right)$ a.s.

The solution is a Markov process. Moreover if $\xi \in \mathbb{R}$, then the solution is also stationary.

- Example: For $\sigma$ constant, $\mathrm{d}\mathcal{E}_t = \sigma \mathcal{E}_t\,\mathrm{d}B_t, \mathcal{E}_0 = 1$ is solved by $\mathcal{E}_t = \exp\left(\sigma B_t - \frac{1}{2}\sigma^2 t\right)$.

# Multiple Wiener–Itô integrals

# § 2
# Generalization of Itô calculus

# Motivation

- Iterated integrals: Consider the iterated integral $\int_0^T \int_0^T \mathrm{d}B_s\, \mathrm{d}B_t = \int_0^T B_T\, \mathrm{d}B_t \stackrel{?}{=} B_T B_t$.
- Note that $\mathbb{E}(B_T B_t) = T \wedge t = t \neq 0$, so no martingale property ☹.
- Stochastic differential equations with anticipation

$$\begin{aligned} \mathrm{d}X_t &= X_t\, \mathrm{d}B_t & \qquad \mathrm{d}Y_t &= B_T\, \mathrm{d}B_t \\ X_0 &= B_1 & Y_0 &= 1 \end{aligned}$$

- Problem: We want to define $\int_0^T X_t\, \mathrm{d}B_t$, where $X_\cdot$ is not adapted (anticipating).
- Some approaches
  - ⋆ Itô's decomposition of integrand $B_t = \left(B_t - \int_0^t \frac{B_T - B_s}{T-s}\, \mathrm{d}s\right) + \int_0^t \frac{B_T - B_s}{T-s}\, \mathrm{d}s$
  - ⋆ Enlargement of filtration
  - ⋆ White noise theory
  - ⋆ ...

# The new integral [AK08; AK10]: Idea

- A process $\mathcal{Y}^{\cdot}$ and filtration $\mathcal{F}_{\cdot}$ are called instantly independent if $\mathcal{Y}^{t}$ and $\mathcal{F}_{t}$ are independent $\forall t$.
- Ideas
  1. Decompose the integrand into adapted and instantly independent parts.
  2. Evaluate the adapted and the instantly independent parts at the left and right endpoints.
- Consider two continuous stochastic processes, $X_t$ adapted and $\mathcal{Y}^t$ instantly independent w.r.t. $\mathcal{F}_{\cdot}$. Then the integral $\int_0^T X_t \mathcal{Y}^t \, \mathrm{d}B_t$ is defined as

$$\int_0^T X_t \mathcal{Y}^t \, \mathrm{d}B_t \triangleq \lim_{\|\Delta_n\| \to 0} \sum_{j=0}^{n-1} X_{t_j} \mathcal{Y}^{t_{j+1}} \Delta B_j,$$

  provided that the limit exists in probability.
- Now, for any stochastic process $Z(t) = \sum_{k=1}^{n} X_t^{(k)} \mathcal{Y}_{(k)}^{t}$ we extend the definition by linearity.
- This is well-defined [HKS+16].

# A simple example

- In the following, denote $\Delta B_j = B_{t_{j+1}} - B_{t_j}$ and lim is the limit in $L^2$.

$$\int_0^t B_T \,\mathrm{d}B_t = \int_0^t (B_t + (B_T - B_t))\,\mathrm{d}B_t = \int_0^t B_t \,\mathrm{d}B_t + \int_0^t (B_T - B_t)\,\mathrm{d}B_t$$

$$= \lim \sum_{j=0}^{n-1} B_{t_j} \Delta B_j + \lim \sum_{j=0}^{n-1} (B_T - B_{t_{j+1}})\Delta B_j$$

$$= \lim \sum_{j=0}^{n-1} \left(B_T - \Delta B_j\right) \Delta B_j$$

$$= B_T \lim \sum_{j=0}^{n-1} \Delta B_j - \lim \sum_{j=0}^{n-1} (\Delta B_j)^2 = B_T B_t - t$$

- Note that $\mathbb{E}(B_T B_t - t) = 0$.
- In general, $\mathbb{E} \int_0^t Z(t)\,\mathrm{d}B_t = 0$. ☺

# Generalized Itô formula [HKS+16]

- Let $\mathrm{d}X_t = m(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B_t$ be an $d$-dimensional Itô process, $Y^t = \tilde{m}(t)\,\mathrm{d}t + \tilde{\sigma}(t)\,\mathrm{d}B_t$ be a $\tilde{d}$-dimensional instantly independent process, $f(x,y) \in C^2(\mathbb{R}^2)$. Then

$$\begin{aligned}\mathrm{d}f(X_t, Y^t) = {} & \langle (\mathrm{D}_x f)(X_t, Y^t), \mathrm{d}X_t \rangle + \frac{1}{2}\Big\langle \mathrm{d}X_t, (D_x^2 f)(X_t, Y^t)\ \mathrm{d}X_t \Big\rangle \\ & + \Big\langle (\mathrm{D}_y f)(X_t, Y^t), \mathrm{d}Y^t \Big\rangle - \frac{1}{2}\Big\langle \mathrm{d}Y^t, (D_y^2 f)(X_t, Y^t)\ \mathrm{d}Y^t \Big\rangle,\end{aligned}$$

  where we use the rule $\mathrm{d}B_t \otimes \mathrm{d}B_t = I_d\,\mathrm{d}t$.

- Example: TODO

# Iterated integrals

**Theorem** ([Itô51]) Let $f \in L^2([0,T]^n)$ and $\hat{f}$ be its symmetrization. Then

$$\int_{[0,T]^n} f(t_1,\dots,t_n)\, \mathrm{d}B_{t_1}\dots\mathrm{d}B_{t_n} = n! \int_0^T \cdots \int_0^{t_{n-1}} \hat{f}(t_1,\dots,t_n)\, \mathrm{d}B_{t_n}\dots\mathrm{d}B_{t_1},$$

**Theorem** ([AK10]) Let $f \in L^2([0,T]^n)$. Then

$$\int_{[0,T]^n} f(t_1,\dots,t_n)\, \mathrm{d}B_{t_1}\dots\mathrm{d}B_{t_n} = \int_0^T \cdots \int_0^T f(t_1,\dots,t_n)\, \mathrm{d}B_{t_n}\dots\mathrm{d}B_{t_1}.$$

# Near-martingale property [HKS+17]

- Question: What are the analogues of the martingale property and the Markov property?
- Partial answer: near-martingales
- Let $Z(t)$ be a stochastic process such that $\mathbb{E}\,|Z(t)| < \infty\ \forall t$, and $0 \leq s \leq t \leq T$. Then, with respect to $\mathcal{F}_{\cdot}$, the process $Z(t)$ is called a
  - ⋆ near-martingale if $\mathbb{E}(Z(t) - Z(s) \mid \mathcal{F}_s) = 0$,
  - ⋆ near-submartingale if $\mathbb{E}(Z(t) - Z(s) \mid \mathcal{F}_s) \geq 0$, and
  - ⋆ near-supermartingale if $\mathbb{E}(Z(t) - Z(s) \mid \mathcal{F}_s) \leq 0$.

# § 3
# Large deviations theory

# Motivation: an example

1. Setup. Let the following hold:

   - $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.
   - $(X_n)$ is a sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite moment generating function $M$.
   - $\mathbb{E}X_1 = m$, $\mathbb{V}X_1 = \sigma^2$, and $X_1 \sim \mu$.
   - $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$.

2. Asymptotic behavior of $\bar{X}_n$:

   - Weak law of large numbers: $\bar{X}_n \xrightarrow{\mathbb{P}} m$.
   - Central limit theorem: $\sqrt{n}\bar{X}_n \xrightarrow{w^*} \sqrt{n}m + \mathcal{N}(0, \sigma^2)$.

3. But at what speed?

4. We want to *control large deviations from the mean*.

# Example: large deviation bounds

1. Fixing $x > m$ and forcing the exponential with a free parameter $\theta > 0$, we get

$$\mathbb{P}\left\{\bar{X}_n \geq x\right\} = \mathbb{P}\left\{e^{\theta n \bar{X}_n} \geq e^{\theta n x}\right\} \leq e^{-\theta n x}\mathbb{E}\left(e^{\theta n \bar{X}_n}\right) = e^{-\theta n x} M_X(\theta)^n = e^{-n(\theta x - \log M_X(\theta))}$$

2. Since $\theta$ was arbitrary, we have

$$\mathbb{P}\left\{\bar{X}_n \geq x\right\} \leq \inf_\theta e^{-n(\theta x - \log M_X(\theta))} = e^{-n \sup_\theta(\theta x - \log M_X(\theta))} =: e^{-nI(x)}.$$

3. Generalizing, we get the large deviation upper bound

$$\overline{\lim} \frac{1}{n} \log \mathbb{P}\left\{\bar{X}_n \in E\right\} \leq -\inf_{\bar{E}} I \qquad \forall E \in \mathscr{B}.$$

4. We can also obtain a lower bound too using an exponential change of measure

$$\underline{\lim} \frac{1}{n} \log \mathbb{P}\left\{\bar{X}_n \in E\right\} \geq -\inf_{\mathring{E}} I \qquad \forall E \in \mathscr{B}.$$

5. So informally, we get $\mathbb{P}\left\{\bar{X}_n = x\right\} \asymp e^{-nI(x)}$ for $x \in \mathbb{R}$.

# Definitions

- The setup: $(X_n)$ is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in a Polish space $(\mathcal{X}, d)$.
- A function $I : \mathcal{X} \to [0, \infty]$ is called a rate function if it has compact level sets.
- $I$ is lower semicontinuous and attains its infimum on a nonempty closed set.
- For any Borel set $E$, denote $I(E) = \inf_{x \in E} I(x)$.

**Definition** $(X_n)$ is said to satisfy the large deviation principle on $\mathcal{X}$ with rate function $I$ if the following two conditions hold.

$$\begin{array}{lll} \text{(upper bound)} & \overline{\lim} \dfrac{1}{n} \log \mathbb{P}\left\{\bar{X}_n \in F\right\} \leq -I(F) & \forall F \text{ closed} \\ \text{(lower bound)} & \underline{\lim} \dfrac{1}{n} \log \mathbb{P}\left\{\bar{X}_n \in E\right\} \geq -I(G) & \forall G \text{ open} \end{array}$$

# Cramér theorem

**Theorem ([Cra38])** Let $(X_n)$ be a sequence of i.i.d. real random variables with finite moment generating function $M$. Then $(X_n)$ follows large deviation principle with rate function $I(x) = \sup_\theta (\theta x - \log M(\theta))$.

Rate function for some common distributions for $X$.

| Distribution | $M(\theta)$ | $I(x)$ |
|---|---|---|
| $Bern(p)$ | $1 - p + pe^\theta$ | $x \log \frac{x}{1-p} + (x-1) \log \frac{p}{x-1}$ |
| $Pois(\lambda)$ | $e^{\lambda(e^\theta - 1)}$ | $\lambda - x + x \log \frac{x}{\lambda}$ |
| $Exp(\lambda)$ | $(1 - \theta\lambda^{-1})^{-1}$ | $\lambda x - 1 + x \log(\lambda x)$ |
| $\mathcal{N}(m, \sigma^2)$ | $e^{m\theta + \frac{1}{2}\sigma^2\theta^2}$ | $\frac{(x-m)^2}{2\sigma^2}$ |
| $\chi^2(k)$ | $(1 - 2\theta)^{-\frac{k}{2}}$ | $\frac{1}{2}\left(x - k + k \log \frac{k}{x}\right)$ |

# Sanov theorem

# LD in $\infty$-dimensions — Schilder theorem

**Aim**: Estimate the probability that a scaled-down sample path of a Brownian motion will stray far from the mean path (the 0 function).

**Setup**

- Let $B_\cdot$ be a $d$-dimensional Brownian motion, so $B_\cdot \in C_0 = C_0([0,T];\mathbb{R}^d)$
- $\forall \varepsilon > 0$, let $W_\varepsilon$ denote the law of $\sqrt{\varepsilon} B_\cdot$
- Let $\mathrm{CM} = \left\{\omega \in C_0 : \omega \in \mathrm{AC},\ \text{and } \dot{\omega}_t \in L^2[0,T]\right\}]$

**Theorem** On the Banach space $(C_0, \|\cdot\|_\infty)$, the family of probability measures $\{W_\varepsilon : \varepsilon > 0\}$ satisfy the large deviations principle with the rate function $I : C_0 \to \overline{\mathbb{R}}$ given by

$$I(\omega) = \left(\frac{1}{2}\int_0^T |\dot{\omega}(t)|^2 \, \mathrm{d}t\right) \mathbb{1}_{\mathrm{AC}}(\omega) + \infty \mathbb{1}_{\mathrm{AC}^\complement}(\omega)$$

# Freidlin–Wentzell theorem

**Aim**: Estimate the probability that a scaled-down sample path of an Itô diffusion will stray far from the mean path.

**Setup**

- Let $B_{\cdot}$ be a $d$-dimensional Brownian motion, so $B_{\cdot} \in C_0 = C_0([0,T];\mathbb{R}^d)$
- $\forall \varepsilon > 0$, let $X^{(\varepsilon)}$ be a $\mathbb{R}^d$-valued Itô diffusion solving an Itô SDE of the form

$$\mathrm{d}X_t^{(\varepsilon)} = b(X_t^{(\varepsilon)})\,\mathrm{d}t + \sigma(X_t^{(\varepsilon)})\sqrt{\varepsilon}\,\mathrm{d}B_t, \quad X_0^{(\varepsilon)} = 0.$$

- $\forall \varepsilon > 0$, let $W_\varepsilon$ denote the law of $X_{\cdot}^{(\varepsilon)}$.

**Theorem (Freidlin, Wentzell (year?))** On the Banach space $(C_0, \|\cdot\|_\infty)$, the family of probability measures $\{W_\varepsilon : \varepsilon > 0\}$ satisfy the large deviations principle with the rate function $I : C_0 \to \overline{\mathbb{R}}$ given by

$$I(\omega) = \left(\frac{1}{2}\int_0^T |\dot{\omega}_t - b(\omega_t)|^2\,\mathrm{d}t\right)\mathbb{1}_{H^1([0,T];\mathbb{R}^d)}(\omega) + \infty\mathbb{1}_{H^1([0,T];\mathbb{R}^d)^\complement}(\omega)$$

# Applications

•

# § 4
# Conclusion

# Open areas for research

* ⋆ Extension to SDEs with anticipating coefficients
* ⋆ Near-Markov property
* ⋆ Girsanov theorem for anticipating integrals
* ⋆ Freidlin-Wintzell type result for SDEs with anticipation

> The Earth, as a habitat for animal life, is in old age and has a fatal illness. Several, in fact. It would be happening whether humans had ever evolved or not. But our presence is like the effect of an old-age patient who smokes many packs of cigarettes per day—and we humans are the cigarettes.

# Bibliography


Ayed, W. & Kuo, H. H. (2008). An extension of the Itô integral. *Communications on Stochastic Analysis*, *2*(3). doi:10.31390/cosa.2.3.05

——— (2010). An extension of the Itô integral: Toward a general theory of stochastic integration. *Theory of Stochastic Processes*, *16*(32), 17–28. Retrieved from http://mi.mathnet.ru/thsp56

Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, *736*, 5–23.

Hwang, C. R., Kuo, H. H., Saitô, K., & Zhai, J. (2016). A general Itô formula for adapted and instantly independent stochastic processes. *Communications on Stochastic Analysis*, *10*(3). doi:10.31390/cosa.10.3.05

——— (2017). Near-martingale Property of Anticipating Stochastic Integration. *Communications on Stochastic Analysis*, *11*(4). doi:10.31390/cosa.11.4.06

Itô, K. (1944). Stochastic integral. *Proc. Imp. Acad.*, *20*(8), 519–524. doi:10.3792/pia/1195572786

——— (1951). Multiple Wiener Integral. *J. Math. Soc. Japan*, *3*(1), 157–169. doi:10.2969/jmsj/00310157