STOCHASTIC PROCESSES : AN INTRODUCTION

박수찬

21st KIAS-APCTP Winter School on Statistical Physics
2024년 1월 8일∼ 1월 12일

가톨릭대학교 물리학과

Table of contents

Brief Introduction to Probability Theory

Sample space (표본 공간), sample points, and events (사건)

- sample space Ω : a set of all possible outcomes (결과)
  - toss a coin $\Omega = \{\text{H, T}\}$
  - cast a die $\Omega = \{\omega_1, \omega_2, \cdots, \omega_6\}$
  - Maxwell velocity distribution $\Omega = \{(v_1, v_2, v_3) : -\infty < v_i < \infty\}$
  - Wiener Process $\Omega = \{W(t) : W \in C^0, W(0) = 0, 0 \leq t \leq T\}$
- $\omega \in \Omega$: outcome, sample point, sample path
- an event is a subset of the sample space.
  - Ω : a sure (or certain) event
  - $\emptyset$: an event that never happens.
- two events A and B are called *mutually exclusive*, if $A \cap B = \emptyset$.
- if $A_1, A_2, A_3, \cdots$ are events, then we expect

$$\bigcup_{i=1}^{\infty} A_i \text{ and } \bigcap_{i=1}^{\infty} A_i \text{ are events.}$$

Probability (measure)

- for an event A, $0 \le \mathbb{P}(A) \le 1$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ (합의 법칙).
- if $A_1, A_2, \ldots$ are mutually exclusive,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \text{ (countable union).}$$

- why countable?
  consider $\Omega = \{x|0 \le x \le 1\}$, $\mathbb{P}(\{x|a \le x \le b\}) = b - a$.
  let $A_c = \{c\}$. $\mathbb{P}(A_c) = 0$, but $\mathbb{P}(\bigcup_{0 \le c \le 1} A_c) = 1$.
- 참고 : Banach-Tarski paradox (수학적 세계와 물리적 세계의 차이)

Random variable (확률 변수) and stochastic process (확률 과정)

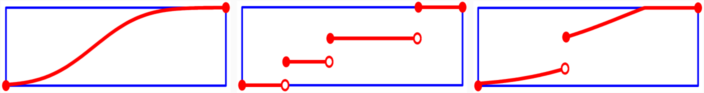
- A random variable (RV) is a function $X : \Omega \mapsto \mathbb{R}$.
- Unlike its name, RV is a function of outcomes, not a variable (in a usual sense).
- RV's are generally denoted by uppercase letters (X, Y, Z).
- Outcomes are generally denoted by lowercase letters (x, y, z).
- X is not necessarily a one-to-one function.
- for example,
  - cast a die $X(\omega_n) = n$.
  - cast a die $X(\text{Even}) = 1,\ X(\text{Odd}) = -1$.
  - Maxwell velocity distribution $X(\omega) = v_1$
  - Wiener Process $X(\omega) = W(t)$ at "time" t
- stochastic process : random variables indexed by "time."

Probability distribution

- (cumulative) distribution function

$$F(x) = \mathbb{P}(X \leq x)$$

- càdlàg (or RCLL) function : right continous with left limits



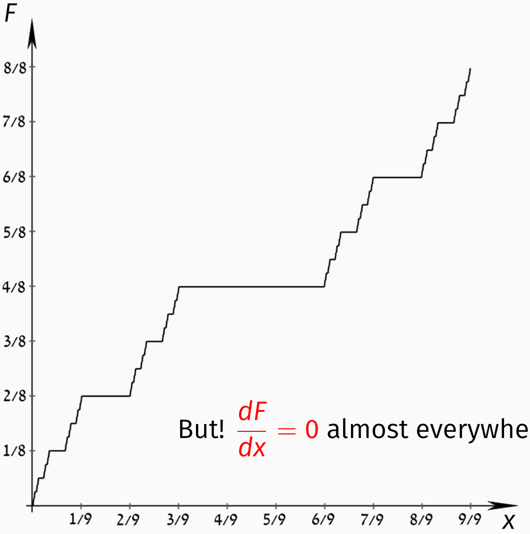
- French: “continue à droite, limite à gauche”
- Skorokhod space : a set of all càdlàg functions

- If F is absolutely continuous, there is P such that

$$F(x) = \int_{-\infty}^{x} P(y) dy.$$

- $P(x)$ is called a *probability density function* (math, physics) or a *probability distribution* (physics).
- For a discrete RV, it is convenient to set $P(x) = \sum_n p_n \delta(x - x_n)$.

a continuous distribution function without a density



But! $\frac{dF}{dx} = 0$ almost everywhere.

Characteristic function

- average $\langle f(X)\rangle = \int_{-\infty}^{\infty} f(x)P(x)dx$
- characteristic function (CF)

$$G(k) \equiv \left\langle e^{ikX} \right\rangle = \int e^{ikx} P(x)dx$$

- $G(k)$ exists for all real k with $G(0) = 1$.
- In general (when?), $|G(k)| < 1$ for nonzero k.
- inverse formula

$$P(x) = \frac{1}{2\pi} \int G(k)e^{-ikx}dk$$

$G(k)$ characterizes $P(x)$.

Moments and moment-generating function

- m-th Moment $\mu_m \equiv \langle X^m \rangle$
- If $G(k)$ is analytic at $k = 0$,

$$G(k) = 1 + \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} \mu_m \to \mu_m = (-i)^m \frac{\partial^m}{\partial k^m} G(k) \bigg|_{k=0}$$

(Moment-)Generating Function (GF)

- CF always exists, but GF may not. For example, $G(k) = \exp(-|k|)$.
- In most cases in physics, CF is identical to GF.
- If CF's of P_X and P_Y are identical, then $P_X = P_Y$ (almost everywhere).

- If X only assumes integral values, it is convenient to introduce

$$\mathscr{G}(z) \equiv \sum_{n=-\infty}^{\infty} z^n P_n, \quad P_n = \frac{1}{2\pi i} \oint_{|z|=1} \frac{\mathscr{G}(z)}{z^{n+1}} dz.$$

- In this case, we define the *factorial moments*

$$\phi_m \equiv \langle X(X-1)\cdots(X-m+1)\rangle, \quad \phi_0 = 1.$$

Sometimes $\langle X^m \rangle_f$ is used to denote ϕ_m.

$$\left.\frac{d^m}{dz^m}\mathscr{G}(z)\right|_{z=1} = \phi_m.$$

- Gaussian

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) \to G(k) = \exp\left(ik\mu_1 - \frac{1}{2}\sigma^2 k^2\right)$$

- Lorentzian (Cauchy distribution)

$$P(x) = \frac{1}{\pi}\frac{1}{x^2+1} \to G(k) = e^{-|k|}$$

No moments exist. (even average does not exist.)

- Poisson distribution

$$P_n = \frac{\lambda^n}{n!} e^{-\lambda} \to \mathscr{G}(z) = e^{(z-1)\lambda}, \quad \phi_m = \lambda^m.$$

- Can we conclude $P_X = P_Y$ if all moments are the same?
- log-normal distribution : $\ln X$ is normal-distributed.

$$P(x) = \Theta(x)\frac{1}{x\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\ln x)^2\right] \Rightarrow \mu_m = \exp(m^2/2).$$

$$P_\varepsilon(x) := P(x)\left[1+\varepsilon \sin(2\pi \ln x)\right]$$

Since, for any non-negative integer n (using $\ln x = y + n$),

$$\int_0^\infty x^n P(x) \sin(2\pi \ln x) dx = 0,$$

$\mu_n = \exp(m^2/2)$ for any ε with $|\varepsilon| < 1$.

- NB: Since $\lim\limits_{m\to\infty} \frac{|k|^m}{m!}\mu_m = \infty$, $\sum\limits_n^\infty \frac{(ik)^m}{m!}\mu_m$ is ill-defined.

Cumulant generating function

$$\ln G(k) = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} \kappa_m, \rightarrow \kappa_m = (-i)^m \frac{\partial^m}{\partial k^m} \ln G(k) \Big|_{k=0}$$

- $\kappa_1 = \mu_1$: mean
- $\kappa_2 = \mu_2 - \mu_1^2 = \left\langle (X - \langle X \rangle)^2 \right\rangle$: variance, centered 2nd moment
- $\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \left\langle (X - \langle X \rangle)^3 \right\rangle$: centered 3rd moment
- skewness = $\frac{\kappa_3}{\kappa_2^{3/2}}$
- $\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^6 \neq \left\langle (X - \langle X \rangle)^4 \right\rangle$
- kurtosis = $\frac{\kappa_4}{\kappa_2^2}$

Examples

- Gaussian

$$\ln G(k) = ik\mu_1 + \frac{(ik)^2}{2}\sigma^2$$

$$\kappa_1 = \mu_1, \quad \kappa_2 = \sigma^2, \quad \kappa_m = 0 \text{ for } m > 2.$$

- Does $P(x)$ exist whose $\ln G(k)$ is a polynomial of order $n > 2$? No! (Marcinkiewicz theorem)
- Poisson distribution

$$G(k) = \mathscr{G}(e^{ik}) = \exp\left[\left(e^{ik} - 1\right)\lambda\right].$$

$$\ln G(k) = \left(e^{ik} - 1\right)\lambda = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!}\lambda \rightarrow \kappa_m = \lambda \text{ for all } m.$$

Multivariate random variables or random vector

- Let $\boldsymbol{X}$ be a random vector with components $X_1, \cdots, X_r$.
- joint probability distribution (결합확률분포) of r variables

$$P_r(\boldsymbol{x}) = P_r(x_1, x_2, \cdots, x_r)$$

- marginal probability distribution (주변확률분포)

$$P_s(x_1, \cdots, x_s) \equiv \int P_s(x_1, \cdots, x_s, x_{s+1}, \cdots x_r) dx_{s+1} \cdots dx_r$$

- conditional probability

$$P(x_1, \cdots, x_s | x_{s+1}, \cdots, x_{s+k}) = \frac{P_s(x_1, \cdots, x_r)}{P_k(x_{s+1}, \cdots, x_{s+k})}$$

Remove delta functions, if any, in the denominator.

- Two sets of RV's $(X_1, \cdots, X_s)$ and $(X_{s+1}, \cdots, X_{s+k})$ are statistically independent if

  $$P_{s+k}(x_1, \cdots, x_{s+k}) = P_s(x_1, \cdots, x_s) P_k(x_{s+1}, \cdots, x_{s+k}).$$

  Accordingly,

  $$P(x_1, \cdots, x_s | x_{s+1}, \cdots, x_{s+k}) = P_s(x_1, \cdots, x_s).$$

- Random variables $X_1, \cdots, X_r$ are called independent and identically distributed (i.i.d.) if
  - $P_r(x_1, \cdots, x_r) = P_1(x_1) \cdots P(x_r)$,
  - $P_1(X_i = x) = P_1(X_j = x)$ for all pairs of i, j.

- pairwise independence : For any pair i, j, $P(x_i, x_j) = P(x_i)P(x_j)$.
- Does pairwise independence imply statistical independence?
- counterexample
  - sample space $\Omega = \{(1,1,1),(1,0,0),(0,1,0),(0,0,1)\}$.
    $\boldsymbol{X} = (X_1, X_2, X_3), \quad P_3(\boldsymbol{X}) = 1/4.$

    $$P_1(X_i = 1) = P_1(X_i = 0) = \frac{1}{2}$$
  - It is easy to prove pairwise independence.

    $$P_2(X_1, X_2) = P_1(X_1)P_1(X_2)$$

    However,
    $P_3(X_1 = 1, X_2 = 1, X_3 = 1) \neq P_1(X_1 = 1)P_1(X_2 = 1)P_1(X_3 = 1)$.

- If X_1 and X_2 are independent,

  $$\langle f_1(X_1) f_2(X_2)\rangle = \langle f_1(X_1)\rangle \langle f_2(X_2)\rangle .$$

  In particular, the characteristic function of $Y = X_1 + X_2$ is

  $$G(k) \equiv \left\langle e^{ikY} \right\rangle = \left\langle e^{ik(X_1+X_2)} \right\rangle = G_{X_1}(k) G_{X_2}(k).$$

- covariance

  $$\langle X_1, X_2\rangle \equiv \langle (X_1 - \langle X_1\rangle)(X_2 - \langle X_2\rangle)\rangle = \langle X_1 X_2\rangle - \langle X_1\rangle\langle X_2\rangle .$$

  If X_1 and X_2 are independent, $\langle X_1, X_2\rangle = 0$.

Law of large numbers

Markov's inequality

If X is a positive RV, then $\mathbb{P}(X \geq a > 0) \leq \frac{\mu_1}{a}$ (trivial if $\mu_1 = \infty$).

Proof.

$\mu_1 = \sum_{x<a} xP(x) + \sum_{x \geq a} xP(x) \geq a \sum_{x \geq a} P(x) = a\mathbb{P}(X \geq a)$. □

Chebyshev's inequality

If X is a RV with finite mean μ_1, then $\mathbb{P}(|X - \mu_1| \geq a > 0) \leq \frac{\kappa_2}{a^2}$

Proof.

$$\mathbb{P}(|X - \mu_1| \geq a) = \mathbb{P}((X - \mu_1)^2 \geq a^2) \leq \frac{\kappa_2}{a^2}$$

by the Markov's inequality. □

(weak) law of large numbers

Let $X_1, \ldots, X_n$ be i.i.d. RV's with a common probability density $P(x)$. If the mean μ_1 and variance κ_2 of $P(x)$ exist, then for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu_1\right| < \varepsilon\right) = 1.$$

Proof.
Let $Y_n = (X_1 + \cdots + X_n)/n$. Since $\langle Y_n \rangle = \mu_1$ and $\langle (Y_n - \mu_1)^2 \rangle = \kappa_2/n$, Chebyshev's inequality gives

$$1 - \frac{\kappa_2}{n\varepsilon^2} \leq \mathbb{P}(|Y_n - \mu_1| < \varepsilon) \leq 1.$$

Hence we get the desired limit. □

- If a sequence of RV's $(Y_n)_{n=1}^{\infty}$ satisfies

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - a| < \varepsilon) = 1$$

for any $\varepsilon > 0$, we write $Y_n \overset{P}{\to} a$ (convergence *in probability*).

- statistical probability (통계적 확률)

$$X_i = \begin{cases} 1 & \text{if an event } A \text{ happens}, \\ 0 & \text{otherwise}, \end{cases} \qquad P(x) = \begin{cases} \mathbb{P}(A) & x = 1, \\ 1 - \mathbb{P}(A) & x = 0. \end{cases}$$

Since $\mu_1 = \mathbb{P}(A)$ and $\kappa_2 < \infty$,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mathbb{P}(A).$$

Central limit theorem (CLT)

Let $X_1, \dots, X_n, \dots$ be a sequence of independent RV's. Assume $m_i := \langle X_i \rangle$ and $\sigma_i^2 := \langle (X_i - m_i)^2 \rangle$ exist for all i. Let $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m_i}{\sigma_i}$. Then,

$$\lim_{n\to\infty} \mathbb{P}\left(S_n < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}y^2\right) dy$$

NB: the CLT implies the (weak) law of large numbers, because for any $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(|Y_n - \mu_1| < \varepsilon\right) &= \mathbb{P}\left(|S_n| < \frac{\sqrt{n}\varepsilon}{\kappa_2}\right) \\ &\xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1. \end{aligned}$$

Taylor expansion of the cumulant generating function for a RV $(X_i - m_i)/\sigma_i$ is $\ln G_i(k) = -\frac{k^2}{2} + R_i(k)$, where $R_i(x)/x^2 \to 0$ as $x \to 0$. Now consider the characteristic function for S_n

$$\langle e^{ikS_n} \rangle = \prod_{i=1}^{n} G_i\left(\frac{k}{\sqrt{n}}\right) \Rightarrow \ln\langle e^{ikY_n}\rangle = -\frac{k^2}{2} + \sum_{i=1}^{n} R_i\left(\frac{k}{\sqrt{n}}\right).$$

Since, for any $k \neq 0$,

$$\left|\sum_{i=1}^{n} R_i\left(\frac{k}{\sqrt{n}}\right)\right| \leq n \max\left\{\left|R_i\left(\frac{k}{\sqrt{n}}\right)\right|\right\} \to 0,$$

we have the desired result.

Stable distributions

stable distribution

Assume X_1, X_2, X are i.i.d. RV's with $P(x)$. If $aX_1 + bX_2$ is identical (in distribution) to $cX + d$ with suitable c and d, then $P(x)$ is called a stable distribution.

- Gaussian (assignment 1)

  Let X_i's are i.i.d. Gaussian RV with mean 0 and variance 1. Let $Y = (X_1 + \cdots + X_n)/\sqrt{n}$. Then, $P_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$.

- Lorentzian (cf: the law of large numbers)

$$G(k) = e^{-|k|}$$

Let $Y = (X_1 + \cdots + X_n)/n$. $G_Y(k) \equiv \langle e^{ikY} \rangle = \left(G\left(\frac{k}{n}\right) \right)^n = G(k)$

- Lévy distribution (assignment 2)

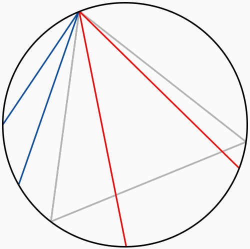
$$\left\langle e^{-kx} \right\rangle = e^{-\sqrt{2k}} \quad (k > 0)$$

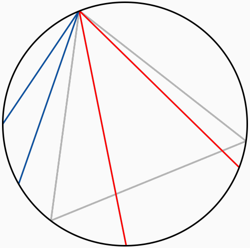
To do list

1. postulating probability: equal *a priori* probability, for instance
2. performing suitable mathematical and/or numerical analysis
3. comparing the *a posteriori* distribution with observation

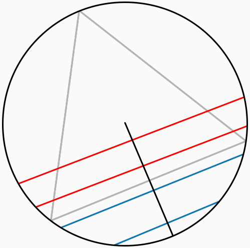
Equal *a priori* probability : importance of ensemble

- principle of insufficient reason (Laplace)
- Bertrand's circle with "random" chords.
  If a chord is chosen "at random" in a circle of radius 1, what is the probability that the length of the chord is larger than $\sqrt{3}$?

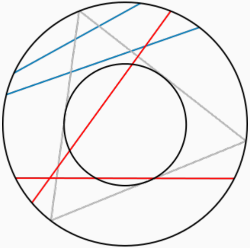




$\frac{1}{3}$



$\frac{1}{2}$



$\frac{1}{4}$

For a detail, see

`en.wikipedia.org/wiki/Bertrand_paradox_(probability)`

Markov Processes

Stochastic process

- stochastic process $\boldsymbol{X}(t)$: a collection of RV's indexed by t.
- In physics, the index is usually interpreted as "time."
- The range (possible values) of random variables is called a *state space* to be denoted in this lecture by $\mathfrak{X}$.
- Elements of $\mathfrak{X}$ can be "vectors." (boldface X)
- "Space" (state space) and "time" (index) can respectively be either continuous or discrete.
- sample path : a realization of $\boldsymbol{X}(t)$ in a probability experiment.
- We can measure values $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$ at times $t_1, t_2, \cdots$.
- Stochastic process is fully determined by $P(\boldsymbol{x}_1, t_1; \cdots; \boldsymbol{x}_n, t_n; \cdots)$

a) complete independence

$$P(x_1, t_1; x_2, t_2; \cdots) = \prod_i P(x_i, t_i)$$

b) Bernoulli (or binomial) process: complete independence and time-independent success (=1) probability $P(x_i, t_i) = P(x_i)$ $(\mathfrak{X} = \{0, 1\})$. Discrete space, discrete time.

c) For a Bernoulli process X_t, set $S_t = X_1 + \cdots + X_t$ (random walk).

d) martingales (fair games):

$$\langle \boldsymbol{X}(t) | [\boldsymbol{x}_0, t_0] \rangle \equiv \int d\boldsymbol{x}\, \boldsymbol{x}\, P(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) = \boldsymbol{x}_0$$

We have defined conditional average

e) Markov processes: Present determines future.

Markov process

Markov assumption ($t_1 > t_2 > \cdots > t_n > \tau_1 > \tau_2 > \cdots$)

$$P(\boldsymbol{x}_1, t_1; \cdots; \boldsymbol{x}_n, t_n | \boldsymbol{y}_1, \tau_1; \boldsymbol{y}_2, \tau_2; \cdots) = P(\boldsymbol{x}_1, t_1; \cdots; \boldsymbol{x}_n, t_n | \boldsymbol{y}_1, \tau_1)$$

- Only in this section, $t_i \geq \tau_j$ (for all i, j) is assumed.
- $P(\boldsymbol{x}, t | \boldsymbol{y}, \tau)$ is called the *transition probability*.
- $P(\boldsymbol{x}, t | \boldsymbol{y}, \tau)$ completely defines the Markov process.

$$\begin{aligned}
&P(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \cdots; \boldsymbol{x}_n, t_n) \\
&= P(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2; \cdots; \boldsymbol{x}_n, t_n) P(\boldsymbol{x}_2, t_2; \cdots; \boldsymbol{x}_n, t_n) \\
&= P(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2) P(\boldsymbol{x}_2, t_2 | \boldsymbol{x}_3, t_3) \cdots P(\boldsymbol{x}_{n-1}, t_{n-1} | \boldsymbol{x}_n, t_n) P(\boldsymbol{x}_n, t_n),
\end{aligned}$$

provided $t_1 > t_2 > \cdots > t_n$.

- Does the Markov assumption impose time direction?

Present determines past, too.

$$P(\boldsymbol{y}_1, \tau_1 | \boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) = P(\boldsymbol{y}_1, \tau_1 | \boldsymbol{x}_2, t_2) \text{ if } t_1 > t_2 > \tau_1.$$

$$\begin{aligned}
P(y_1, \tau_1 | x_1, t_1; x_2, t_2) &= \frac{P(x_1, t_1; x_2, t_2; y_1, \tau_1)}{P(x_1, t_1; x_2, t_2)} \\
&= \frac{P(x_1, t_1 | x_2, t_2; y_1, \tau_1) P(x_2, t_2; y_1, \tau_1)}{P(x_1, t_1; x_2, t_2)} \\
&= \underbrace{\frac{P(x_1, t_1 | x_2, t_2)}{P(x_1, t_1; x_2, t_2)}}_{=1/P(x_2, t_2)} P(x_2, t_2; y_1, \tau_1) \\
&= \frac{P(x_2, t_2; y_1, \tau_1)}{P(x_2, t_2)} = P(y_1, \tau_1 | x_2, t_2)
\end{aligned}$$

Provided $t_1 > t_2 > \cdots > t_n$, we use $P(A|B) = P(B|A)P(A)/P(B)$ to get

$$\begin{aligned} P(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2; \cdots; \boldsymbol{x}_n, t_n) &= \left[\prod_{i=1}^{n-1} P(\boldsymbol{x}_i, t_i|\boldsymbol{x}_{i+1}, t_{i+1})\right] P(\boldsymbol{x}_n, t_n) \\ &= \left[\prod_{i=1}^{n-1} P(\boldsymbol{x}_{i+1}, t_{i+1}|\boldsymbol{x}_i, t_i) \frac{P(\boldsymbol{x}_i, t_i)}{P(\boldsymbol{x}_{i+1}, t_{i+1})}\right] P(\boldsymbol{x}_n, t_n), \\ &= \left[\prod_{i=1}^{n-1} P(\boldsymbol{x}_{i+1}, t_{i+1}|\boldsymbol{x}_i, t_i)\right] p(\boldsymbol{x}_1, t_1), \end{aligned}$$

$P(\boldsymbol{y}, \tau|\boldsymbol{x}, t)$ also determines the stochastic process to the past.

- a question : for a Markov process

$$\mathbb{P}(a_3 \leq X_3 \leq b_3|a_2 \leq X_2 \leq b_2; a_1 \leq X_1 \leq b_1) \overset{?}{=} \mathbb{P}(a_3 \leq X_3 \leq b_3|a_2 \leq X_2 \leq b_2)$$

Chapman-Kolmogorov equation

- two identities (vaild to all stochastic processes)

$$
\begin{aligned}
P(\boldsymbol{x}_1, t_1) &= \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2) P(\boldsymbol{x}_2, t_2), \\
P(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_3, t_3) &= \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2 | \boldsymbol{x}_3, t_3) \\
&= \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1 | \boldsymbol{x}_2, t_2; \boldsymbol{x}_3, t_3) P(\boldsymbol{x}_2, t_2 | \boldsymbol{x}_3, t_3)
\end{aligned}
$$

- Time ordering is not assumed.
- If $t_1 > t_2 > t_3$ and the Markov assumption is introduced, we have

Chapman-Kolmogorov (CK) equation

$$P(\boldsymbol{x}_1, t_1|\boldsymbol{x}_3, t_3) = \int d\,\boldsymbol{x}_2\, P(\boldsymbol{x}_1, t_1|\boldsymbol{x}_2, t_2)P(\boldsymbol{x}_2, t_2|\boldsymbol{x}_3, t_3).$$

- consistency check

$$\begin{aligned} P(\boldsymbol{x}_1, t_1) &= \int d\boldsymbol{x}_3\; P(\boldsymbol{x}_1, t_1; \boldsymbol{x}_3, t_3) = \int d\boldsymbol{x}_3\; P(\boldsymbol{x}_1, t_1|\boldsymbol{x}_3, t_3)P(\boldsymbol{x}_3, t_3) \\ &= \int d\boldsymbol{x}_3\; d\boldsymbol{x}_2\; P(\boldsymbol{x}_1, t_1|\boldsymbol{x}_2, t_2)P(\boldsymbol{x}_2, t_2|\boldsymbol{x}_3, t_3)P(\boldsymbol{x}_3, t_3) \\ &= \int d\boldsymbol{x}_2\; P(\boldsymbol{x}_1, t_1|\boldsymbol{x}_2, t_2)P(\boldsymbol{x}_2, t_2) = \int d\boldsymbol{x}_2\; P(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) \end{aligned}$$

Is a solution of the CK equation a Markov process?

$\Omega = \{(1,1,1), (1,0,0), (0,1,0), (0,0,1)\}$

$$\begin{aligned} P(1_3|1_1) &= \frac{1}{2} = P(1_3|0_2)P(0_2|1_1) + P(1_3|1_2)P(1_2|1_1), \\ P(0_3|1_1) &= \frac{1}{2} = P(0_3|0_2)P(0_2|1_1) + P(0_3|1_2)P(1_2|1_1), \\ P(1_3|0_1) &= \frac{1}{2} = P(1_3|0_2)P(0_2|0_1) + P(1_3|1_2)P(1_2|0_1), \\ P(0_3|0_1) &= \frac{1}{2} = P(0_3|0_2)P(0_2|0_1) + P(0_3|1_2)P(1_2|0_1), \end{aligned}$$

Hence, $P(x_3|x_1) = \sum_{x_2=0}^{1} P(x_3|x_2)P(x_2|x_1)$. But, $P(1_3|1_2;1_1) = 1 \neq P(1_3|1_2)$.

The CK equation is a necessary condition for the Markov property.

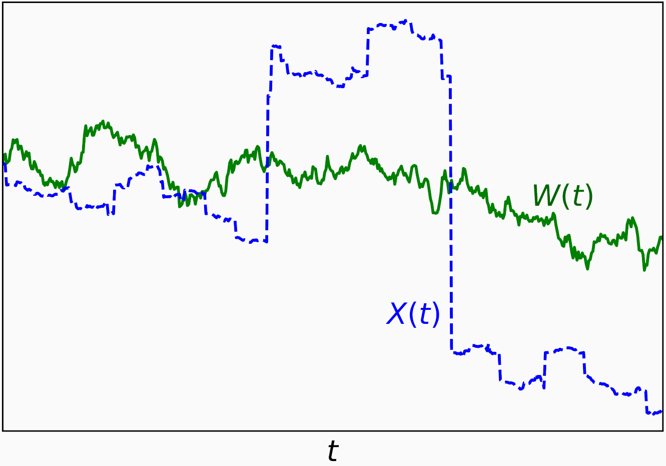
Continuity of stochastic processes

- Lindberg condition

  For a Markov process with continuous space-time, the sample paths are continuous function of t with probability one, if, for any $\varepsilon > 0$,

$$\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{y}|>\varepsilon} d\boldsymbol{x}\, P(\boldsymbol{x}, t+\Delta t | \boldsymbol{y}, t) = 0.$$

- examples (assignment 3)
  - $P(x, t+\Delta t | y, t) = \frac{1}{\sqrt{4\pi D \Delta t}} \exp\left(-\frac{(x-y)^2}{4D\Delta t}\right)$: continuous
  - $P(x, t+\Delta t | y, t) = \frac{\Delta t}{\pi\left[(x-y)^2 + \Delta t^2\right]}$: discontinuous



$X(t)$: Cauchy process (Lorentzian), $W(t)$: Wiener Process (Gaussian).

Differential Chapman-Kolmogorov equation

$$\left.\begin{aligned}\frac{\partial P(\boldsymbol{x},t|\boldsymbol{y},t_0)}{\partial t} &= -\sum_i \frac{\partial}{\partial x_i}\left[A_i(\boldsymbol{x},t)P(\boldsymbol{x},t|\boldsymbol{y},t_0)\right] \\ &\quad +\frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\left[B_{ij}(\boldsymbol{x},t)P(\boldsymbol{x},t|\boldsymbol{y},t_0)\right]\end{aligned}\right\} \text{continuous (Fokker-Planck equation)}$$

$$+\underbrace{\int d\boldsymbol{z}\left[W(\boldsymbol{x}|\boldsymbol{z},t)P(\boldsymbol{z},t|\boldsymbol{y},t_0) - W(\boldsymbol{z}|\boldsymbol{x},t)P(\boldsymbol{x},t|\boldsymbol{y},t_0)\right],}_{\text{discontinuous (master equation)}}$$

$$W(\boldsymbol{x}|\boldsymbol{z},t) \equiv \lim_{\Delta t\to 0} P(x,t+\Delta t|\boldsymbol{z},t)/\Delta t,$$

$$A_i(z,t) = \lim_{\varepsilon\to 0}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}(x_i - z_i)p(\boldsymbol{x},t+\Delta t|\boldsymbol{z},t),$$

$$B_{ij}(z,t) = \lim_{\varepsilon\to 0}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} d\boldsymbol{x}(x_i - z_i)(x_j - z_j)p(\boldsymbol{x},t+\Delta t|\boldsymbol{z},t).$$

Markovian versus non-Markovian

- In mechanics, present always dictates future.
- Then, is every physically relevant process a Markovian?
    - Yes, if we can track down every degree of freedom in the universe!
    - No, if we only look at small part of the universe and this part is not independent of the other part of the universe.
- However, we can come up with an effective Markovian theory, which should be compared with experiments (for example, Einstein, Langevin). This is what physicists have been doing (square well potential in QM, for example).

- Certain non-Markovian process can be cast into a Markovian by extending the state space.
- for example, persistent random walks in one dimension.
  - probability p of continuing same direction. ($p \neq \frac{1}{2}$)
  - If $\mathfrak{X}_0 = (\dots, -2, -1, 0, 1, 2, \dots)$, non-Markovian
  - If $\mathfrak{X} = \{(1, -1)\} \otimes \mathfrak{X}_0 = \{\dots, [(1, 0), (-1, 0)], \dots\}$, Markovian.

$$\mathbb{P}((s, n), t) = p\mathbb{P}((s, n - s), t) + (1 - p)\mathbb{P}((-s, n - s), t)$$

- Likewise, a Markovian becomes non-Markovian if we limit ourselves to a smaller state space (coarse-graining, if you like).

$$\mathbb{P}(n, t) = \sum_{s=\pm 1} \mathbb{P}((s, n), t).$$

- Question: Can you write an evolution equation for $\mathbb{P}(n, t)$ only using $\mathbb{P}(n, t')$? (Don't take it seriously)

Master Equation

Discrete space-time

- Markov Process in discrete space-time : CK equation

$$P(\boldsymbol{n}_1, m+1|\boldsymbol{n}', 0) = \sum_{\boldsymbol{n}_2} P(\boldsymbol{n}_1, m+1|\boldsymbol{n}_2, m) P(\boldsymbol{n}_2, m|\boldsymbol{n}', 0).$$

- matrix representation
  Let $\Psi(m) = (P(\boldsymbol{n}_1, m|\boldsymbol{n}', 0), P(\boldsymbol{n}_2, m|\boldsymbol{n}', 0), \ldots)^\dagger$ and
  $T(m)_{\boldsymbol{n}_1\boldsymbol{n}_2} \equiv P(\boldsymbol{n}_1, m+1|\boldsymbol{n}_2, m)$, then

$$\Psi(m+1) = T(m)\Psi(m).$$

- stochastic matrix
  - Its elements are all non-negative.
  - Each column adds up to unity.
  - $(\cdots,1,1,1,\cdots)$ is the left eigenstate of T with eigenvalue 1.

- homogeneous Markov process: If we assume $T(m) = T$ (time independent), then $\Psi(m) = T^m \Psi(0)$ and

$$P(\boldsymbol{n}_1, m|\boldsymbol{n}_2, m') = (T^{m-m'})_{\boldsymbol{n}_1\boldsymbol{n}_2} = P(\boldsymbol{n}_1, m - m'|\boldsymbol{n}_2, 0)$$

- cf: *stationary process* $P(\boldsymbol{x}, t) = P_s(\boldsymbol{x})$
- Markov chains : discrete 'space-time' and (time) homogeneous
- Existence of stationary state for finite system (by Perron-Frobenius theorem)

$$\lim_{m\to\infty} \Psi(m) = \lim_{m\to\infty} T^m \Psi(0) = \Psi_s,$$

  where Ψ_s is the right eigenstate of T with eigenvalue 1.
- cf: random walk in one dimension: no stationary state.

Galton-Watson branching process: an example

- discrete ‘generation’ (time) model
- $p(k)$: probability of each individual’s having k offspring
- X_m : number of individuals at m-th generation
- $\mathfrak{X} = \{0, 1, 2, \ldots\}$: state space
- What is the extinction probability $\mathbb{P}(\lim_{m\to\infty} X_m = 0)$, if $X_0 = 1$?
- i.i.d. RV η_j^{m+1} : number of offspring of j-th individual at generation m

$$X_{m+1} = \sum_{j=1}^{X_m} \eta_j^{m+1}$$

$$T_{ki} \equiv P(X_{m+1} = k | X_m = i) = [p(k)]^{*i} = \sum_{k_1 + \cdots k_i = k} p(k_1) \cdots p(k_i),$$

i-fold convolution of $p(k)$ with itself.

- CK equation

$$P_k(m) \equiv \mathbb{P}(X_m = k) = \sum_{i=0}^{\infty} T_{ki} P_i(m-1)$$

- Generating function

$$\mathscr{G}_m(z) \equiv \langle z^{X_m} \rangle = \sum_{k=0}^{\infty} z^k P_k(m)$$

- Evolution equation for $\mathscr{G}_m$

$$\mathscr{G}_{m+1}(z) = \left\langle z^{X_{m+1}} \right\rangle = \left\langle z^{\sum_{j=1}^{X_m} \eta_j^{m+1}} \right\rangle = \langle \mathscr{G}(z)^{X_m} \rangle = \mathscr{G}_m(\mathscr{G}(z)),$$

where $\mathscr{G}(z) = \sum_{k=0}^{\infty} z^k p(k)$.

- Solution by iteration

$$\begin{aligned}\mathscr{G}_m(z) &= \mathscr{G}_{m-1}(\mathscr{G}(z)) = \mathscr{G}_{m-2}(\mathscr{G}(\mathscr{G}(z)))\\ &= \mathscr{G}_1(\underbrace{\mathscr{G}(\mathscr{G}(\ldots))}_{m-1}) = \mathscr{G}^{(m)}(z) = \mathscr{G}\left(\mathscr{G}^{(m-1)}(z)\right)\\ &= \mathscr{G}(\mathscr{G}_{m-1}(z)),\end{aligned}$$

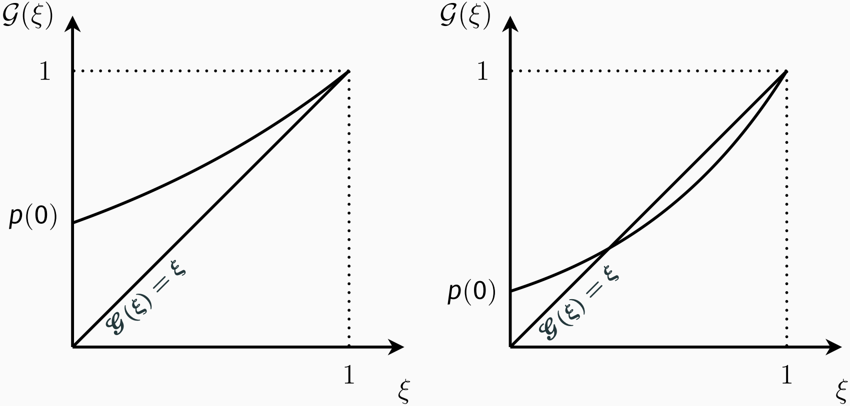
where $\mathscr{G}_1(z) = \mathscr{G}(z)$ (because $X_0 = 1$).

- Extinction probability, ξ

Since $\xi_m \equiv P(X_m = 0) = \mathscr{G}_m(z = 0),\ \xi_m = \mathscr{G}(\xi_{m-1})$. Thus, ξ is the smallest solution of

$$\xi = \mathscr{G}(\xi)(0 \leq \xi \leq 1).$$

$$\mu \equiv \sum_k k p(k) = \mathscr{G}'(z=1), \quad \mathscr{G}(z=1) = 1$$



$\mu < 1$

$\mu > 1$

- mean

$$\begin{aligned}\mu_1(m) \equiv \sum_k k P_k(m) = \mathscr{G}_m'(1) = &\ \frac{\partial}{\partial z}\mathscr{G}\left(\mathscr{G}_{m-1}(z)\right)\Big|_{z=1} \\ = \mathscr{G}'(1)\left(\frac{\partial}{\partial z}\mathscr{G}_m(z)\Big|_{z=1}\right) &= \mu\,\mu_1(m-1)\end{aligned}$$

  Hence, $\mu_1(m) = \mu^m$.

- variance (check it)

$$\sigma(m)^2 = \mathscr{G}_m''(1) + \mathscr{G}_m'(1) - \left[\mathscr{G}_m'(1)\right]^2 = \begin{cases} \sigma^2 \mu^m \dfrac{\mu^m - 1}{\mu^2 - \mu} & \mu \neq 1, \\ m\sigma^2 & \mu = 1. \end{cases}$$

- assignment 4, 5

Markov chains in the continuous time limit

- fixed $t = m\tau$ with $\tau \to 0$.

$$T_{\boldsymbol{n}_1\boldsymbol{n}_2} = \left(1 - \sum_{\boldsymbol{n}_3 \neq \boldsymbol{n}_2} P(\boldsymbol{n}_3, t+\tau|\boldsymbol{n}_2, t)\right)\delta_{\boldsymbol{n}_1,\boldsymbol{n}_2} + P(\boldsymbol{n}_1, t+\tau|\boldsymbol{n}_2, t)(1-\delta_{\boldsymbol{n}_1,\boldsymbol{n}_2})$$

$$\begin{aligned}\frac{dP(\boldsymbol{n}_1, t|\boldsymbol{n}', 0)}{dt} &\equiv \lim_{\tau\to 0} \frac{P(\boldsymbol{n}_1, t+\tau|\boldsymbol{n}', 0) - P(\boldsymbol{n}_1, t|\boldsymbol{n}', 0)}{\tau} \\ &= \sum_{\boldsymbol{n}_2 \neq \boldsymbol{n}_1} [W_{\boldsymbol{n}_1\boldsymbol{n}_2} P(\boldsymbol{n}_2, t|\boldsymbol{n}', 0) - W_{\boldsymbol{n}_2\boldsymbol{n}_1} P(\boldsymbol{n}_1, t|\boldsymbol{n}', 0)],\end{aligned}$$

with *transition rate* $W_{\boldsymbol{n}_1\boldsymbol{n}_2} \equiv \lim_{\tau\to 0} \frac{P(\boldsymbol{n}_1, t+\tau|\boldsymbol{n}_2, t)}{\tau}$.

master equation

$$\frac{dP(\boldsymbol{n}_1, t)}{dt} = \sum_{\boldsymbol{n}_2 \neq \boldsymbol{n}_1} [W_{\boldsymbol{n}_1\boldsymbol{n}_2} P(\boldsymbol{n}_2, t) - W_{\boldsymbol{n}_2\boldsymbol{n}_1} P(\boldsymbol{n}_1, t)]$$

Imaginary-time Schrödinger equation

- ‘ket’ state and the projection vector
  $|\Psi\rangle_t \equiv \sum_{\boldsymbol{n}} P(\boldsymbol{n},t)|\boldsymbol{n}\rangle, \quad \langle\cdot| \equiv \sum_{\boldsymbol{n}} \langle\boldsymbol{n}|, \quad \{|\boldsymbol{n}\rangle\}:$ orthonormal basis.
- Normalization $\langle\cdot|\Psi\rangle_t = 1$ (in QM $\langle\Psi|\Psi\rangle = 1$)
- ‘Hamiltonian’
  $\langle\boldsymbol{n}_1|\hat{H}|\boldsymbol{n}_2\rangle = -W_{\boldsymbol{n}_1,\boldsymbol{n}_2}, \langle\boldsymbol{n}_1|\hat{H}|\boldsymbol{n}_1\rangle = \sum_{\boldsymbol{n}_2\neq\boldsymbol{n}_1} W_{\boldsymbol{n}_2,\boldsymbol{n}_1}.$

imaginary-time Schrödinger equation

$$\frac{\partial}{\partial t}|\Psi\rangle = -\hat{H}|\Psi\rangle \Rightarrow |\Psi\rangle_t = e^{-\hat{H}t}|\Psi\rangle_0.$$

- Due to the normalization $\langle\cdot|\hat{H} = 0$.
- Stationary state (if exists) is the right eigenstate of $\hat{H}$ with eigenvalue 0.

Stationary state and detailed balance

- stationary state $P_s(\boldsymbol{n})$

$$0 = \frac{dP_s(\boldsymbol{n})}{dt} = \sum_{\boldsymbol{n}_1 \neq \boldsymbol{n}} \left[W_{\boldsymbol{n}\boldsymbol{n}_1} P_s(\boldsymbol{n}_1) - W_{\boldsymbol{n}_1\boldsymbol{n}} P_s(\boldsymbol{n})\right] = \langle \boldsymbol{n}|\hat{H}|\Psi\rangle_s \text{ for all } \boldsymbol{n}.$$

- $|\Psi\rangle_s$ is the right eigenstate of $\hat{H}$ for eigenvalue 0.
- In the long time limit, $P(\boldsymbol{n}, t|\boldsymbol{n}_0, 0) \to P_s(\boldsymbol{n})$, irrespective of $\boldsymbol{n}_0$.
- Detailed balance (characteristic of the equilibrium distribution)

$$W_{\boldsymbol{n}_1\boldsymbol{n}_2} P_s(\boldsymbol{n}_2) = W_{\boldsymbol{n}_2\boldsymbol{n}_1} P_s(\boldsymbol{n}_1)$$

- Can we know if the detailed balance is satisfied even though we do not know what $P_s(\boldsymbol{n})$ is? In principle, yes (assignment 6).

Increase of ‘entropy’

Assume that the stationary distribution $P_s(n)$ exists with $P_s(\boldsymbol{n}) > 0$. Define

$$H(t) \equiv \sum_n P_s(n) f\left(\frac{P(n,t)}{P_s(n)}\right) \equiv \sum_n P_s(n) f(x_n),$$

where $f''(x) > 0$ for $0 \le x < \infty$. Then we get

$$\frac{dH(t)}{dt} = \sum_{nm} W_{nm} P_s(m) \left[x_m f'(x_n) - x_m f'(x_m)\right]$$

Note that, for any ψ,

$$\sum_{nm} W_{nm} P_s(m)(\psi_n - \psi_m) = \sum_n \psi_n \sum_m (W_{nm} P_s(m) - W_{mn} P_s(n)) = 0$$

Set $\psi_n = f(x_n) - x_n f'(x_n)$ and add the above two equations.

$$H'(t) = \sum_{nm} W_{nm} P_s(m) \left[(x_m - x_n) f'(x_n) + f(x_n) - f(x_m) \right].$$

Since $f''(x) > 0$ (convex function), $H'(t) < 0$ and, therefore, $H(t)$ monotously decreases to $f(1)$ with t.

Kullback-Leibler divergence

If we choose $f(x) = x \ln x$, we have $H = \sum_n P(n,t) \ln \dfrac{P(n,t)}{P_s(n)}$.

One dimensional random walks : example

- For brevity, we do not write the initial condition.
- $P(n, t)$: prob. that a walker is located at $x = n$ at time t.
- CK equation

$$P(n, t+\tau) = pP(n-1, t) + qP(n+1, t) + (1-p-q)P(n, t).$$

- (naive) continuum limit

$$\frac{P(n, t+\tau) - P(n, t)}{\tau} = \frac{p}{\tau}P(n-1, t) + \frac{q}{\tau}P(n+1, t) - \frac{p+q}{\tau}P(n, t),$$

$$\frac{dP(n, t)}{dt} = w_+P(n-1, t) + w_-P(n+1, t) - (w_+ + w_-)P(n, t),$$

where $p/\tau \to w_+$ and $q/\tau \to w_-$.

Time between jumps

Let $Q(\boldsymbol{n}_1, t, t_0)$ be the probability that we are “still” at point $\boldsymbol{n}_1$ at t, provided we start from $\boldsymbol{n}_1$ at t_0.

$$Q(\boldsymbol{n}_1, t+dt, t_0) = \left(1 - \sum_{\boldsymbol{n}_2 \neq \boldsymbol{n}_1} W_{\boldsymbol{n}_2 \boldsymbol{n}_1} dt\right) Q(\boldsymbol{n}_1, t, t_0),$$

$$\frac{\partial}{\partial t} Q(\boldsymbol{n}_1, t, t_0) = -\sum_{\boldsymbol{n}_2 \neq \boldsymbol{n}_1} W_{\boldsymbol{n}_2 \boldsymbol{n}_1} Q(\boldsymbol{n}_1, t, t_0) \equiv -\lambda Q(\boldsymbol{n}_1, t, t_0),$$

where $\lambda \equiv \sum_{\boldsymbol{n}_2 \neq \boldsymbol{n}_1} W_{\boldsymbol{n}_2 \boldsymbol{n}_1}$. Thus, $Q(\boldsymbol{n}_1, t, t_0) = e^{-\lambda(t-t_0)}$.

For any $\lambda_0 \geq 0$, we have

$$
\begin{aligned}
& e^{-(\lambda+\lambda_0)(t-t_0)} + \sum_{n=1}^{\infty} \int_{t_0}^{t} \lambda_0 dt_1 e^{-(\lambda+\lambda_0)(t_1-t_0)} \int_{t_1}^{t} \lambda_0 dt_2 e^{-(\lambda+\lambda_0)(t_2-t_1)} \times \\
& \cdots \int_{t_{n-2}}^{t} \lambda_0 dt_{n-1} e^{-(\lambda+\lambda_0)(t_{n-1}-t_{n-2})} \int_{t_{n-1}}^{t} \lambda_0 dt_n e^{-(\lambda+\lambda_0)(t_n-t_{n-1})} e^{-(\lambda+\lambda_0)(t-t_n)} \\
& = e^{-(\lambda+\lambda_0)(t-t_0)} \left(1 + \sum_{n=1}^{\infty} \lambda_0^n \int_{t_0}^{t} dt_1 \int_{t_1}^{t} dt_2 \cdots \int_{t_{n-1}}^{t} dt_n \right) \\
& = e^{-(\lambda+\lambda_0)(t-t_0)} \left(1 + \sum_{n=1}^{\infty} \lambda_0^n \frac{(t-t_0)^n}{n!} \right) = e^{-\lambda(t-t_0)} = Q(\boldsymbol{n}_1, t, t_0)
\end{aligned}
$$

See “Merged Poisson processes” below.

Properties of the exponential distribution

- lack of memory or Markov property ($U(t) \equiv \mathbb{P}(\Delta t > t)$)

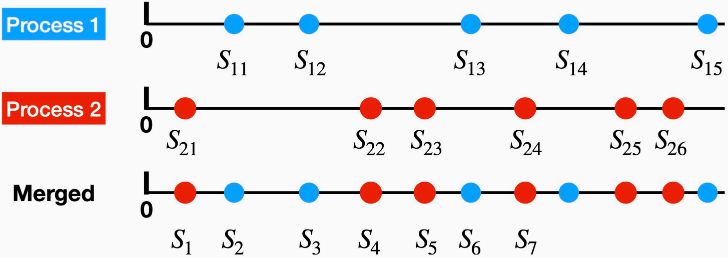
$$\mathbb{P}(\Delta t > s+t | \Delta t > t) = \frac{\mathbb{P}(\Delta t > s+t)}{\mathbb{P}(\Delta t > t)} = \exp(-\lambda s) = \mathbb{P}(\Delta t > s),$$

$$U(t+s) = U(t)U(s).$$

- unique solution of $U(t+s) = U(t)U(s)$ for bounded $U(t)$.
- cf. Cauchy's equation $f(s+t) = f(s) + f(t)$
- Poisson process (assignment 7)
  Let $X_1, \cdots, X_n$ be i.i.d. RV with the exponential distribution. Let $S_n = X_1 + \cdots + X_n$ with $S_0 \equiv 0$. Let $N(t)$ be the number of indices $k \geq 1$ such that $S_k \leq t$, then

$$P(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Merged Poisson processes



Consider n independent Poisson processes with transition rate λ_i each. By S_{ij} we denote j-th occurrence in i-th process (with $S_{i0} = 0$). Then, $X_{ij} := S_{ij} - S_{i,j-1}$ is an i.i.d. RV for each i.

Let $S_1 = \min\{S_{i1} : 1 \leq i \leq n\}$, $S_{k+1} = \min\{S_{ij} : S_{ij} > S_k\}$ $(k \geq 1)$, $Y_{ik} = \min\{S_{ij} - S_{k-1} : \text{ for all } j \text{ with } S_{ij} > S_{k-1}\}$, and $\lambda = \sum_i \lambda_i$.

- $p_i(t) := \mathbb{P}(\text{first event occurs in } i\text{-th process and } S_1 > t)$

$$\begin{aligned} p_i(t) &= \mathbb{P}(X_{j1} > X_{i1} \text{ for all } j \neq i, X_{i1} > t) \\ &= \int_t^\infty \lambda_i e^{-\lambda_i t_i} dt_i \prod_{j \neq i} \int_{t_i}^\infty e^{-\lambda_j t_j} \lambda_j dt_j = \frac{\lambda_i}{\lambda} e^{-\lambda t} \end{aligned}$$

- $\mathbb{P}(S_1 > t) = \sum_i p_i(t) = e^{-\lambda t}$
- $\mathbb{P}(\text{first event occurs in } i\text{-th process}) = p_i(0) = \dfrac{\lambda_i}{\lambda}$
- Thus, two events $\{S_1 > t\}$ and {first event occurs in i-th process} are mutually independent.
- By the Markov property, $\mathbb{P}(Y_{ik} > t) = e^{-\lambda_i t}$ for all i and k. And, accordingly, $\mathbb{P}(k\text{-th event occurs in } i\text{-th process}) = \dfrac{\lambda_i}{\lambda}$ for all i, k.
- Let $X_{k+1} := S_{k+1} - S_k$. Since $X_k = \min\{Y_{ik} : 1 \leq i \leq n\}$, $P(X_k > t) = e^{-\lambda t}$ by the Markov property.

- one way to simulate the merged process
    1. Choose Δt from the exponential distribution $e^{-\lambda t}$. This can be done by $\Delta t = -\ln(1 - Z_1)/\lambda$, where $Z_1 \sim U(0,1)$.
    2. i is chosen with probability λ_i/λ, which can be done by finding i such that $Q_{i-1} < Z_2 < Q_i$, where $Z_2 \sim U(0,1)$ and $Q_i = \sum_{k=1}^{i} \lambda_i/\lambda$ with $Q_0 = 0$.
- $Z \sim U(0,1)$ means that Z is a random number from the unifrom distribution in $(0,1)$.
- Even if the transition rates depend on k in S_k, the above formulae and the simulation method are intact once λ_i is replaced by $\lambda_i(k)$.

Waiting time paradox

- Buses arrive in accordance with a Poisson process with expected time between consecutive buses to be λ^{-1}. I arrive at t. What is the expectation $\langle W_t \rangle$ of my waiting time for the next bus?

  Solution 1 The lack of memory implies $\langle W_t \rangle = \lambda^{-1}$.

  Solution 2 My arrival time is chosen “at random” between two consecutive buses. So due to the symmetry, $\langle W_t \rangle = \lambda^{-1}/2$.

- See assignment 7 for the resolution of the paradox.

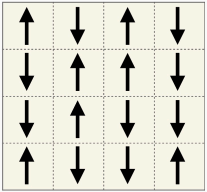
How to simulate a sample path of the master equation

- Assume we are at $\boldsymbol{n}_1$ at time t. Then,
  1. calculate $\lambda = \sum_{\boldsymbol{n}_2 \neq \boldsymbol{n}_1} W_{\boldsymbol{n}_2, \boldsymbol{n}_1}$.
  2. choose Δt from $\mathbb{P}(\Delta t > \tau) = \exp(-\lambda\tau)$.
     If $Z \sim U(0,1)$, we set $\Delta t = -\ln(1-Z)/\lambda$.
  3. choose $\boldsymbol{n}_2$ from $\mathbb{P}(\boldsymbol{n}_2) = W_{\boldsymbol{n}_2 \boldsymbol{n}_1}/\lambda$. We are now at $\boldsymbol{n}_2$ at $t + \Delta t$.
- Since $\langle -\ln(1-Z)\rangle = 1$ and simulation time is much larger than $1/\lambda$ in practice, one usually sets $\Delta t = 1/\lambda$ without generating a random number Z. 주의 : 항상 맞는 결과를 주는 것은 아니다.

$$\frac{\partial P(n,t)}{\partial t} = P(n-1,t) - P(n,t).$$

- Sometimes, calculating λ and $W_{\boldsymbol{n}_2\boldsymbol{n}_1}/\lambda$ can be time consuming. In this case, we choose a convenient λ_0 that normally depends on $\boldsymbol{n}_1$.
    1. Replace λ by $\lambda' := \lambda + \lambda_0$ at step 1. λ_0 should be chosen such that λ' is easy to calculate.
    2. At step 2, choose $\Delta t'$ with λ'.
    3. At step 3, allow $\boldsymbol{n}_2$ to be $\boldsymbol{n}_1$ with probability $\lambda_0/(\lambda + \lambda_0)$. Time increases by $\Delta t'$. At this step, implementing $W_{\boldsymbol{n}_2,\boldsymbol{n}_1}/\lambda'$ should be easy.
- If you like, λ_0 can be regarded as ‘transition rate’ being at $\boldsymbol{n}_1$ $(\lambda_0 = W_{\boldsymbol{n}_1,\boldsymbol{n}_1})$.

Application : spin-flip dynamics of the 2D Ising model

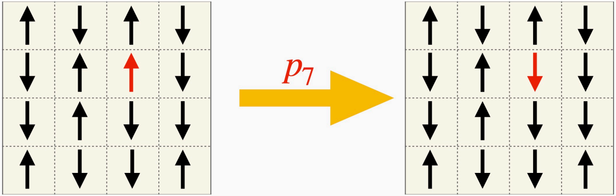


$\boldsymbol{n}_1$

p_1	p_2	p_3	p_4
p_5	p_6	p_7	p_8
p_9	p_{10}	p_{11}	p_{12}
p_{13}	p_{14}	p_{15}	p_{16}

$W_{\boldsymbol{n}_2,\boldsymbol{n}_1}$

$$p_i = \begin{cases} 1 & \Delta E < 0 \\ e^{-\beta\Delta E} & \Delta E > 0 \end{cases}, \qquad \beta\Delta E = 2K s_i \sum_j{}' s_j$$

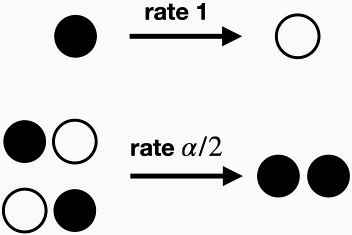


- $\lambda = \sum_i p_i$. $\mathbb{P}(\uparrow_7 \to \downarrow_7) = p_7/\lambda$.
- Let $\lambda_0 = \sum_i (1 - p_i)$, which gives $\lambda' := \lambda + \lambda_0 = N = 16$
  - next 'jump' occurs at $t + 1/N$.
  - probability of staying $\boldsymbol{n}_1$ is $\sum_i (1/N) \times (1 - p_i)$.
  - probability of the flip of the 7th spin is $(1/N) \times p_7$.
- How to do
  1. choose a site i at random with equal probability $1/N$
  2. flip the spin with probability p_i. But with probability $1 - p_i$, nothing happens.
  3. Regardless of the result, time increases by $1/N$. Note that we do not need to calculate λ' and $W_{\boldsymbol{n}_2,\boldsymbol{n}_1}/\lambda'$.
- For an 'exact' simulation, time increment should be $-\ln(1 - Z)/N$ with $Z \sim U(0, 1)$.

- In case certain p_i's are zero, set $\lambda_0 = \sum_i'(1-p_i)$ (sum over nonzero p_i). Then λ' is the number N' of sites with nonzero p_i.
    1. Choose one site i among sites with nonzero p_i with equal probability $1/N'$.
    2. spin-flip dynamics as in step 2 above.
    3. time increases by $1/N'$.
- example : contact process in one dimension

$$\lambda = 3 + 2\alpha$$

$$\lambda' = 3(1+\alpha), \quad p := \frac{1}{1+\alpha}$$



- method with λ'
    - Set $\Delta t = -\log(1 - Z)/\lambda'$ (or simply $1/\lambda'$)
    - Choose one of particles with equal probability.
        - With probability p, this particle is removed.
        - With probability $1 - p$, choose one of its nearest neighbor with equal probability. If the chosen neighbor is empty, add a particle there. Otherwise, nothing happens.
        - Time increases by Δt.
- inefficient but right method
    - Let $\lambda'' = N(1 + \alpha)$, where N is the number of sites.
    - Set $\Delta t = -\log(1 - Z)/\lambda''$.
    - Choose one of sites with equal probability.
        - If the chosen site is empty, nothing happens.
        - If the chosen site is occupied, do the same with p as above.
        - Time increases by Δt.

Some tips for simulations and numerical analysis

- periodic boundary conditions (in 1 D)
  - Assume the system size is `L`.
  - Set site index as `i=0,1,2,`$\cdots$`, L-1`.
  - In general `(i+r)%L` and `(i+L-r)%L` can be used ($\texttt{r} > 0$).
  - In case `L=(1<<n)` (shift operator), define `Lm1=L-1` and use `(i+r)&Lm1` (`bitwise AND`) for any `r`, which is possible because $-k \equiv 2^{32} - k$ for an `int` variable.
- long time power-law behavior $\rho(t) = At^{a}\left(1 + Bt^{-\chi} + o(t^{-\chi})\right)$
  - A fitting gives a good guide for a, but can be error-prone because of the (unknown) corrections-to-scaling $Bt^{-\chi}$.
  - For a systematic analysis, use an effective exponent (or successive slope).

- effective exponent

$$a_{\mathrm{eff}}(t;b) := \frac{\ln \rho(t) - \ln \rho(t/b)}{\ln b} \simeq a - \frac{b^{\chi} - 1}{\ln b} B t^{-\chi} + o(t^{-\chi}),$$

  where b is a fixed constant.

- my tip for measurement time (integer).
  - Up to 50 measurements, set $T_i =$`1,2,`$\cdots$`, 50`.
  - For another 30 measurements, set $T_{k+50} =$`(int)`$\left(50 \times 10^{0.01k}\right)$
  - Up to maximum measurement time, set $T_i = 2T_{i-30}$.
  - with this choice, use $b = 2, 4, 8, 16, 32, \ldots$.
- my tip for measurement time (real).
  - Set $T_i = 10^{0.01\times(i-1)}$
  - One can use $b = 10^{0.01n}$ for any integer n.
- Now plot a_{eff} as a function of $t^{-\chi}$. If χ is correct, a_{eff} approaches the y-axis with a finite slope.

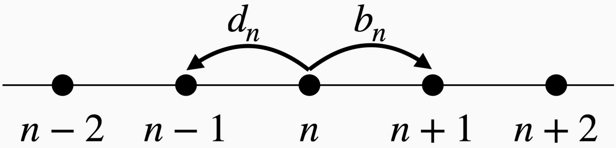
- How to find χ?
  - trial-and-error until a_{eff} looks straight (but too subjective).
  - Analyse corrections-to-scaling function (Park 2013, 2014).

$$Q(t;b) = \log \frac{\rho(t)\rho(t/b^2)}{\rho(t/b)^2} \simeq (b^\chi - 1)^2 B t^{-\chi} + o(t^{-\chi})$$

  - Find Q for various b ($2,4,8,16,32,\cdots$)
  - If χ is correct, the asymptotic behavior of $Q(t;b)/(b^\chi - 1)^2$ should be the same for any b. (at this point, we cannot help using trial-and-error).
  - Since it is a second derivative, real data for Q can be very noisy unless the number of independent runs is large enough.

Birth-and-death (Jump, One-step) processes

- transition rates (integer space)



- master equation

$$\frac{\partial}{\partial t}P_n(t) = d_{n+1}P_{n+1}(t) + b_{n-1}P_{n-1}(t) - (d_n + b_n)P_n(t).$$

- state space
  - infinite : $\mathfrak{X} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
  - half-infinite : $\mathfrak{X} = \{0, 1, 2, \ldots\}$ $(b_{-1} = d_0 = 0)$.
  - finite: $\mathfrak{X} = \{0, 1, 2, \ldots, N\}$ $(b_{-1} = d_0 = b_N = d_{N+1} = 0)$.

- equation for the generating function $\mathscr{G}(z,t) \equiv \sum_n z^n P_n(t)$

$$\begin{aligned}\frac{\partial \mathscr{G}}{\partial t} &= \sum_n \left[z^n d_{n+1} P_{n+1} - z^n d_n P_n + z^n b_{n-1} P_{n-1} - z^n b_n P_n\right] \\ &= \sum_n \left[\left(z^{n-1} - z^n\right) d_n P_n + \left(z^{n+1} - z^n\right) b_n P_n\right].\end{aligned}$$

- mean $\langle n \rangle$

$$\frac{d\langle n\rangle}{dt} = \frac{d}{dt}\left(\left.\frac{\partial \mathscr{G}}{\partial z}\right|_{z=1}\right) = \sum_n (b_n - d_n) P_n = \langle b_n \rangle - \langle d_n \rangle.$$

- Poisson process $b_n = \lambda$, $P(n,0) = \delta_{n0}$.

$$\dot{P}_n = -\lambda P_n + \lambda P_{n-1}, \quad \dot{P}_0 = -\lambda P_0 \Rightarrow P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

- linear birth-and-death process ($d_n = \mu n$, $b_n = \lambda(n+b)$)

$$\frac{\partial \mathscr{G}}{\partial t} = \sum_{n=0}^{\infty} \left[\left(z^{n-1} - z^{n}\right) \underbrace{\mu\, n}_{\text{death}} P_n + \left(z^{n+1} - z^{n}\right) \underbrace{\lambda(n+b)}_{\text{birth}} P_n \right]$$

$$= (1-z)(\mu - \lambda z)\frac{\partial \mathscr{G}}{\partial z} + \lambda b(z-1)\mathscr{G},$$

with $\mathscr{G}(z,0) = \sum_n z^n P_n(0) = z^m$ $(P_n(0) = \delta_{nm})$.

- how to solve: the method of characteristics

The method of characteristics

To solve partial differential equations

$$\frac{\partial \psi}{\partial t} + c(t,x)\frac{\partial \psi}{\partial x} = f(t,x,\psi).$$

Assume x is a function of t (with unknown constant of integration C_0) satisfying

$$\frac{dx}{dt} = c(t,x),$$

whose solution is called a characteristic. Now we have an ODE

$$\frac{d\psi}{dt} = \frac{\partial \psi}{\partial t} + \frac{dx}{dt}\frac{\partial \psi}{\partial x} = f(t,x(t),\psi).$$

$\psi(x(t),t) = \psi(C_0,\psi_0,t)$ with another constant of integration ψ_0.

Determine $\psi_0 := \psi(s(0),0) = \psi_0(C_0)$ and put $C_0 = C_0(x,t)$.

Solution of the birth-and-death process

$$\text{Equation}: \frac{\partial \mathcal{G}}{\partial t} + (z-1)(\mu - \lambda z)\frac{\partial \mathcal{G}}{\partial z} = \lambda b (z-1)\mathcal{G}.$$

- method of characteristics

  Assume z is a function of t $[z = z(t)]$ with $z_0 = z(t=0)$.

  Choose a characteristic curve such that

  $$\frac{dz}{dt} = (z-1)(\mu - \lambda z) \to \frac{1-z}{\mu - \lambda z} e^{(\lambda-\mu)t} = C_0 \text{ (constant)},$$

  $$\frac{d\mathcal{G}}{dt} = \lambda b (z(t)-1)\mathcal{G} \to \ln\left(\frac{\mathcal{G}}{\mathcal{G}_0}\right) = \lambda b \int_0^t (z-1)dt' = \int_{z_0}^{z(t)} \frac{\lambda b dz'}{\mu - \lambda z}$$

  $$= -b \ln\left|\frac{\mu - \lambda z}{\mu - \lambda z_0}\right| \Rightarrow \mathcal{G} = \mathcal{G}_0 \left(\frac{\mu - \lambda z}{\mu - \lambda z_0}\right)^{-b}.$$

Since $\mathscr{G}_0 = \mathscr{G}(z(0), 0) = z_0^m$,

$$C_0 = \frac{1-z}{\mu - \lambda z} e^{(\lambda-\mu)t} = \frac{1-z_0}{\mu - \lambda z_0} \to z_0 = \frac{\mu(1-\varepsilon) + z(\mu\varepsilon - \lambda)}{\mu - \lambda\varepsilon - \lambda(1-\varepsilon)z},$$

$$\mathscr{G}(z,t) = \left(\frac{\mu(1-\varepsilon) + z(\mu\varepsilon - \lambda)}{\mu - \lambda\varepsilon - \lambda(1-\varepsilon)z} \right)^m \left(\frac{\mu - \lambda\varepsilon - \lambda(1-\varepsilon)z}{\mu - \lambda} \right)^{-b},$$

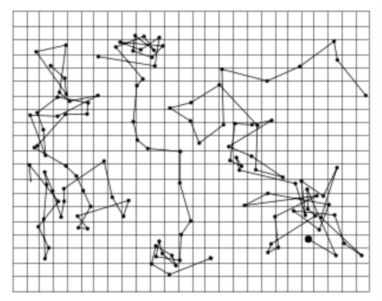
where $\varepsilon \equiv e^{(\lambda-\mu)t}$.

- $b = 0$: (continuous time) branching process (assignment 8)
- $\lambda = 0$: pure death process.
- $\mu = 0$: pure birth process

Fokker-Planck Equation

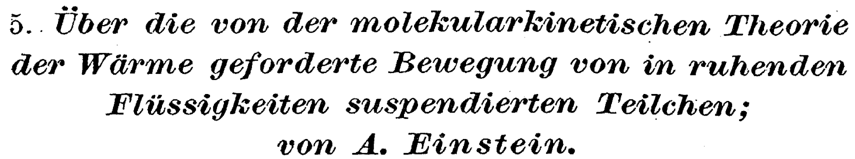
Brownian motion (R. Brown, 1827)

- pollen grains (꽃가루) in water : manifest of life?



- But, any fine particles exhibit such a motion.
- For a nice introduction to the history of Brownian motion, E. Nelson, *Dynamical Theories of Brownian Motion* (1967). http://www.math.princeton.edu/~nelson/books.html

Einstein’s contribution (1905)



5. *Über die von der molekularkinetischen Theorie*
der Wärme geforderte Bewegung von in ruhenden
Flüssigkeiten suspendierten Teilchen;
von A. Einstein.

열의 분자운동이론이 필요한, 고요한 액체 속에 있는 작은 입자의 운동에 대하여

A. Einstein, Annalen der Physik **17**, 549 (1905)

beginning of stochastic modelling of natural phenomena

Einstein’s prediction and experimental confirmation

- Einstein’s prediction

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}, \quad \langle x(t)^2 \rangle = 2Dt, \quad D = \frac{k_B T}{6\pi\eta a}$$

  a : radius of the suspended particle, η : viscosity,
  T: temperature.

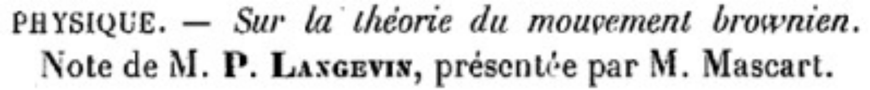
- Smoluchowski’s independent work (1906).
- Jean Baptiste Perrin’s experiment (Avogadro number)

The Nobel Prize in Physics 1926
Jean Baptiste Perrin

  http://nobelprize.org/nobel_prizes/physics/laureates/1926

- triumph of the atomic theory!

Langevin’s contribution (1908)



PHYSIQUE. — *Sur la théorie du mouvement brownien.*
Note de M. P. **Langevin**, présentée par M. Mascart.

브라운 운동 이론에 대하여

P. Langevin, C. R. Acad. Sci. (Paris) **146**, 530 (1908).

English translation: D. S. Lemons and A. Gythiel, Am. J. Phys. **65**, 1079 (1997).

“infinitely more simple”

foundation of the *stochastic differential equation*

- viscous friction and random force (fluctuation)

$$m\frac{d^2x}{dt^2} = -6\pi\eta a\frac{dx}{dt} + X.$$

- multiply x on both sides of the equation

$$\frac{m}{2}\frac{d^2}{dt^2}x^2 - mv^2 = -3\pi\eta a\frac{d}{dt}x^2 + xX.$$

- average and equipartition theorem

$$\frac{m}{2}\frac{d^2}{dt^2}\langle x^2\rangle + 3\pi\eta a\frac{d}{dt}\langle x^2\rangle = \underbrace{\langle mv^2\rangle}_{=k_BT} + \underbrace{\langle xX\rangle}_{=0}$$

$$\frac{d}{dt}\langle x^2\rangle = \frac{k_BT}{3\pi\eta a} + C\exp\left(-\frac{6\pi\eta a}{m}t\right) \xrightarrow{t\to\infty} \frac{k_BT}{3\pi\eta a}.$$

$$\langle x^2\rangle = 2Dt = \frac{k_BT}{3\pi\eta a}t.$$

Fokker-Planck Equation

differential CK equation with $W(\boldsymbol{x}|\boldsymbol{y},t) = 0$

Fokker-Planck equation (FPE)

$$\begin{aligned}\frac{\partial P(\boldsymbol{x},t|\boldsymbol{y},t_0)}{\partial t} =& -\sum_i \frac{\partial}{\partial x_i}\left[A_i(\boldsymbol{x},t)P(\boldsymbol{x},t|\boldsymbol{y},t_0)\right] \\ &+ \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\left[B_{ij}(\boldsymbol{x},t)P(\boldsymbol{x},t|\boldsymbol{y},t_0)\right]\end{aligned}$$

- sample paths are almost surely continuous.
- $\boldsymbol{A}(\boldsymbol{x})$: drift vector
- $\boldsymbol{B}(\boldsymbol{x})$: diffusion matrix
- Initial condition : $\lim_{t\to t_0^+} P(\boldsymbol{x},t|\boldsymbol{y},t_0) = \delta(\boldsymbol{x}-\boldsymbol{y})$.

short time behavior

- If Δt is small, (cf: Langevin Equation)

$$P(\boldsymbol{x}, t + \Delta t | \boldsymbol{y}, t) \approx \left\{ (2\pi)^N \det[\boldsymbol{B}(\boldsymbol{y}, t)\Delta t] \right\}^{-1/2} \times$$

$$\times \exp \left\{ -\frac{1}{2} \frac{[\boldsymbol{x} - \boldsymbol{y} - \boldsymbol{A}(\boldsymbol{y}, t)\Delta t]^\mathsf{T} [\boldsymbol{B}(\boldsymbol{y}, t)]^{-1} [\boldsymbol{x} - \boldsymbol{y} - \boldsymbol{A}(\boldsymbol{y}, t)\Delta t]}{\Delta t} \right\}$$

$$\boldsymbol{x}(t + \Delta t) = \boldsymbol{x}(t) + \boldsymbol{A}(\boldsymbol{x}(t), t)\Delta t + \boldsymbol{\eta}(t)\Delta t^{1/2},$$

  where $\langle \boldsymbol{\eta}(t) \rangle = 0, \quad \langle \boldsymbol{\eta}(t)\boldsymbol{\eta}(t)^\mathsf{T} \rangle = \boldsymbol{B}(\boldsymbol{y}, t).$

- Sample paths are continuous with probability one.
- Sample paths are nowhere differentiable because of $\Delta t^{1/2}$.

The Wiener Process

FPE for the Wiener Process

$$\frac{\partial}{\partial t} P(w,t|w_0,t_0) = \frac{1}{2}\frac{\partial^2}{\partial w^2} P(w,t|w_0,t_0),$$
$$\lim_{t\to t_0^+} P(w,t|w_0,t_0) = \delta(w-w_0)$$

- generating function

$$\phi(s,t) \equiv \int dw P(w,t|w_0,t_0)\exp(isw), \quad \phi(s,t_0) = \exp(isw_0).$$
$$\frac{\partial\phi}{\partial t} = \int dw \frac{\partial}{\partial t} P(w,t|w_0,t_0)\exp(isw)$$
$$= \int dw \frac{1}{2}\frac{\partial^2}{\partial w^2} P(w,t|w_0,t_0)\exp(isw) = -\frac{1}{2}s^2\phi.$$
$$\phi(s,t) = \exp\left(-\frac{1}{2}s^2(t-t_0) + isw_0\right).$$

- Fourier transform

$$P(w,t|w_0,t_0) = \frac{1}{2\pi}\int ds\ \phi(s,t)\exp(-isw)$$
$$= \frac{1}{\sqrt{2\pi(t-t_0)}}\exp\left(-\frac{1}{2}\frac{(w-w_0)^2}{t-t_0}\right).$$

- mean and variance

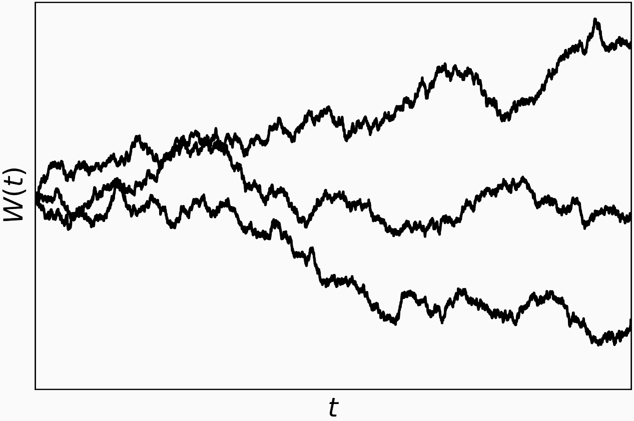
$$\langle W(t)\rangle = w_0 \text{ (martingale) },\quad \langle (W(t)-w_0)^2\rangle = t-t_0.$$

- cf. cumulant generating funcion

$$\ln\phi(s,t) = isw_0 - \frac{1}{2}s^2(t-t_0) = is\langle W(t)\rangle + \frac{1}{2}(is)^2\langle (W(t)-w_0)^2\rangle$$

Properties of the Wiener process

- Irregularity of Sample Paths
  $\langle W(t) \rangle$ remains constant, but the variance diverges : Sample paths are *variable and irregular*.



- continuous everywhere but nowhere differentiable

$$\mathbb{P}\left\{\left|\frac{W(t+h)-W(t)}{h}\right| > k\right\} = 2\int_{kh}^{\infty} dw \frac{1}{\sqrt{2\pi h}}\exp\left(-\frac{w^2}{2h}\right)$$

$$= 2\int_{k\sqrt{h}}^{\infty} dx \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \xrightarrow{h\to 0} 1 \text{ for any } k > 0.$$

Thus, $\frac{dW(t)}{dt}$ does not exist, as anticipated $(dW \sim \sqrt{dt})$.

- cf. Weierstrass function.
- 참고
  - cardinality of continuous functions = $\mathfrak{c}$.
  - cardinality of everywhere differentiable functions = $\mathfrak{c}$.
  - cardinality of continuous nowhere differentiable functions = $\mathfrak{c}$.

- independence of increments

  joint probability (due to the Markov property)

$$P(w_n, t_n; w_{n-1}, t_{n-1}; \cdots ; w_0, t_0) = \prod_{i=0}^{n-1} P(w_{i+1}, t_{i+1} | w_i, t_i) P(w_0, t_0)$$

  Let $\Delta W_i = W(t_i) - W(t_{i-1})$ (new r.v.), $\Delta t_i = t_i - t_{i-1}$,

$$\begin{aligned} &P(\Delta w_n; \Delta w_{n-1}; \cdots ; \Delta w_1; w_0) \\ &\qquad = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \Delta t_i}} \exp\left(-\frac{\Delta w_i^2}{2\Delta t_i} \right) \right\} P(w_0, t_0). \end{aligned}$$

  r.v.'s ΔW_i are independent of each other and of $W(t_0)$.

- autocorrelation function (cf. covariance)

$$\langle W(t)W(s)|[w_0,t_0]\rangle = \int dw_1\,dw_2\,w_1 w_2 p(w_1,t;w_2,s|w_0,t_0)$$

Assuming $t > s$ and using the independence of increment,

$$\begin{aligned}\langle W(t)W(s)|[w_0,t_0]\rangle &= \langle [W(t)-W(s)]\,W(s)\rangle + \langle W(s)^2\rangle \\ &= s - t_0 + w_0^2\end{aligned}$$

In general,

$$\begin{aligned}\langle W(t)W(s)|[w_0,t_0]\rangle &= \min(t-t_0, s-t_0) + w_0^2, \\ \langle W(t),W(s)|[w_0,t_0]\rangle &= \min(t-t_0, s-t_0).\end{aligned}$$

- If $W(t)$ is a Wiener process, so are $(a > 0, s > 0)$

$$-W(t), \quad W(t+s) - W(s), \quad aW(t/a^2).$$

- For any $\varepsilon > 0$, a sample path of the Wiener process meets zero infinitely many times in the open interval $(0, \varepsilon)$ with probability one.

Donsker's theorem or functional CLT

Let X_i's are i.i.d. RV's with zero mean and unit variance. Let $S_n = X_1 + \cdots + X_n$ (random walk). Let $W^{(n)}(t) = \dfrac{S_{\lfloor nt \rfloor}}{\sqrt{n}}$ $(t \in [0,1])$. Then, $W(t) = \lim_{n\to\infty} W^{(n)}(t)$ is the Wiener process (in distribution).

The Ornstein-Ulhenbeck Process

FPE for the Ornstein-Ulhenbeck Process

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0)=\frac{\partial}{\partial x}\left(kxP(x,t|x_0,t_0)\right)+\frac{1}{2}D\frac{\partial^2}{\partial x^2}P(x,t|x_0,t_0)$$

- generating function

$$\phi(s,t)\equiv\int dx\,P(x,t|x_0,t_0)\exp(isx),\quad \phi(s,t_0)=\exp(isx_0)$$

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

- the method of characteristics

$$\frac{ds}{dt} = ks \to se^{-kt} = A(\text{constant of integration}),$$
$$\frac{d\phi}{dt} = -\frac{1}{2}Ds(t)^2\phi(t) = -\frac{1}{2}DA^2e^{2kt}\phi(t)$$
$$\to \phi = \phi_0 \exp\left(-\frac{DA^2}{4k}\left(e^{2kt}-1\right)\right) = \phi_0 \exp\left(-\frac{Ds^2}{4k}\left(1-e^{-2kt}\right)\right).$$

When $t = 0, \phi_0 = e^{is(t=0)x_0} = e^{iAx_0}$. Hence,

$$\phi(s,t) = \exp\left(ise^{-kt}x_0 - \frac{D}{4k}s^2\left(1-e^{-2kt}\right)\right).$$

- mean and variance

$$\langle X(t)\rangle = x_0e^{-kt}, \ \ \text{var}(X(t)) = \frac{D}{2k}\left(1-e^{-2kt}\right)$$

- stationary solution

$$\phi(s,\infty) = \exp\left(-\frac{Ds^2}{4k}\right),\ P_s(x) = \sqrt{\frac{k}{\pi D}}\exp\left(-\frac{kx^2}{D}\right)$$

Note that $P_s(x)$ is the solution of the stationary FPE

$$\partial_x\left[kxP + \frac{1}{2}D\partial_x P\right] = 0$$

- If $P(x_0) = P_s(x_0)$, then

$$\phi_0 = \exp\left(-\frac{A^2}{2}\left\langle x_0^2\right\rangle\right) = \exp\left(-\frac{Ds^2}{4k}e^{-2kt}\right),$$

which gives $\phi(s,t) = \phi(s,\infty)$ for all t.

- time correlation function at stationarity

  $$\langle X(t)X(s)|[x_0,t_0]\rangle = \int dx_1 dx_2 x_1 x_2 P(x_1,t|x_2,s)P(x_2,s|x_0,t_0),$$

  where $t \geq s \geq t_0$ is assumed. Take $t_0 \to -\infty$, we get

  $$\langle X(t),X(s)\rangle_s = \langle X(t)X(s)\rangle_s = \frac{D}{2k}\exp\left(-k|t-s|\right)$$

  The Ornstein-Uhlenbeck process in its stationary state models a realistic noise signal with correlation time $1/k = \tau$.

- how to generate a sample path $X(t)$? Langevin equation

Langevin Equation

White noise in the Langevin equation

- Langevin equation

  $$\frac{dx}{dt} = a(x,t) + b(x,t)\xi(t),$$

  where $\xi(t)$ is the *rapidly fluctuating random term* called the *white noise*.
- $\langle \xi(t) \rangle = 0$, $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$ (no correlation at different times)
- But, what is $\xi(t)$, mathematically?

Assume $\xi(t)$ is well-defined. Let $u(t) = \int_0^t \xi(t')dt'$ (continuous stochastic process).

- $u(t)$ is a Markov process.

$$u(t') = \underbrace{\lim_{\varepsilon \to 0} \left[\int_0^{t-\varepsilon} ds \xi(s) \right]}_{U_1 = u(t)} + \underbrace{\int_t^{t'} \xi(s)}_{U_2}$$

U_1 is independent of U_2.
Thus, $u(t)$ and $u(t') - u(t)$ are statistically independent.
Furthermore, $u(t') - u(t)$ is independent of $u(t'')$ for $t'' < t$.

- FPE for $u(t)$.

$$\begin{aligned}\langle u(t+\Delta t)-u_0|[u_0,t]\rangle &= \left\langle \int_t^{t+\Delta t} \xi(s)ds \right\rangle = 0,\\ \langle (u(t+\Delta t)-u_0)^2|[u_0,t]\rangle &= \int_t^{t+\Delta t}\int_t^{t+\Delta t} dsds' \left\langle \xi(s)\xi(s')\right\rangle\\ &= \int_t^{t+\Delta t}\int_t^{t+\Delta t} dsds'\delta(s-s') = \Delta t.\end{aligned}$$

Hence, $A(u_0,t)=0$, $B(u_0,t)=1:$ the Wiener Process.

- $\xi(t) = \dfrac{dW(t)}{dt}$: ill-defined!

Stochastic Integral

- Mathematically speaking, the Langevin equation does not exist.
- However, the *integral equation* can be interpreted consistently.

$$\begin{aligned} x(t) - x(0) &= \int_0^t a[x(s), s] ds + \int_0^t b[x(s), s] dW(s) \\ &\doteq \int_0^t a[x(s), s] ds + \int_0^t b[x(s), s] \xi(s) ds, \end{aligned}$$

which is a kind of a (stochastic) Stieltjes integral w.r.t. a sample path $W(t)$.

- (naive) definition

$$\int_{t_0}^{t} G(t')dW(t') \overset{?}{=} \lim_{n\to\infty} \underbrace{\left\{ \sum_{i=1}^{n} G(\tau_i)\left[W(t_i) - W(t_{i-1})\right] \right\}}_{\equiv S_n}$$

where $t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_{n-1} \leq t_n = t$ and $t_{i-1} \leq \tau_i \leq t_i$.

- But S_n depends on the choice of τ's. Take $G(t) = W(t)$,

$$\begin{aligned} \langle S_n \rangle &= \sum_i \left\langle W(\tau_i)\left[W(t_i) - W(t_{i-1})\right] \right\rangle \\ &= \sum_{i=1}^{n} \left[\min(\tau_i, t_i) - \min(\tau_i, t_{i-1})\right] = \sum_{i=1}^{n} (\tau_i - t_{i-1}) \end{aligned}$$

Choose $\tau_i = \alpha t_i + (1-\alpha)t_{i-1}$ $(0 \leq \alpha \leq 1)$, then $\langle S_n \rangle = \alpha(t - t_0)$.

Itô stochastic integral

the *Itô stochastic integral* **(**$\alpha = 0$ **or** $\tau_i = t_{i-1}$**)**

$$\int_{t_0}^{t} G(t')dW(t') \equiv \underset{n\to\infty}{\text{ms-lim}} \left\{ \sum_{i=1}^{n} G\left(t_{i-1}\right)\left[W\left(t_i\right) - W\left(t_{i-1}\right)\right] \right\}$$

$G(t)$ is assumed not to be affected by the “future” : causality.

Such a $G(t)$ is called a *nonanticipating function*.

- examples of nonanticipating functions
  1. $W(t)$
  2. $\int^t F[W(t')]dt'$
  3. $\int^t F[W(t')]dW(t')$
  4. $\int^t G(t')dt'$ ($G(t)$ itself is a nonanticipating function)
  5. $\int^t G(t')dW(t')$

Example $\int W(t)dW(t)$

$$\begin{aligned} S_n &= \sum_{i=1}^{n} W_{i-1}\left(W_i - W_{i-1}\right) = \frac{1}{2}\sum_i \left[W_i^2 - W_{i-1}^2 - \Delta W_i^2\right] \\ &= \frac{1}{2}\left[W(t)^2 - W(t_0)^2\right] - \frac{1}{2}\underbrace{\sum_i \Delta W_i^2}_{\equiv U}. \end{aligned}$$

- $\langle U\rangle = \sum_i \langle \Delta W_i^2\rangle = \sum_i (t_i - t_{i-1}) = t - t_0$.
- $\langle (U - (t-t_0))^2\rangle = 2\sum_i (t_i - t_{i-1})^2 \to 0$ as $n \to \infty$ (check it!)

Hence (mean square limit: assignment 9),

$$\int_{t_0}^{t} W(t')dW(t') = \underset{n\to\infty}{\text{ms-lim}}\, S_n = \frac{1}{2}\left[W(t)^2 - W(t_0)^2\right] - \frac{1}{2}(t - t_0).$$

- $\left\langle \int_{t_0}^{t} W(t')dW(t') \right\rangle = \left\langle \frac{1}{2}\left[W(t)^2 - W(t_0)^2\right] - \frac{1}{2}(t - t_0) \right\rangle = 0.$
- ΔW_i^2 is not negligible $[\sim O(dt)]$.
- the Stratonovich integral

$$
\begin{aligned}
(\mathrm{S}) \int_{t_0}^{t} W(t')dW(t') &= \operatorname*{ms-lim}_{n\to\infty} \sum_{i=1}^{n} \frac{W_i + W_{i-1}}{2}(W_i - W_{i-1}) \\
&= \frac{1}{2}\left[W(t)^2 - W(t_0)^2\right],
\end{aligned}
$$

    - similar to the ordinary calculus.
    - midpoint rule with $\alpha = \frac{1}{2}$ is equivalent.

- $dW(t)^2 = dt$ and $dW(t)^{2+N} = dW(t)dt = 0$ in the sense that

$$\int_{t_0}^{t} \left[dW(t')\right]^{2+N} G(t') \equiv \operatorname*{ms-lim}_{n\to\infty} \sum_i G_{i-1}\Delta W_i^{2+N}$$

$$= \begin{cases} \displaystyle\int_{t_0}^{t} G(t')dt' & \text{for } N = 0, \\ 0 & \text{for } N > 0. \end{cases}$$

- existence: $\int G(t')dW(t')$ exists whenever $G(t)$ is *continuous and nonanticipating*

- integration of Polynomials:

  Since

$$\begin{aligned} d[W(t)]^n &= [W(t) + dW(t)]^n - W(t)^n \\ &= \sum_{r=1}^{n} \binom{n}{r} W(t)^{n-r} dW(t)^r \leftarrow dW(t)^2 = dt \\ &= nW(t)^{n-1} dW(t) + \frac{n(n-1)}{2} W^{n-2} dt, \end{aligned}$$

  we get

$$\begin{aligned} \int_{t_0}^{t} W(t')^n dW(t') =& \frac{1}{n+1} \left[W(t)^{n+1} - W(t_0)^{n+1} \right] \\ & - \frac{n}{2} \int_{t_0}^{t} W(t')^{n-1} dt'. \end{aligned}$$

- general differentiation rule (keep terms up to dW^2)

$$\begin{aligned} df[W(t),t] &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial W}dW(t) + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}dW(t)^2 \\ &= \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right)dt + \frac{\partial f}{\partial W}dW(t) \end{aligned}$$

- mean value formula

$$\left\langle \int G(t')dW(t') \right\rangle = 0.$$

- correlation formula

$$\left\langle \int_{t_0}^{t} G(t')dW(t') \int_{t_0}^{t} H(t')dW(t') \right\rangle = \int_{t_0}^{t} \langle G(t')H(t') \rangle \, dt'.$$

Stochastic differential equation (SDE)

- Itô SDE : definition

$$dx(t) = a\left[x(t), t\right] dt + b\left[x(t), t\right] dW(t),$$

if for all t and t_0,

$$x(t) = x(t_0) + \int_{t_0}^{t} a\left[x(t'), t'\right] dt' + \int_{t_0}^{t} b\left[x(t'), t'\right] dW(t').$$

- $x(t)$ is a Markov process.
- *additive noise* if $b[x(t), t]$ does not depend on x.
- *multiplicative noise* if $b[x(t), t]$ does depend on x.

Itô’s Formula

$$\begin{aligned} df[x(t)] &= f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)]dx(t)^2 \\ &= f'[x(t)]\left\{a[x(t),t]dt + b[x(t),t]dW(t)\right\} + \frac{1}{2}f''[x(t)]b[x(t),t]^2 dW(t)^2 \\ &= \left\{a[x(t),t]f'[x(t)] + \frac{1}{2}f''[x(t)]b[x(t),t]^2\right\}dt + f'[x(t)]b[x(t),t]dW(t) \end{aligned}$$

where we have used,

$$dx(t) = a\left[x(t),t\right]dt + b\left[x(t),t\right]dW(t),$$

$dW(t)^2 = dt$, and $dW(t)^{2+N} = dW(t)dt = 0$.

$$\frac{d\langle f[x(t)]\rangle}{dt} = \left\langle a[x(t),t]f'[x(t)] + \frac{1}{2}f''[x(t)]b[x(t),t]^2 \right\rangle$$

Connection between the FPE and the SDE

$$\begin{aligned}\frac{d}{dt}\langle f[x(t)]\rangle &= \int dx\, f(x)\, \frac{\partial}{\partial t} P(x,t|x_0,t_0) \\ &= \int dx \left\{ a(x,t) f'(x) + \frac{1}{2} b(x,t)^2 f''(x) \right\} P(x,t|x_0,t_0) \\ &= \int dx\, f(x) \left\{ -\frac{\partial}{\partial x}(a(x,t)P) + \frac{1}{2}\frac{\partial^2}{\partial x}(b(x,t)^2 P) \right\}.\end{aligned}$$

Since f is arbitrary,

FPE-SDE connection (assignment 10)

$$dx = a dt + b dW \Leftrightarrow \frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}(aP) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b^2 P).$$

Stratonovich SDE

- Stratonovich integral : definition

$$
\begin{aligned}
&(\mathrm{S}) \int_{t_0}^{t} G\left[x(t'), t'\right] dW(t') \\
&= \operatorname*{ms-lim}_{n \to \infty} \sum_{i=1}^{n} G\left[\frac{1}{2}\left(x(t_i) + x(t_{i-1})\right), t_{i-1}\right]\left(W(t_i) - W(t_{i-1})\right).
\end{aligned}
$$

- Stratonovich SDE

$$(\mathrm{S}) dx(t) = a\left[x(t), t\right] dt + b\left[x(t), t\right] dW(t),$$

if for all t and t_0,

$$x(t) = x(t_0) + \int_{t_0}^{t} a\left[x(t'), t'\right] dt' + (\mathrm{S}) \int_{t_0}^{t} b\left[x(t'), t'\right] dW(t').$$

- change of variables (same as the ordinary calculus rule)

$$(\mathrm{S})df[x(t)] = f'[x(t)]\left\{a\left[x(t),t\right]dt + b\left[x(t),t\right]dW(t)\right\}.$$

- From Stratonovich to Itô

$$[(\mathrm{S})]dx = x_i - x_{i-1} = \beta\left(\frac{x_i + x_{i-1}}{2}, t_{i-1}\right)\left(W(t_i) - W(t_{i-1})\right)$$

$$= \left[\beta\left(x_{i-1}\right) + \frac{dx}{2}\beta'(x_{i-1})\right]\Delta W_i = \beta(x_{i-1})dW_i + \frac{1}{2}\beta(x_{i-1})\beta'(x_{i-1})dt.$$

Stratonovich vs. Itô

$$
\begin{aligned}
&[\mathrm{S}]dx = \alpha(x,t)dt + \beta(x,t)dW \\
&\Leftrightarrow dx = \left(\alpha(x,t) + \frac{1}{2}\beta(x,t)\frac{\partial\beta}{\partial x}\right)dt + \beta(x,t)dW
\end{aligned}
$$

$$
\langle\beta(x,t)dW\rangle = \begin{cases} 0, & \text{Itô}, \\ \\ \dfrac{1}{2}\beta(x,t)\dfrac{\partial\beta}{\partial x}dt, & \text{Stratonovich}. \end{cases}
$$

Example I : Geometric Brownian Motion

- $dx = cx dW(t)$.
- change of variable : $y = \ln x$

$$\begin{aligned} dy &= \frac{dx}{x} - \frac{1}{2x^2}(dx)^2 = c dW(t) - \frac{1}{2}c^2 dt \\ &\Rightarrow y(t) = y(t_0) + c\left[W(t) - W(t_0)\right] - \frac{1}{2}c^2(t - t_0) \\ &\Rightarrow x(t) = x(t_0)\exp\left\{c\left[W(t) - W(t_0)\right] - \frac{1}{2}c^2(t - t_0)\right\} \end{aligned}$$

- mean and autocorrelation function (assignment 11)

$$\begin{aligned} \langle x(t) \rangle &= \langle x(t_0) \rangle, \\ \langle x(t), x(s) \rangle &= \langle x(t_0)^2 \rangle \exp\left\{c^2 \text{min}(t - t_0, s - t_0)\right\} \end{aligned}$$

Exmaple 2: Ornstein-Uhlenbeck Process

- $dx = -kxdt + \sqrt{D}dW$
- change of variable : $y = xe^{kt}$

$$\begin{aligned} dy &= (dx)e^{kt} + xd(e^{kt}) = \left[-kxdt + \sqrt{D}dW\right] e^{kt} + kxe^{kt}dt \\ &= \sqrt{D}e^{kt}dW \end{aligned}$$

$$x(t) = x(0)e^{-kt} + \sqrt{D}\int_0^t e^{-k(t-t')}dW(t').$$

- mean and autocorrelation function

$$\begin{aligned} \langle x(t)\rangle &= \langle x(0)\rangle e^{-kt}, \\ \langle x(t)x(s)\rangle &= \langle x(0)^2\rangle e^{-k(s+t)} + D\int_0^{\min(t,s)} e^{-k(t+s-2t')}dt' \\ \langle x(t), x(s)\rangle &= \left[\operatorname{var}\{x(0)\} - \frac{D}{2k}\right] e^{-k(t+s)} + \frac{D}{2k}e^{-k|t-s|}. \end{aligned}$$

The white noise limit

- In real physical systems, noise should be correlated.
- We are interested in a limit of a differential equation

$$\frac{dx}{dt} = a(x) + b(x)\xi_\gamma(t),$$

where $\xi_\gamma(t)$ is a stochastic source with nonzero correlation time.

- If $\langle \xi_\gamma(t) \rangle = 0$ and

$$\lim_{\gamma \to \infty} \langle \xi_\gamma(t)\xi_\gamma(t') \rangle_s = \delta(t - t'),$$

the above differential equation becomes

The white noise limit (assignment 12)

$$(\mathrm{S}) dx = a(x)dt + b(x)dW(t)$$

Positive Poisson Representation

Theorem
For any P_n, a positive $f(\alpha)$ exists such that

$$P_n = \int d^2\alpha \left(e^{-\alpha} \frac{\alpha^n}{n!} \right) f(\alpha),$$

where $\alpha = \alpha_x + i\alpha_y$ and $d^2\alpha = d\alpha_x d\alpha_y$.

- $1 = \sum_{n=0}^{\infty} P_n = \int d^2\alpha \left(\sum_n e^{-\alpha} \frac{\alpha^n}{n!} \right) f(\alpha) = \int d^2\alpha \, f(\alpha).$
  Hence, $f(\alpha)$ is a probability density.
- If $f(\alpha) = \delta(\alpha - \rho)$ for real ρ, $P_n = e^{-\rho} \rho^n / n!$.
- $\langle n \rangle = \langle \alpha \rangle$ always.
- $\langle n^m \rangle_f \equiv \langle n(n-1) \cdots (n-m+1) \rangle = \langle \alpha^m \rangle$

SDE using Poisson representation

$$X \underset{k_3}{\overset{k_1}{\rightleftarrows}} 0 \ , \qquad X \underset{k_4}{\overset{k_2}{\rightleftarrows}} 2X$$

$$d\alpha = \left[k_3 + (k_2 - k_1)\alpha - k_4\alpha^2\right] dt + \left[2(k_2\alpha - k_4\alpha^2)\right]^{1/2} dW(t).$$

- If $k_4 = 0$, the SDE is exactly solvable (linear birth-death).
- If $k_2 = k_4 = 0$, the dynamics is deterministic.
- If $k_2 = 0$, α should be complex.
- cf. Path integral approach (Cardy, cond-mat/9607163)
- assignment 14