AM216 Stochastic Differential Equations

Lecture 01

List of topics in this lecture

- The framework of repeated experiments for probability
- Outcome, sample space, random variable, event, probability of an event, union, intersection, complement, mutually exclusive events
- Conditional probability, independent events, law of total probability
- Expected value of a random variable, probability mass function (PMF), probability density function (PDF), joint probability density of two random variables
- Conditional probability density, independent random variables, conditional expectation, law of total expectation

---

Review of probability theory

Question: What is probability?

Example 1: Flip a fair coin

Prob(head) = 50%

What is the exact meaning of this statement?

While most people won't be struggling with this statement, the next example is somewhat more puzzling when we think about it carefully.

Example 2:

I go to see my doctor. Before doing any test, she tells me

Prob(I have cancer) = 15%

What is the exact meaning of this statement?

I either have or not have cancer. The answer is deterministic, which is unknown at the moment and which theoretically will be known after a sequence of comprehensive tests. When we restrict the scope of consideration to one person (me), the cancer status is not uncertain. It can be fully determined. It is just unknown for the time being.

So the probability "Prob(I have cancer)" has to be interpreted in a proper framework …

<u>The framework of repeated experiments</u> (with uncertain outcomes):

Probability of an event

= fraction of repeated experiments with the event occurring

$$= \frac{\text{\# of repeats with the event occurring}}{\text{\# of repeats}}$$

<u>Example 1:</u> Flip a fair coin. Repeat this M times.

$$\text{Prob(head)} = \lim_{M\to\infty} \frac{\text{\# of heads}}{M \text{ repeats}}$$

<u>Example 2:</u> Select a subject randomly from a sub-population S. Repeat this M times.

$$\text{Prob(cancer)} = \lim_{M\to\infty} \frac{\text{\# of subjects having cancer}}{M \text{ repeats}}$$

<u>Question:</u> What is the sub-population S?

There are many possibilities.

S = { all men }

S = { all persons over 50 years old }

S = { all Asian Americans }

S = { all foreign-born }

S = { all university professors }

S = { all persons with BMI in the normal range (18.5 ≤ BMI ≤ 24.9) }

Different doctors may view a given patient as a member of <u>different sub-populations</u>. As a result, different doctors may have different interpretation of Prob(cancer) for the same patient. <u>This is why probability can be subjective</u>. The subjectiveness is in specifying how to repeat the experiment.

So the probability "Prob(I have cancer)" makes sense only when the subject (me) is viewed as a member of a subpopulation. The selection of a different subpopulation leads to a different value of the probability, which is mathematically correct. The full statement should be something like the one below:

"When viewed as a random member of the subpopulation of all Asian Americans, the probability that I have cancer is … "

<u>Observation:</u>

Without specifying how the experiment is repeated, the concept of probability does not make sense.

With the concept of probability established in the framework of repeated experiments, we introduce terminology associated with probability.

Outcome of an experiment:

= Full description of the relevant result

Example: Flip a coin n times and view the sequence of n flips as ONE experiment.

Outcome: $\omega = x_1 x_2 \cdots x_n$

$x_i = 1$ (H, head) or 0 (T, tail)

This form of description is adequate for answering most questions.

However, if we want to study the possible connection between the height of toss and the landing result, this form of outcome is inadequate. For that purpose, we have to include the heights of n tosses in the outcome.

Outcome: $\omega = v_1 v_2 \cdots v_n$

$$v_i = \begin{pmatrix} x_i \\ h_i \end{pmatrix}, \qquad x_i = \text{side facing up}, \quad h_i = \text{height}$$

Sample space of an experiment

Ω = { all possible outcomes }

Example: Flip a coin 3 times and view the sequence of 3 flips as ONE experiment.

The sample space is

Ω = { TTT, TTH, THT, THH, HTT, HTH, HHT, HHH }

Random variable:

= A function of outcpme.

Full notation: $X(\omega)$

Short notation: X

Example: Flip a coin n times and view the sequence of n flips as ONE experiment.

Let N = # of heads in n flips. N is a random variable.

Full notation: $N(\omega) = N(x_1 x_2 \cdots x_n) = \Sigma_j x_j$

Event:

= A subset of sample space

Example: Flip a coin 3 times and view the sequence of 3 flips as ONE experiment

A = "exactly 2 heads in 3 flips"

= {THH, HTH, HHT}

Full notation: A = { ω | N(ω) = 2 } = { ω | ω contains exactly 2 heads }

In this example, event A is conveniently described by random variable $N(\omega)$.

Not all events are conveniently described by a random variable.

Example:

B = { ω | no consecutive heads in ω }

= {HTT, THT, TTH, HTH, TTT}

Event B is not conveniently described by a random variable.

Probability of an event:

$$\Pr(A) = \Pr(\text{outcome } \omega \in A) = \lim_{M\to\infty} \frac{\#\text{ of } \omega \in A}{M \text{ repeats}}$$

Remark: Mathematically, Pr(A) is the measure of set A in space Ω.

Example: Flip a fair coin 3 times and view the sequence of 3 flips as ONE experiment.

Pr(exactly 2 heads in 3 flips)

= 3/8 {THH, HTH, HHT}

Pr(no consecutive heads in 3 flips)

= 5/8 {HTT, THT, TTH, HTH, TTT}

Special cases:

Pr(Ω) = 1 (ω ∈ Ω is always true)

Pr(∅) = 0 (ω ∈ ∅ is never true where ∅ ≡ empty set)

Intersection of two events A and B: both A and B are true

AB = { ω | ω ∈ A and ω ∈ B } **Draw a Venn diagram to show it.**

Alternative notation for intersection: A ∩ B

Union of two events A and B: at least one of A and B is true

A+B = { ω | ω ∈ A or ω ∈ B or both } **Draw a Venn diagram to show it.**

Alternative notation for union: A ∪ B

Complement of event A: A is false

A^C = { ω | ω ∉ A } **Draw a Venn diagram to show it.**

For complement, we always have

$\Pr(A^C) = 1 - \Pr(A)$

Conditional probability: $\Pr(A \mid B)$

Repeat the experiment M times.

Consider only those repeats with $\omega \in B$.

$$\Pr(A|B) = \lim_{M\to\infty} \frac{\#\text{ of } (\omega \in A \text{ and } \omega \in B)}{\#\text{ of } \omega \in B}$$

$$= \lim_{M\to\infty} \frac{\dfrac{\#\text{ of } \left(\omega \in A \text{ and } \omega \in B\right)}{M}}{\dfrac{\#\text{ of } \omega \in B}{M}} = \frac{\Pr(AB)}{\Pr(B)}$$

Thus, we obtain

$$\boxed{\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}}$$

Example:

Pr(exactly 2 heads in 3 flips AND no consecutive heads)

= 1/8 { HTH } out of all 8 outcomes

This is the probability of an intersection.

Pr(exactly 2 heads in 3 flips | no consecutive heads)

= 1/5 { HTH } out of {HTT, THT, TTH, HTH, TTT}

This is a conditional probability.

Independent events:

Intuition:

$\Pr(A|B) = \Pr(A)$, probability of A is not affected by the occurrence of B

<==> $\dfrac{\Pr(AB)}{\Pr(B)} = \Pr(A)$

<==> $\Pr(AB) = \Pr(A)\Pr(B)$

Definition (independent events):

Events A and B are called independent if

$\Pr(AB) = \Pr(A)\Pr(B)$.

Mutually exclusive events:

Definition (mutually exclusive events):

Events A and B are called mutually exclusive if $AB = \emptyset$.

Draw a Venn diagram to show it.

Note: Two mutually exclusive events are definitely not independent.

Law of total probability

Definition: (Partition of sample space Ω)

If $\{B_n, n = 1, 2, \ldots\}$ satisfies

i) $B_1 + B_2 + \cdots = \Omega$

ii) $B_i B_j = \emptyset$ for all $i \neq j$

then $\{B_n, n = 1, 2, \ldots\}$ is called a partition of Ω.

Draw a Venn diagram to show it.

Example: A simple partition of Ω.

$\Omega = A + A^C$

Theorem (the law of total probability)

Suppose $\{B_n, n = 1, 2, \ldots\}$ is a partition of Ω. Then we have

$$\boxed{\Pr(A) = \sum_n \Pr(A \mid B_n)\Pr(B_n)}$$

Proof: $\Pr(A) = \sum_n \Pr(AB_n) = \sum_n \Pr(A \mid B_n)\Pr(B_n)$.

This is called the law of total probability. This law is useful for calculating probability when conditional probabilities are easy to find.

Expected value of a random variable

Random variable: $X(\omega)$

Notation for expected value:

$E(X)$, $E[X]$, $\langle X \rangle$

Repeat the experiment M times. Collect M outcomes, $\{\omega_j, j = 1, 2, \ldots, M\}$

$$E(X)=\lim_{M\to\infty}\frac{\sum_{j=1}^{M}X(\omega_j)}{M}$$

Properties of expected value:

- $E(\alpha X)=\alpha E(X)$
- $E(X+Y)=E(X)+E(Y)$ for all X and Y

  Proof: $\sum_{j=1}^{M}\left(X(\omega_j)+Y(\omega_j)\right)=\sum_{j=1}^{M}X(\omega_j)+\sum_{j=1}^{M}Y(\omega_j)$

Note: Operationally, this definition is not practical for calculating $E(X)$. To establish a convenient mathematical formulation we introduce probability mass function.

Probability mass function (PMF) (of a **discrete** random variable)

Random variable: $N(\omega)$

PMF of random variable $N(\omega)$:

$p_N(k) \equiv \Pr(N(\omega) = k)$

Note: The statistical behavior of a discrete random variable is completely described by its probability mass function (PMF).

Example:

Let N = # of heads in n flips of a fair coin

$$\Pr(N=0)=\left(1/2\right)^n$$

$$\Pr(N=1)=n\left(1/2\right)^n$$

$$\Pr(N=k)=C(n,k)\left(1/2\right)^n, \qquad k=0,1,2,\ldots,n$$

where $C(n, k)$ = # of ways of choosing an unordered subset of k elements from n elements. $C(n, k)$ has the expression

$$C(n,k)=\frac{k!(n-k)!}{n!}$$

$C(n, k)$ is called the binomial coefficient because it appears in …

$$(a+b)^n=a^n+C(n,1)a^{n-1}b^1+C(n,2)a^{n-2}b^2+\cdots$$

Expected value (of a discrete random variable) in terms of PMF

$$E(N)=\lim_{M\to\infty}\frac{\sum_{j=1}^{M}N(\omega_j)}{M}=\lim_{M\to\infty}\frac{\sum_{k}k\times\left(\#\text{ of }N(\omega_j)=k\right)}{M}$$

$$= \sum_k k \times \left(\lim_{M\to\infty} \frac{\#\text{ of } N(\omega_j)=k}{M} \right) = \sum_k k \Pr\left(N(\omega)=k\right) = \sum_k k\, p_N(k)$$

We obtain:

$$\boxed{E(N) = \sum_k k\, p_N(k)}$$

Expected value of $f(N)$

$$E(f(N)) = \sum_k f(k) p_N(k)$$

Probability density function (PDF) (of a **continuous** random variable)

Random variable: $X(\omega)$

$$\rho_X(x) = \lim_{\Delta x\to 0} \frac{\Pr\left(x < X(\omega) \le x+\Delta x\right)}{\Delta x}$$

Note: Here we use $\Pr\left(x < X(\omega) \le x+\Delta x\right)$ for its simplicity. To properly accommodate both continuous and discrete random variables, we should use

$\Pr\left(x-\Delta x/2 < X(\omega) \le x+\Delta x/2\right)$.

Cumulative distribution function (CDF)

$$F_X(x) = \Pr\left(X(\omega) \le x\right)$$

Connection between CDF and PDF:

$$\rho_X(x) = \lim_{\Delta x\to 0} \frac{\Pr\left(x < X(\omega) \le x+\Delta x\right)}{\Delta x} = \lim_{\Delta x\to 0} \frac{F_X(x+\Delta x)-F_X(x)}{\Delta x} = \frac{d}{dx} F_X(x)$$

We obtain:

$$\boxed{\rho_X(x) = \frac{d}{dx} F_X(x)}$$

Expected value (of a continuous random variable) in terms of PDF

$$E(X) = \lim_{M\to\infty} \frac{\sum_{j=1}^{M} X(\omega_j)}{M}$$

We divide the real axis into bins of size Δx. Let $x_i = x_0 + i\,\Delta x$.

$$E(X)=\lim_{\Delta x\to 0}\lim_{M\to\infty}\frac{\sum_i x_i\left(\#\text{ of }\omega_j\ \text{ satisfying } x_i<X(\omega_j)\le x_{i+1}\right)}{M}$$

$$=\lim_{\Delta x\to 0}\sum_i x_i \underbrace{\Pr\left(x_i<X(\omega)\le x_{i+1}\right)}_{\approx\rho_X(x_i)\Delta x}=\lim_{\Delta x\to 0}\sum_i x_i\,\rho_X(x_i)\Delta x=\int x\rho_X(x)dx$$

We obtain:

$$\boxed{E(X)=\int x\rho_X(x)dx}$$

Expected value of $f(X)$

$$E(f(X))=\int f(x)\rho_X(x)dx$$

Remark: With the notation of δ function, we can treat a discrete random variable as continuous where $\rho_X(x)$ is a sum of δ functions.

Example:

$$X=\begin{cases}1, & \text{prob}=0.6\\ 0, & \text{prob}=0.4\end{cases},\qquad \rho_X(x)=0.4\delta(x)+0.6\delta(x-1)$$

Joint density of two random variables (X, Y)

$$\rho_{(X,Y)}(x,y)=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{\Pr\left(x<X(\omega)\le x+\Delta x\ \text{ AND }\ y<Y(\omega)\le y+\Delta y\right)}{(\Delta x)(\Delta y)}$$

Conditional probability density: $\rho_X(x|Y=y)$

$$\rho_X(x|B)=\lim_{\Delta x\to 0}\frac{\Pr\left(x<X(\omega)\le x+\Delta x\middle|B\right)}{\Delta x}$$

$$=\lim_{\Delta x\to 0}\frac{\Pr\left(x<X(\omega)\le x+\Delta x\ \text{ AND }\ \omega\in B\right)}{(\Delta x)}\cdot\frac{1}{\Pr(B)}$$

The above works when $\Pr(B) > 0$.

When $\Pr(B) = 0$, we find a way to work around it.

$$\rho_X\left(x\middle|Y=y\right)=\lim_{\Delta y\to 0}\rho_X\left(x\middle|y<Y\le y+\Delta y\right)$$

(again, we should use $\rho_X\left(x\middle|y-\Delta y/2<Y\le y+\Delta y/2\right)$).

$$= \lim_{\Delta x \to 0} \frac{\dfrac{\Pr\left(x < X(\omega) \le x + \Delta x \text{ AND } y < Y \le y + \Delta y\right)}{(\Delta x)(\Delta y)}}{\dfrac{\Pr\left(y < Y \le y + \Delta y\right)}{(\Delta y)}} = \frac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)}$$

We obtain

$$\boxed{\rho_X\left(x \middle| Y = y\right) = \frac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)}}$$

<u>Independent random variables</u>

<u>Intuition:</u>

$\rho_X\left(x \middle| Y = y\right) = \rho_X(x)$, density of X is not affected by the value of Y.

<==> $\dfrac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)} = \rho_X(x)$

<==> $\rho_{(X,Y)}(x,y) = \rho_X(x)\rho_Y(y)$

<u>Definition</u> (independent random variables):

Random variables X and Y are called independent if

$$\rho_{(X,Y)}(x,y) = \rho_X(x)\rho_Y(y) .$$

<u>Conditional expectation:</u> E(X|B), E(X|Y=y)

We first study E(X|B).

Repeat the experiment M times. Collect M outcomes, $\{\omega_j, j = 1, 2, \ldots, M\}$

Consider only those repeats with $\omega_j \in B$.

$$E\left(X \middle| B\right) = \lim_{M \to \infty} \frac{\sum_{\omega_j \in B} X(\omega_j)}{\#\text{ of } \omega_j \in B}$$

(We divide the real axis into bins of size Δx. Let $x_i = x_0 + \Delta x$).

$$= \lim_{\substack{M \to \infty \\ \Delta x \to 0}} \frac{\sum_i x_i \dfrac{\#\text{ of } \left(x_i < X(\omega_j) \le x_{i+1} \text{ AND } \omega_j \in B\right)}{M}}{\dfrac{\#\text{ of } \omega_j \in B}{M}}$$

$$= \lim_{\Delta x \to 0} \frac{\sum_i x_i \Pr\left(x_i < X(\omega) \le x_{i+1} \text{ AND } \omega \in B\right)}{\Pr(B)} = \int x \rho_X(x|B) dx$$

We obtain

$$\boxed{E(X|B) = \int x \rho_X(x|B) dx}$$

The above works when Pr(B) > 0.

Next we study E(X|Y=y). When Pr(B) = 0, we find a way to work around it.

$$E\left(X\middle|Y = y\right) = \lim_{\Delta y \to 0} E\left(X\middle|y < Y \le y + \Delta y\right)$$

$$= \lim_{\Delta y \to 0} \int x \rho_X\left(x\middle|y < Y \le y + \Delta y\right) dx = \int x \rho_X\left(x\middle|Y = y\right) dx$$

We obtain

$$\boxed{E(X|Y = y) = \int x \rho_X(x|Y = y) dx}$$

Remarks:

i) $\rho_X(x|Y = y) = \lim_{\Delta y \to 0} \rho_X\left(x\middle|y < Y \le y + \Delta y\right)$

$$E(X|Y = y) = \lim_{\Delta y \to 0} E\left(X\middle|y < Y \le y + \Delta y\right)$$

ii) E(X|Y=y) is a function of y. When we apply this function to random variable Y, we get a derived random variable

$E(X|Y) \equiv E(X|Y = y)\Big|_{y=Y}$ is a function of Y, a derived random variable.

We can consider the expected value of random variable E(X|Y).

Law of total expectation

Theorem (the law of total expectation):

$$\boxed{E(X) = E\left(E(X|Y)\right)}$$

This is called the law of total expectation.

Proof:

$$E\left(E(X|Y)\right) = \int E(X|Y = y) \rho_Y(y) dy = \int \left(\int x \rho_X(x|Y = y) dx \right) \rho_Y(y) dy$$

$$= \int x \Big(\int \underbrace{\rho_X(x|Y = y) \rho_Y(y)}_{\rho_{(X,Y)}(x,y)} dy \Big) dx = \int x \Big(\int \rho_{(X,Y)}(x, y) dy \Big) dx = E(X)$$

A special case of the law of total expectation:

Suppose $\{B_n, n = 1, 2, \ldots\}$ is a partition of Ω. We define random variable Y as

$$Y(\omega) = n \text{ where } \omega \in B_n$$

Recall that $E(X|Y)$ is a function of discrete random variable Y. We calculate its expected value directly using $E\big(f(Y)\big) = \sum_n f(n) p_Y(n)$.

$$E\big(E(X|Y)\big) = \sum_n E(X|Y=n)\Pr(|Y=n) = \sum_n E(X|B_n)\Pr(B_n)$$

From the law of total expectation, we obtain

$$\boxed{E(X) = \sum_n E(X|B_n)\Pr(B_n)}$$

AM216 Stochastic Differential Equations

Lecture 02

List of topics in this lecture

- Variance, properties of expectation and variance
- Bernoulli distribution, binomial distribution, memoryless process, derivation of PDF of exponential distribution, normal distribution
- CDF of normal distribution, error function, confidence interval for the mean
- Interpretation of confidence interval

Recap

The framework of repeated experiments for probability: without specifying how the experiment is repeated, the concept of probability does not make sense.

Properties of expected value:

$E(\alpha X) = \alpha E(X)$

$E(X+Y) = E(X) + E(Y)$ for all X and Y

Law of total probability: $\Pr(A) = \sum_n \Pr(A|B_n)\Pr(B_n)$

Law of total expectation: $E(X) = E\left(E(X|Y)\right)$, $E(X) = \sum_n E(X|B_n)\Pr(B_n)$

Review of probability theory (continued)

Variance:

$$\text{var}(X) \equiv E\left((X-E(X))^2\right) = E\left(X^2 - 2E(X)X + (E(X))^2\right)$$

Recall that $E(X)$ is a deterministic number

$$= E(X^2) - 2E(X)E(X) + \left(E(X)\right)^2 = E(X^2) - \left(E(X)\right)^2$$

We obtain:

$$\boxed{\text{var}(X) = E(X^2) - \left(E(X)\right)^2}$$

Standard deviation:

$$std(X) = \sqrt{\text{var}(X)}$$

Properties of $E(X)$

i) $E(aX + bY) = aE(X) + bE(Y)$

This is valid for all X and Y. In particular, X and Y do not need to be independent.

ii) If X and Y are independent, then we have

$$E(X\,Y) = E(X)\,E(Y)$$

Proof:

Independence implies

$$\rho_{(X,Y)}(x,y) = \rho_X(x)\rho_Y(y)$$

Using the independence in the calculation of $E(XY)$, we get

$$E(XY) = \int xy\rho_{(X,Y)}(x,y)\,dx\,dy = \int xy\rho_X(x)\rho_Y(y)\,dx\,dy$$

$$= \left(\int x\rho_X(x)dx\right)\left(\int y\rho_Y(y)dy\right) = E(X)E(Y)$$

Caution:

- $E(X\,Y) = E(X)\,E(Y)$ may not be true if X and Y are not independent.

  Example:

  Let $X = Y = \begin{cases} 2\,, & \Pr = 0.5 \\ 0\,, & \Pr = 0.5 \end{cases}$.

  We have $E(X) = E(Y) = 2\times 0.5 = 1$, $\quad E(X\,Y) = 4\times 0.5 = 2$

  ==> $E(X\,Y) \neq E(X)\,E(Y)$

- $E(X\,Y) = E(X)\,E(Y)$ does not imply that X and Y are independent.

  Example:

  Let $(X,Y) = \begin{cases} (0,1)\,, & \Pr = 0.25 \\ (0,-1)\,, & \Pr = 0.25 \\ (1,0)\,, & \Pr = 0.25 \\ (-1,0)\,, & \Pr = 0.25 \end{cases}$.

  We have $E(X) = 0$, $\quad E(Y) = 0$, $\quad E(X\,Y) = 0$

  ==> $E(X\,Y) = E(X)\,E(Y)$

But $Y^2 = 1 - X^2$. So X and Y are definitely not independent of each other.

Properties of var(X)

iii) $\mathrm{var}(\alpha X) = \alpha^2 \mathrm{var}(X)$

Proof is in your homework.

iv) If X and Y are independent, then we have

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$$

Proof:

$$\mathrm{var}(X+Y) = \mathrm{E}((X+Y)^2) - (\mathrm{E}(X+Y))^2 = \cdots$$

Complete the proof in your homework.

Examples of distributions:

1) Bernoulli distribution

Consider the number of success in ONE trial with success probability p

$$X = \begin{cases} 1, & \Pr = p \\ 0, & \Pr = 1-p \end{cases}$$

We say random variable X has the Bernoulli distribution with parameter p.

Notation:

$X \sim \mathrm{Bern}(p)$

Range = {0, 1}.

Example: Flip a coin

1: head, success

0: tail, failure

Expected value and variance:

$\mathrm{E}(X) = 0\times(1-p)+1\times p = p, \qquad \mathrm{E}(X^2) = p$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = p(1-p)$

2) Binomial distribution

Consider the number of successes in a sequence of n independent trials, each with success probability p.

N = sum of n independent Bernoulli random variables with parameter p.

$$N = \sum_{i=1}^{n} X_i\,, \qquad X_i \sim \text{(iid) Bern}(p)$$

iid = independently and identically distributed

We say random variable N has the binomial distribution with parameters (n, p).

Notation:

$N \sim \text{Bino}(n, p)$ or simply $N \sim \text{B}(n, p)$

Range = $\{0, 1, 2, \ldots, n\}$.

PMF (probability mass function):

$$\Pr(N=k) = C(n,k)p^k(1-p)^{n-k}\,, \qquad k = 0,1,2,\ldots,n$$

Example: # of heads in n flips of a coin

Expected value and variance:

$$E(N) = E\left(X_1 + X_2 + \cdots + X_n\right) = np$$

$$\text{var}(N) = \text{var}\left(X_1 + X_2 + \cdots + X_n\right) = n\,\text{var}(X_1) = np(1-p)$$

3) Exponential distribution

Example: (Escape problem)

T = time until escape from a deep potential well by thermal fluctuations

PDF of T has the form

$$\rho_T(t) = \begin{cases} \lambda \exp\left(-\lambda t\right), & t \ge 0 \\ 0\,, & t < 0 \end{cases}$$

We say random variable T has the exponential distribution with parameter λ.

Notation:

$T \sim \text{Exp}(\lambda)$

Range = $(0, +\infty)$.

Mathematical definition of exponential distribution:

T = time from $t = 0$ until occurrence of an event in a memoryless system.

Derivation of PDF of T based on the "memoryless" property:

Recall that T = time until occurrence. "Memoryless" means

"Given that the event has not occurred at t_0, the additional time until occurrence is not affected by t_0 no matter how large or how small t_0 is."

==> $\Pr(\underbrace{(T-t_0)}_{\text{additional time}} \le t \,|\, T>t_0) = \Pr(T \le t)$

Consider the complementary cumulative distribution function (CCDF)

$$G(t) \equiv \Pr\left(T>t\right) = \int_t^\infty \rho_T(t')dt'$$

$$G(0) = \Pr(T > 0) = 1$$

We re-write the memoryless property in terms of $G(t)$.

$$\frac{\Pr\left((T-t_0)\le t \text{ AND } T>t_0\right)}{\Pr\left(T>t_0\right)} = \Pr\left(T \le t\right)$$

==> $\Pr\left(t_0 < T \le t_0 + t\right) = \Pr\left(T \le t\right)\Pr\left(T > t_0\right)$

==> $G(t_0) - G(t_0+t) = \left(1-G(t)\right)G(t_0)$

Replace t with Δt, divide by Δt, and take the limit as $\Delta t \to 0$, we get

$$\frac{G(t_0)-G(t_0+\Delta t)}{\Delta t} = \frac{G(0)-G(\Delta t)}{\Delta t}G(t_0)$$

==> $G'(t_0) = \underbrace{G'(0)}_{-\lambda} G(t_0)$

Let $\lambda \equiv -G'(0)$. We obtain an initial value problem (IVP) for $G(t_0)$

$$\begin{cases} G'(t_0) = -\lambda G(t_0), & t_0 > 0 \\ G(0) = 1 \end{cases}$$

The solution is $\quad G(t) = \exp(-\lambda t), \quad t > 0$.

Differentiate $G(t) \equiv \int_t^\infty \rho_T(t')dt'$, we obtain

$$\rho_T(t) = -\frac{d}{dt}G(t) = \begin{cases} \lambda \exp(-\lambda t), & t \ge 0 \\ 0, & t < 0 \end{cases}$$

<u>Remark:</u> The time until occurrence of an event in a memoryless system must have a PDF of the form given above.

Expected value and variance:

$$E(T) = \int t \rho_T(t) dt = \int_0^{+\infty} t \lambda \exp(-\lambda t) dt = \frac{1}{\lambda}$$

$$E(T^2) = \int_0^{+\infty} t^2 \lambda \exp(-\lambda t) dt = \frac{2}{\lambda^2} \qquad \text{(see Appendix A for the calculation)}$$

$$\text{var}(T) = E(T^2) - E(T)^2 = \frac{1}{\lambda^2}$$

CDF:

$$F_T(t) = \Pr\left(T \le t\right) = 1 - \exp(-\lambda t) \qquad \text{for } t \ge 0$$

4) Normal distribution

PDF:

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

We say random variable X has the normal distribution with parameters (μ, σ^2).

Notation:

$X \sim N(\mu, \sigma^2)$

Range = $(-\infty, +\infty)$

Example: (*Central Limit Theorem*)

Suppose $\{X_1, X_2, \ldots, X_M\}$ are iid (independently and identically distributed).

When M is large, $X = \sum_{j=1}^{M} X_j$ approximately has a normal distribution.

Expected value and variance:

$$E(X) = E(X-\mu) + \mu = \underbrace{\int (x-\mu)\rho_X(x) dx}_{=0 \text{ because of symmetry}} + \mu = \mu$$

$$\text{var}(X) = E\left((X-\mu)^2\right) = \int (x-\mu)^2 \rho_X(x) dx = \sigma^2 \qquad \text{(see Appendix A)}$$

CDF of normal distribution:

$$F_X(x) = \Pr\left(X \le x\right) = \int_{-\infty}^{x} \rho_X(x)dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x'-\mu)^2}{2\sigma^2}\right) dx'$$

Change of variables: $s = \dfrac{x'-\mu}{\sqrt{2\sigma^2}}$, $dx' = \sqrt{2\sigma^2}\,ds$

$$F_X(x) = \int_{-\infty}^{\frac{x-\mu}{\sqrt{2\sigma^2}}} \frac{1}{\sqrt{\pi}} \exp(-s^2)ds = \frac{1}{2} + \int_{0}^{\frac{x-\mu}{\sqrt{2\sigma^2}}} \frac{1}{\sqrt{\pi}} \exp(-s^2)ds$$

We write the CDF in terms of the error function.

<u>The error function:</u>

$$\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z \exp(-s^2)ds$$

<u>Properties of erf(z):</u>

i) erf(0) = 0

ii) erf(+∞) = 1

iii) erf(−z) = −erf(z)

In terms of erf(z), the CDF of normal distribution has the expression

$$\boxed{F_X(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right)}$$

<u>Example:</u>

$$\Pr\left(X \le \mu + \eta\sigma\right) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\mu + \eta\sigma - \mu}{\sqrt{2\sigma^2}}\right)\right) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\eta}{\sqrt{2}}\right)\right)$$

<u>Interval containing 95% probability</u>

We like to find η such that

$$\Pr\left(\left|X-\mu\right| \le \eta\sigma\right) = 0.95 \qquad (95\%)$$

We express this probability in terms of CDF, and then in terms of erf( ).

$$\Pr\left(\left|X-\mu\right| \le \eta\sigma\right) = \Pr\left(\mu - \eta\sigma \le X \le \mu + \eta\sigma\right)$$

$$= F_X\left(\mu + \eta\sigma\right) - F_X\left(\mu - \eta\sigma\right) = \quad \cdots \quad = \operatorname{erf}\left(\frac{\eta}{\sqrt{2}}\right)$$

Setting $\mathrm{erf}\left(\frac{\eta}{\sqrt{2}}\right)=0.95$, we calculate η using the inverse error function

$$\eta=\mathrm{erfinv}(0.95)\sqrt{2}=1.96$$

We obtain

$$\boxed{\Pr\left(\left|X-\mu\right|\le 1.96\sigma\right)=95\%}$$

Similarly, we can obtain

$$\boxed{\Pr\left(\left|X-\mu\right|\le 2.5758\sigma\right)=99\%}$$

<u>Confidence interval:</u>

Suppose we are given a data set of n independent samples of $X \sim N(\mu, \sigma^2)$.

$\{X_j, j = 1, 2, \ldots, n\}$

Suppose we don't know μ and we want to estimate μ from the data.

<u>Question:</u> How to estimate μ from data?

We can use the sample mean.

$$\hat{\mu}=\frac{1}{n}\sum_{j=1}^{n}X_j$$

<u>Question:</u> How to estimate the uncertainty/error in $\hat{\mu}$?

First we recognize that $\hat{\mu}$ is a random variable, derived from random variables (X_1, X_2, …, X_n). Each data set gives a (potentially) different value of $\hat{\mu}$.

$$E(\hat{\mu})=E\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)=\frac{1}{n}E\left(X_1+\cdots+X_n\right)=\frac{1}{n}n\mu=\mu$$

$$\mathrm{var}(\hat{\mu})=\mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)=\frac{1}{n^2}\mathrm{var}\left(X_1+\cdots+X_n\right)=\frac{1}{n^2}n\,\mathrm{var}(X_1)=\frac{\sigma^2}{n}$$

Here we used the independence of $\{X_j\}$.

<u>Theorem:</u>

Sum of independent normal random variables is a normal random variable.

<u>Proof:</u> It will be proved in the discussion of characteristic functions.

It follows from the theorem that $\hat{\mu}$ is normal.

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$ This is called a standard normal.

The interval containing 95% probability is

$$\Pr\left(\left|\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}\right| \le 1.96\right) = 95\%$$

Case 1: Suppose the value of σ is given.

$$\left|\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}\right| \le 1.96 \quad \Longleftrightarrow \quad \mu \in \left(\hat{\mu}-1.96\frac{\sigma}{\sqrt{n}},\ \hat{\mu}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

which is called the 95% confidence interval (CI) for the mean.

Example:

We are given a data set of 100 independent samples of $X \sim N(\mu, \sigma^2)$:

{3.0811, 0.7589, 1.9611, 0.3050, 0.3887, 1.4971, 1.3225, −0.8563, … }

We are given σ = 1.3. We estimate μ using the sample mean

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j = 0.475$$

$$1.96\frac{\sigma}{\sqrt{n}} = 0.2548$$

The 95% confidence interval for the mean is (0.2202, 0.7298)

Interpretation of the confidence interval

Question: What is the meaning of the 95% confidence interval for the mean?

μ is fixed, although unknown. μ is not random.

For the given data set, the 95% confidence interval is determined: (0.2202, 0.7298).

We have either $\mu \in (0.2202,\ 0.7298)$ or $\mu \notin (0.2202,\ 0.7298)$.

It is not uncertain. It is just unknown to us (because μ is unknown).

"$\Pr(\mu \in (0.2202,\ 0.7298)) = 95\%$" does not make sense.

Two key components in interpreting the confidence interval:

i) The confidence interval is an algorithm/function that maps a data set $\{X_j\}$ to an interval

$$\{X_j\} \longrightarrow \left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$$

where $\quad \hat{\mu}_L(\{X_j\}) = \hat{\mu}(\{X_j\}) - 1.96\frac{\sigma}{\sqrt{n}}, \quad \hat{\mu}_H(\{X_j\}) = \hat{\mu}(\{X_j\}) + 1.96\frac{\sigma}{\sqrt{n}}$

It is important to notice that $\left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$ varies with data set $\{X_j\}$.

For a random data set, $\left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$ is a random variable, derived from the random data set.

ii) We view it in the framework of repeated experiments.

Draw a data set of n independent samples of $X \sim N(\mu, \sigma^2)$.

Repeat the drawing M times (M is large).

When we go over M data sets and estimate the confidence interval for each data set, for 95% of data sets, the estimated confidence interval contains μ.

$$\Pr\left(\underbrace{\mu}_{\text{Fixed}} \in \underbrace{\left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)}_{\substack{\text{Random}\\ \text{variable}}}\right) = 0.95$$

In summary, the two key components for interpreting the confidence interval are

i) the confidence interval is an algorithm mapping a data set to an interval; and

ii) the 95% probability is in the framework of hypothetically drawing a large number of data sets and applying the algorithm to each data set.

Case 2: σ is unknown

Recall the definition of standard deviation.

$$\sigma = \sqrt{\text{var}(X)} = \sqrt{E\left((X-\mu)^2\right)}$$

From the given samples, we can calculate the sample standard deviation

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \hat{\mu})^2}, \qquad \hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

Note: The denominator is $(n-1)$ instead of n. This modification is to make the sample variance unbiased: $E(\hat{\sigma}^2) = \sigma^2$.

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \hat{\mu})^2\right) = \frac{1}{(n-1)}\sum_{j=1}^{n} E\left((X_j - \hat{\mu})^2\right), \qquad \hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Let $Y_j \equiv X_j - \mu$. We have

$$X_j = \mu + Y_j\,,\quad \hat{\mu} = \mu + \frac{1}{n}\sum_{k=1}^{n} Y_k\,,\quad E(Y_k) = 0 \text{ and } E(Y_k^{\,2}) = \sigma^2$$

$$E\left((X_1-\hat{\mu})^2\right) = E\left((Y_1 - \frac{1}{n}\sum_{k=1}^{n} Y_k)^2\right) = E\left(\left(\frac{n-1}{n}Y_1 - \frac{1}{n}Y_2 - \cdots - \frac{1}{n}Y_n\right)^2\right)$$

$$= E\left(\frac{(n-1)^2}{n^2}Y_1^{\,2} + \frac{1}{n^2}Y_2^{\,2} + \cdots + \frac{1}{n^2}Y_n^{\,2}\right) = \left(\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right)\sigma^2 = \frac{n-1}{n}\sigma^2$$

$$E(\hat{\sigma}^2) = \frac{1}{(n-1)}\sum_{j=1}^{n} E\left((X_j - \hat{\mu})^2\right) = \frac{1}{(n-1)} n \frac{(n-1)}{n}\sigma^2 = \sigma^2$$

Using $\hat{\sigma}$, we write out an approximate 95% confidence interval

$$\left(\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

A better solution for case 2 (optional):

When σ is unknown, we use $\hat{\sigma}$ to replace σ. $\dfrac{\hat{\mu}-\mu}{\hat{\sigma}/\sqrt{n}}$ is not exactly a normal distribution (it is approximately a normal distribution).

$\dfrac{\hat{\mu}-\mu}{(\hat{\sigma}/\sqrt{n})}$ is exactly a Student's t-distribution with $(n-1)$ degrees of freedom.

From the inverse CDF of the t-distribution, we can find the exact value of η such that

$$\Pr\left(\left|\frac{\hat{\mu}-\mu}{(\hat{\sigma}/\sqrt{n})}\right| \le \eta\right) = 95\%$$

<==> $F_t\left(\eta, (n-1)\right) = 0.975$

<==> $\eta = F_t^{(\mathrm{inv})}\left(0.975, (n-1)\right)$

The 95% confidence interval is $\left(\hat{\mu} - \eta\dfrac{\hat{\sigma}}{\sqrt{n}},\ \hat{\mu} + \eta\dfrac{\hat{\sigma}}{\sqrt{n}}\right)$.

Appendix A: An alternative way of calculating some integrals

<u>Integral 1:</u> $I_1 = \int_0^{+\infty} t^2 \lambda \exp(-\lambda t) dt$

To calculate I_1, we consider

$$G(\lambda) \equiv \int_0^{+\infty} \exp(-\lambda t) dt = \frac{1}{\lambda}, \qquad \frac{dG(\lambda)}{d\lambda} = -\int_0^{+\infty} t \exp(-\lambda t) dt = \frac{-1}{\lambda^2}$$

We write I_1 as

$$I_1 = \lambda \int_0^{+\infty} t^2 \exp(-\lambda t) dt = \lambda \frac{d^2 G(\lambda)}{d\lambda^2} = \lambda \frac{2}{\lambda^3} = \frac{2}{\lambda^2}$$

<u>Integral 2:</u> $I_2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx$

To calculate I_2, we consider

$$G(\sigma) \equiv \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}, \qquad \frac{dG(\sigma)}{d\sigma} = \frac{1}{\sigma^3} \int_{-\infty}^{+\infty} x^2 \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \sqrt{2\pi}$$

We write I_2 as

$$I_2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \frac{\sigma^2}{\sqrt{2\pi}} \frac{dG(\sigma)}{d\sigma} = \sigma^2$$

AM216 Stochastic Differential Equations

Lecture 03

List of topics in this lecture

- Characteristic function (CF) of a RV, relation with Fourier transform
- Properties of CF, CF of sum of two independent RVs, CF of a normal RV
- Sum of independent normal RVs is a normal RV.
- Monty Hall's game, incomplete description of a game
- Stochastic process, the Wiener process $W(t)$

Recap

Bernoulli distribution, binomial distribution, exponential distribution, memoryless process, normal distribution, error function, confidence interval

Short notations:

RV = random variable

PDF = probability density function

CDF = cumulative distribution function

FT = Fourier transform

CF = characteristic function

Review of probability theory (Continued)

We now develop tools to show that

$\boxed{\text{Sum of independent normal RVs is a normal RV}}$

Characteristic function (CF) of a random variable

Random variable: $X(\omega)$

PDF of X: $\rho_X(x)$

The characteristic function of X is defined as

$$\phi_X(\xi) \equiv E\Big(\exp(i\xi X)\Big) = \int_{-\infty}^{+\infty} \exp(i\xi x)\rho_X(x)dx$$

This is very similar to the Fourier transform (FT) of $\rho_X(x)$

Fourier transform (FT): $f(x) \to \hat{f}(\xi)$

$$\hat{f}(\xi) = \int_{-\infty}^{+\infty} \exp(-i2\pi\xi x) f(x)\, dx$$

Inverse transform: $\hat{f}(\xi) \to f_2(x)$

$$f_2(x) = \int_{-\infty}^{+\infty} \exp(i2\pi\xi x) \hat{f}(\xi)\, d\xi$$

Theorem:

$$f_2(x) = f(x)$$

This theorem justifies the name "inverse transform".

Relation between the characteristic function (CF) and the Fourier transform (FT)

$$\phi_X(\xi) = \int_{-\infty}^{+\infty} \exp(i\xi x) \rho_X(x)\, dx$$

$$\hat{\rho}_X(\xi') = \int_{-\infty}^{+\infty} \exp(-i2\pi\xi' x) \rho_X(x)\, dx$$

==> $$\boxed{\phi_X(\xi) = \hat{\rho}_X(\xi')\Big|_{\xi' = \frac{-\xi}{2\pi}}}$$

Theorem (Properties of CF):

- $\phi_X(\xi)\Big|_{\xi=0} = 1$

  Proof: $\phi_X(\xi)\Big|_{\xi=0} = E\left(\exp(i\xi X)\right)\Big|_{\xi=0} = E(1) = 1$

- CF and the first moment

$$\frac{d}{d\xi}\phi_X(\xi)\Bigg|_{\xi=0} = iE(X)$$

Proof:

$$\frac{d}{d\xi}\phi_X(\xi) = \frac{d}{d\xi}E\left(\exp(i\xi X)\right) = E\left(\frac{d}{d\xi}\exp(i\xi X)\right) = E\left(iX\exp(i\xi X)\right)$$

==> $$\frac{d}{d\xi}\phi_X(\xi)\Bigg|_{\xi=0} = iE(X)$$

- CF and the second moment

$$\left.\frac{d^2}{d\xi^2}\phi_X(\xi)\right|_{\xi=0}=-E(X^2)$$

- Expansion of CF around $\xi = 0$

$$\phi_X(\xi)=1+iE(X)\xi-\frac{E(X^2)}{2}\xi^2+\cdots$$

- Mapping from PDF to CF is invertible:

  If $\phi_X(\xi)=\phi_Y(\xi)$, then $\rho_X(s)=\rho_Y(s)$.

  Proof: this property follows from the invertibility of FT.

- CF of the sum of two independent RVs.

  If random variables X and Y are independent, then we have

$$\phi_{(X+Y)}(\xi)=\phi_X(\xi)\cdot\phi_Y(\xi)$$

  Proof:

$$\phi_{(X+Y)}(\xi)=E\Big(\exp(i\xi(X+Y))\Big)=E\Big(\exp(i\xi X)\cdot\exp(i\xi Y)\Big)$$

(using the independence)

$$=E\Big(\exp(i\xi X)\Big)\cdot E\Big(\exp(i\xi Y)\Big)=\phi_X(\xi)\cdot\phi_Y(\xi)$$

- CF of a shifted RV.

  Let $Y = \mu + X$. The CFs of the two are related by

$$\phi_Y(\xi)=\exp(i\xi\mu)\phi_X(\xi)$$

  Proof:

$$\phi_Y(\xi)=E\Big(\exp(i\xi Y)\Big)=E\Big(\exp(i\xi(\mu+X))\Big)=\exp(i\xi\mu)E\Big(\exp(i\xi X)\Big)=\exp(i\xi\mu)\phi_X(\xi)$$

- CF of a scaled RV.

  Let $Y = \sigma X$. The CFs of the two are related by

$$\phi_Y(\xi)=\phi_X(\sigma\xi)$$

  Proof:

$$\phi_Y(\xi)=E\Big(\exp(i\xi Y)\Big)=E\Big(\exp(i\xi\sigma X)\Big)=E\Big(\exp\big(i(\sigma\xi)X\big)\Big)=\left.\phi_X(\xi')\right|_{\xi'=\sigma\xi}$$

<u>CF of a normal random variable</u>: Let $X \sim N(\mu, \sigma^2)$

PDF: $\rho_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\dfrac{-(x-\mu)^2}{2\sigma^2} \right)$

Characteristic function:

$$\phi_X(\xi) \equiv E\big(\exp(i\xi X)\big) = \int \exp(i\xi x)\rho_X(x)dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(\frac{-(x-\mu)^2 + i2\sigma^2\xi x}{2\sigma^2} \right) dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(\frac{-(x-\mu)^2 + i2\sigma^2\xi(x-\mu) - (i\sigma^2\xi)^2}{2\sigma^2} + i\mu\xi - \frac{\sigma^2\xi^2}{2} \right) dx$$

$$= \exp\left(i\mu\xi - \frac{\sigma^2\xi^2}{2} \right) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(\frac{-[(x-\mu) - i\sigma^2\xi]^2}{2\sigma^2} \right) dx$$

change of variables: $z = (x-\mu) - i\sigma^2\xi$

$$= \exp\left(i\mu\xi - \frac{\sigma^2\xi^2}{2} \right) \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(\frac{-z^2}{2\sigma^2} \right) dz = \exp\left(i\mu\xi - \frac{\sigma^2\xi^2}{2} \right)$$

<u>Theorem</u> (CF of a normal RV):

$X \sim N(\mu, \sigma^2)$ if and only if $\phi_X(\xi) = \exp\left(i\mu\xi - \dfrac{\sigma^2\xi^2}{2} \right)$.

We apply the theorem to the <u>sum of two independent normal RVs</u>.

<u>Theorem</u> (sum of two independent normal RVs)

Suppose X and Y are independent, and $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$.

Then we have

$(X + Y) \sim N(\mu_1+\mu_2, \sigma_1^2 + \sigma_2^2)$.

<u>Proof:</u>

$$\phi_{(X+Y)}(\xi) = \phi_X(\xi)\cdot\phi_Y(\xi) = \exp\left(i\mu_1\xi - \frac{\sigma_1^2\xi^2}{2} \right) \cdot \exp\left(i\mu_2\xi - \frac{\sigma_2^2\xi^2}{2} \right)$$

$$= \exp\left(i(\mu_1 + \mu_2)\xi - \frac{\left(\sigma_1^2 + \sigma_2^2\right)\xi^2}{2} \right)$$

which is the CF of $N(\mu_1+\mu_2, \sigma_1^2 + \sigma_2^2)$.

Remark: The collection of normal distributions has the property that it is closed with respect to the operation of summation.

Conversely, we can use this property to derive the PDF of normal distribution.

Let $\{\Theta(\mu, \sigma^2)\}$ denote a family of distributions parameterized by mean = μ and variance = σ^2. Let $f(x; \mu, \sigma^2)$ be the PDF of distribution $\Theta(\mu, \sigma^2)$.

Theorem:

Suppose the distribution family $\{\Theta(\mu, \sigma^2)\}$ is closed to i) translation, ii) scalar multiplication and iii) summation of independent RVs. Then the PDF of $Z \sim \Theta(\mu, \sigma^2)$ must have the expression

$$f(z;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(z-\mu)^2}{2\sigma^2}\right)$$

Proof: see Appendix A

Monty Hall's game:

Before we start the discussion of stochastic differential equations, let us look at another example to demonstrate the importance of the framework of repeated experiments.

A possible background:

Your group and Mike's group will do summer camping together. But each group has a different itinerary in mind. To decide on a joint itinerary, you and Mike play a game ONCE with Mike hosting.

Specifications of the simple game:

1) The host (Mike) puts a card in one of the 3 boxes without you looking (so he knows which box has the card but you don't know).
2) You select a box and the host (Mike) opens it.
3) If the box you pick contains the card, you win (and you will have priority in the itinerary planning).

   Otherwise, you lose (and Mike will have priority in the itinerary planning).
4) At the end, all boxes are opened to verify that the host is not cheating.

The incident:

After you select box #1, before opening your selection, the host (Mike) says

"Let us play the Monty Hall style".

- He opens box #2 to show that it is empty.
- Then without opening box #1, he offers you the option of switching to box #3.

Question: Should you switch?

<u>Answer:</u> The behavior of the host (Mike) is incompletely specified for repeating the experiment. There are many possible ways the game can be repeated.

<u>Version 1:</u> Theoretical Monty Hall's game

This is Mathematicians' definition of Monty Hall's game. The real game show, hosted by Monty Hall, did not actually follow these mathematical rules.

- Upon your initial selection, before opening your selection, the host <u>must</u> open one of the two remaining boxes.
- The host <u>must</u> open an empty box to show it is empty.
- The host <u>must</u> offer you the option of switching to the other remaining box.

For this game, we have

Pr(winning | not switching) = 1/3

Pr(winning | switching) = 2/3

See Appendix B1 for derivation.

<u>Version 2:</u> The greedy host

The greedy host wants to lure you away from the correct box.

- Upon your initial selection, the greedy host will open one of the two remaining boxes <u>if and only if</u> your initial selection is correct (containing the card).
- If the host opens a remaining box, he must offer you the option of switching to the other remaining box.

For this game, we have

Pr(winning | not switching if offered) = 1/3

Pr(winning | switching if offered) = 0

See Appendix B2 for derivation.

<u>Caution:</u> the condition in the first conditional probability is "not switching if offered" which includes two cases:

i) you are not offered the option of switching, and

ii) you are offered the option but you do not switch.

The same description applies to the second probability.

<u>Version 3:</u> The less greedy host

The less greedy host still wants to lure you away from the correct box. But he wants to avoid this behavior being easily recognized in <u>repeated games</u>.

- Upon your initial selection, before opening your selection, the host may or may not open one of the two remaining boxes. The less greedy host adds some randomness to the decision on whether or not to open a box.

  Pr(opening a box| your initial selection is incorrect) = p_1

  Pr(opening a box| your initial selection is correct) = p_2

- If the host opens a box, he must open an empty box to show it is empty.
- If the host opens a box, he must offer you the option of switching to the other remaining box.

Version 2 is a special case of Version 3 with $p_1 = 0$ and $p_2 = 1$.

Version 1 is a special case of Version 3 with $p_1 = 1$ and $p_2 = 1$.

For this game, we have

Pr(winning | not switching if offered) = 1/3

Pr(winning | switching if offered) = $(1+2p_1-p_2)/3$

For example, for $p_1 = 0.25$ and $p_2 = 0.75$

Pr(winning | switching if offered) = 0.25

See Appendix B3 for derivation.

Key observation:

When you encounter an incompletely specified game only ONCE you have to make a model perceiving how the game is repeated. The model is subjective.

Stochastic differential equation

$$dX(t) = b(X(t),t)dt + \sqrt{a(X(t),t)}\,dW(t)$$

Or in a more concise form

$$dX = b(X,t)dt + \sqrt{a(X,t)}\,dW$$

Notations: $dW \equiv W(t+dt) - W(t)$, $dX \equiv X(t+dt) - X(t)$

We need to introduce $W(t)$.

Definition:

A random variable maps ω to a number or a vector, $X(\omega)$.

A stochastic process maps ω to a function of time, $F(t;\omega)$.

Remark: The collection of all functions is infinite dimensional. In $F(t;\omega)$, we need ω to be infinite dimensional, which conceptually is a bit challenging.

The Wiener process (Brownian motion)

Definition 1:

The Wiener process, denoted by $W(t)$, is a stochastic process satisfying

1) $W(0) = 0$
2) For $t \geq 0$, $W(t) \sim N(0, t)$
3) For $t_4 \geq t_3 \geq t_2 \geq t_1 \geq 0$,

   increments $W(t_2)-W(t_1)$ and $W(t_4)-W(t_3)$ are independent.

Definition 2:

1) $W(0) = 0$
2) For $t_2 \geq t_1 \geq 0$, $W(t_2)-W(t_1) \sim N(0, t_2-t_1)$
3) For $t_4 \geq t_3 \geq t_2 \geq t_1 \geq 0$,

   increments $W(t_2)-W(t_1)$ and $W(t_4)-W(t_3)$ are independent.

Definition 2 appears to be stronger than Definition 1.

Question: Are these two definitions equivalent?

Answer: Yes.

Theorem:

Suppose X and Y are independent, and $X \sim N(\mu_1, \sigma_1^2)$ and $(X+Y) \sim N(\mu_2, \sigma_2^2)$.

Then we have

$$Y \sim N(\mu_2-\mu_1, \sigma_2^2-\sigma_1^2)$$

Proof: Homework problem.

Remark:

Suppose X and Y are independent.

A previous theorem: $X \sim N(\ )$ and $Y \sim N(\ )$ ==> $X+Y \sim N(\ )$.

The current theorem: $X \sim N(\ )$ and $X+Y \sim N(\ )$ ==> $Y \sim N(\ )$.

Using this theorem, we show that Definition 1 is as strong as Definition 2.

We start with Definition 1 and derive Definition 2.

Definition 1:

==> $W(t_1) \sim N(0, t_1)$ and $W(t_2) \sim N(0, t_2)$

We write $W(t_2)$ as a sum

$$W(t_2) = W(t_1) + (W(t_2)-W(t_1))$$

For $t_2 \ge t_1 \ge 0$, $W(t_1)$ and $(W(t_2)-W(t_1))$ are independent.

Applying the theorem above, we conclude

$$W(t_2)-W(t_1) \sim N(0, t_2-t_1)$$

which is Definition 2.

Appendix A

Theorem:

Suppose the distribution family $\{\Theta(\mu, \sigma^2)\}$ is closed to i) translation, ii) scalar multiplication and iii) summation of independent RVs. Then the PDF of $Z \sim \Theta(\mu, \sigma^2)$ must have the expression

$$f(z;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(z-\mu)^2}{2\sigma^2} \right)$$

Proof:

Consider independent RVs X and Y with $X \sim \Theta(0, 1)$ and $Y \sim \Theta(0, \varepsilon)$.

$Y \sim \Theta(0, \varepsilon)$ and $\sqrt{\varepsilon}\, X \sim \Theta(0, \varepsilon)$ have the same PDF and thus the same CF.

$$\phi_Y(\xi) = \phi_{\sqrt{\varepsilon} X}(\xi) \rightarrow \phi_X(\sqrt{\varepsilon}\,\xi) \qquad \text{(E01)}$$

$\sqrt{1+\varepsilon}\, X \sim \Theta(0, 1+\varepsilon)$ and $X + Y \sim \Theta(0, 1+\varepsilon)$ have the same PDF and thus the same CF.

$$\phi_X(\sqrt{1+\varepsilon}\,\xi) \leftarrow \phi_{\sqrt{1+\varepsilon} X}(\xi) = \phi_{(X+Y)}(\xi) \rightarrow \phi_X(\xi)\phi_Y(\xi)$$

Using the expression of $\phi_Y(\xi)$ in (E01), we obtain

$$\phi_X(\sqrt{1+\varepsilon}\,\xi) = \phi_X(\xi)\phi_X(\sqrt{\varepsilon}\,\xi) \qquad \text{(E02)}$$

We expand the LHS and RHS of (E02) in terms of ε for small ε.

$$\text{LHS} = \phi_X(\sqrt{1+\varepsilon}\,\xi) = \phi_X(\xi + (\varepsilon/2)\xi + \cdots) = \phi_X(\xi) + \phi_X'(\xi)\frac{\xi}{2}\varepsilon + \cdots$$

Recall $E(X) = 0$, $E(X^2) = 1$, and $\phi_X(\delta) = 1 + E(X)\delta - \frac{E(X^2)}{2}\delta^2 + \cdots$

$$\text{RHS} = \phi_X(\sqrt{\varepsilon}\,\xi) = 1 - \frac{\xi^2}{2}\varepsilon + \cdots$$

Substituting the expansions into (E02) yields

$$\phi_X(\xi)+\phi_X'(\xi)\frac{\xi}{2}\varepsilon+\cdots=\phi_X(\xi)-\phi_X(\xi)\frac{\xi^2}{2}\varepsilon+\cdots$$

Equating the coefficients of corresponding ε terms on both sides, we get

$$\phi_X'(\xi)=-\phi_X(\xi)\xi$$

$$\Longrightarrow \quad \frac{d}{d\xi}\ln\phi_X(\xi)=-\xi$$

This is an ODE on $\phi_X(\xi)$. Solving it with condition $\phi_X(0) = 1$ gives us

$$\phi_X(\xi)=\exp\left(\frac{-\xi^2}{2}\right)$$

For a general $Z \sim \Theta(\mu, \sigma^2)$, we notice that Z and $(\mu + \sigma X) \sim \Theta(\mu, \sigma^2)$ have the same PDF and thus the same CF.

Recall the scaling and translation properties: $\phi_{\sigma X}(\xi)=\phi_X(\sigma\xi)$, $\phi_{\mu+X}(\xi)=\exp(i\xi\mu)\phi_X(\xi)$

$$\phi_Z(\xi)=\phi_{\mu+\sigma X}(\xi)=\exp\left(i\xi\mu-\frac{\sigma^2\xi^2}{2}\right)$$

Mapping the CF to the PDF, we arrive at

$$f(z;\mu,\sigma^2)=\rho_Z(z)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(z-\mu)^2}{2\sigma^2}\right)$$

Appendix B1

Suppose you never switch.

Pr(winning | not switching)

= Pr(your initial selection is correct) = 1/3

Suppose you always switch. Recall that the host must open an empty box.

You will switch to the correct box if and only if you initial section is incorrect.

Pr(winning | switching)

= Pr(your initial selection is incorrect) = 2/3

Appendix B2

Suppose you never switch if offered.

Pr(winning | not switching if offered)

= Pr(your initial selection is correct) = 1/3

Suppose you always switch if offered. Let

C = "your initial selection is correct"

O = "the host opens an empty box and offers you the option "

S = "switching if offered"

W = "winning"

The greedy host opens a box and offers you the option of switching if and only if your initial selection is correct.

O <==> C

We use the law of total probability.

Pr(winning | switching if offered) = Pr(W | S)

= Pr(W | C and S) Pr(C) + Pr(W| C^C and S) Pr(C^C)

= 0 × (1/3) + 0× (2/3) = 0

Appendix B3

Suppose you never switch if offered.

Pr(winning | not switching if offered)

= Pr(your initial selection is correct) = 1/3

Suppose you always switch if offered. Let

C = "your initial selection is correct"

O = "the host opens an empty box and offers you the option "

S = "switching if offered"

W = "winning"

The host decides whether or not to open a box with probabilities

Pr(O | C^C) = p_1

Pr(O | C) = p_2

We use the law of total probability.

Pr(winning | switching if offered) = Pr(W | S)

= Pr(W | C and O and S) Pr(C and O)

+ Pr(W | C and O^C and S) Pr(C and O^C)

+ Pr(W | C^C and O and S) Pr(C^C and O)

+ Pr(W | C^C and O^C and S) Pr(C^C and O^C)

We first calculate the various terms used in the law of total probability.

Pr(C and O) = Pr(O | C) Pr(C) = p_2*(1/3)

Pr(C and O^C) = Pr(O^C | C) Pr(C) = (1− p_2)*(1/3)

Pr(C^C and O) = Pr(O | C^C) Pr(C^C) = p_1*(2/3)

Pr(C^C and O^C) = Pr(O^C | C^C) Pr(C^C) = (1− p_1)*(2/3)

Pr(W | C and O and S) = 0

Pr(W | C and O^C and S) = 1

Pr(W | C^C and O and S) = 1

Pr(W | C^C and O^C and S) = 0

Substituting these terms into the law of total probability, we obtain

Pr(winning | switching if offered)

= 0 + (1− p_2)*(1/3) + p_1*(2/3) + 0

= (1+2p_1−p_2)/3

AM216 Stochastic Differential Equations

Lecture 04

List of topics in this lecture

- Properties of Wiener process, $dW = O(\sqrt{dt})$
- Discrete version of $W(t)$; arc length of $W(t)$ over finite time is infinity!
- Ito's lemma; $(dW)^2$ can be replaced by dt.
- The gambler's ruin problem; applications of Ito's lemma, law of total probability, law of total expectation; survival probability as a function of (initial cash, time)

Recap

Translation and scaling of normal RVs

If $X \sim N(\mu, \sigma^2)$, then $\dfrac{X-\mu}{\sigma} \sim N(0,1)$, which is called a standard normal RV.

Theorem:

Sum of independent normal RVs is a normal RV.

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then $(X+Y) \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Stochastic differential equation (SDE)

$$dX = b(X,t)dt + \sqrt{a(X,t)}\,dW$$

Notations: $dW \equiv W(t+dt) - W(t)$, $dX \equiv X(t+dt) - X(t)$

The Wiener process, denoted by $W(t)$, satisfies

1) $W(0) = 0$
2) For $t_2 \ge t_1 \ge 0$, $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$
3) For $t_4 \ge t_3 \ge t_2 \ge t_1 \ge 0$,

   increments $W(t_2) - W(t_1)$ and $W(t_4) - W(t_3)$ are independent.

Note: $W(t)$ is a stochastic process. The full notation is $W(t, \omega)$.

Complication of SDE

Ordinary Difference Equation:

$$\Delta X = b(X,t)\Delta t + o(\Delta t)$$

==> $$\lim_{\Delta t \to 0} \frac{\Delta X}{\Delta t} = b(X,t) + \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t}$$

==> $$\frac{dX}{dt} = b(X,t)$$

We can work with derivatives, instead of differences.

Stochastic Difference Equation:

$$\Delta X = b(X,t)\Delta t + \sqrt{a(X,t)}\,\Delta W + o(\Delta t)$$

==> $$\lim_{\Delta t \to 0} \frac{\Delta X}{\Delta t} = b(X,t) + \sqrt{a(X,t)} \lim_{\Delta t \to 0} \frac{\Delta W}{\Delta t} + \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t}$$

Unfortunately $\frac{\Delta W}{\Delta t}$ does not exist as a regular function.

We have to work with differences and finite Δt.

The general approach:

In the discussion of stochastic differential equations, we work with finite dt. Then at the end, we take the limit as $dt \to 0$.

Properties of Wiener process:

1) $dW \sim N(0, dt)$ ==> $dW = \sqrt{dt}\, X$ where $X \sim N(0,1)$
2) $\mathrm{E}(dW) = 0$
3) $\mathrm{E}((dW)^2) = dt$
4) $dW(t_1)$ and $dW(t_2)$ are independent if the time intervals are disjoint.
5) $dW = O(\sqrt{dt})$ in the statistical sense.

   The RMS (root mean square) of dW is

$$\mathrm{RMS}(dW) = \sqrt{E\left((dW)^2\right)} = \sqrt{dt}$$

A discrete version of $W(t)$

The discrete version is conceptually easy to understand, and is computationally practical to work with in simulations.

Consider $W(t)$ on a grid over time interval $[0, t_f]$.

Grid points: $\left\{(j\Delta t),\quad j=0,1,\ldots,n\right\},\qquad \Delta t=\dfrac{t_f}{n}$

$W(t)$ on the grid: $\left\{W_j=W(j\Delta t),\quad j=0,1,\ldots,n\right\}$

<u>Question:</u> How to generate a discrete sample path $\{W_j, j = 0, 1, \ldots, n\}$?

<u>Answer:</u> By the definition of $W(t)$, we have

$$\Delta W_j \equiv (W_{j+1}-W_j)=\sqrt{\Delta t}\,X_j\,,\qquad X_j\sim N(0,1)\,,\qquad j=0,1,\ldots,n-1$$

ΔW_j and ΔW_k are independent for $j \neq k$.

<u>Method:</u>

Generate n independent samples of $N(0, 1)$.

$$\left\{X_j,\ j=0,1,\ldots,n-1\right\}\sim \text{(iid)}\,N(0,1)$$

(In Matlab, "randn(1, n)" generates n independent samples of $N(0, 1)$.)

Calculate $\{W_j, j = 0, 1, \ldots, n\}$ as a cumulative sum.

$$W_0=0\,,\qquad W_j=\sqrt{\Delta t}\sum_{k=0}^{j-1}X_k\,,\qquad j=1,2,\ldots,n$$

(In Matlab, "cumsum(X)" calculates the cumulative sum of array X.)

<u>Remarks:</u>

- This method completely specifies the random experiment for generating a discrete sample path $\{W_j, j = 0, 1, \ldots, n\}$.
- On the grid, discrete sample $\{W_j, j = 0, 1, \ldots, n\}$ is exactly the same as the underlying full sample path $W(t)$ (i.e., no approximation error).
- Given a coarse-grid sample $\{W_j, j = 0, 1, \ldots, n\}$, it is desirable to refine it to obtain a fine-grid sample $\{W_k, k = 0, 1, \ldots, 2n\}$ of the same underlying full sample path $W(t)$. The problem of refining a given discrete sample path will be discussed after introducing Bayes' theorem.

If you think $W(t)$ is somewhat unusual and different from the functions we are familiar with, it indeed is. Below we illustrate one peculiar feature of $W(t)$.

<u>A peculiar feature of $W(t)$</u>:

The arc length of $W(t)$ over $[0, t_f]$ is infinity.

<u>Derivation:</u>

We start with the arc length of discrete sample path $\{W_j, j = 0, 1, 2, \ldots\}$.

$$\text{Discrete arc length} = \sum_{j=0}^{n-1}\left|(t_{j+1}, W_{j+1}) - (t_j, W_j)\right| = \sum_{j=0}^{n-1}\left|(\Delta t,\ \Delta W_j)\right|$$

$$= \sum_{j=0}^{n-1}\sqrt{(\Delta t)^2 + (\sqrt{\Delta t}\,X_j)^2}\,, \qquad \{X_j\} \sim \text{iid } N(0,1)$$

$$> \sum_{j=0}^{n-1}\sqrt{\Delta t}\,|X_j| = n\sqrt{\Delta t}\left(\frac{1}{n}\sum_{j=0}^{n-1}|X_j|\right)$$

$$\text{Use } \Delta t = \frac{t_f}{n} \quad \text{and} \quad \frac{1}{n}\sum_{j=0}^{n-1}\left|X_j\right| \approx E(|X|) = \sqrt{\frac{2}{\pi}}$$

$$= \sqrt{n t_f}\sqrt{\frac{2}{\pi}} \ \to +\infty \qquad \text{as } n \to +\infty$$

Here we used $E(|X|) = \sqrt{\frac{2}{\pi}}$ for $X \sim N(0, 1)$ (Homework problem)

Therefore, we conclude that the arc length of $W(t)$ over $[0, t_f]$ is infinity!

<u>Ito's lemma</u>:

Suppose $f(t, w)$ is a smooth function of two variables t and w.

Replacing w by $W(t)$ gives us $f(t, W(t))$, a non-smooth random function of single variable t. The randomness comes from the Wiener process $W(t, \omega)$.

We examine the increment of $f(t, W(t))$ corresponding to dt.

$$df(t, W(t)) \equiv f(t + dt, W + dW) - f(t, W).$$

First we expand $f(t, w)$ as a smooth 2-variable function.

$$f(t+dt, w+dw) = f(t,w) + f_t dt + f_w dw$$
$$+\frac{1}{2}\left[f_{tt}(dt)^2 + 2f_{tw}(dt)(dw) + f_{ww}(dw)^2\right]$$
$$+O\left((dt)^3 + (dt)^2(dw) + (dt)(dw)^2 + (dw)^3\right)$$

We apply the expansion to $f(t+dt, W+dW)$, use $dW = O(\sqrt{dt})$, and neglect $o(dt)$ terms.

$$df(t, W(t)) = f_t dt + f_w dW + \frac{1}{2} f_{ww}(dW)^2 + o(dt) \qquad \text{(E01)}$$

<u>Claim:</u> we can replace $(dW)^2$ with dt and write df as

$$df(t,W(t)) = f_t dt + f_w dW + \frac{1}{2} f_{ww} dt + o(dt) = \left(f_t + \frac{1}{2} f_{ww} \right) dt + f_w dW + o(dt)$$

<u>Theorem (Ito's lemma):</u>

Given $f(0, 0)$, at any $t_f > 0$, the two SDEs below give the same $f(t_f, W(t_f))$.

$$df(t,W(t)) = f_t dt + f_w dW + \frac{1}{2} f_{ww} (dW)^2 + o(dt)$$

$$df(t,W(t)) = \left(f_t + \frac{1}{2} f_{ww} \right) dt + f_w dW + o(dt)$$

<u>Outline of proof:</u>

Let $dt = t_f/n$ and $t_j = jdt$. We calculate $\{f(t_j, W(t_j)), j = 1, 2, \ldots, n\}$ sequentially.

In one step of dt, the error of replacing $(dW)^2$ with dt is

$$\text{err}_j = \frac{1}{2} f_{ww} \left((dW_j)^2 - dt \right), \qquad dW_j = \sqrt{dt}\, X_j\,, \quad X_j \sim N(0,1)$$

The total error at t_f is

$$\text{err}_{\text{tot}} = \sum_{j=0}^{n-1} \text{err}_j$$

In the simple case of $f_{ww} \equiv 2$, we have

$$E(\text{err}_j) = E\left((dW_j)^2 - dt \right) = 0$$

$$\text{var}(\text{err}_j) = \text{var}\left((dW_j)^2 \right) = 2(dt)^2 \qquad \text{(Homework problem)}$$

Since $\{dW_j, j = 0, 1, 2, \ldots\}$ are independent, we obtain

$$E(\text{err}_{\text{tot}}) = \sum_{j=0}^{n-1} E(\text{err}_j) = 0$$

$$\text{var}(\text{err}_{\text{tot}}) = \sum_{j=0}^{n-1} \text{var}(\text{err}_j) = 2n(dt)^2 = 2t_f(dt) \to 0 \quad \text{as } dt \to 0$$

We will look at related materials in subsequent lectures/assignments.

<u>The mean-value version of Ito's lemma:</u>

The mean of $df(t, W(t))$ can be calculated exactly using $E(dW)=0$ and $E((dW)^2)=dt$.

$$E_{dW}\left(f(t+dt, W+dW) \right) = f(t,W) + f_t dt + \frac{1}{2} f_{ww} dt + o(dt)$$

We will use this version of Ito's lemma to study the Gambler's ruin problem.

Another version of law of total probability

We start with a unified view of probability and expectation.

Key observation: The probability of an event can be written in terms of the expectation of a random variable.

Given event A, we define random variable X as

$$X(\omega)=\begin{cases}1, & \text{if } \omega\in A\\ 0, & \text{otherwise}\end{cases}$$

==> $\Pr(A) = E(X)$

$E(X) = E(E(X|Y))$ (The law of total expectation)

==> $\boxed{\Pr(A) = E(\Pr(A|Y))}$ (The law of total probability)

Gambler's ruin (applications of Ito's lemma)

Notation and modeling approach:

C: total cash = the sum of your cash and casino's cash (assuming you are the only one playing with the casino).

$X(t)$: your cash at time t. In practice, $C \gg X(0)$.

"Breaking the bank" means "$X(t)$ hits C before hitting 0".

Case 1: we first consider a fair game

$$dX = dW$$

which means $X(t+dt)=X(t)+dW$

It is a fair game because

$$E_{dW}(dX)=E(dW)=0$$

We study the two questions below.

Question #1: How long can you play?

Question #2: What is the chance that you break the bank?

Answer to Question #2 (we address Question #1 after this)

Let $u(x)=\Pr\left(A\middle| X(0)=x\right)$, $A\equiv\{X(t) \text{ hits } C \text{ before } 0\}$.

Strategy:

Find a boundary value problem (BVP) governing $u(x)$.

Boundary condition:

$u(C) = 1$ and $u(0) = 0$.

Differential equation:

Start with $X(0) = x \in (0, C)$. After a short time dt, we have

$X(dt) = x + dW$

Recall that $dW = O(\sqrt{dt})$. For a fixed $x \in (0, C)$, when dt is small enough, the probability of $X(t)$ hitting 0 or C in time interval $[0, dt]$ is exponentially small. Here the magnitude of dt depends on how close x is to the two boundaries.

For a fixed $x \in (0, C)$, when dt is small enough (depending on x), we have

$$u(x) = \Pr(A) = E(\underbrace{\Pr(A|X(dt) = x + dW)}_{u(x+dW)}) + o(dt), \qquad A = \{X(t) \text{ hits } C \text{ before } 0\}$$

$$= E_{dW}\left(u(x+dW)\right) + o(dt)$$

Here we used the law of total probability $\boxed{\Pr(A) = E(\Pr(A|Y))}$ (draw a diagram).

Expanding $u(x+dW)$ inside E(), we get

$$u(x) = E_{dW}\left(u(x) + u_x dW + \frac{1}{2}u_{xx}(dW)^2\right) + o(dt)$$

$$= u(x) + \frac{1}{2}u_{xx}dt + o(dt)$$

Divide by dt and then take the limit as $dt \to 0$, we obtain

$$u_{xx} = 0$$

This is the differential equation governing $u(x)$. Thus, function $u(x)$ satisfies the boundary value problem (BVP)

$$\begin{cases} u_{xx}(x) = 0 & \text{differential equation} \\ u(0) = 0, \quad u(C) = 1 & \text{boundary conditions} \end{cases}$$

Solving the differential equation: $u(x) = c_1 + c_2 x$

Enforcing the boundary conditions: $u(x) = \dfrac{x}{C}$

The probability of breaking the bank is proportional to your initial cash and inversely proportional to the total cash.

Answer to Question #1

Let $T(x) = E\left(Z \middle| X(0) = x\right), \quad Z \equiv \left(\text{time from 0 until } X(t) = C \text{ or } X(t) = 0\right)$

Strategy:

Find a boundary value problem (BVP) governing $T(x)$.

Boundary condition:

$T(0) = 0$ and $T(C) = 0$.

Differential equation:

Start with $X(0) = x \in (0, C)$. After a short time dt, we have

$X(dt) = x + dW$

For a fixed $x \in (0, C)$, when dt is small enough (depending on x), we have

$$T(x) = E(Z) = E(E(Z \mid X(dt) = x + dW)) + o(dt), \quad Z = \binom{\text{time from 0 until}}{X(t)=C \text{ or } X(t)=0}$$

$$= E(\underbrace{E((Z+dt) \mid X(0) = x + dW)}_{T(x+dW)}) = dt + E_{dW}\left(T(x+dW)\right) + o(dt)$$

Here we used the law of total expectation $\boxed{E(Z) = E(E(Z|Y))}$ (draw a diagram).

Expanding $T(x+dW)$ inside E(), we get

$$T(x) = dt + E_{dW}\left(T(x) + T_x dW + \frac{1}{2}T_{xx}(dW)^2\right) + o(dt)$$

$$= dt + T(x) + \frac{1}{2}T_{xx}dt + o(dt)$$

Divide by dt and then take the limit as $dt \to 0$, we obtain

$$T_{xx} = -2$$

This is the differential equation governing $T(x)$. Thus, function $T(x)$ satisfies the boundary value problem (BVP)

$$\begin{cases} T_{xx}(x) = -2 & \text{differential equation} \\ T(0) = 0, \quad T(C) = 0 & \text{boundary conditions} \end{cases}$$

A particular solution of DE: $T(x) = -x^2$

The general solution of DE: $T(x) = c_1 + c_2 x - x^2$

Enforcing the BCs: $T(x) = x(C - x)$

Remark:

The average does not give us the full picture!

$T(x)$ is the average time until going bankrupt or breaking the bank. However, this average does not give us the full picture of how long we can play with initial cash x.

In particular, when $C = \infty$ (when the casino has infinite amount of cash), we have

$T(x) = x(C-x) = \infty$.

This certainly does not mean we can play forever with initial cash x.

<u>A more detailed answer to Question #1:</u>

We look at the probability of surviving beyond time t.

Assume $C = \infty$. We consider a function of two variables

$$P(x,t) = \Pr\Big(A(t)\,\Big|\,X(0)=x\Big), \qquad A(t) \equiv \Big\{X(\tau)>0 \text{ for } \tau\in[0,t]\Big\}$$

$P(x, t)$ is the conditional probability of surviving beyond time t given $X(0) = x$.

<u>Strategy:</u>

Find an initial boundary value problem (IBVP) governing $P(x, t)$.

<u>Initial and boundary conditions:</u>

Initial condition: $P(x, 0) = 1$ for $x > 0$

(with $x > 0$, we can certainly survive beyond time 0)

Boundary condition: $P(0, t) = 0$ for $t > 0$

(with $x = 0$, we cannot survive beyond time 0)

<u>Differential equation:</u>

Start with $X(0) = x > 0$. After a short time dt, we have

$X(dt) = x + dW$

For a fixed $x > 0$, when dt is small enough (depending on x), we have

$$P(x,t) = \Pr(A(t)) = E(\Pr(A(t)|X(dt)=x+dW)) + o(dt), \quad A(t) = \Big\{X(\tau)>0 \text{ for } \tau\in[0,t]\Big\}$$

$$= E(\underbrace{\Pr(A(t-dt)|X(0)=x+dW)}_{P(x+dW,t-dt)}) + o(dt) = E_{dW}\Big(P(x+dW,t-dt)\Big) + o(dt)$$

Here we used the law of total probability $\boxed{\Pr(A) = E(\Pr(A|Y))}$ (draw a diagram).

Expanding $P(x+dW, t-dt)$ inside E(), we get

$$P(x,t) = E_{dW}\left(P(x,t) + P_t(-dt) + P_x dW + \frac{1}{2}P_{xx}(dW)^2\right) + o(dt)$$

$$= P(x,t) + P_t(-dt) + \frac{1}{2}P_{xx}dt + o(dt)$$

Divide by dt and then take the limit as $dt \to 0$, we obtain

$$P_t = \tfrac{1}{2}P_{xx}$$

This is the PDE governing $P(x, t)$. Thus, function $P(x, t)$ satisfies the initial boundary value problem (IBVP)

$$\begin{cases} P_t = \frac{1}{2} P_{xx} & \text{partial differential equation} \\ P(0,t) = 0 & \text{boundary condition} \\ P(x,0) = 1 & \text{initial condition} \end{cases}$$

We use the odd extension to convert it to an IVP. The odd extension satisfies the zero-value boundary condition automatically.

Odd extension:

$P(-x, t) = -P(x, t)$

The extended function $P(x, t)$ is governed by the IVP

$$\begin{cases} P_t = \frac{1}{2} P_{xx} \\ P(x,0) = \begin{cases} 1, & x>0 \\ -1, & x<0 \end{cases} \end{cases} \qquad \text{(E02)}$$

Solution of a general IVP of the heat equation:

$$\begin{cases} u_t = a u_{xx} \\ u(x,0) = f(x) \end{cases} \qquad \text{(E03)}$$

The solution of IVP (E03) has the expression:

$$u(x,t) = \frac{1}{\sqrt{4\pi a t}} \int_{-\infty}^{+\infty} \exp\left(\frac{-\xi^2}{4at} \right) f(x-\xi) d\xi$$

Solution of IVP (E02).

Applying the general formula to (E02), we identify

$$a = \frac{1}{2}, \qquad f(x-\xi) = \begin{cases} 1, & \xi < x \\ -1, & \xi > x \end{cases}$$

We write out $P(x, t)$, the solution of (E02).

$$P(x,t) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \exp\left(\frac{-\xi^2}{2t} \right) f(x-\xi) d\xi$$

$$= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{x} \exp\left(\frac{-\xi^2}{2t} \right) d\xi - \frac{1}{\sqrt{2\pi t}} \int_{x}^{\infty} \exp\left(\frac{-\xi^2}{2t} \right) d\xi$$

$$= \frac{2}{\sqrt{2\pi t}} \int_{0}^{x} \exp\left(\frac{-\xi^2}{2t} \right) d\xi$$

Change of variables: $\xi = \sqrt{2t}\, s$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{2t}}} \exp(-s^2) ds = \operatorname{erf}\left(\frac{x}{\sqrt{2t}} \right)$$

Thus, probability $P(x, t)$ has the expression

$$P(x,t) = \operatorname{erf}\left(\frac{x}{\sqrt{2t}} \right)$$

Scaling property of $P(x, t)$:

Start with initial cash x. The survival probability p and the time t are related by

$$p = \operatorname{erf}\left(\frac{x}{\sqrt{2t}} \right)$$

==> $$\frac{x}{\sqrt{2t}} = \operatorname{erfinv}(p)$$

==> $$t = \frac{x^2}{2\operatorname{erfinv}(p)^2}$$

Given a prescribed threshold p, the maximum time t with surviving probability $\geq p$ is proportional to x^2 with the coefficient depending on p.

A few example values of the coefficient:

$p = 0.1$ ==> $t = 63.33\, x^2$

$p = 0.3$ ==> $t = 6.735\, x^2$

$p = 0.5$ ==> $t = 2.198\, x^2$

$p = 0.7$ ==> $t = 0.931\, x^2$

$p = 0.9$ ==> $t = 0.370\, x^2$

AM216 Stochastic Differential Equations

Lecture 04 Supplemental
Copyright by Hongyun Wang, UCSC

Computational demonstration of Ito's lemma

The exact and two approximations of $f(t_f, W(t_f))$

Let $f(t, w)$ be a smooth function of (t, w). Replacing w with $W(t)$ yields $f(t, W(t))$, a non-smooth random function of t. We select a numerical grid on t.

$$\Delta t = \frac{t_f}{N}, \qquad t_j = j\Delta t, \qquad W_j = W(t_j), \qquad \Delta W_j = W_{j+1} - W_j$$

Consider $\Delta f_j \equiv f(t_{j+1}, W(t_{j+1})) - f(t_j, W(t_j))$, the increment of $f(t, W(t))$. We study two approximations of Δf_j.

- Approximate increment based on Taylor expansion

$$\left(\Delta f_{\text{Taylor}}\right)_j = (f_t)_j \Delta t + (f_w)_j \Delta W_j + \frac{1}{2}(f_{ww})_j (\Delta W_j)^2$$

- Approximate increment based on Ito's lemma

$$\left(\Delta f_{\text{Ito}}\right)_j = (f_t + \frac{1}{2} f_{ww})_j \Delta t + (f_w)_j \Delta W_j$$

These two approximations of Δf_j correspond to two stochastic differential equations.

Given the initial value $f(0, 0)$, using the two approximations of Δf_j, we calculate the corresponding two approximations of $f(t_f, W(t_f))$.

- Approximation based on Taylor expansion

$$\left(f_{\text{Taylor}}\right)\Big|_{t_f} = f(0,0) + \sum_{j=0}^{N-1} \left(\Delta f_{\text{Taylor}}\right)_j$$

- Approximation based on Ito's lemma

$$\left(f_{\text{Ito}}\right)\Big|_{t_f} = f(0,0) + \sum_{j=0}^{N-1} \left(\Delta f_{\text{Ito}}\right)_j$$

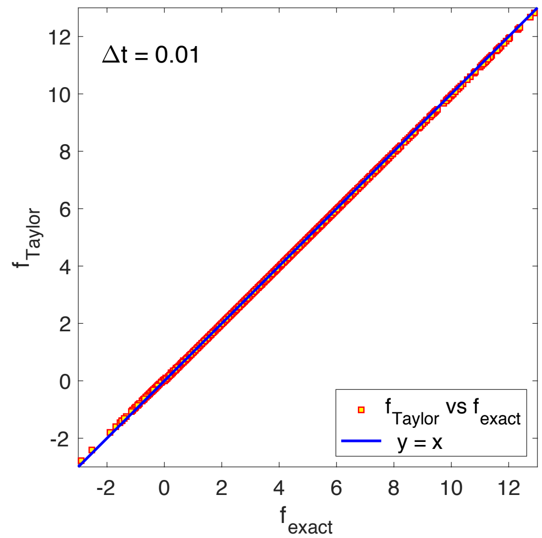
Taylor's theorem implies $\quad \lim\limits_{dt \to 0} \left(f_{\text{Taylor}}\right)\Big|_{t_f} = \left(f_{\text{exact}}\right)\Big|_{t_f}$.

Ito's lemma tells us that $\quad \lim\limits_{dt \to 0} \left(f_{\text{Ito}}\right)\Big|_{t_f} = \left(f_{\text{exact}}\right)\Big|_{t_f}$

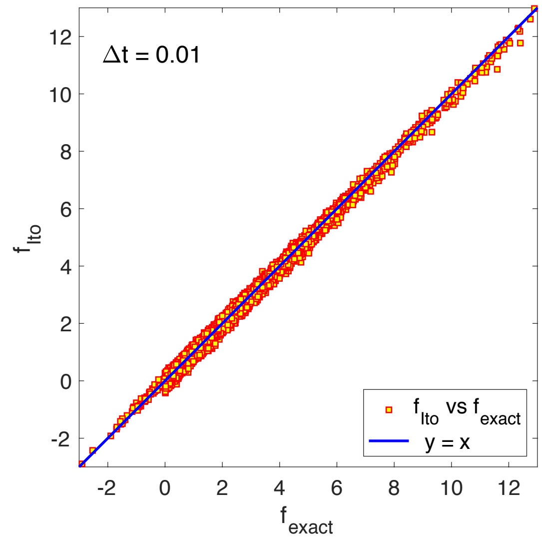
Below we use numerical simulations to confirm these two assertions.

Numerical example:

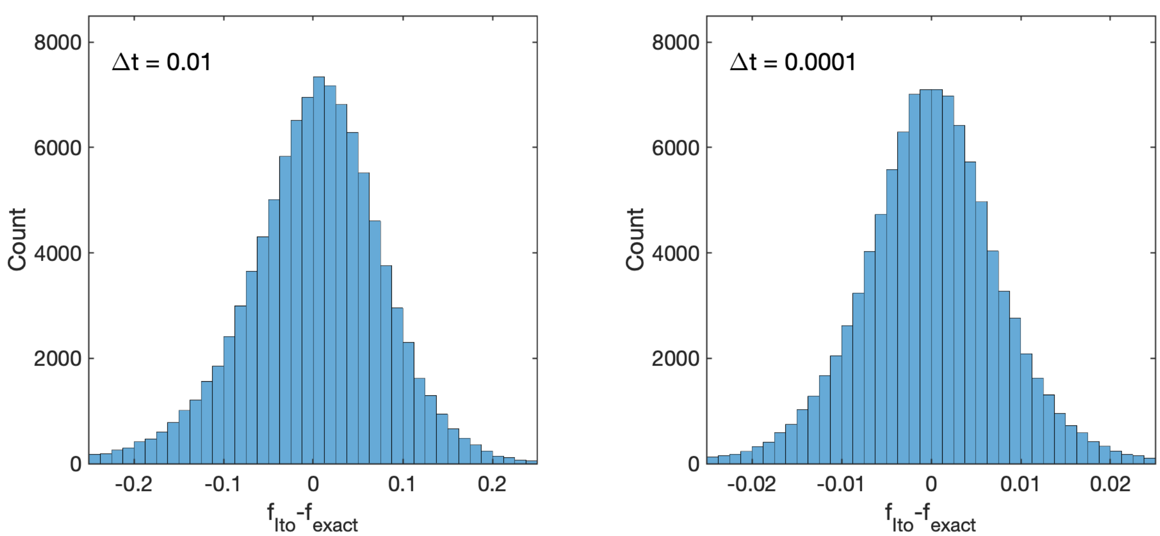
We test Ito's lemma on $f(t,w)=\dfrac{1}{2}w^2+\dfrac{t}{6}w^3$. In simulations, we use $t_f = 1$ and $\Delta t = 0.01$, and we generate 100,000 independent samples of $(f_{\text{exact}}, f_{\text{Taylor}}, f_{\text{Ito}})$.



f_{Taylor} vs f_{exact}



f_{Ito} vs f_{exact}



Histogram of $(f_{\text{Ito}} - f_{\text{exact}})$ for $\Delta t = 0.01$ (left) and $\Delta t = 0.0001$ (right).

Multivariate Normal Distribution

1 Definition of multivariate normal

Recall that a random variable is completely described by its probability density function (PDF). $X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^n$ is a multivariate normal random variable if its PDF is

$$\rho_X(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

where $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is the independent variable (vector) of the PDF, $\mu = (\mu_j) \in \mathbb{R}^n$ is the mean vector, and $\Sigma = (\sigma_{ij}) \in \mathbb{S}^n_{++}$ is the covariance matrix. Here $\mathbb{S}^n_{++}$ represents the set of all real symmetric positive definite matrices. We need to justify several items.

- We need to connect it to the 1D normal distribution we are familiar with.
- We need to justify the name of density: $\displaystyle\int_{\mathbb{R}^n} \rho_X(x; \mu, \Sigma)dx = 1$.
- We need to justify the names of mean vector and covariance matrix.

$$E(X_j) = \mu_j, \qquad E\Big((X_i - \mu_i)(X_j - \mu_j)\Big) = \sigma_{ij}$$

2 Connection to 1D independent Gaussians

Review of linear algebra

A real square matrix U is called orthogonal if $U^T U = UU^T = I$.
For an orthogonal matrix $U \in O(n)$, we have $U^{-1} = U^T$ and $(U^T)^{-1} = U$.
Any real symmetric matrix A is orthogonally diagonalizable. That is, for $A \in \mathbb{S}^n$, there exists an orthogonal matrix $U \in O(n)$ such that

$$A = U\Lambda U^T, \qquad \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

Meaning of PDF

Recall the connection between PDF and probability in an infinitesimal region.

$$\Pr(X \in \delta V) = \int_{\delta V} \rho_X(x)dx \approx \text{Vol}(\delta V)\rho_X(x)$$

Here $\text{Vol}(\delta V)$ is the volume of δV.

PDF of a transformed X

Since Σ is symmetric and positive definite, we write Σ and Σ^{-1} as

$$\begin{aligned}\Sigma &= U\Lambda U^T, \qquad \Lambda = \mathrm{diag}(d_1^2, d_2^2, \ldots, d_n^2)\\ \Sigma^{-1} &= U\Lambda^{-1}U^T\end{aligned}$$

Note that since $\Sigma \in \mathbb{S}_{++}^n$ (positive definite), we can write eigenvalues as $\{d_j^2\}$. Let $Y \equiv U^T(X-\mu)$ where U is from the diagonalization of Σ. We write $X = UY + \mu$ and

$$\begin{aligned}\Pr(Y \in \delta V) &= \Pr\Big(X \in U(\delta V){+}\mu\Big)\\ \mathrm{Vol}(\delta V)\rho_Y(y) &= \mathrm{Vol}(U\delta V{+}\mu)\,\rho_X(x)\Big|_{x=Uy+\mu}\end{aligned}$$

Note that the volume is invariant under a rigid body transformation. We obtain

$$\begin{aligned}&\mathrm{Vol}(U\delta V{+}\mu) = \mathrm{Vol}(\delta V)\\ &\rho_Y(y) = \rho_X(x)\Big|_{x=Uy+\mu} = \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}}\exp\Big(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)\Big|_{x=Uy+\mu}\end{aligned}$$

In the PDF of Y above, we have

$$\begin{aligned}&\det\Sigma = (\det U)(\det\Lambda)(\det U^T) = \det\Lambda = \prod_{j=1}^{n} d_j^2\\ &(x-\mu)^T\Sigma^{-1}(x-\mu)\Big|_{x=Uy+\mu} = (Uy)^T U\Lambda^{-1}U^T(Uy) = y^T\Lambda^{-1}y = \sum_{j=1}^{n}\frac{y_j^2}{d_j^2}\end{aligned}$$

Using these results, we write out the PDF of Y.

$$\rho_Y(y) = \frac{1}{(2\pi)^{n/2}(\det\Lambda)^{1/2}}\exp\Big(\frac{-1}{2}y^T\Lambda^{-1}y\Big) = \prod_{j=1}^{n}\frac{1}{\sqrt{2\pi d_j^2}}\exp\Big(\frac{-y_j^2}{2d_j^2}\Big)$$

This is a product of n functions, each a 1D normal density. We conclude

$$\begin{aligned}&Y \sim N(0,\Lambda) \in \mathbb{R}^n, \qquad \Lambda = \mathrm{diag}(d_1^2, d_2^2, \ldots, d_n^2)\\ &Y_j \sim N(0, d_j^2) \in \mathbb{R}, \quad Y_i \text{ and } Y_j \text{ are independent } (i \neq j).\end{aligned}$$

This leads to $\displaystyle\int_{\mathbb{R}^n}\rho_X(x;\mu,\Sigma)dx = \int_{\mathbb{R}^n}\rho_Y(y)dy = 1$, which justifies the name of density.

Standard isotropic normal

$Z \sim N(0, I_n) \in \mathbb{R}^n$ is called the standard isotropic normal, in which

$$Z_j \sim N(0,1) \in \mathbb{R}, \quad Z_i \text{ and } Z_j \text{ are independent } (i \neq j).$$

In terms of standard isotropic normal, we write $Y \equiv U^T(X-\mu) \sim N(0,\Lambda)$ as

$$Y = \Lambda^{1/2}Z, \qquad Z \sim N(0, I_n), \qquad \Lambda^{1/2} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$$

Finally, we write X in terms of standard isotropic normal: $X = U\Lambda^{1/2}Z + \mu$.

Theorem 1. *(Multivariate Gaussian as an affine mapping of standard isotropic normal)*
For $X \sim N(\mu, \Sigma) \in \mathbb{R}^n$, we can write is as

$$\boxed{X = U\Lambda^{1/2}Z + \mu, \qquad Z \sim N(0, I_n), \qquad \Sigma = U\Lambda U^T}$$

We make a few observations:

- Any multivariate normal $X \sim N(\mu, \Sigma)$ can be viewed as an affine mapping of a standard isotropic normal Z.
- This makes sense even when $\Sigma \in \mathbb{S}^n_+$ (when it is only positive <u>semi</u>-definite). When $d_j = 0$, we simply take the limit as $d_j \to 0_+$; everything makes sense.

3 Partition function and a key result

$$\rho_X(x) \propto \exp(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)) \quad \longleftarrow \quad \text{energy form of density}$$

$$Z \equiv \int_{\mathbb{R}^n} \exp(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx \quad \longleftarrow \quad \text{definition of partition function}$$

Theorem 2. *(a key result on partition function)*

$$\boxed{Z \equiv \underbrace{\int_{\mathbb{R}^n} \exp(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx = (2\pi)^{n/2}(\det\Sigma)^{1/2}}_{\textit{key result}}}$$

This result is valid even when μ is a complex vector.

4 Characteristic function of a multivariate normal

For $X \sim N(\mu, \Sigma) \in \mathbb{R}^n$, its characteristic function (CF) is

$$\begin{aligned}\phi_X(\xi) &= E\Big(\exp(i\xi^T X)\Big), \qquad \xi \in \mathbb{R}^n \\ &= \frac{1}{Z}\int_{\mathbb{R}^n} \exp(i\xi^T x - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx\end{aligned}$$

In the exponent, we complete the square (homework).

$$\begin{aligned}& i\xi^T x - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \\ &= -\frac{1}{2}(x-\mu-i\Sigma\xi)^T\Sigma^{-1}(x-\mu-i\Sigma\xi) + \underbrace{(i\xi^T\mu - \frac{1}{2}\xi^T\Sigma\xi)}_{\text{does not contain } x}\end{aligned}$$

Apply the result of completing the square in the expression of CF, we obtain

$$\phi_X(\xi) = \underbrace{[\frac{1}{Z}\int_{\mathbb{R}^n} \exp(-\frac{1}{2}(x-\mu-i\Sigma\xi)^T\Sigma^{-1}(x-\mu-i\Sigma\xi))dx]}_{=1,\text{ from the key result}} \exp(i\xi^T\mu - \frac{1}{2}\xi^T\Sigma\xi)$$

Theorem 3. *(Characteristic function of multivariate Gaussian)*

$$\boxed{\begin{aligned} X \sim N(\mu, \Sigma) \quad &\Longleftrightarrow \quad \phi_X(\xi) = \exp(i\xi^T\mu - \frac{1}{2}\xi^T\Sigma\xi) \\ &\Longleftrightarrow \quad \phi_{(X-\mu)}(\xi) = \exp(-\frac{1}{2}\xi^T\Sigma\xi) \end{aligned}}$$

Below, we use the expression of CF to derive other results.

5 Justifying the names of μ and Σ

We show $E(X_j) = \mu_j$ and $E\big((X_i - \mu_i)(X_j - \mu_j)\big) = \sigma_{ij}$.
Differentiating $\phi_{(X-\mu)}(\xi)$ with respect to ξ_j gives

$$E\big(i(X_j - \mu_j)\big) = \frac{\partial \phi_{(X-\mu)}(\xi)}{\partial \xi_j}\Big|_{\xi=0} = \frac{\partial \exp(-\frac{1}{2}\xi^T\Sigma\xi)}{\partial \xi_j}\Big|_{\xi=0} = 0$$

$$\Longrightarrow \quad E(X_j) = \mu_j$$

Differentiating $\phi_{(X-\mu)}(\xi)$ with respect to ξ_i and ξ_j leads to

$$E\big(-(X_i - \mu_i)(X_j - \mu_j)\big) = \frac{\partial^2 \phi_{(X-\mu)}(\xi)}{\partial \xi_i \partial \xi_j}\Big|_{\xi=0} = \frac{\partial^2 \exp(-\frac{1}{2}\xi^T\Sigma\xi)}{\partial \xi_i \partial \xi_j}\Big|_{\xi=0} = -\sigma_{ij}$$

$$\Longrightarrow \quad E\big((X_i - \mu_i)(X_j - \mu_j)\big) = \sigma_{ij}$$

6 Affine mapping of a Gaussian

Theorem 4. *(An affine mapping of a Gaussian is a Gaussian)*
Let $X \sim N(\mu, \Sigma) \in \mathbb{R}^n$. Consider $Y \equiv AX + b$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. We have

$$\boxed{Y \sim N(\mu_Y, \Sigma_{YY}), \quad \mu_Y = A\mu + b, \quad \Sigma_{YY} = A\Sigma A^T}$$

Proof. We write $Y = A(X-\mu) + A\mu+b$ and find its CF.

$$\phi_Y(\xi) = E\Big(\exp(i\xi^T Y)\Big) = E\Big(\exp[i\xi^T A(X-\mu) + i\xi^T(A\mu+b)]\Big), \qquad \xi \in \mathbb{R}^m$$

$$= \exp[i\xi^T(A\mu+b)]E\Big(\exp[i(A^T\xi)^T(X-\mu)]\Big), \qquad \tilde{\xi} \in \mathbb{R}^n$$

$$= \exp[i\xi^T(A\mu+b)]\phi_{(X-\mu)}(\tilde{\xi})\Big|_{\tilde{\xi}=A^T\xi} = \exp[i\xi^T(A\mu+b)]\exp(-\frac{1}{2}\tilde{\xi}^T\Sigma\tilde{\xi})\Big|_{\tilde{\xi}=A^T\xi}$$

$$= \exp[i\xi^T \underbrace{(A\mu+b)}_{\mu_Y} - \frac{1}{2}\xi^T \underbrace{(A\Sigma A^T)}_{\Sigma_Y Y}\xi] = \exp[i\xi^T\mu_Y - \frac{1}{2}\xi^T\Sigma_{YY}\xi]$$

Since the CF is reversible, we conclude $Y \sim N(\mu_Y, \Sigma_{YY})$. □

Special case 4.1 (Sum of independent Gaussians). Let

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix}), \qquad X, Y \in \mathbb{R}^n$$

Then we have

$$(X+Y) \sim N(\mu_X + \mu_Y,\ \Sigma_{XX} + \Sigma_{YY})$$

Derivation: In Theorem 4, pick $A = [I \ \ I]$ and $b = 0$.

$$A \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \mu_X + \mu_Y, \qquad A \begin{bmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix} A^T = \Sigma_{XX} + \Sigma_{YY}$$

Special case 4.2 (Marginal distribution of Gaussian). Let

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}), \qquad X \in \mathbb{R}^m, \quad Y \in \mathbb{R}^n$$

Here m and n may be different. Then we have

$$X \sim N(\mu_X,\ \Sigma_{XX}), \qquad Y \sim N(\mu_Y,\ \Sigma_{YY})$$

Derivation: In Theorem 4, pick $A = [I \ \ 0]$ and $b = 0$.

$$A \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \mu_X, \qquad A \begin{bmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix} A^T = \Sigma_{XX}$$

Special case 4.3 (Independent Gaussians based on the standard isotropic normal).

Let $A \in \mathbb{R}^{m\times n}$ be a matrix with orthogonal rows. In the matrix form, A satisfies

$$AA^T = \Lambda = \text{diag}(d_1^2, d_2^2, \ldots, d_m^2), \qquad d_i = \|a_{i,:}\|$$

Here we do not require $\|a_{i,:}\| = 1$. Then for $Z \sim N(0, I_n)$, we have

$$X = AZ \sim N(0, \Lambda), \qquad \Lambda = \text{diag}(d_1^2, d_2^2, \ldots, d_m^2)$$

That is, the components of $X = AZ$ are independent Gaussians. This result is practically useful.

7 Conditional distribution of Gaussian

Theorem 5. *(Conditional distribution of X when Y is fixed). Let*

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}), \qquad X \in \mathbb{R}^m, \quad Y \in \mathbb{R}^n$$

Here m and n may be different. Then we have

$$\boxed{\begin{aligned} &(X|Y=y) \sim N(\mu_{X|Y},\ \Sigma_{X|Y}) \\ &\qquad \mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y) \\ &\qquad \Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{aligned}}$$

Proof. For finding the conditional distribution, the characteristic function is not very helpful. We work directly with density. The conditional density of $(X|Y=y)$ is

$$\rho_{(X|Y=y)}(x) = \frac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)} \propto \rho_{(X,Y)}(x,y)$$

$$\propto \exp(\frac{-1}{2}[(x-\mu_X)^T \ \ (y-\mu_Y)^T] \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix})$$

Note that we examine $\rho_{(X|Y=y)}(x)$ as a function of x. The denominator $\rho_Y(y)$ is independent of x and is viewed as a part of the normalizing factor. To proceed, we write Σ^{-1} as

$$\begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}^{-1} = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

where the needed properties of A, B and C are to be determined. The full expressions of A, B and C are neither necessary nor sufficient! We write $\rho_{(X|Y=y)}(x)$ as

$$\rho_{(X|Y=y)}(x) \propto \exp(\frac{-1}{2}[(x-\mu_X)^T \ \ (y-\mu_Y)^T] \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix})$$

$$\propto \exp(\frac{-1}{2}[(x-\mu_X)^T A(x-\mu_X) + 2(x-\mu_X)^T B(y-\mu_y)])$$

Again, any term independent of x in the exponent contributes only to the normalizing factor. In the exponent, we complete the square (homework).

$$(x-\mu_X)^T A(x-\mu_X) + 2(x-\mu_X)^T B(y-\mu_y)$$

$$= (x-\mu_X + A^{-1}B(y-\mu_Y))^T A(x-\mu_X + A^{-1}B(y-\mu_Y)) + \underbrace{G(y)}_{\text{does not contain } x}$$

Apply the result of completing the square in conditional density, we obtain

$$\rho_{(X|Y=y)}(x) \propto \exp\Big(\frac{-1}{2}(x-\mu_X + A^{-1}B(y-\mu_Y))^T (A^{-1})^{-1}(x-\mu_X + A^{-1}B(y-\mu_Y))\Big)$$

$$\Longrightarrow \quad (X|Y=y) \sim N(\mu_{X|Y},\, \Sigma_{X|Y}), \quad \mu_{X|Y} = \mu_X - A^{-1}B(y-\mu_Y), \quad \Sigma_{X|Y} = A^{-1}$$

Lemma 1. *(expression of $A^{-1}B$ and A^{-1})*

$$\begin{cases} A^{-1}B = -\Sigma_{XY}\Sigma_{YY}^{-1} \\ A^{-1} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{cases}$$

The proof of Lemma is in your homework.
Substituting the result of Lemma into the expression of $\mu_{X|Y}$ and $\Sigma_{X|Y}$, we obtain

$$\begin{cases} \mu_{X|Y} = \mu_X - A^{-1}B(y-\mu_Y) = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y-\mu_Y) \\ \Sigma_{X|Y} = A^{-1} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{cases}$$

This concludes the proof of Theorem 5. □

Sometimes we can use an ad hoc way to find $(X|Y)$

An ad hoc method (conditional distribution of combinations of standard isotropic normal)

Let $Z \sim N(0, I_n)$. We have

- $(a_1 Z_1 + a_2 Z_2)$ and $(a_2 Z_1 - a_1 Z_2)$ are independent.

$$X \equiv \begin{bmatrix} (a_2 Z_1 - a_1 Z_2) \\ (a_1 Z_1 + a_2 Z_2) \end{bmatrix} = \begin{bmatrix} a_2 & -a_1 \\ a_1 & a_2 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N(0, \begin{bmatrix} a_1^2 + a_2^2 & 0 \\ 0 & a_1^2 + a_2^2 \end{bmatrix})$$

  Note that the matrix above has orthogonal rows. It follows that

$$\Big(a_2 Z_1 - a_1 Z_2 \Big| a_1 Z_1 + a_2 Z_2 = x_2\Big) \sim \Big(a_2 Z_1 - a_1 Z_2\Big)$$

- Conditional distributions involving (Z_1, Z_2) are independent of $\{Z_j,\ j = 3, \ldots n\}$.

$$\Big(b_1 Z_1 + b_2 Z_2 \Big| a_1 Z_1 + a_2 Z_2 = x_2,\ Z_j = z_j,\ j = 3, \ldots n\Big)$$
$$\sim \Big(b_1 Z_1 + b_2 Z_2 \Big| a_1 Z_1 + a_2 Z_2 = x_2\Big)$$

- An example:

$$\Big(a_1 Z_1 \Big| a_1 Z_1 + a_2 Z_2 = x_2,\ Z_j = z_j,\ j = 3, \ldots n\Big) \sim \Big(a_1 Z_1 \Big| a_1 Z_1 + a_2 Z_2 = x_2\Big)$$
$$\sim \Big(a_1 \underbrace{\frac{1}{a_1^2 + a_2^2}\big[a_2(a_2 Z_1 - a_1 Z_2) + a_1(a_1 Z_1 + a_2 Z_2)\big]}_{Z_1} \Big| a_1 Z_1 + a_2 Z_2 = x_2\Big)$$
$$\sim \Big(\frac{a_1 a_2 X_1 + a_1^2 X_2}{a_1^2 + a_2^2} \Big| X_2 = x_2\Big), \qquad X_1 \sim N(0, a_1^2 + a_2^2)$$
$$\sim N(\frac{a_1^2 x_2}{a_1^2 + a_2^2}, \ \frac{a_1^2 a_2^2}{a_1^2 + a_2^2})$$

  In particular, for $a_1 = a_2 = a$ we have

$$\boxed{\Big(a Z_1 \Big| a Z_1 + a Z_2 = x_2\Big) \sim N(\frac{x_2}{2},\ \frac{a^2}{2})}$$

  This result is very useful in the discussion of constrained Wiener process.

AM216 Stochastic Differential Equations

Lecture 05

List of topics in this lecture

- Non-dimensionalization, advantage of working with $\dfrac{dW(t)}{\sqrt{dt}}$
- Gambler's ruin, Case 2: a biased game
- White noise dW/dt is not a regular function in the conventional sense
- Interpretation of the delta function: taking the limit after operations
- Fourier transform (FT), properties of FT

Recap

Properties of $W(t)$:

- $dW = O(\sqrt{dt})$,
- Arclength of $W(t)$ over finite time is infinity
- Ito's lemma: $(dW)^2$ can be replaced with dt.

The Gambler's ruin problem (Continued)

Case 2: a biased game

$$dX = -m dt + dW, \qquad m < 0$$

It is a biased game: $E_{dW}(dX) = -m dt < 0$; on average, you are losing money.

Before we solve case 2, let us study scaling and non-dimensionalization, which is a key component in modeling, analysis and simulations.

Scaling and non-dimensionalization

We look at the dimensions of various physical quantities

$[X]$ = \$, X and dX have the dimension of money

$[t]$ = time, t and dt have the dimension of time

$E(W^2) = t$

==> $[W]=[\sqrt{t}\,]=\sqrt{\text{time}}$, W and dW have the dimension of $\sqrt{\text{time}}$

The original physical equation of case 1 (before non-dimensionalization) should be

$$dX=\sqrt{\sigma^2}\,dW$$

where the dimension of σ is

$$[\sigma]=\frac{[dX]}{[dW]}=\frac{\$}{\sqrt{\text{time}}}.$$

The original physical equation of case 2 is

$$dX=-m\,dt+\sqrt{\sigma^2}\,dW \qquad \text{(E01)}$$

where

$$[X]=\$,\quad [dt]=\text{time},\quad [m]=\frac{\$}{\text{time}},\quad [\sigma]=\frac{\$}{\sqrt{\text{time}}},\quad [dW]=\sqrt{\text{time}}$$

We re-write the physical equation as

$$dX=-m\,dt+\sqrt{\sigma^2 dt}\,\frac{dW(t)}{\sqrt{dt}} \qquad \text{(E02)}$$

Advantage of working with $\dfrac{dW(t)}{\sqrt{dt}}$

$$dW(t)\sim N(0,dt)=\sqrt{dt}\,N(0,1)$$

==> $\dfrac{dW(t)}{\sqrt{dt}}\sim N(0,1)$ is dimensionless and independent of dt and t.

This property is especially useful in non-dimensionalization!

Caution: $\dfrac{dW(t)}{\sqrt{dt}}$ is not $\dfrac{dW(t)}{dt}$, which we will discuss later.

Notation for dimensionless

$$\left[\frac{X}{\$}\right]=\text{one},\qquad \left[\frac{dW}{\sqrt{dt}}\right]=\text{one}$$

Caution: "one" means "dimensionless". It does not mean numerical value 1.

$$\frac{\$8}{\$}=8,\qquad \left[\frac{\$8}{\$}\right]=\text{one}$$

Objectives of non-dimensionalizing (E02)

- A dimensionless equation
- Getting rid of parameter σ.

<u>Scales</u> for various physical quantities

<u>Time scale:</u> $[t_0]$ = time

In this problem, we select the time scale t_0, for example, t_0 = 1 minute.

<u>Money scale:</u> $\left[\sqrt{\sigma^2 t_0}\right] = \$$

The money scale is derived from the given σ and the selected time scale t_0.

<u>Non-dimensional quantities</u>

Non-dimensional time: $t_{\text{ND}} = \dfrac{t}{t_0}$

Non-dimensional money: $X_{\text{ND}} = \dfrac{X}{\sqrt{\sigma^2 t_0}}$

<u>Non-dimensional equation</u>

Start with the physical equation: $dX = -m\,dt + \sqrt{\sigma^2 dt}\,\dfrac{dW(t)}{\sqrt{dt}}$

We write all physical quantities in terms of non-dimensional quantities and then substitute into the physical equation

$$t = t_0\, t_{\text{ND}}\,, \qquad X = \sqrt{\sigma^2 t_0}\, X_{\text{ND}}$$

$$dt = t_0\, dt_{\text{ND}}\,, \qquad dX = \sqrt{\sigma^2 t_0}\, dX_{\text{ND}}\,, \qquad \frac{dW(t)}{\sqrt{dt}} = \frac{dW(t_{\text{ND}})}{\sqrt{dt_{\text{ND}}}}$$

Recall $\boxed{\dfrac{dW(t)}{\sqrt{dt}} \sim N(0,1) \text{ is dimensionless and independent of } dt \text{ and } t.}$

==> $$\sqrt{\sigma^2 t_0}\, dX_{\text{ND}} = -m t_0 dt_{\text{ND}} + \sqrt{\sigma^2 t_0 dt_{\text{ND}}}\,\frac{dW(t_{\text{ND}})}{\sqrt{dt_{\text{ND}}}}$$

Divide the equation by $\sqrt{\sigma^2 t_0}$, we obtain

$$dX_{\text{ND}} = -m\frac{t_0}{\sqrt{\sigma^2 t_0}}dt_{\text{ND}} + \sqrt{dt_{\text{ND}}}\,\frac{dW(t_{\text{ND}})}{\sqrt{dt_{\text{ND}}}}$$

Re-writing it in terms of dW, we arrive at the dimensionless equation

$$dX_{\rm ND} = -m_{\rm ND} dt_{\rm ND} + dW(t_{\rm ND}), \qquad m_{\rm ND} \equiv m\sqrt{\frac{t_0}{\sigma^2}}$$

Once we have the dimensionless equation, we can drop the subscript "ND" and revert back to the simple notation $(X,\ t,\ C,\ m)$.

$$dX = -mdt + dW$$

Summary

When we work with the dimensionless equation, $dX = -mdt + dW$,

we need to keep in mind that

$$X = \frac{X_{\rm phy}}{\sqrt{\sigma^2 t_0}}, \qquad C = \frac{C_{\rm phy}}{\sqrt{\sigma^2 t_0}}, \qquad t = \frac{t_{\rm phy}}{t_0}, \qquad m = m_{\rm phy}\sqrt{\frac{t_0}{\sigma^2}}$$

where the subscript $_{\rm phy}$ denotes the physical quantity before non-dimensionalization.

Solutions of case 2

For the biased game, we again study the two questions.

Question #1: How long can you play?

Question #2: What is the chance that you break the bank?

Answer to Question #2

The strategy we use is the same as that in case 1.

Let $\quad u(x) = \Pr\left(A \middle| X(0) = x\right), \quad A \equiv \{X(t) \text{ hits C before } 0\}$

Strategy:

Find a boundary value problem (BVP) governing $u(x)$.

Boundary condition:

$u(C) = 1$ and $u(0) = 0$.

Differential equation:

Start with $X(0) = x \in (0, C)$. After a small time step, $X(dt)$ has the expression

$X(dt) = x + dX, \qquad dX = -m\, dt + dW$

We need to calculate moments of dX.

$$E_{dW}(dX) = E_{dW}(-mdt + dW) = -mdt$$

$$E_{dW}\left((dX)^2\right) = E_{dW}\left((-mdt)^2 + 2(-mdt)dW + (dW)^2\right) = dt + o(dt)$$

For a fixed $x \in (0, C)$, when dt is small enough (depending on x), we have

$$u(x) = E_{dW}\big(u(x+dX)\big) + o(dt) \qquad \text{(The law of total probability)}$$

$$= E_{dW}\left(u(x) + u_x dX + \frac{1}{2} u_{xx} (dX)^2 \right) + o\big(dt\big)$$

(This is where we need moments of dX.)

$$= u(x) - u_x m dt + \frac{1}{2} u_{xx} dt + o\big(dt\big)$$

Divide by dt and then take the limit as $dt \to 0$, we obtain a 2nd order ODE for $u(x)$

$$u_{xx} - 2m u_x = 0$$

Function $u(x)$ satisfies the boundary value problem (BVP)

$$\begin{cases} u_{xx} - 2m u_x = 0 & \text{differential equation} \\ u(0) = 0, \quad u(C) = 1 & \text{boundary conditions} \end{cases}$$

<u>Solution of the BVP:</u>

$$u(x) = \frac{e^{-2mC}(e^{2mx} - 1)}{1 - e^{-2mC}} \qquad \text{(homework problem)}$$

When mC is moderately large (for example, $mC \geq 5$), we have

$$u(x) = \frac{e^{-2mC}(e^{2mx} - 1)}{1 - e^{-2mC}} \approx e^{-2mC}(e^{2mx} - 1)$$

which decays exponentially with the factor e^{-2mC}.

<u>Comparison of fair game vs biased game</u>

We look at $u(C/2)$, the probability of breaking the bank when you and the casino start with the same amount of cash, $(C/2)$.

Fair game: $u(x) = x/C$ ==> $u\left(\frac{C}{2}\right) = \frac{1}{2}$

Biased game: $u\left(\frac{C}{2}\right) \approx e^{-2mC}(e^{2mC/2} - 1) \approx e^{-mC}$

which decays exponentially with mC.

For example, $mC = 5$ ==> $e^{-mC} = e^{-5} = 0.0067$.

<u>Answer to Question #1</u>

Let $\quad T(x) = E\Big(Z \Big| X(0) = x\Big), \quad Z \equiv \big(\text{time from 0 until } X(t) = C \text{ or } X(t) = 0\big)$.

Strategy:

Find a boundary value problem (BVP) governing $T(x)$.

Boundary condition:

$T(C) = 0$ and $T(0) = 0$.

Differential equation:

Start with $X(0) = x \in (0, C)$. After a small time step, $X(dt)$ has the expression

$X(dt) = x + dX, \qquad dX = -m\,dt + dW$

The moments of dX are

$$E_{dW}(dX) = -mdt\,, \qquad E_{dW}\Big((dX)^2\Big) = dt + o(dt)$$

For a fixed $x \in (0, C)$, when dt is small enough (depending on x), we have

$$T(x) = dt + E_{dW}\Big(T(x+dX)\Big) + o(dt) \qquad \text{(The law of total expectation)}$$

$$= dt + E_{dW}\left(T(x) + T_x dX + \frac{1}{2}T_{xx}(dX)^2\right) + o(dt)$$

(This is where we need moments of dX.)

$$= dt + T(x) - T_x m dt + \frac{1}{2}T_{xx} dt + o(dt)$$

Divide by dt and then take the limit as $dt \to 0$, we obtain an ODE for $T(x)$

$$T_{xx} - 2mT_x = -2$$

Function $T(x)$ satisfies the boundary value problem (BVP)

$$\begin{cases} T_{xx} - 2mT_x = -2 & \text{differential equation} \\ T(0) = 0, \quad T(C) = 0 & \text{boundary conditions} \end{cases}$$

The solution of the BVP:

$$T(x) = \frac{x}{m} - \frac{C}{m}\left(\frac{e^{2mx} - 1}{e^{2mC} - 1}\right) \qquad \text{(homework problem)}$$

When mC is moderately large (for example, $mC \geq 5$), we have

$$T(x) = \frac{x}{m} \cdot \left(1 - \frac{C}{x}\left(\frac{e^{2mx} - 1}{e^{2mC} - 1}\right)\right) \approx \frac{x}{m} \qquad \text{for } x \leq \frac{C}{2}$$

Here we have used $\frac{C}{x}\left(\frac{e^{2mx}-1}{e^{2mC}-1}\right) << 1$, which is derived in Appendix A.

The result, $T(x) \approx x/m$, is consistent with the intuitive picture that if your cash decreases with speed m, then your initial cash x will last a time period of (x/m).

Meaning of $mC \geq 5$ in terms of physical quantities:

$$m = m_{\text{phy}}\sqrt{\frac{t_0}{\sigma^2}}, \qquad C = \frac{C_{\text{phy}}}{\sqrt{\sigma^2 t_0}}, \qquad x = \frac{x_{\text{phy}}}{\sqrt{\sigma^2 t_0}}$$

$$==> \qquad mC = \frac{m_{\text{phy}} C_{\text{phy}}}{\sigma^2}$$

$mC \geq 5$ corresponds to $\frac{m_{\text{phy}} C_{\text{phy}}}{\sigma^2} \geq 5$.

An example (with physical parameters)

Consider a biased game with physical parameters below.

$$\sigma = 5\frac{\$}{\sqrt{\text{min}}}, \qquad m_{\text{phy}} = 0.25\frac{\$}{\text{min}}$$

$$C_{\text{phy}} = 1000\$, \qquad x_{\text{phy}} = 500\$$$

The scales are

t_0 = 1 min (we select t_0), $\qquad \sqrt{\sigma^2 t_0} = 5\$$

The dimensionless quantities are

$$m = m_{\text{phy}}\sqrt{\frac{t_0}{\sigma^2}} = 0.05, \qquad C = \frac{C_{\text{phy}}}{\sqrt{\sigma^2 t_0}} = 200, \qquad mC = 10$$

$$x = \frac{x_{\text{phy}}}{\sqrt{\sigma^2 t_0}} = 100, \qquad mx = 5$$

Since mC = 10, the approximate expressions for $u(x)$ and $T(x)$ are valid.

Probability of breaking the bank:

$$u(x) \approx e^{-2mC}(e^{2mx}-1) = e^{-20}(e^{10}-1) \approx e^{-10} = 4.54\times 10^{-5}$$

The chance of breaking the bank is virtually zero even though you and the casino start with the same amount \$500.

Average time until the end of game:

$$T(x) \approx \frac{x}{m} = \frac{100}{0.05} = 2000$$

The physical time until the end of game is

$T_{\text{phy}} = T\, t_0 = 2000$ minutes.

White noise $\dfrac{dW}{dt}$

Consider the stochastic differential equation (SDE)

$$dX = -m dt + dW$$

We write the "formal" derivative of X as

$$\frac{dX}{dt} = -m + \frac{dW}{dt}$$

Recall that in SDE, dt is finite until we take the limit as $dt \to 0$.

Here $\lim\limits_{dt\to 0} \dfrac{dW}{dt}$ does not exist in the conventional sense.

Key strategy: We take the limit AFTER its interactions with other entities.

The short story of white noise

1) $Z(t) \equiv \dfrac{dW}{dt} = \dfrac{1}{\sqrt{dt}} \cdot \dfrac{dW}{\sqrt{dt}}, \qquad \dfrac{dW}{\sqrt{dt}} \sim N(0,1)$

   $Z(t)$ diverges to $\pm\infty$ as $dt \to 0$. $Z(t)$ is not a regular function.

2) $E\big(Z(t)Z(s)\big) = \delta(t-s)$

3) $\int e^{-i2\pi\xi t} E\big(Z(t)Z(0)\big) dt = 1$

4) $Z(t)$ is a white noise (we will clarify what that means).

Before we discuss the details in the long story of white noise, we review some of the mathematical tools/methods we will use.

Mathematical preparations

Delta function (Dirac's delta function):

Definition 1:

Consider the limit of a boxcar function.

$$\lim_{d\to 0} \Pi_d(x)$$

where $\Pi_d(x) = \begin{cases} \dfrac{1}{d}, & \text{for } x \in \left(\dfrac{-d}{2}, \dfrac{d}{2}\right) \\ 0, & \text{otherwise} \end{cases}$

This limit does not exist in the conventional sense.

However, for any smooth function $g(x)$, we have

$$\lim_{d\to 0} \int \Pi_d(x) g(x) dx = g(0)$$

We "formally" denote $\lim\limits_{d\to 0} \Pi_d(x)$ by $\delta(t)$.

$$\delta(x) = \lim_{d\to 0} \Pi_d(x)$$

$\delta(x)$ "formally" satisfies

$$\int \delta(x) g(x) dx = g(0) \qquad \text{for all smooth functions } g(x).$$

The true meaning of the LHS is $\int \delta(x) g(x) dx \xrightarrow{\text{Defined as}} \lim\limits_{d\to 0} \int \Pi_d(x) g(x) dx$

The <u>key strategy</u> in making sense of $\lim\limits_{d\to 0} \Pi_d(x)$ is that we take the limit AFTER integrating $\Pi_d(x)$ with smooth function $g(x)$.

<u>Definition 2:</u>

In a similar way, we can define $\delta(x)$ as the limit of a normal distribution

$$\delta(x) = \lim_{\sigma\to 0} \rho_{N(0,\sigma^2)}(x) = \lim_{\sigma\to 0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

These two definitions are equivalent. We use whichever is convenient.

<u>Remark:</u>

Both definitions are based on the probability density of a scaled random variable with the multiplier converging to zero.

<u>Definition 1:</u> $\delta(x) = \lim\limits_{\sigma\to 0} \rho_{\sigma X}(x)$, $\quad X \sim \text{uniform}\left(\dfrac{-1}{2}, \dfrac{1}{2}\right)$

<u>Definition 2:</u> $\delta(x) = \lim\limits_{\sigma\to 0} \rho_{\sigma X}(x)$, $\quad X \sim N(0,1)$

<u>Fourier transform</u> (FT):

Forward transform:

$$\underbrace{\hat{y}(\xi)}_{\text{Notation}} \equiv \underbrace{F[y(t)]}_{\text{Operator notation}} \equiv \int_{-\infty}^{+\infty} \exp(-i2\pi\xi t)\, y(t)\, dt$$

Inverse transform:

$$y(t) = F^{-1}[\hat{y}(\xi)] \equiv \int_{-\infty}^{+\infty} \exp(i2\pi\xi t)\, \hat{y}(\xi)\, d\xi$$

Remark: There are several versions of FT. They are all equivalent by a scaling.

Alternative FT 1: $\quad F[y(t)] = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{+\infty} \exp(-i\omega t)\, y(t)\, dt$

Alternative FT 2: $\quad F[y(t)] = \displaystyle\int_{-\infty}^{+\infty} \exp(-i\omega t)\, y(t)\, dt$

Properties of Fourier transform:

1) Fourier transform of a normal PDF

$$F\left[\rho_{N(0,\sigma^2)}(t)\right] = F\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-t^2}{2\sigma^2}\right)\right] = \exp\left(-2\pi^2\sigma^2\xi^2\right)$$

Proof:

Use the characteristic function (CF) of a normal RV we derived in Lecture 3. Then use the connection between the CF and the Fourier transform of PDF.

2) Fourier transform of the delta function

$F[\delta(t)] = 1$

Proof: We view the delta function as the limit of normal distribution

$$\delta(t) = \lim_{\sigma\to 0} \rho_{N(0,\sigma^2)}(t)$$

We apply the Fourier transform and then take the limit as $\sigma \to 0$.

$$F\left[\delta(t)\right] = \lim_{\sigma\to 0} F\left[\rho_{N(0,\sigma^2)}(t)\right] = \lim_{\sigma\to 0} \exp\left(-2\pi^2\sigma^2\xi^2\right) = 1$$

3) Fourier transform of $y(t) \equiv 1$

$F[1] = \delta(\xi)$

Proof: $\quad F[1] = \displaystyle\int_{-\infty}^{+\infty} \exp(-i2\pi\xi t)\, dt$ does not converge in the conventional sense!

We view 1 as the limit of

$$1 = \lim_{\sigma\to\infty} \exp\left(\frac{-t^2}{2\sigma^2}\right)$$

We apply the Fourier transform and then take the limit as $\sigma \to \infty$.

$$F[1] = \lim_{\sigma\to\infty} F\left[\exp\left(\frac{-t^2}{2\sigma^2}\right)\right], \qquad \rho_{N(0,\sigma^2)}(t) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-t^2}{2\sigma^2}\right)$$

$$= \lim_{\sigma\to\infty}\sqrt{2\pi\sigma^2}\,F\left[\rho_{N(0,\sigma^2)}(t)\right] = \lim_{\sigma\to\infty}\sqrt{2\pi\sigma^2}\exp\left(-2\pi^2\sigma^2\xi^2\right)$$

$$= \lim_{\sigma\to\infty}\frac{1}{\sqrt{2\pi\left(\frac{1}{4\pi^2\sigma^2}\right)}}\exp\left(\frac{-\xi^2}{2\left(\frac{1}{4\pi^2\sigma^2}\right)}\right) = \lim_{\sigma\to\infty}\rho_{N\left(0,\frac{1}{4\pi^2\sigma^2}\right)}(\xi)$$

$$= \lim_{s\to 0}\rho_{N(0,s^2)}(\xi) = \delta(\xi)$$

Key observation:

If an operator acting on the limit of a function is invalid in the conventional sense, we can try to make sense of it by delaying taking the limit. That is, we first apply the operator and then we take the limit afterwards. That is why in the discussion of stochastic differential equations, dt is finite until we take the limit at the end.

Example:

"Formally" we can conveniently write

$$\int \underbrace{\lim_{\sigma\to 0}\rho_{N(0,\sigma^2)}(t)}_{\text{Not a regular function}} g(t)\,dt = \int \delta(t)g(t)\,dt = g(0)$$

The true mathematical meaning is

$$\lim_{\sigma\to 0}\int \rho_{N(0,\sigma^2)}(t)g(t)\,dt = g(0)$$

which makes sense and is mathematically rigorous.

Appendix A

Theorem: When mC is moderately large and $x \le C/2$, we have

$$\frac{C}{x}\left(\frac{e^{2mx}-1}{e^{2mC}-1}\right) << 1$$

Proof: We first introduce a lemma.

Lemma: Function $f(s) \equiv \dfrac{e^{s}-1}{s}$ increases monotonically for $s > 0$.

Proof:

$$\frac{e^{s}-1}{s} = \frac{1}{s}\left(\sum_{n=0}\frac{1}{n!}s^{n}-1\right) = \sum_{n=1}\frac{1}{n!}s^{n-1}$$

Each term in the summation is positive, and increases monotonically for $s > 0$.

End of proof

Apply the lemma to $\dfrac{e^{2mx}-1}{x}$ for $x \le C/2$,we get

$$\frac{e^{2mx}-1}{x} = 2m\cdot\underbrace{\frac{(e^{2mx}-1)}{2mx}}_{f(2mx)} \le 2m\cdot\underbrace{\frac{(e^{2mC/2}-1)}{2mC/2}}_{f(2mC/2)} = \frac{2(e^{mC}-1)}{C}$$

Using this inequality, we write $\dfrac{C}{x}\left(\dfrac{e^{2mx}-1}{e^{2mC}-1}\right)$ as

$$\frac{C}{x}\left(\frac{e^{2mx}-1}{e^{2mC}-1}\right) = \frac{C}{(e^{2mC}-1)}\left(\frac{e^{2mx}-1}{x}\right) \le \frac{C}{(e^{2mC}-1)}\frac{2(e^{mC}-1)}{C}$$

$$= \frac{2(e^{mC}-1)}{(e^{2mC}-1)} = \frac{2}{e^{mC}+1} \approx 2e^{-mC} << 1 \qquad \text{for moderately large } mC$$

Therefore, we conclude $\dfrac{C}{x}\left(\dfrac{e^{2mx}-1}{e^{2mC}-1}\right) << 1$ when mC is moderately large and $x \le C/2$.

AM216 Stochastic Differential Equations

Lecture 06

List of topics in this lecture

- Energy spectrum density (ESD), power spectrum density (PSD)
- Stationary stochastic process, auto-correlation function (ACF)
- Wiener-Khinchin theorem: PSD is Fourier transform of ACF
- Definition of white noise: PSD is constant in frequency domain
- Calculating ACF and PSD of $Z(t) \equiv dW/dt$
- Constrained Wiener process, Bayes Theorem

Recap

Gambler's ruin problem:

Methodology of deriving BVPs for $u(x)$ and $T(x)$

Scaling and non-dimensionalization, advantage of working with $\dfrac{dW(t)}{\sqrt{dt}}$

Short story of white noise ...

Fourier transform: $F[y(t)] \equiv \int_{-\infty}^{+\infty} \exp(-i2\pi\xi t)\, y(t)\, dt$

Properties of Fourier transform (continued)

1) $F\left[\rho_{N(0,\sigma^2)}(t)\right] = F\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-t^2}{2\sigma^2}\right)\right] = \exp\left(-2\pi^2\sigma^2\xi^2\right)$

2) $F[\delta(x)] = 1$

3) $F[1] = \delta(\xi)$

---

4) Parseval's theorem

$$\int \left|y(t)\right|^2 dt = \int \left|\hat{y}(\xi)\right|^2 d\xi$$

Proof:

$$\int\left|y(t)\right|^2 dt = \int y(t)\overline{y(t)}dt = \int\left(\int \exp(i2\pi\xi t)\hat{y}(\xi)d\xi \int \exp(-i2\pi\eta t)\overline{\hat{y}(\eta)}d\eta\right)dt$$

$$= \int\left(\int\int \exp\left(-i2\pi(\eta-\xi)t\right)\hat{y}(\xi)\overline{\hat{y}(\eta)}d\xi d\eta\right)dt$$

Change the order of integration

$$= \int\int \hat{y}(\xi)\overline{\hat{y}(\eta)}\underbrace{\left(\int \exp\left(-i2\pi(\eta-\xi)t\right)dt\right)}_{F[1]=\delta(\eta-\xi)}d\eta d\xi$$

$$= \int\int \hat{y}(\xi)\overline{\hat{y}(\eta)}\delta(\eta-\xi)d\eta d\xi = \int \hat{y}(\xi)\overline{\hat{y}(\xi)}d\xi = \int\left|\hat{y}(\xi)\right|^2 d\xi$$

A rigorous proof: $\int\left|y(t)\right|^2 dt = \lim_{\sigma\to 0}\int y(t)\overline{y(t)}e^{-\sigma^2t^2}dt = \cdots$

Recall the short story of white noise:

1) $Z(t) \equiv \dfrac{dW}{dt}$ is not a regular function.
2) $E\left(Z(t)Z(s)\right) = \delta(t-s)$
3) $\int \exp(-i2\pi\xi t)E\left(Z(t)Z(0)\right)dt = 1$
4) $Z(t)$ is a white noise (we will clarify what that means).

The long story of white noise

We follow the steps listed below.

- Energy $\propto \int_{-T}^{T}\left|y(t)\right|^2 dt$ --> Energy spectrum density (ESD)
- Power $\propto \dfrac{1}{T}\int_{-T}^{T}\left|y(t)\right|^2 dt$ --> Power spectrum density (PSD)
- Relation between PSD and auto-correlation function (ACF)
- Definition of white noise based on PSD
- Calculating ACF and PSD of $Z(t) \equiv dW/dt$

Energy spectrum density (ESD)

In many physics problems,

$$\text{Energy} \propto \int_{-\infty}^{+\infty} \left| y(t) \right|^2 dt$$

Examples:

$y(t)$ = electric current

$$\text{Energy} = \int_{-\infty}^{+\infty} R \cdot y(t)^2 \, dt\,, \qquad R = \text{electrical resistance}$$

Here "energy" refers to the dissipated energy.

$y(t)$ = velocity

$$\text{Energy} = \int_{-\infty}^{+\infty} b \cdot y(t)^2 \, dt\,, \qquad b = \text{viscous drag coefficient}$$

For mathematical convenience, we define

$$\text{Energy} \equiv \underbrace{\int_{-\infty}^{+\infty} \left| y(t) \right|^2 dt = \int_{-\infty}^{+\infty} \left| \hat{y}(\xi) \right|^2 d\xi}_{\text{Parseval's theorem}} = \int_{-\infty}^{+\infty} \left| \int_{-\infty}^{+\infty} \exp(-i2\pi\xi t) y(t) dt \right|^2 d\xi$$

We like to know how the energy is distributed in the frequency domain.

Definition of energy spectrum density (ESD)

$$\text{ESD} \equiv \left| \hat{y}(\xi) \right|^2 = \left| \int_{-\infty}^{+\infty} \exp(-i2\pi\xi t) y(t) dt \right|^2$$

Caution: $\left| \hat{y}(\xi) \right|^2$ is an unnormalized density.

$$\int_{-\infty}^{+\infty} \left| \hat{y}(\xi) \right|^2 d\xi = \int_{-\infty}^{+\infty} \left| y(t) \right|^2 dt = \text{Energy} \neq 1$$

Other examples of unnormalized density:

Population density: X number of persons per square mile

Pollution density: X amount of chemicals per unit volume of air or water

Car density: X number of cars per mile of highway

Caution: (slightly different definitions of ESD)

In electrical engineering (EE), energy spectrum density is defined as

$$\text{ESD} \equiv \Phi(\omega) = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-i\omega t) y(t) dt \right|^2 \qquad \omega\text{: angular frequency}$$

$\Phi(\omega)$ and $\left| \hat{y}(\xi) \right|^2$ are related by:

$$\Phi(\omega)=\frac{1}{2\pi}\left|\hat{y}(\xi)\right|^2,\qquad \xi=\frac{\omega}{2\pi}:$$ ordinary frequency

Power spectrum density (PSD)

Energy spectrum density is meaningful only when $\int_{-\infty}^{+\infty}\left|y(t)\right|^2 dt=\text{finite}$.

Example: $y(t)$ = electric current = y_0 = constant in time

$$\text{Energy}=\int_{-\infty}^{+\infty}R\cdot y_0^2\,dt=\infty$$

When the total energy is unbounded, we look at the energy per time.

$$\int_{-T}^{T}\left|y(t)\right|^2 dt=\int_{-\infty}^{\infty}\left|\int_{-T}^{T}\exp(-i2\pi\xi t)y(t)dt\right|^2 d\xi \qquad \text{(Parseval's theorem)}$$

$$\frac{1}{2T}\int_{-T}^{T}\left|y(t)\right|^2 dt=\int_{-\infty}^{\infty}\frac{1}{2T}\left|\int_{-T}^{T}\exp(-i2\pi\xi t)y(t)dt\right|^2 d\xi$$

Definition of power spectrum density (PSD)

$$\text{PSD}\equiv\lim_{T\to\infty}\frac{1}{2T}\left|\int_{-T}^{T}\exp(-i2\pi\xi t)y(t)dt\right|^2$$

Expression of power spectrum density (PSD)

We write PSD into a more workable expression.

$$\text{PSD}=\lim_{T\to\infty}\frac{\int_{-T}^{T}\exp(-i2\pi\xi t)y(t)dt\int_{-T}^{T}\exp(i2\pi\xi s)\overline{y(s)}ds}{2T}$$

$$=\lim_{T\to\infty}\frac{\int_{-T}^{T}\int_{-T}^{T}\exp(-i2\pi\xi(t-s))y(t)\overline{y(s)}dt\,ds}{2T}$$

change of variable $\tau = t - s$

$$=\lim_{T\to\infty}\frac{\int_{-T}^{T}\int_{-T-s}^{T-s}\exp(-i2\pi\xi\tau)y(\tau+s)\overline{y(s)}d\tau\,ds}{2T}$$

Draw the integration region in s-τ plane.

For each s, the range for τ is $[-T-s, T-s]$.

For each τ, the range for s is $[a(\tau), b(\tau)]$ where

$$a(\tau)=\begin{cases}-T-\tau, & \tau\in[-2T,\,0]\\ -T, & \tau\in[0,\,2T]\end{cases},\qquad b(\tau)=\begin{cases}T, & \tau\in[-2T,\,0]\\ T-\tau, & \tau\in[0,\,2T]\end{cases}$$

Change the order of integration

$$\text{PSD} = \lim_{T\to\infty} \frac{\int_{-2T}^{2T} \exp(-i2\pi\xi\tau)\int_{a(\tau)}^{b(\tau)} y(\tau+s)\overline{y(s)}ds\,d\tau}{2T} \qquad \text{(PSD01)}$$

So far, we worked with deterministic process $y(t)$.

Next we introduce stochastic process and stationary stochastic process.

Definition of stochastic process

A stochastic process is a function of time that varies with ω.

$$\underbrace{y(t)}_{\text{Short notation}} = \underbrace{y(t,\omega)}_{\text{Full notation}} \qquad \omega = \text{random outcome of an experiment}$$

Definition of stationary stochastic process

Let $y(t)$ be a stochastic process. We say $y(t)$ is stationary if for any set of time instances $(t_1, t_2, \ldots, t_k)$, the joint distribution of $(y(t+t_1), y(t+t_2), \ldots, y(t+t_k))$ is independent of t.

Examples:

- $W(t)$ is a stochastic processes. It is not stationary:
  $\text{var}(W(t)) = t$ varies with t.
- $Z(t) = \frac{dW(t)}{dt} \sim \frac{1}{\sqrt{dt}} N(0,dt)$ is a well defined stochastic process for finite dt.

  It is stationary: the joint distribution is invariant under a shift.

Properties of stationary stochastic process

For a stationary stochastic process, we have

- $E(y(t)) = E(y(0))$
- $\text{var}(y(t)) = \text{var}(y(0))$
- $E\left(y(s+\tau)\overline{y(s)}\right) = E\left(y(\tau)\overline{y(0)}\right)$

Definition of auto-correlation function (ACF)

For a stationary stochastic process $y(t)$, the auto-correlation function (ACF) is

$$R(\tau) \equiv E\left(y(s+\tau)\overline{y(s)}\right) = E\left(y(\tau)\overline{y(0)}\right)$$

Note: $R(\tau)$ is independent of s (for a stationary process).

Caution: be careful with the term "auto-correlation"

Auto-correlation coefficient is defined as

$$\rho(\tau) \equiv \frac{E\left(\left[y(\tau) - E(y(0))\right]\left[\overline{y(0) - E(y(0))}\right]\right)}{\mathrm{var}(y(0))}$$

<u>Auto-correlation function</u> (ACF) is defined as

$$R(\tau) \equiv E\left(y(\tau)\overline{y(0)}\right)$$

<u>Relation between PSD and ACF</u>

For a stationary stochastic process, the power spectrum density (PSD) is

$$\underbrace{s(\xi)}_{\substack{\text{New notation}\\ \text{for PSD}}} \equiv \mathrm{PSD} \equiv \lim_{T\to\infty} \frac{E\left(\left|\int_{-T}^{T} \exp(-i2\pi\xi t) y(t) dt\right|^2\right)}{2T}$$

We use (PSD01), obtained above for a deterministic process, to rewrite $s(\xi)$

$$\mathrm{s}(\xi) = \lim_{T\to\infty} \frac{E\left(\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) \int_{a(\tau)}^{b(\tau)} y(\tau+s)\overline{y(s)} ds\, d\tau\right)}{2T}$$

Change the order of integration and expectation

$$= \lim_{T\to\infty} \frac{\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) \int_{a(\tau)}^{b(\tau)} E\left(y(\tau+s)\overline{y(s)}\right) ds\, d\tau}{2T}$$

$$= \lim_{T\to\infty} \frac{\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) \int_{a(\tau)}^{b(\tau)} R(\tau) ds\, d\tau}{2T} \qquad R(\tau) \text{ is independent of } s.$$

$$= \lim_{T\to\infty} \frac{\int_{-2T}^{2T} \exp(-i2\pi\xi\tau) R(\tau)(b(\tau) - a(\tau)) d\tau}{2T}$$

The term $(b(\tau) - a(\tau))$ has the expression:

$$b(\tau) - a(\tau) = \begin{cases} 2T + \tau, & \tau \in [-2T,\ 0] \\ 2T - \tau, & \tau \in [0,\ 2T] \end{cases} = 2T - |\tau|$$

Substituting it into the expression of $s(\xi)$ yields

$$s(\xi) = \lim_{T\to\infty} \int_{-2T}^{2T} \exp(-i2\pi\xi\tau) R(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau$$

Taking the limit as $T \to \infty$, we arrive at

$$s(\xi) = \int_{-\infty}^{+\infty} \exp(-i2\pi\xi\tau) R(\tau) d\tau$$

We just derived the Wiener-Khinchin theorem.

Wiener-Khinchin theorem (relation between PSD and ACF)

For a stationary stochastic process $y(t)$, the power spectrum density, $s(\xi)$, and the auto-correlation function, $R(t)$, are related by

$$s(\xi) = \int_{-\infty}^{+\infty} \exp(-i2\pi\xi t) R(t) dt$$

In other words, the PSD is the Fourier transform of ACF.

Definition of white noise

Let $y(t)$ be a stationary stochastic process. We say $y(t)$ is a white noise if

$s(\xi) = const$

In other words, the power of a white noise is uniformly distributed in the frequency domain. The Wiener-Khinchin theorem tells us that

$$s(\xi) = const \qquad \Longleftrightarrow \qquad R(t) \equiv E\left(y(t)\overline{y(0)} \right) \propto \delta(t)$$

Working out items in the short story

We re-write the short story in terms of the auto-correlation function $R(\tau)$ and power spectrum density $s(\xi)$.

1) $Z(t) \equiv \dfrac{dW}{dt}$ is not a regular function.

2) $R(\tau) = E\left(Z(s+\tau) Z(s) \right) = \delta(\tau)$

3) $s(\xi) = \int \exp(-i2\pi\xi t) R(t) dt = 1$

4) $Z(t)$ is a white noise.

- To show $Z(t)$ is a white noise (item 4), we only need $s(\xi)$ = const (item 3).
- To show $s(\xi) = 1$(item 3), we only need $R(t) = \delta(t)$ (item 2)

Thus, the remaining task is to show item 2, which we do now.

Derivation of $R(t) = \delta(t)$ for $Z(t) \equiv dW/dt$

Here we present a "formal" derivation. A rigorous derivation is in Appendix A.

We first calculate $E(W(t)W(s))$ for $t \geq s$.

$$E\left(W(t)W(s) \right) = E\left((W(t) - W(s) + W(s)) W(s) \right)$$

$$= E\Big((W(t)-W(s))W(s)\Big)+E\Big(W(s)^2\Big)=0+s=s$$

Since $E(W(t)W(s)) = E(W(s)W(t))$, we obtain

$$E\Big(W(t)W(s)\Big)=\min(t,s)$$

Next, in the calculation of $E(Z(t)Z(s))$, we "formally" exchange the order of differentiation and expectation.

$$E\Big(Z(t)Z(s)\Big)=E\left(\frac{\partial}{\partial s}\frac{\partial}{\partial t}\Big(W(t)W(s)\Big)\right)$$

$$=\frac{\partial}{\partial s}\frac{\partial}{\partial t}E\Big(W(t)W(s)\Big)=\frac{\partial}{\partial s}\frac{\partial}{\partial t}\min(t,s)$$

As a function of t, we have

$$\min(t,s)=\begin{cases} t, & t<s \\ s, & t>s \end{cases}$$

Differentiating with respect to t, and then writing it as a function of s, we get

$$\frac{\partial}{\partial t}\min(t,s)=\begin{cases} 1, & t<s \\ 0, & t>s \end{cases} \qquad \text{(as a function of } t\text{)}$$

$$=\begin{cases} 0, & s<t \\ 1, & s>t \end{cases} \qquad \text{(as a function of } s\text{)}$$

Differentiating with respect to s, we arrive at

$$\frac{\partial}{\partial s}\frac{\partial}{\partial t}\min(t,s)=\delta(s-t)$$

Therefore, we conclude

$$E\Big(Z(t)Z(s)\Big)=\frac{\partial}{\partial s}\frac{\partial}{\partial t}\min(t,s)=\delta(s-t)$$

$$\Longrightarrow \qquad R(\tau)=E\Big(Z(s+\tau)Z(s)\Big)=\delta(\tau)$$

<u>A class of colored noise</u>:

In the subsequent discussion of Ornstein-Uhlenbeck process (OU), we will see that its auto-correlation has the form:

$$R(t) = E\left(y(t)\overline{y(0)}\right) \propto \exp(-\beta|t|)$$

The corresponding power spectrum density is

$$s(\xi) = \int \exp(-i2\pi\xi t)R(t)dt \propto \int \exp(-i2\pi\xi t)\exp(-\beta|t|)dt = \frac{2\beta}{\beta^2 + 4\pi^2\xi^2}$$

End of discussion of white noise

Constrained Wiener process

For an unconstrained Wiener process, we have

$W(0) = 0$ and $W(t_1) \sim N(0, t_1)$

<u>Question:</u> What happens if it is constrained by $W(t_1+t_2) = y$?

We like to know the conditional distribution $(W(t_1) \mid W(t_1+t_2) = y)$.

To answer this question, we need to introduce Bayes theorem.

<u>Bayes Theorem</u>

Consider two events A and B. We write Pr(A and B) in two ways.

Pr(A and B) = Pr(A | B) Pr(B)

Pr(A and B) = Pr(B | A) Pr(A)

Equating the two, we get

Pr(A | B) Pr(B) = Pr(B | A) Pr(A)

Express Pr(A | B) in terms of Pr(B | A), we arrive at

<u>Bayes Theorem for events:</u>

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

To derive Bayes theorem for densities, we consider

A = "$x < X \le x+\Delta x$"

B = "$y < Y \le y+\Delta y$"

We write probabilities in terms of densities

$$\Pr(A|B) \approx \rho(X = x | Y = y)\Delta x$$

$$\Pr(B|A) \approx \rho(Y = y | X = x)\Delta y$$

$\Pr(A) \approx \rho(X=x)\Delta x$

$\Pr(B) \approx \rho(Y=y)\Delta y$

Substituting these terms into Bayes theorem, we obtain…

Bayes theorem for densities

$$\rho(X=x \mid Y=y) = \frac{\rho(Y=y \mid X=x)\cdot\rho(X=x)}{\rho(Y=y)}$$

A useful trick:

In density $\rho(X=x|Y=y)$, x is the independent variable and y is a parameter. On the RHS of Bayes theorem, $\rho(Y=y)$ has no dependence on x and serves as a normalizing factor.

Thus, we don't need to explicitly keep track of $\rho(Y=y)$. We can write Bayes theorem conveniently as

$$\boxed{\rho(X=x \mid Y=y) \propto \rho(Y=y \mid X=x)\cdot\rho(X=x)}$$

where the RHS needs a proper normalizing factor to make it integrate to 1.

This trick is especially convenient for normal distributions:

$$X \sim N(\mu, \sigma^2) \qquad \Longleftrightarrow \qquad \rho(X=x) \propto \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right),$$

Conditional density $\rho(W(t_1) = x \mid W(t_1+t_2) = y)$

To use the Bayes theorem, we first find $\rho(W(t_1) = x)$ and $\rho(W(t_1+t_2) = y \mid W(t_1) = x)$.

$W(t_1) \sim N(0, t_1)$

$$\Longrightarrow \quad \rho\big(W(t_1)=x\big) = \rho_{N(0,t_1)}(x) \propto \exp\left(\frac{-x^2}{2t_1}\right)$$

$$W(t_1+t_2) = W(t_1) + \underbrace{\big(W(t_1+t_2)-W(t_1)\big)}_{\sim N(0,t_2)}$$

$$\Longrightarrow \quad \big(W(t_1+t_2) \mid W(t_1)=x\big) \sim N(x, t_2)$$

$$\Longrightarrow \quad \rho\big(W(t_1+t_2)=y \mid W(t_1)=x\big) = \rho_{N(x,t_2)}(y) \propto \exp\left(\frac{-(y-x)^2}{2t_2}\right)$$

The Bayes theorem gives us

$$\rho\Big(W(t_1)=x\,\Big|\,W(t_1+t_2)=y\Big) \propto \rho\Big(W(t_1+t_2)=y\,\Big|\,W(t_1)=x\Big)\cdot\rho\Big(W(t_1)=x\Big)$$

(We don't need to keep track of factors that are independent of x!)

$$\propto \exp\left(\frac{-(y-x)^2}{2t_2}\right)\exp\left(\frac{-x^2}{2t_1}\right) \propto \exp\left(-\left(\left(\frac{1}{2t_1}+\frac{1}{2t_2}\right)x^2-2\frac{y}{2t_2}x\right)\right)$$

(Completing the square)

$$\propto \exp\left(\frac{-\left(x-\dfrac{t_1 y}{t_1+t_2}\right)^2}{2\dfrac{t_1 t_2}{t_1+t_2}}\right) \sim N\left(\frac{t_1 y}{t_1+t_2},\ \frac{t_1 t_2}{t_1+t_2}\right)$$

We conclude

$$\boxed{\rho\Big(W(t_1)=x\,\Big|\,W(t_1+t_2)=y\Big) \sim N\left(\frac{t_1 y}{t_1+t_2},\ \frac{t_1 t_2}{t_1+t_2}\right)}$$

<u>For the general case</u>, we have

$$\boxed{\rho\Big(W(a+t_1)=x\,\Big|\,W(a)=y_a \text{ and } W(a+t_1+t_2)=y_b\Big) \sim N\left(\frac{t_1 y_b+t_2 y_a}{t_1+t_2},\ \frac{t_1 t_2}{t_1+t_2}\right)}$$

<u>A special case:</u> $t_1 = t_2 = h/2$

$$\rho\left(W(a+\frac{h}{2})=x\,\Big|\,W(a)=y_a \text{ and } W(a+h)=y_b\right) \sim N\left(\frac{y_a+y_b}{2},\ \frac{h}{4}\right)$$

This is very useful in refining a discrete sample path of $W(t)$.

Appendix A: A rigorous derivation of $R(t)$ and $s(\xi)$ for $Z(t) \equiv dW/dt$

First, we work with finite dt. Let $\Delta t \equiv dt$. We have

$$Z(t) = \frac{W(t+\Delta t)-W(t)}{\Delta t} \qquad \text{a well defined stationary stochastic process}$$

$$E\Big(Z(t)Z(s)\Big) = E\left(\frac{W(t+\Delta t)-W(t)}{\Delta t}\cdot\frac{W(s+\Delta t)-W(s)}{\Delta t}\right)$$

$$= \begin{cases} 0, & |t-s| > \Delta t \\ \dfrac{\Delta t - |t-s|}{(\Delta t)^2}, & |t-s| \le \Delta t \end{cases} \qquad \text{(derivation not included)}$$

$$R(\tau) = E\Big(Z(s+\tau)Z(s)\Big) = \begin{cases} 0, & |\tau| > \Delta t \\ \dfrac{\Delta t - |\tau|}{(\Delta t)^2}, & |\tau| \le \Delta t \end{cases}$$

Taking the Fourier transform of $R(t)$, we obtain

$$s(\xi) = \int \exp(-i2\pi\xi t)R(t)dt = \int_{-\Delta t}^{\Delta t} \exp(-i2\pi\xi t)\frac{\Delta t - |t|}{(\Delta t)^2}dt$$

$$= 2\frac{\cosh(i2\pi\xi\Delta t)-1}{(i2\pi\xi\Delta t)^2} \qquad \text{(derivation not included)}$$

Finally, we take the limit as $\Delta t \to 0$. At any fixed ξ, as $\Delta t \to 0$, we have

$$\lim_{\Delta t\to 0} s(\xi) = \lim_{\Delta t \to 0} 2\frac{\cosh(i2\pi\xi\Delta t)-1}{(i2\pi\xi\Delta t)^2} = 1$$

<u>Observation:</u>

- Mathematically, working with finite dt until taking the limit at the end is a rigorous approach in which every step is properly justified.
- “Formal” derivations are not rigorous but are much simpler.

More on Multivariate Gaussian

1 Revisit a Key Result

Theorem 1. *(An affine mapping of a Gaussian is a Gaussian)*
Let $X \sim N(\mu, \Sigma) \in \mathbb{R}^n$. *Consider* $Y \equiv AX + b$ *where* $A \in \mathbb{R}^{m\times n}$ *and* $b \in \mathbb{R}^m$. *We have*

$$\boxed{Y \sim N(\mu_Y, \Sigma_{YY}), \quad \mu_Y = A\mu + b, \quad \Sigma_{YY} = A\Sigma A^T}$$

2 Orthogonality and independence

Corollary 1.1. *(a strong result based on a strong condition.)*

Let $A \in \mathbb{R}^{m\times n}$ *be a matrix with orthogonal rows. In the matrix form, A satisfies*

$$AA^T = \Lambda = \text{diag}(d_1^2, d_2^2, \ldots, d_m^2), \qquad d_i = \|a_{i,:}\|$$

Here we do not require $\|a_{i,:}\| = 1$. *Then for* $Z \sim N(0, I_n)$, *we have*

$$X = AZ \sim N(0, AA^T) = N(0, \Lambda)$$

That is, the components of $X = AZ$ *are independent Gaussians. In particular*

$$\boxed{\Big((X_1, X_2, \ldots, X_{m-1})\Big| X_m = x_m\Big) \sim (X_1, X_2, \ldots, X_{m-1})}$$

If we just want the independence between $(X_1, X_2, \ldots, X_{m-1})$ and X_m, a much weaker condition will do, which is much easier to check practically.

Corollary 1.2. *(A useful corollary based on a weak condition.)*

Let $A \in \mathbb{R}^{m\times n}$, $Z \sim N(0, I_n) \in \mathbb{R}^n$ *the standard isotropic normal, and* $X = AZ \sim N(0, AA^T) \in \mathbb{R}^m$. *Suppose the covariance between* $(X_1, X_2, \ldots, X_{m-1})$ *and* X_m *is zero.*

$$Cov\big(X_j X_m\big) = E\Big((X_j - E(X_j))(X_m - E(X_m))\Big) = 0, \quad 1 \leqslant j \leqslant (m-1)$$

Then $(X_1, X_2, \ldots, X_{m-1})$ *is independent of* X_m.

$$\boxed{\Big((X_1, X_2, \ldots, X_{m-1})\Big| X_m = x_m\Big) \sim (X_1, X_2, \ldots, X_{m-1})}$$

<u>Example:</u> Let $Z = (Z_1, Z_2) \sim N(0, I_2)$, $X_1 = aZ_1 - aZ_2$ and $X_2 = aZ_1 + aZ_2$. We have

$$X_1 \sim N(0, 2a^2), \qquad X_2 \sim N(0, 2a^2), \qquad \text{Cov}\big(X_1 X_2\big) = 0$$

Then X_1 and X_2 are independent. We can use it to calculate conditional distribution of Z_1.

$$\big(X_1 \big| X_2 = x_2\big) \sim X_1 \sim N(0, 2a^2),$$

$$\Big(aZ_1 \Big| X_2 = x_2\Big) \sim \Big(\frac{1}{2}(X_1 + X_2)\Big| X_2 = x_2\Big) \sim \frac{1}{2}(X_1 + x_2) \sim N(\frac{x_2}{2}, \frac{\text{Var}(X_1)}{4})$$

$$\boxed{\big(aZ_1 \big| aZ_1 + aZ_2 = x_2\big) \sim N(\frac{x_2}{2}, \frac{a^2}{2})}$$

We apply it to $X_1 = \Delta W_0 - \Delta W_1$ and $X_2 = \Delta W_0 + \Delta W_1$.

$$\text{Var}(X_1) = 2\Delta t, \qquad \text{Cov}(X_1 X_2) = 0 \;\; \leftarrow \;\; \text{This is the key step}$$

$$\boxed{\big(\Delta W_0 \big| \Delta W_0 + \Delta W_1 = x_2\big) \sim N(\frac{x_2}{2}, \frac{\Delta t}{2})}$$

This result is useful in constrained sample refinement of $W(t)$. Note that when checking the key step above, we don't need to write out (X_1, X_2) explicitly in terms of $Z_1 = \frac{1}{\sqrt{\Delta t}}\Delta W_0$ and $Z_2 = \frac{1}{\sqrt{\Delta t}}\Delta W_1$. The isotropic $(Z_1, Z_2) \sim N\big(0, (\Delta t) I_2\big)$ stays in the background.

3 Brownian bridge

A Brownian bridge is a constrained Wiener process.

$$\Big(\{W(t),\ 0 < t < T\}\Big| W(T) = w_T\Big)$$

Remark. *A constrained Wiener process is a stochastic process. It has randomness.*

Theorem 2. *(Brownian bridge)*

$$\underbrace{\Big(\{W(t),\ 0 < t < T\}\Big| W(T) = w_T\Big)}_{W(t)\text{ is constrained.}} \sim \underbrace{\{W(t) + \frac{t}{T}(w_T - W(T)),\ 0 < t < T\}}_{W(t)\text{ is unconstrained.}}$$

Remark. *For a random vector in $\mathbb{R}^N$, its distribution is completely described by the (joint) density, which is a function of N variables. A stochastic process is infinite dimensional. To show that two stochastic processes $X(t)$ and $Y(t)$ have the same distribution, we need to show that for any $n > 0$ and any set of $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$, the two random vectors below have the same distribution.*

$$(X(t_1), X(t_2), \ldots, X(t_n)) \sim (Y(t_1), Y(t_2), \ldots, Y(t_n))$$

Proof. To prove the expression of Brownian bridge given in the theorem, we need to show that for any $n > 1$ and any set of $0 = t_0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n = T$, we have

$$\boxed{\begin{aligned}&\underbrace{\Big((W(t_1), W(t_2), \ldots, W(t_{n-1}))\Big| W(t_n) = w_T\Big)}_{W(t)\text{ is constrained.}} \\ &\quad \sim (B(t_1), B(t_2), \ldots, B(t_{n-1})), \quad \underbrace{B(t) = W(t) + \frac{t}{T}(w_T - W(T))}_{W(t)\text{ is unconstrained.}}\end{aligned}}$$

The background isotropic $(Z_1, Z_2, \ldots, Z_n)$ are defined as

$$\Delta t_j = t_{j+1} - t, \qquad \Delta W_j = W(t_{j+1}) - W(t_j)$$
$$Z_{j+1} = \frac{1}{\sqrt{\Delta t_j}} \Delta W_j \sim N(0,1), \quad 0 \leqslant j \leqslant (n-1)$$

We find $(n-1)$ linear combinations of $\{Z_j\}$ that are independent of $W(t_n)$. As pointed out previously, we do not need to explicitly work with $(Z_1, Z_2, \ldots, Z_n)$.

$$W(t_n) = \underbrace{W(t_j) + \big(W(t_n) - W(t_j)\big)}_{\text{valid for } 1 \leqslant j \leqslant (n-1)} \text{ different ways of writing } W(t_n)$$
$$X_j = (T - t_j) W(t_j) - t_j \big(W(t_n) - W(t_j)\big), \quad 1 \leqslant j \leqslant (n-1)$$
$$\mathrm{Cov}\big(X_j W(t_n)\big) = E\big(X_j W(t_n)\big) = 0, \quad 1 \leqslant j \leqslant (n-1)$$

It follows that

$$\Big((X_1, X_2, \ldots, X_{n-1})\Big| W(t_n) = w_T\Big) \sim (X_1, X_2, \ldots, X_{n-1})$$

We rewrite the relation between $\{W(t_j),\ 1 \leqslant j \leqslant (n-1)\}$, $W(t_n)$ and $\{X_j,\ 1 \leqslant j \leqslant (n-1)\}$.

$$X_j = T\, W(t_j) - t_j W(t_n), \quad 1 \leqslant j \leqslant (n-1)$$
$$W(t_j) = \frac{1}{T} X_j + \frac{t_j}{T} W(t_n), \quad 1 \leqslant j \leqslant (n-1)$$

Under the constraint $W(t_n) = w_T$, we have $W(t_j) = \frac{1}{T} X_j + \frac{t_j}{T} w_T$ and

$$\underbrace{\Big((W(t_1), W(t_2), \ldots, W(t_{n-1}))\Big| W(t_n) = w_T\Big)}_{W(t)\text{ is constrained.}} \sim (B_1, B_2, \ldots, B_{n-1}), \quad \underbrace{B_j = \frac{1}{T} X_j + \frac{t_j}{T} w_T}_{\{X_j\}\text{ are unconstrained.}}$$

Writing $\{X_j\}$ in terms of unconstrained $\{W_j\}$ gives $X_j = T\, W(t_j) - t_j W(t_n)$ and

$$B_j = W(t_j) + \frac{t_j}{T}\big(w_T - W(t_n)\big) = B(t_j), \quad B(t) = W(t) + \frac{t}{T}(w_T - W(T))$$

Therefore, we arrive at the expression of Brownian bridge. $\square$

Gaussian Process

1 Preliminaries

Definition 1. *(Gaussian process)*
A stochastic process $\{X(t) : t \in (-\infty, +\infty)\}$ is called a Gaussian process if for any $n > 0$ and any set of $\{t_1, t_2, \ldots, t_n\} \in (-\infty, +\infty)$, the random variable vector $\big(X(t_1), X(t_2), \ldots, X(t_n)\big)$ has a multivariate Gaussian distribution.

$$\big(X(t_1), X(t_2), \ldots, X(t_n)\big) \sim N(\mu_{(t_1,t_2,\ldots,t_n)}, \Sigma_{(t_1,t_2,\ldots,t_n)})$$

where the mean vector μ and the covariance matrix Σ vary with $(t_1, t_2, \ldots, t_n)$.

A Gaussian process is complete specified by two functions below.

- *The mean function: $m(t) = E\big(X(t)\big)$.*
- *The covariance function (kernel): $k(t,s) = Cov\big(X(t), X(s)\big)$.*

The short notation for a Gaussian process is

$$X(t) \sim \mathcal{GP}\big(m(t), k(t,s)\big)$$

where $X(t)$ represent a random sample/realization of the Gaussian process.

Definition 2. *(Stationary process)*
A stochastic process $\{X(t)\}$ is called stationary in strong sense if

$$\big(X(t_1), X(t_2), \ldots, X(t_n)\big) \sim \big(X(t_1+t), X(t_2+t), \ldots, X(t_n+t)\big) \quad \textit{for all } t$$

That is, the distribution of $\big(X(t_1+t), X(t_2+t), \ldots, X(t_n+t)\big)$ is independent of t.

A stochastic process $\{X(t)\}$ is called stationary in weak sense if

$$E\big(X(t)\big) = \mu, \quad Cov(X(t), X(t+\tau)) = R(\tau) \quad \textit{independent of } t$$

Remark. *For a Gaussian process, "weak sense" is equivalent to "strong sense'.*

<u>Examples:</u>

- The Wiener process is a (non-stationary) Gaussian process

$$m(t) = 0, \qquad k(t,s) = \min(t,s)$$

- The white noise is a Gaussian process (in the distribution sense)

$$m(t) = 0, \qquad k(t,s) = \delta(t-s)$$

- A Gaussian process with the exponential kernel

$$m(t) = 0, \qquad k(t,s) = \sigma^2 \exp(\frac{-|t-s|}{\ell})$$

  We will see that the Ornstein-Uhlenbeck process, defined by a SDE, is such a process.

- A Gaussian process with the squared exponential kernel

$$m(t) = 0, \qquad k(t,s) = \sigma^2 \exp(\frac{-(t-s)^2}{2\ell^2})$$

  The squared exponential kernel is also known as the Radial Basis Function (RBF) kernel. RBF kernel is widely used in machine learning.

2 Revisit a Key Result

Theorem 1. *(Conditional distribution of X when Y is fixed). Let*

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}), \qquad X \in \mathbb{R}^m, \quad Y \in \mathbb{R}^n$$

Here m and n may be different. Then we have

$$\boxed{\begin{aligned} &(X|Y=y) \sim N(\mu_{X|Y},\, \Sigma_{X|Y}) \\ &\qquad \mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y-\mu_Y) \\ &\qquad \Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{aligned}} \tag{1}$$

This is the foundation for Gaussian process inference.

3 Framework of Gaussian Process Inference

We consider two situations.

1. Suppose $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ are jointly Gaussian. When Y is observed, we can incorporate the observed Y to calculate the conditional distribution $(X|Y=y)$.

2. Suppose $X \in \mathbb{R}^m \sim N(\mu_X, \Sigma_{XX})$. Let $Y = WX \in \mathbb{R}^n$, $W \in \mathbb{R}^{n\times m}$. Y contains n linear combinations of X. It follows that X and Y are jointly Gaussian.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \underbrace{\begin{bmatrix} I \\ W \end{bmatrix}}_{A} X \sim N(A\mu_X,\; A\Sigma_{XX}A^T) = N(\begin{bmatrix} \mu_X \\ W\mu_X \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XX}W^T \\ W\Sigma_{XX} & WW^T \end{bmatrix})$$

When Y is observed, we can incorporate the observed Y to calculate the conditional distribution $(X|Y=y)$. That is, we do not need to observe X directly. Any observation related to X can be used to narrow the distribution of X.

4 Problem 1: reconstructing function $y = f(t)$ based on observed $\{y_j\}$

Let $f(t)$ be a smooth unknown function that is somewhat vertically centered around the t-axis (if we have data points or some knowledge about $f(t)$, we can shift $f(t)$ vertically to make it reasonably centered). Consider the Gaussian process with zero mean and the RBF kernel.

$$\mathcal{GP}\big(0, k_{\text{RBF}}(t,s)\big), \quad k_{\text{RBF}}(t,s) = \sigma^2 \exp(\frac{-(t-s)^2}{2\ell^2})$$

We view the underlying unknown function $f(t)$ as a random realization of $\mathcal{GP}\big(0, k_{\text{RBF}}(t,s)\big)$.

$$\underbrace{f(t) \sim \mathcal{GP}\big(0, k_{\text{RBF}}(t,s)\big)}_{\text{distribution over functions}}$$

<u>Caution:</u> Here $f(t)$ represents both the underlying unknown function (which has no randomness) and a random sample of $\mathcal{GP}\big(0, k_{\text{RBF}}(t,s)\big)$. This view corresponds to the framework of repeated experiments for the situation where we assign a probability to a deterministic but unknown result. (For example, a given person having cancer or not at the present time is deterministic). At any t_*, $f(t_*)$ is a random variable; the marginal distribution of $f(t_*)$ is Gaussian.

$$\underbrace{f(t_*) \sim N\big(0, \sigma^2\big)}_{\text{distribution on } \mathbb{R}}$$

We use the mean of $f(t_*)$ to predict the underlying unknown realization. The uncertainty of the prediction is described by twice the standard deviation (95% CI).

$$f_{\text{pred}}(t_*) = E\big(f(t_*)\big) \pm 2\sqrt{\text{Var}\,\big(f(t_*)\big)} = 0 \pm 2\sigma$$

Here $f(t_*)$ on the RHS is a random variable; $f_{\text{pred}}(t_*)$ on the LSH is the prediction of the underlying unknown function. In the absence of data, this prediction is reasonable.

Now suppose we have a data set containing observed $\{y_j = f(t_j),\ 1 \leqslant j \leqslant n\}$. We consider a discrete version of stochastic process $f(t)$ on $\{\tau_j,\ 1 \leqslant j \leqslant m\}$. Let

$$\begin{aligned} X &= \big(f(\tau_1), f(\tau_2), \ldots, f(\tau_m)\big) \in \mathbb{R}^m \\ Y &= \big(f(t_1), f(t_2), \ldots, f(t_n)\big) \in \mathbb{R}^n, \quad y = \big(y_1, y_2, \ldots, y_n\big) \in \mathbb{R}^n \end{aligned}$$

Before we have data, the joint distribution of X and Y is given by $\mathcal{GP}\big(0, k_{\text{RBF}}(t,s)\big)$.

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(0, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix})$$

$$(\Sigma_{XX})_{(i,j)} = k_{\text{RBF}}(\tau_i, \tau_j), \quad (\Sigma_{XY})_{(i,j)} = k_{\text{RBF}}(\tau_i, t_j)$$
$$(\Sigma_{YY})_{(i,j)} = k_{\text{RBF}}(t_i, t_j)$$

The conditional distribution of X given the data $Y = y$ is described by the theorem.

$$\boxed{\begin{aligned} &(X|Y = y) \sim N(\mu_{X|Y},\ \Sigma_{X|Y}) \\ &\qquad \mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \quad \mu_X = 0, \quad \mu_Y = 0 \\ &\qquad \Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{aligned}}$$

From this conditional distribution, we can do two things.

- We draw a discrete sample of stochastic process $f(t)$ on $\{\tau_j,\ 1 \leqslant j \leqslant m\}$, constrained by data. The sample gives a possible realization of the underlying unknown function.

- At any t_*, the marginal distribution of $f(t_*)$ conditional on the data is

$$\big(f(t_*)\big|Y = y\big) \sim N(\mu_{t_*|Y}, \sigma^2_{t_*|Y})$$

  We use the marginal distribution to predict the underlying unknown function at t_*.

$$f_{\text{pred}}(t_*) = E\big(f(t_*)\big|Y = y\big) \pm 2\sqrt{\text{Var}\left(f(t_*)\big|Y = y\right)} = \mu_{t_*|Y} \pm 2\sigma_{t_*|Y}$$

  Note that $\mu_{t_*|Y}$ is a function of t_*. It is the result of training Gaussian process using the data $Y = y$. It is the reconstructed function $f(t_*)$ based on the data.

Observation. *Why do we use Gaussian process instead of a conventional interpolation, for example, linear interpolation? Answer: Gaussian process is very versatile.*

- *Gaussian process works when we have data on an irregular grid $\{t_j, 1 \leqslant j \leqslant n\}$. This is especially convenient in 2D and 3D. Think about the task of reconstructing a 2D function from data on a set of random points.*

- *Gaussian process works even if we do not have direct observations of $f(t)$. We can use any data to narrow the distribution to reconstruct $f(t)$ (see Problems 2 and 3 below).*

- *Gaussian process is a flexible framework for extracting information from various data.*

5 Problem 2: reconstructing function $y = f(t)$ based on observed Wy

Recall that in Problem 1, $y = \big(y_1, y_2, \ldots, y_n\big) \in \mathbb{R}^n$ is observed where $y_j = f(t_j)$. Now suppose we don't have direct observation on y. Instead, we have data on $\tilde{n}$ linear combination of $\{y_j\}$.

$$\tilde{y} = Wy, \quad W \in \mathbb{R}^{\tilde{n} \times n}$$

The joint distribution of X and $\tilde{Y} = WY$ is

$$\begin{bmatrix} X \\ \tilde{Y} \end{bmatrix} = \underbrace{\begin{bmatrix} I & \\ & W \end{bmatrix}}_{A} \begin{bmatrix} X \\ Y \end{bmatrix} \sim N(0, A \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} A^T) \equiv N(0, \begin{bmatrix} \Sigma_{XX} & \Sigma_{X\tilde{Y}} \\ \Sigma_{\tilde{Y}X} & \Sigma_{\tilde{Y}\tilde{Y}} \end{bmatrix})$$

$$\Sigma_{X\tilde{Y}} = \Sigma_{XY} W^T, \quad \Sigma_{\tilde{Y}X} = W\Sigma_{YX}, \quad \Sigma_{\tilde{Y}\tilde{Y}} = W\Sigma_{YY}W^T$$

Note that W (the coefficient matrix in linear combinations) is given as part of the data. The conditional distribution of X given the data $\tilde{Y} = \tilde{y}$ is described by the theorem.

$$\boxed{\begin{aligned} & (X|\tilde{Y} = \tilde{y}) \sim N(\mu_{X|\tilde{Y}}, \, \Sigma_{X|\tilde{Y}}) \\ & \qquad \mu_{X|\tilde{Y}} = \mu_X + \Sigma_{X\tilde{Y}} \Sigma_{\tilde{Y}\tilde{Y}}^{-1} (\tilde{y} - \mu_{\tilde{Y}}), \quad \mu_X = 0, \quad \mu_{\tilde{Y}} = 0 \\ & \qquad \Sigma_{X|\tilde{Y}} = \Sigma_{XX} - \Sigma_{X\tilde{Y}} \Sigma_{\tilde{Y}\tilde{Y}}^{-1} \Sigma_{\tilde{Y}X} \end{aligned}}$$

Vector X is a discrete version of the stochastic process $f(t)$. Once we obtain the conditional distribution of X given the data, we can do predictions as described in Problem 1.

6 Several related issues

6.1 Numerical singularity

For a large data set, the covariance matrix Σ_{YY} is ill-conditioned. Practically, Σ_{YY} is rank deficient in the standard IEEE double precision ($\varepsilon_{\text{mach}} = 10^{-16}$). To avoid numerical singularity, we need to "regulate" Σ_{YY} when we calculating its inverse. The most straightforward regulation is the ridge method. We simply add a small positive multiple of identity I_n to the covariance matrix Σ_{YY} to make it soundly positive definite even in finite precision.

$$\text{Ridge method:} \ \ \Sigma_{YY} \xrightarrow{\text{adding a ridge}} (\Sigma_{YY} + \lambda I_n), \qquad \lambda = 10^{-6} \text{ to } 10^{-8}$$

Mathematically, the added ridge corresponds to the situation where the observed value of Y contains an i.i.d point-wise Gaussian noise of variance λ. Let Y be the true value and Y_{obs} the observed value of Y. Y_{obs} and Y are related by

$$Y_{\text{obs}} = Y + \varepsilon, \qquad \varepsilon \sim N(0, \lambda I_n)$$

The covariance matrix of Y_{obs} is

$$\begin{aligned} & \text{Cov}(Y_{\text{obs}}) = \text{Cov}(Y) + \text{Cov}(\varepsilon) \\ \Longrightarrow \quad & \Sigma_{Y_{\text{obs}}Y_{\text{obs}}} = \Sigma_{YY} + \lambda I_n \end{aligned}$$

6.2 Hyper-parameters in Gaussian process

When applying a Gaussian process, we need to know the mean $m(t)$ and the kernel $k(t,s)$. In most situations, we set $m(t) \equiv 0$. This works well if the underlying function $f(t)$ is roughly centered vertically. Alternatively, if we know $f(t)$ is off center and has a linear trend, we can consider a tilted function $\tilde{f}(t) \equiv f(t) - (c_0 + c_1 t)$ and still work with $m(t) \equiv 0$. Coefficients c_0 and c_1 may be from our prior knowledge on $f(t)$ or from a linear regression on the data $\{(t_j, y_j = f(t_j))\}$.

For all smooth functions, we select the RBF kernel which has two hyper-parameters.

$$f(t) \sim \mathcal{GP}\big(0, k_{\mathrm{RBF}}(t,s)\big), \qquad k_{\mathrm{RBF}}(t,s) = \sigma^2 \exp\Big(\frac{-(t-s)^2}{2\ell^2}\Big)$$

Here σ represents the amplitude of $f(t)$ and ℓ the characteristic time scale in $f(t)$. We may set (σ, ℓ) based on our prior knowledge on $f(t)$. Alternatively, we can estimate (σ, ℓ) from a MLE inference on the data $\{(t_j, y_j = f(t_j))\}$.

$$\Big(\{y_j\}\Big|(\sigma,\ell),\{t_j\}\Big) \sim N\big(0,\ \{k(t_i,t_j|\sigma,\ell)\}\big)$$

$$(\sigma_{(\mathrm{MLE})}, \ell_{(\mathrm{MLE})}) = \arg\max_{(\sigma,\ell)} \rho\Big(\{y_j\}\Big|(\sigma,\ell),\{t_j\}\Big)$$

7 Problem 3: reconstructing $\{x_t\}$ of a dynamical system based on observed Wx_t

Consider the situation where the system state x_t is governed by a discrete time dynamical system that is known but the full state x_t is not directly observed. Instead, a lower dimensional projection, $y_t = Wx_t$, is observed. The mathematical formulation is as follows.

$$\begin{cases} x_{t+1} = Ax_t + b, \quad x_t \in \mathbb{R}^d, \quad A \in \mathbb{R}^{d\times d}, \quad b \in \mathbb{R}^d \\ y_t = Wx_t, \quad W \in \mathbb{R}^{p\times d}, \quad p < d \end{cases} \tag{2}$$

where matrix A and vector b are known, the system state x_t is unknown, and the output $\{y_t\}$ is observed. The goal is to estimate the full state trajectory $\{x_t\}$ from the observed $\{y_t\}$. System (2) is a simplified formulation. A more realistic version in application is

$$\begin{cases} x_{t+1} = A(u_t,t)x_t + b(u_t,t) \\ y_t = Wx_t \\ u_t = G\big(\{y_\tau\}_{\tau\leqslant t}, t\big) \end{cases}$$

where the control input u_t is calculated based the observed $\{y_\tau\}_{\tau\leqslant t}$ up to the current time t. The calculation in function $G(\ )$ involves estimating the state trajectory $\{x_\tau\}_{\tau\leqslant t}$. Here we illustrate the basic idea of estimating $\{x_t\}$ in the simplified system (2).

Since $\{x_t\}$ is unknown, we view it as a random realization of a Gaussian process. In this problem, before we have any data, we have more prior knowledge than we do in Problems 1 and 2 above. In Problems 1 and 2, we have very little information about function $f(t)$ before we have data. Here we know a very important property of $\{x_t\}$: it is governed by a given dynamical system. To comply with the dynamical system, we do not set the mean function $m(t)$ or the kernel function $k(t, s)$. Instead, we let the dynamical system evolve the Gaussian distribution forward in time, from t to $t+1$. We first introduce notations.

$$\big(x_t \big| \{y_\tau\}_{\tau \leqslant (t-1)}\big) \sim N(\mu_{t|(t-1)}, \Sigma_{t|(t-1)}) :$$

distribution of x_t given observation up to $(t-1)$

$$\big(x_t \big| \{y_\tau\}_{\tau \leqslant t}\big) \sim N(\mu_{t|t}, \Sigma_{t|t}) :$$

distribution of x_t given observation up to t

We start $\big(x_t \big| \{y_\tau\}_{\tau \leqslant (t-1)}\big)$ at time $t = 0$ with an isotropic Gaussian distribution.

$$\boxed{\mu_{0|(-1)} = 0, \qquad \Sigma_{0|(-1)} = \sigma^2 I_d} \tag{3}$$

Note that observation starts at time 0. There is no observation for $\tau \leqslant (-1)$.

At time t, we start the computational cycle with $\big(\mu_{t|(t-1)}, \Sigma_{t|(t-1)}\big)$. For $t = 0$, this is already specified in (3) in the prior of x_0. Given the observation up to time $(t-1)$, before we obtain the new observation at time t, the conditional distribution of x_t is

$$\big(x_t \big| \{y_\tau\}_{\tau \leqslant (t-1)}\big) \sim N(\mu_{t|(t-1)}, \Sigma_{t|(t-1)})$$

Using $y_t = W x_t$, we write out the conditional joint distribution of (x_t, y_t).

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} I_d \\ W \end{bmatrix} x_t \quad \leftarrow \text{ This is a linear mapping.}$$

$$\left(\begin{bmatrix} x_t \\ y_t \end{bmatrix} \middle| \{y_\tau\}_{\tau \leqslant (t-1)} \right) \sim N\left(\begin{bmatrix} \mu_{t|(t-1)} \\ W\mu_{t|(t-1)} \end{bmatrix}, \begin{bmatrix} \Sigma_{t|(t-1)} & \Sigma_{t|(t-1)} W^T \\ W\Sigma_{t|(t-1)} & W\Sigma_{t|(t-1)} W^T \end{bmatrix} \right)$$

With the new observation y_t. The conditional distribution of x_t is updated as follows.

$$\big(x_t \big| \{y_\tau\}_{\tau \leqslant t}\big) \sim N(\mu_{t|t},\ \Sigma_{t|t})$$

$$\boxed{\begin{cases} \mu_{t|t} = \mu_{t|(t-1)} + \big(\Sigma_{t|(t-1)} W^T\big)\big(W\Sigma_{t|(t-1)} W^T\big)^{-1}\big(y_t - W\mu_{t|(t-1)}\big) \\ \Sigma_{t|t} = \Sigma_{t|(t-1)} - \big(\Sigma_{t|(t-1)} W^T\big)\big(W\Sigma_{t|(t-1)} W^T\big)^{-1}\big(W\Sigma_{t|(t-1)}\big) \end{cases}} \tag{4}$$

In the above, we have used the expression for conditional distribution $(X|Y = y)$.

$$\mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y),$$

$$\Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$$

$\mu_{t|t}$ and $\Sigma_{t|t}$ obtained in (4) describe the mean estimate of x_t and its uncertainty.

$$x_{t,\text{est}} = \mu_{t|t} + (\text{elliptic confidence region based on } \Sigma_{t|t})$$

In particular the root mean squared error in $x_{t,\text{est}}$ is predicted to be

$$\text{RMSE}_t = \sqrt{\text{trace}(\Sigma_{t|t})}$$

In feedback control, the estimated state trajectory up to time t, $\{x_\tau\}_{\tau \leqslant t}$, is used in the calculation of control input u_t. From time t to $(t+1)$, the dynamical system gives

$$\underbrace{\left(x_{t+1}\big|\{y_\tau\}_{\tau\leqslant t}\right)}_{\sim N(\mu_{(t+1)|t},\, \Sigma_{(t+1)|t})} = A \underbrace{\left(x_t\big|\{y_\tau\}_{\tau\leqslant t}\right)}_{\sim N(\mu_{t|t},\, \Sigma_{t|t})} + b \quad \leftarrow \text{ This is an affine mapping.}$$

$$\boxed{\begin{cases} \mu_{(t+1)|t} = A\mu_{t|t} + b \\ \Sigma_{(t+1)|t} = A\Sigma_{t|t}A^T \end{cases}} \tag{5}$$

(5) is the end of the cycle at time t. With the $(\mu_{(t+1)|t}, \Sigma_{(t+1)|t})$ obtained in (5), we start the new cycle at time $(t+1)$. Note that this process estimates the system state x_t as new observation $y_t = Wx_t$ arrives. It has the behavior of an on-line filter. It is the basic idea of Kalman filter. We summarize the iterative process of estimating x_t as follows.

0. At $t = 0$, $\left(\mu_{0|(-1)}, \Sigma_{0|(-1)}\right)$ of distribution $\left(x_0\big|\{y_\tau\}_{\tau\leqslant(-1)}\right)$ is given in (3).
1. At time t, start a new cycle with $\left(\mu_{t|(t-1)}, \Sigma_{t|(t-1)}\right)$ of distribution $\left(x_t\big|\{y_\tau\}_{\tau\leqslant(t-1)}\right)$.
2. Apply (4) to calculate $\left(\mu_{t|t}, \Sigma_{t|t}\right)$ of distribution $\left(x_t\big|\{y_\tau\}_{\tau\leqslant t}\right)$.
3. Apply (5) to calculate $\left(\mu_{(t+1)|t}, \Sigma_{(t+1)|t}\right)$ of distribution $\left(x_{t+1}\big|\{y_\tau\}_{\tau\leqslant t}\right)$.
4. Go back to Step 1 to start a new cycle at time $(t+1)$.

Below we write one cycle of Kalman filter as an algorithm

Algorithm 1 One cycle of Kalman filter for estimating x_t

function KF_1CYC$(\mu_{t|(t-1)}, \Sigma_{t|(t-1)}, y_t)$ ▻ y_t is the observation at t.

Update $\mu_{t|t}$ and $\Sigma_{t|t}$

$\quad \mu_{t|t} \leftarrow \mu_{t|(t-1)} + \left(\Sigma_{t|(t-1)}W^T\right)\left(W\Sigma_{t|(t-1)}W^T\right)^{-1}\left(y_t - W\mu_{t|(t-1)}\right)$

$\quad \Sigma_{t|t} \leftarrow \Sigma_{t|(t-1)} - \left(\Sigma_{t|(t-1)}W^T\right)\left(W\Sigma_{t|(t-1)}W^T\right)^{-1}\left(W\Sigma_{t|(t-1)}\right)$

Update $\mu_{(t+1)|t}$ and $\Sigma_{(t+1)|t}$

$\quad \mu_{(t+1)|t} \leftarrow A\mu_{t|t} + b$

$\quad \Sigma_{(t+1)|t} \leftarrow A\Sigma_{t|t}A^T$

Estimate x_t and uncertainty

$\quad x_{t,\text{est}} \leftarrow \mu_{t|t}$ ▻ Estimated x_t

$\quad \text{RMSE}_t \leftarrow \sqrt{\text{trace}(\Sigma_{t|t})}$ ▻ Predicted error

$\quad$ **return** $x_{t,\text{est}},\ \text{RMSE}_t,\ \mu_{(t+1)|t},\ \Sigma_{(t+1)|t}$

AM216 Stochastic Differential Equations

Lecture 07

List of topics in this lecture

- Wiener process is continuous in probability
- Ornstein-Uhlenbeck process (OU), Stokes law, thermal excitations
- Solution of particle velocity in OU, colored noise, convergence to a white noise
- Fluctuation-dissipation theorem, Maxwell-Boltzmann distribution
- Solution of particle position in OU

Recap

Energy spectrum density (ESD), power spectrum density (PSD)

Stationary stochastic process, auto-correlation function (ACF)

Wiener-Khinchin theorem: PSD is Fourier transform of ACF

Definition of white noise: PSD is constant in frequency domain

Calculating ACF and PSD of $Z(t) \equiv dW/dt$

Constrained Wiener process (Bayes theorem)

$$\rho\left(W(a+\frac{h}{2})=x \middle| W(a)=y_a \text{ and } W(a+h)=y_b \right) \sim N\left(\frac{y_a+y_b}{2}, \frac{h}{4} \right)$$

This is very useful in refining a discrete sample path of $W(t)$.

The Wiener process is continuous in probability

Recall the continuity of a regular function. $f(t)$ is continuous at t if for any $\varepsilon > 0$,

$\left| f(t+h) - f(t) \right| \geq \varepsilon$ is impossible when h is small enough.

Definition (continuity of $F(t, \omega)$ in probability)

Intuitively, $F(t, \omega)$ is continuous in probability if for any $\varepsilon > 0$,

$\left|F(t+h,\omega)-F(t,\omega)\right| \geq \varepsilon$ is almost impossible when h is small enough.

More precisely, $F(t, \omega)$ is continuous in probability if for any $\varepsilon>0$,

$$\lim_{h\to 0} \Pr\left(\left|F(t+h,\omega)-F(t,\omega)\right| \geq \varepsilon\right) = 0$$

Theorem: $W(t)$ is continuous in probability.

Proof: To prove the theorem, we need

Chebyshev-Markov inequality:

Let X be a random variable X. For $\alpha > 0$, we write $E(|X|^\alpha)$ as

$$E\left(|X|^\alpha\right) = \int |x|^\alpha \rho(x)dx \geq \int_{|x|\geq\varepsilon} |x|^\alpha \rho(x)dx \geq$$

$$\geq \varepsilon^\alpha \int_{|x|\geq\varepsilon} \rho(x)dx = \varepsilon^\alpha \Pr\left(|X| \geq \varepsilon\right)$$

==> $\boxed{\Pr\left(|X|\geq\varepsilon\right) \leq \frac{1}{\varepsilon^\alpha} E\left(|X|^\alpha\right)}$ This is valid for any $\alpha > 0$.

This is called the Chebyshev-Markov inequality.

We apply the Chebyshev-Markov inequality to $X = W(t+h)-W(t)$ with $\alpha = 2$.

$$\Pr\left(|W(t+h)-W(t)|\geq\varepsilon\right) \leq \frac{E\left(|W(t+h)-W(t)|^2\right)}{\varepsilon^2}$$

$$= \frac{h}{\varepsilon^2} \to 0 \quad \text{as } h\to 0$$

Thus, the Wiener process $W(t)$ is continuous in probability.

Similar to the continuity of $F(t, \omega)$ in probability, we can define the convergence of sequence $\{X_n(\omega)\}$ in probability.

Definition (convergence of $\{X_n(\omega)\}$ in probability)

As $n \to +\infty$, $\{X_n(\omega)\}$ converges to q in probability if for any $\varepsilon>0$,

$$\lim_{n\to+\infty} \Pr\left(\left|X_n(\omega)-q\right| \geq \varepsilon\right) = 0$$

Theorem (a sufficient condition for convergence in probability)

Suppose $\lim\limits_{n\to 0} E\left(X_n(\omega)\right) = q$ and $\lim\limits_{n\to 0} \mathrm{var}\left(X_n(\omega)\right) = 0$.

Then $\{X_n(\omega)\}$ converges to q in probability as $n \to +\infty$.

Proof: (homework problem)

Ornstein-Uhlenbeck Process

Consider the stochastic motion of a small particle in water (as Robert Brown observed the motion of pollen particles in water under a microscope).

For simplicity, we discuss the one-dimensional motion. Let

X = position of the particle

Y = velocity of the particle

m = mass of the particle

Newton's second law

$$m\frac{dY}{dt} = \text{viscous drag} + \text{Brownian force}$$

We discuss these two forces.

Stokes law (for the viscous drag)

viscous drag = $-b\,Y$

where b is the drag coefficient. For a spherical particle, the drag coefficient is

$b = 6\pi\,\eta\,a$

a = radius of the particle

η = viscosity of the fluid media

A short digression: Pollution particles suspended in air

When an object is dropped in mid-air, it first accelerates downward, driven by the gravity. Then it reaches a steady velocity when the drag force balances the gravitational force. This steady velocity is called the terminal velocity for large objects (such as a spacecraft returning to Earth) or the settling velocity for small particles. Here we focus on small particles. The settling velocity satisfies

$$\underbrace{(6\pi\eta a)V_{\text{settling}}}_{\text{Drag force}} = \underbrace{\left(\frac{4}{3}\pi a^3 \rho_{\text{mass}}\right)g}_{\text{Gravity}}$$

$$\text{==>} \qquad V_{\text{settling}} = \left(\frac{2\rho_{\text{mass}}g}{9\eta}\right)a^2 \propto a^2$$

where the air viscosity is $\eta = 1.8\times10^{-4}\ \text{g(cm)}^{-1}\text{s}^{-1}$.

Consider BUD (budesonide), a drug used in treating asthma. It has $\rho_{\text{mass}} = 1.26\ \text{g/cm}^3$.

For a BUD particle of 0.1 mm in diameter

a = 50 μm ==> V_{settling} = 38 cm/s

For a particle of 10 μm in diameter (PM_{10} particles)

a = 5 μm ==> V_{settling} = 0.38 cm/s

For a particle of 2.5 μm in diameter ($PM_{2.5}$ particles)

a = 1.25 μm ==> V_{settling} = 0.0238 cm/s

With this tiny settling velocity, it takes more than 1 hour for a 2.5 μm particle to descend 1 meter with respect to the surrounding air.

$$T_{1\,\text{meter}} = \frac{1\,\text{meter}}{V_{\text{settling}}} = \frac{100\text{ cm}}{0.0238\text{ cm/s}} = 4200\text{ seconds} = 1.17\text{ hours}$$

Remark:

Small pollution particles are more dangerous for two reasons:

- They stay in air much longer (virtually forever)
- They can pass the filtration system of human body and enter the circulatory system (blood circulation).

End of digression

Back to the discussion of forces.

Thermal excitations (Brownian force)

We model the Brownian force as a white noise.

$$\text{Brownian force} = q\frac{dW}{dt}$$

where the coefficient q is to be determined in the fluctuation-dissipation relation.

The governing equation of the particle

$$m\,dY = \underbrace{-bY\,dt}_{\text{dissipation}} + \underbrace{q\,dW}_{\text{fluctuation}}$$

$$dX = Y\,dt$$

This is called the Ornstein-Uhlenbeck process.

Remark:

Both the viscous drag and the Brownian force on the particle are results from the particle colliding with surrounding fluid molecules: the viscous drag is the mean and the Brownian force is the fluctuations of the random colliding force. As a result, the fluctuation coefficient (q) and the dissipation coefficient (b) are related by the fluctuation-dissipation theorem, which we will discuss later.

Four goals in the discussion of the Ornstein-Uhlenbeck process

1) Solve for $Y(t)$
2) Show that

   A) $Y(t)$ is a colored noise and

   B) $Y(t)$ converges to a white noise as "m converges to zero".
3) Relate q to b (fluctuation-dissipation theorem)
4) Study the behavior of $X(t)$

Goal #1: We solve for $Y(t)$.

For mathematical convenience, we divide the equation by m

$$mdY = -bY\,dt + q\,dW$$

==> $\quad dY = -\beta Y\,dt + \gamma\,dW\,, \qquad \beta = \frac{b}{m}\,,\quad \gamma = \frac{q}{m}$

We use the method of integrating factor. Multiply by $e^{\beta t}$, we get

$$e^{\beta t}dY + \beta e^{\beta t}Y\,dt = \gamma e^{\beta t}dW$$

==> $\quad d\left(e^{\beta t}Y(t)\right) = \gamma e^{\beta t}dW$

Note:

$$\Delta\left(e^{\beta t}Y(t)\right) = e^{\beta(t+\Delta t)}Y(t+\Delta t) - e^{\beta t}Y(t) = e^{\beta t}\left(1+\beta\Delta t + o(\Delta t)\right)\left(Y(t)+\Delta Y\right) - e^{\beta t}Y(t)$$

$$= e^{\beta t}\Delta Y + \beta e^{\beta t}\,Y(t)\Delta t + o(\Delta t)$$

Therefore, $d\left(e^{\beta t}Y(t)\right) = e^{\beta t}dY + \beta e^{\beta t}Y\,dt$ is justified.

Summing over all time intervals gives us

$$e^{\beta t}Y(t) - Y(0) = \int_0^t \gamma e^{\beta s}\,dW(s) \equiv G(t)$$

where the integral of dW is defined as the limit of a Riemann sum.

$$G(t) \equiv \int_0^t \gamma e^{\beta s}\,dW(s) \equiv \lim_{N\to\infty}\sum_{j=0}^{N-1}\gamma e^{\beta s_j}\Delta W_j$$

where $\qquad \Delta s = \frac{t}{N}\,, \qquad s_j = j\,\Delta s\,, \qquad \Delta W_j = W(s_{j+1}) - W(s_j)\,.$

Recall that the sum of independent normal RVs is a normal RV.

$\{\Delta W_j,\ \ j = 0, 1, \ldots, N-1\}$ are independent normal RVs.

==> $G(t) \equiv \lim_{N\to\infty} \sum_{j=0}^{N-1} \gamma e^{\beta s_j} \Delta W_j$ is a normal RV.

The mean and variance of $G(t)$ are

$$E\big(G(t)\big) = \lim_{N\to\infty} \sum_j \gamma e^{\beta s_j} E(\Delta W_j) = 0$$

$$\mathrm{var}\big(G(t)\big) = \lim_{N\to\infty} \sum_j (\gamma e^{\beta s_j})^2 \,\mathrm{var}(\Delta W_j) = \lim_{N\to\infty} \sum_j (\gamma e^{\beta s_j})^2 \Delta s$$

$$= \int_0^t (\gamma e^{\beta s})^2 ds = \gamma^2 \int_0^t e^{2\beta s} ds = \frac{\gamma^2}{2\beta}(e^{2\beta t} - 1)$$

<u>Caution:</u>

For $t > 0$, $\int_0^t \gamma e^{\beta s} dW(s)$ is a sum of independent normal RVs.

For $t < 0$, increments $\{\Delta W_j,\ \ j = 0, 1, \ldots, N-1\}$ are backwards in time and are no longer independent. For now, we only consider $t > 0$.

<u>Summary</u> (distribution of $G(t)$)

$$G(t) \equiv \int_0^t \gamma e^{\beta s} dW(s) \sim N\left(0,\ \frac{\gamma^2}{2\beta}(e^{2\beta t} - 1)\right) \qquad \text{for } t > 0$$

In the above, we just derived a theorem.

<u>Theorem:</u>

$$\int_0^L f(t) dW(t) \sim N\left(0,\ \int_0^L f(t)^2 dt\right)$$

We continue solving for $Y(t)$

$$e^{\beta t} Y(t) - Y(0) = G(t), \qquad G(t) \sim N\left(0,\ \frac{\gamma^2}{2\beta}(e^{2\beta t} - 1)\right) \qquad \text{for } t > 0$$

==> $Y(t) = e^{-\beta t} Y(0) + e^{-\beta t} G(t) \qquad \text{for } t > 0$

==> $\big(Y(t) \big| Y(0) = y_0\big) \sim N\left(e^{-\beta t} y_0,\ \frac{\gamma^2}{2\beta}(1 - e^{-2\beta t})\right) \qquad \text{for } t > 0$

<u>Summary</u> (solution of $Y(t)$)

$$\big(Y(t_0 + t) \big| Y(t_0) = y_0\big) \sim N\left(e^{-\beta t} y_0,\ \frac{\gamma^2}{2\beta}(1 - e^{-2\beta t})\right) \qquad \text{for } t > 0$$

Equilibrium state:

For large t, $Y(t)$ reaches an equilibrium normal distribution.

$$\left(Y(t) \middle| Y(0)=y_0\right) \sim N\left(0,\ \frac{\gamma^2}{2\beta}\right) \qquad \text{for large } t>0, \text{ independent of } y_0.$$

Goal #2: we show that

A) $Y(t)$ is a colored noise and

B) $Y(t)$ converges to a white noise as "m converges to zero".

We assume that the equilibrium has been reached long time ago and $Y(t)$ is already a stationary process. Under this assumption, $Y(t)$ has the equilibrium distribution.

$$Y(t) \sim N\left(0,\ \frac{\gamma^2}{2\beta}\right) \qquad \text{for all } t$$

Goal #2A: We show that $Y(t)$ is a colored noise.

We calculate the autocorrelation function.

$$R(t) \equiv E\left(Y(t)Y(0)\right)$$

We use the law of total expectation.

$$E(Z_1) = E\left(E(Z_1 | Z_2)\right)$$

We select $Z_1 = Y(t)\, Y(0)$ and $Z_2 = Y(0)$. We consider the case of $t > 0$.

$$R(t) \equiv E\left(Y(t)Y(0)\right) = E\left(E\left(Y(t)Y(0) \middle| Y(0)\right)\right) = E\left(Y(0)\cdot E\left(Y(t) \middle| Y(0)\right)\right)$$

Using $E\left(Y(t) \middle| Y(0)\right) = e^{-\beta t}Y(0)$ for $t > 0$ and $Y(t) \sim N\left(0,\ \frac{\gamma^2}{2\beta}\right)$ for all t, we write

$$R(t) = E\left(Y(0)\cdot e^{-\beta t}Y(0)\right) = e^{-\beta t}E\left(Y(0)^2\right) = \frac{\gamma^2}{2\beta}e^{-\beta t} \qquad \text{for } t>0$$

From the definition of auto-correlation function, $R(t)$ is an even function of t:

$$R(-t) \equiv E\left(Y(s-t)Y(s)\right) \qquad \text{for all } s$$

$$\xrightarrow{\text{select } s=t} \ = E\left(Y(0)Y(t)\right) = R(t)$$

Therefore, we obtain

$$R(t) = \frac{\gamma^2}{2\beta}\exp(-\beta|t|) \qquad \text{for } t \in (-\infty,\ +\infty)$$

The corresponding power spectrum density (PSD) is

$$s(\xi)=\frac{\gamma^2}{2\beta}F\left[\exp(-\beta|t|)\right]=\frac{\gamma^2}{2\beta}\cdot\frac{2\beta}{\beta^2+4\pi^2\xi^2}=\frac{\gamma^2}{\beta^2+4\pi^2\xi^2}$$

(Homework problem)

In conclusion, $Y(t)$ is a colored noise.

Goal #2B: We show that $Y(t)$ converges to a white noise as "$m \to 0$"

A simplified story: $m \to 0$ (while b and q stay unchanged)

This corresponds to the situation where the mass density of the particle goes to zero while the particle size is fixed.

Recall that $\beta=\frac{b}{m}$, $\gamma=\frac{q}{m}$.

$$R(t)=\frac{\gamma^2}{2\beta}\exp(-\beta|t|)==\frac{q^2}{b^2}\cdot\underbrace{\frac{b}{m}}_{1/h}\cdot\underbrace{\frac{1}{2}\exp\left(-\frac{b}{m}|t|\right)}_{f(t/h)},\qquad h\equiv\frac{m}{b}$$

We write it as a scaled probability density function.

$$R(t)=\frac{q^2}{b^2}\cdot\frac{1}{h}f\left(\frac{t}{h}\right),\qquad h\equiv\frac{m}{b},\qquad f(u)\equiv\frac{1}{2}\exp\left(-|u|\right)$$

$f(u)$ given above is a probability density function (satisfying $\int f(u)du = 1$).

It follows that

$$\lim_{h\to 0}\frac{1}{h}f\left(\frac{t}{h}\right)=\delta(t)\qquad\text{and}\qquad\lim_{m\to 0}R(t)=\frac{q^2}{b^2}\cdot\delta\left(t\right)$$

Therefore, $\lim\limits_{m\to 0}Y(t)$ is a white noise.

The real story:

Mathematically, the limit above is rigorous. The assumption of mass density converging to zero, however, is not a realistic one in physics.

The mass of a spherical particle is

$$m=\frac{4\pi}{3}\rho_{\text{mass}}a^3$$

where ρ_{mass} is the mass density and a the radius of particle.

In physics, we are interested in the situation where radius $a \to 0$ while ρ_{mass} is fixed.

We need to consider the effect of radius a on coefficients m, b and q.

$$m = \frac{4\pi}{3}\rho_{\text{mass}}\, a^3 = O(a^3) \to 0$$

$$b = 6\pi\eta a = O(a) \to 0$$

$$h \equiv \frac{m}{b} = O(a^2) \to 0$$

$$q = \sqrt{2k_B T b} = O(\sqrt{a}) \to 0$$ (we will derive q shortly)

$$\frac{q^2}{b^2} = O(a^{-1}) \to \infty$$

$$a\frac{q^2}{b^2} = O(1)$$ independent of a.

Consider $\sqrt{a}\,Y(t)$. We have

$$R_{\sqrt{a}Y}(t) = E\left(\sqrt{a}\,Y(s+t)\sqrt{a}\,Y(s)\right) = aE\left(Y(s+t)Y(s)\right) = aR_Y(t) = a\frac{q^2}{b^2}\cdot\frac{1}{h}f\left(\frac{t}{h}\right)$$

$$\lim_{a\to 0} R_{\sqrt{a}Y}(t) = \lim_{a\to 0} a\frac{q^2}{b^2}\cdot\frac{1}{h}f\left(\frac{t}{h}\right) = \left(a\frac{q^2}{b^2}\right)\cdot\delta(t) \quad \text{where} \quad \left(a\frac{q^2}{b^2}\right) = O(1)$$

Therefore, $\lim\limits_{a\to 0}\sqrt{a}\,Y(t)$ is a white noise.

In physics, as radius $a \to 0$, $Y(t)$ converges to a white noise of magnitude $O\left(\frac{1}{\sqrt{a}}\right)$.

Goal #3: We relate the fluctuation coefficient q to the drag coefficient b.

To connect b and q, we need the Maxwell-Boltzmann distribution

Maxwell-Boltzmann distribution

For a system in equilibrium with a thermal bath, we have

$$\rho(Y = y) \propto \exp\left(\frac{-\text{Energy}(Y = y)}{k_B T}\right)$$

where k_B is the Boltzmann constant and

T is the absolute temperature of the thermal bath.

Maxwell-Boltzmann distribution is a universal law applicable to all thermodynamic systems. In our system, Y = velocity and

$$\text{Energy}(Y=y) = \frac{1}{2}m\,y^2$$

The Maxwell-Boltzmann distribution gives us

$$\rho(Y=y) \propto \exp\left(\frac{-\text{Energy}}{k_B T}\right) = \exp\left(\frac{-\frac{1}{2}my^2}{k_B T}\right)$$

We write it into the form of a normal density $\exp\left(\frac{-y^2}{2\sigma^2}\right)$

$$\rho(Y=y) \propto \exp\left(\frac{-y^2}{2(k_B T/m)}\right) = \rho_{N\left(0,\,(k_B T/m)\right)}(y)$$

We have two expressions for the equilibrium of $Y(t)$:

- The equilibrium described by the Maxwell-Boltzmann distribution:

$$Y(t) \sim N\left(0,\, \frac{k_B T}{m}\right)$$

- The equilibrium mathematically derived from the OU process:

$$Y(t) \sim N\left(0,\, \frac{\gamma^2}{2\beta}\right)$$

The OU process is a model. To make it consistent with the Maxwell-Boltzmann distribution, we equate these two equilibrium distributions.

$$\frac{\gamma^2}{2\beta} = \frac{k_B T}{m}, \qquad \beta = \frac{b}{m}, \quad \gamma = \frac{q}{m}$$

$$\implies \quad \frac{q^2}{m^2}\cdot\frac{m}{2b} = \frac{k_B T}{m}$$

$$\implies \quad q^2 = 2k_B T b$$

Therefore, we arrive at the fluctuation dissipation theorem.

Theorem (fluctuation dissipation relation):

The fluctuation coefficient q and the drag coefficient b are related by

$$\boxed{q = \sqrt{2k_B T b}}$$

With the fluctuation dissipation relation, the OU process becomes.

$$mdY = \underbrace{-bY\,dt}_{\text{dissipation}} + \underbrace{\sqrt{2k_B T b}\,dW}_{\text{fluctuation}}$$

Remark: Now all coefficients in the governing equation are specified.

Goal #4: we study the behavior of $X(t)$.

First, we solve for $X(t)$.

$$Y(t) = e^{-\beta t}Y(0) + e^{-\beta t}G(t) \qquad \text{for } t > 0, \qquad G(t) \equiv \int_0^t \gamma e^{\beta s}\,dW(s)$$

$$X(t) - X(0) = \int_0^t Y(\tau)d\tau = \int_0^t \left(e^{-\beta\tau}Y(0) + e^{-\beta\tau}\int_0^\tau \gamma e^{\beta s}\,dW(s) \right) d\tau$$

$$= \frac{1}{\beta}(1 - e^{-\beta t})Y(0) + \gamma \int_0^t \int_0^\tau e^{-\beta\tau}e^{\beta s}\,dW(s)d\tau$$

Change the order of integration

$$= \frac{1}{\beta}(1 - e^{-\beta t})Y(0) + \gamma \int_0^t \left(\int_s^t e^{-\beta\tau}\,d\tau \right) e^{\beta s}\,dW(s)$$

$$= \frac{1}{\beta}(1 - e^{-\beta t})Y(0) + \frac{\gamma}{\beta} \cdot \underbrace{\int_0^t (1 - e^{-\beta(t-s)})dW(s)}_{G_2(t)}$$

$G_2(t) \equiv \int_0^t (1 - e^{-\beta(t-s)})dW(s)$ is a sum of independent normal RVs.

==> $G_2(t)$ is a normal RV.

Therefore, $(X(t) - X(0))$ is a normal RV. We will look into it in more detail.

AM216 Stochastic Differential Equations

Lecture 08
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- OU process (continued), solution of particle position $X(t)$
- Behavior of $X(t)$, diffusion coefficient, converging to $W(t)$
- Going backward in time using Bayes theorem
- Time reversibility of an equilibrium system
- Different interpretations of stochastic integrals

Recap

Ornstein-Uhlenbeck process (OU):

$$m\,dY = \underbrace{-bY\,dt}_{\text{dissipation}} + \underbrace{q\,dW}_{\text{fluctuation}}\,, \qquad q=\sqrt{2k_B T b}$$

Four goals of the discussion

Goal 1: Solve for $Y(t)$, the particle velocity

$$\Big(Y(t_0+t)\Big|Y(t_0)=y_0\Big) \sim N\left(e^{-\beta t}y_0,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t})\right) \qquad \text{for } t>0$$

$$\text{Equilibrium:}\quad Y(t) \sim N\left(0,\ \frac{\gamma^2}{2\beta}\right) \qquad \text{for large } t>0$$

Goal 2A: $Y(t)$ is a colored noise

Goal 2B: $Y(t)$ converges to a white noise as "m converges to zero"

Goal 3: Fluctuation-dissipation theorem: $q=\sqrt{2k_B T b}$.

Goal 4: Study the behavior of $X(t)$, the particle position

$$Y(t)=e^{-\beta t}Y(0)+e^{-\beta t}G(t)\,, \qquad G(t)\equiv\int_0^t \gamma e^{\beta s}\,dW(s)$$

$$X(t)-X(0)=\int_0^t Y(s)ds=\frac{1}{\beta}(1-e^{-\beta t})Y(0)+\frac{\gamma}{\beta}G_2(t)$$

where $G_2(t) \equiv \int_0^t (1-e^{-\beta(t-s)})dW(s) \sim$ normal.

Goal 4: (continued): We calculate the mean and variance of $G_2(t)$.

$$E\left(G_2(t)\right) = \int_0^t (1-e^{-\beta(t-s)}) E\left(dW(s)\right) = 0$$

$$\text{var}\left(G_2(t)\right) = \int_0^t (1-e^{-\beta(t-s)})^2 ds = t - \frac{2}{\beta}(1-e^{-\beta t}) + \frac{1}{2\beta}(1-e^{-2\beta t})$$

We write out the distribution of $(X(t) - X(0))$.

$$\left(X(t)-X(0)\right) \sim \frac{(1-e^{-\beta t})}{\beta} Y(0) + \left(\frac{\gamma}{\beta}\right) \underbrace{N\left(0, \left(t - \frac{2(1-e^{-\beta t})}{\beta} + \frac{(1-e^{-2\beta t})}{2\beta}\right)\right)}_{\text{containing } dW\text{'s in } [0,t]} \qquad \text{(E01)}$$

Remark:

We cannot integrate $G(t)$ directly because $G(t_1)$ and $G(t_2)$ are not independent. We need to write the integral as a sum of dW's.

In Goal 4, we discuss two cases for $X(t)$.

Goal 4A: finite m

We show that over long time, $(X(t) - X(0))$ demonstrates a diffusion behavior.

The diffusion coefficient is defined as

$$D \equiv \lim_{t\to\infty} \frac{1}{2t} \text{var}\left(X(t)-X(0)\right)$$

We use (E01) to show the limit exists and to calculate the limit.

$$D \equiv \lim_{t\to\infty} \frac{1}{2t} \text{var}\left(X(t)-X(0)\right) = \frac{1}{2}\left(\frac{\gamma}{\beta}\right)^2$$

Substituting $\beta = \dfrac{b}{m}$, $\gamma = \dfrac{q}{m}$, and $q = \sqrt{2k_B T b}$, we have

$$\left(\frac{\gamma}{\beta}\right)^2 = \frac{q^2}{b^2} = \frac{2k_B T b}{b^2} = \frac{2k_B T}{b} \qquad \text{(E02)}$$

Thus, we arrive at $\boxed{D = \dfrac{k_B T}{b}}$.

This is called the Einstein-Smoluchowski relation.

It relates the drag coefficient to the diffusion coefficient.

Remark: The diffusion coefficient is independent of the mass density of the particle. It is affected by the particle size via the drag coefficient b.

Goal 4B: $m \to 0$ (while b and q stay unchanged)

We show that $(X(t) - X(0))$ converges to $\sqrt{2D}\,W(t)$ on any discrete time grid.

Specifically, we show that for $t_2 > t_1 > 0$, as $m \to 0$, we have

- $X(t_1) - X(0) \to \sqrt{2D}\,N(0, t_1)$
- $X(t_1 + t_2) - X(t_1) \to \sqrt{2D}\,N(0, t_2)$
- $(X(t_1) - X(0))$ and $(X(t_1+t_2) - X(t_1))$ are independent.

Using (E01), we write $(X(t_1) - X(0))$ as

$$\left(X(t_1) - X(0)\right) \sim (1 - e^{-\beta t_1})\frac{Y(0)}{\beta} + \sqrt{2D}\,\underbrace{N\left(0, \left(t_1 - \frac{2(1-e^{-\beta t_1})}{\beta} + \frac{(1-e^{-2\beta t_1})}{2\beta}\right)\right)}_{\text{containing } dW\text{'s in } [0, t_1]}$$

As $m \to 0$, we have

$$\beta = \frac{b}{m} = O(m^{-1}), \qquad \gamma = \frac{q}{m} = O(m^{-1}), \qquad \frac{\gamma}{\beta} = O(1)$$

$$2D \equiv \left(\frac{\gamma}{\beta}\right)^2 = O(1) \quad \text{and} \quad \frac{1}{\beta}(1 - e^{-\beta t_1}) = O(m) \to 0$$

Caution: $\lim_{m \to 0} |Y(0)| = \infty$. The Maxwell-Boltzmann distribution gives

$$Y(0) \sim N\left(0, \frac{\gamma^2}{\beta}\right) = O\left(\sqrt{\frac{\gamma^2}{\beta}}\right) = O(m^{-0.5})$$

$$\Longrightarrow \qquad \frac{Y(0)}{\beta} = O(m^{0.5}) \to 0$$

Taking the limit as $m \to 0$, we obtain

- $\left(X(t_1) - X(0)\right) \xrightarrow{\text{as } m \to 0} \sqrt{2D}\ \underbrace{N(0, t_1)}_{\substack{\text{containing } dW\text{'s} \\ \text{in } [0, t_1]}}$

Similarly, we have

$$\Big(X(t_1+t_2)-X(t_1)\Big) \sim (1-e^{-\beta t_2})\frac{Y(t_1)}{\beta}+\sqrt{2D}\,\underbrace{N\left(0,\left(t_2-\frac{2(1-e^{-\beta t_2})}{\beta}+\frac{(1-e^{-2\beta t_2})}{2\beta}\right)\right)}_{\text{containing } dW\text{'s in } [t_1,t_1+t_2]}$$

- $\Big(X(t_1+t_2)-X(t_1)\Big) \xrightarrow{\text{as } m\to 0} \sqrt{2D}\,\underbrace{N(0,t_2)}_{\substack{\text{containing } dW\text{'s} \\ \text{in}[t_1,t_1+t_2]}}$

Notice that $\Big(X(t_1+t_2)-X(t_1)\Big)-(1-e^{-\beta t_2})\dfrac{Y(t_1)}{\beta}$ contains dW's in $[t_1, t_1+t_2]$.

Since $(1-e^{-\beta t_2})\dfrac{Y(t_1)}{\beta}=O(m^{0.5})\to 0$ as $m \to 0$, we arrive at

- $(X(t_1)-X(0))$ and $(X(t_1+t_2)-X(t_1))$ are independent in the limit of $m \to 0$.

Therefore, as $m \to 0$, $(X(t) - X(0))$ converges to $\sqrt{2D}\,W(t)$ on any discrete time grid.

Remarks:

1. The diffusion coefficient of the standard Wiener process is 1/2 (not 1).

   $$D_{\text{Wiener}} \equiv \frac{1}{2t}\text{var}\Big(W(t)\Big)=\frac{1}{2}$$

2. In the limit of $m \to 0$, $(X(t) - X(0))$ exhibits the behavior of a scaled Wiener process, called the Brownian motion, named after Scottish botanist Robert Brown.
3. The derivation above is for the "simplified story". The real story where radius $a \to 0$ while ρ_{mass} is fixed, is presented in Appendix A.

Going backward in time in an equilibrium OU process

In the discussion of Goals #1–4 above, we focused on going forward in time.

$$E\Big(Y(t)\Big|Y(0)\Big)=e^{-\beta t}Y(0) \qquad \text{for } t>0$$

Question:

What happens for $(-t) < 0$? Do we have

$$E\Big(Y(-t)\Big|Y(0)\Big)=e^{+\beta t}Y(0)\ ?$$

which diverges to infinity as $t \to +\infty$. That seems unreasonable.

Answer: $t_{\text{new}} = -t_{\text{old}}$ does not work in stochastic differential equations.

Recall that when we scale dW, it is best to work with $\dfrac{dW}{\sqrt{dt}}$

$$dW(t) = \sqrt{dt} \cdot \frac{dW(t)}{\sqrt{dt}}, \qquad \frac{dW(t)}{\sqrt{dt}} \sim N(0,1) \quad \text{independent of } t \text{ and } dt$$

It is clear that this works only for $dt > 0$, not for $t_{\text{new}} = -t_{\text{old}}$.

Key point:

In stochastic differential equations, scaling $t_{\text{new}} = -t_{\text{old}}$ does not work!

Bayes theorem describes Pr(A | B) in terms of Pr(B | A). We use Bayes theorem to calculate the backward time evolution based on the forward time evolution.

Bayes theorem for densities:

$$\rho\left(Y(-t) = y_1 \middle| Y(0) = y_2\right) \propto \rho\left(Y(0) = y_2 \middle| Y(-t) = y_1\right) \cdot \rho\left(Y(-t) = y_1\right)$$

Backward time evolution in an equilibrium OU process

We assume that the equilibrium has been reached long time ago (at $t = -\infty$) and $Y(t)$ is already a stationary process for all t (including negative t). In particular, the unconstrained $Y(t)$ has the equilibrium distribution for all t.

$$Y(-t) \sim N\left(0, \frac{\gamma^2}{2\beta}\right)$$

$$\Longrightarrow \quad \rho\left(Y(-t) = y_1\right) \propto \exp\left(\frac{-y_1^{\ 2}}{2\gamma^2/(2\beta)}\right)$$

For the forward time evolution, we already derived

$$\left(Y(t_1 + t) \middle| Y(t_1) = y_1\right) \sim N\left(e^{-\beta t} y_1,\ \frac{\gamma^2}{2\beta}\left(1 - e^{-2\beta t}\right)\right) \qquad \text{for } t > 0 \text{ and any } t_1$$

$$\Longrightarrow \quad \rho\left(Y(0) = y_2 \middle| Y(-t) = y_1\right) \propto \exp\left(\frac{-(y_2 - e^{-\beta t} y_1)^2}{2(1 - e^{-2\beta t})\gamma^2/(2\beta)}\right)$$

Substituting into Bayes theorem, we obtain

$$\rho\left(Y(-t) = y_1 \middle| Y(0) = y_2\right) \propto \rho\left(Y(0) = y_2 \middle| Y(-t) = y_1\right) \cdot \rho\left(Y(-t) = y_1\right)$$

$$\propto \exp\left(\frac{-(y_2 - e^{-\beta t} y_1)^2}{2(1 - e^{-2\beta t})\gamma^2/(2\beta)}\right) \cdot \exp\left(\frac{-y_1^{\ 2}}{2\gamma^2/(2\beta)}\right)$$

Note that here y_1 is the independent variable of the PDF and we only need to keep track factors that depend on y_1.

$$\rho\left(Y(-t) = y_1 \middle| Y(0) = y_2\right) \propto \exp\left(\frac{-\left[e^{-2\beta t} y_1^{\ 2} - 2e^{-\beta t} y_2 \cdot y_1 + (1 - e^{-2\beta t}) y_1^{\ 2}\right]}{2(1 - e^{-2\beta t})\gamma^2/(2\beta)}\right)$$

$$\propto \exp\left(\frac{-\left[y_1^{\,2}-2e^{-\beta t}y_2\cdot y_1\right]}{2(1-e^{-2\beta t})\gamma^2/(2\beta)}\right)\propto \exp\left(\frac{-(y_1-e^{-\beta t}y_2)^2}{2(1-e^{-2\beta t})\gamma^2/(2\beta)}\right)$$

We recognize that this is a normal distribution.

It follows that in an equilibrium system, the backward time evolution is described by

$$\left(Y(-t)\middle|Y(0)=y_2\right)\sim N\left(e^{-\beta t}y_2,\ \frac{\gamma^2}{2\beta}\left(1-e^{-2\beta t}\right)\right)\qquad \text{for } t>0$$

We compare it with the forward time evolution

$$\left(Y(t)\middle|Y(0)=y_2\right)\sim N\left(e^{-\beta t}y_2,\ \frac{\gamma^2}{2\beta}\left(1-e^{-2\beta t}\right)\right)\qquad \text{for } t>0$$

Conclusions/remarks:

- At equilibrium, the evolution of going backward in time is statistically the same as the evolution of going forward in time. This is called the time reversibility of equilibrium.
- The time reversibility of equilibrium is a universal law applicable to all thermodynamic systems.
- The intuitive meaning of time reversibility is that if we are given a time series of a system in equilibrium, we won't be able to tell the direction of the time no matter how long and how detailed the time series is.
- Bayes theorem is very powerful in expressing the backward time evolution in terms of the forward time evolution.

Going backward in time in non-equilibrium OU process (optional)

Suppose the system starts with $Y(0) = 0$.

For $t_1 > 0$ and $t_2 > 0$, we use Bayes theorem to calculate $\rho\left(Y(t_1)=y_1\middle|Y(t_1+t_2)=y_2\right)$.

Bayes theorem for densities:

$$\rho\left(Y(t_1)=y_1\middle|Y(t_1+t_2)=y_2\right)\propto \rho\left(Y(t_1+t_2)=y_2\middle|Y(t_1)=y_1\right)\cdot\rho\left(Y(t_1)=y_1\right)$$

We already derived

- $\left(Y(t_1)\middle|Y(0)=0\right)\sim N\left(0,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t_1})\right)\qquad \text{for } t_1>0$

==> $\rho\left(Y(t_1)=y_1\right)\propto \exp\left(\frac{-y_1^{\,2}}{2(1-e^{-2\beta t_1})\gamma^2/(2\beta)}\right)$

- $\left(Y(t_1+t_2)\middle|Y(t_1)=y_1\right) \sim N\left(e^{-\beta t_2}y_1,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t_2})\right)$ for $t_1>0,\ t_2>0$

$$==> \quad \rho\left(Y(t_1+t_2)=y_2\middle|Y(t_1)=y_1\right) \propto \exp\left(\frac{-(y_2-e^{-\beta t_2}y_1)^2}{2(1-e^{-2\beta t_2})\gamma^2/(2\beta)}\right)$$

Substituting into Bayes theorem, we obtain

$$\rho\left(Y(t_1)=y_1\middle|Y(t_1+t_2)=y_2\right) \propto \rho\left(Y(t_1+t_2)=y_2\middle|Y(t_1)=y_1\right)\cdot\rho\left(Y(t_1)=y_1\right)$$

$$\propto \exp\left(\frac{-(y_2-e^{-\beta t_2}y_1)^2}{2(1-e^{-2\beta t_2})\gamma^2/(2\beta)}\right)\cdot\exp\left(\frac{-y_1{}^2}{2(1-e^{-2\beta t_1})\gamma^2/(2\beta)}\right)$$

(we only need to keep track factors that depend on y_1).

$$\propto \exp\left(\frac{-\left[(1-e^{-2\beta(t_1+t_2)})y_1{}^2-2(1-e^{-2\beta t_1})e^{-\beta t_2}y_2\cdot y_1\right]}{2(1-e^{-2\beta t_1})(1-e^{-2\beta t_2})\gamma^2/(2\beta)}\right)$$

$$\propto \exp\left(\frac{-\left(y_1-\frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}e^{-\beta t_2}y_2\right)^2}{2\frac{(1-e^{-2\beta t_1})(1-e^{-2\beta t_2})}{(1-e^{-2\beta(t_1+t_2)})}\gamma^2/(2\beta)}\right)$$

It follows that

$$\left(Y(t_1)\middle|Y(t_1+t_2)=y_2\right)\sim N\left(\frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}e^{-\beta t_2}y_2,\ \ \frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}\frac{\gamma^2}{2\beta}(1-e^{-2\beta t_2})\right)$$

We discuss two special cases for t_1 and t_2

<u>Case i)</u> $\quad t_1 \to +\infty$ while t_2 = fixed

$$\frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}e^{-\beta t_2}y_2 \to e^{-\beta t_2}y_2 \qquad \text{for large } t_1$$

$$\frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}\frac{\gamma^2}{2\beta}(1-e^{-2\beta t_2}) \to \frac{\gamma^2}{2\beta}(1-e^{-2\beta t_2}) \qquad \text{for large } t_1$$

$$==> \quad \left(Y(t_1)\middle|Y(t_1+t_2)=y_2\right) \sim N\left(e^{-\beta t_2}y_2,\ \ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t_2})\right) \quad \text{for large } t_1$$

This is the same as the equilibrium case, not a surprise at all.

Case ii) $t_1 = t_2 = h$

$$\frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}e^{-\beta t_2}y_2 = \frac{e^{-\beta h}y_2}{1+e^{-2\beta h}}$$

$$\frac{(1-e^{-2\beta t_1})}{(1-e^{-2\beta(t_1+t_2)})}\frac{\gamma^2}{2\beta}(1-e^{-2\beta t_2}) = \frac{\gamma^2}{2\beta}\left(\frac{1-e^{-2\beta h}}{1+e^{-2\beta h}}\right)$$

$$\left(Y(h)\middle|Y(2h)=y_2\right) \sim N\left(\frac{e^{-\beta h}y_2}{1+e^{-2\beta h}},\ \frac{\gamma^2}{2\beta}\left(\frac{1-e^{-2\beta h}}{1+e^{-2\beta h}}\right)\right)$$

We compare it with the forward time evolution

$$\rho\left(Y(2h)\middle|Y(h)=y_1\right) \sim N\left(e^{-\beta h}y_1,\ \frac{\gamma^2}{2\beta}\left(1-e^{-2\beta h}\right)\right)$$

When βh is not large, this case clearly demonstrates the difference between forward time evolution and backward time evolution in a non-equilibrium system.

Different interpretations of stochastic integrals

Beauty of the deterministic calculus

Consider the integral of a deterministic function $f(s)$.

$$\int_0^t f(s)ds = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j)\Delta s$$

where $\Delta s = \dfrac{t}{N}$, $s_j = j\Delta s$, $\tilde{s}_j \in [s_j\ s_{j+1}]$

Note: When $f(s)$ is piecewise continuous, the choice of $\tilde{s}_j \in [s_j, s_{j+1}]$ does not affect the limit. We can use any $\tilde{s}_j \in [s_j, s_{j+1}]$. In particular,

$$\lim_{N\to\infty}\sum_{j=0}^{N-1} f(s_j)\Delta s = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(s_{j+1})\Delta s = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(s_{j+1/2})\Delta s$$

A simple stochastic integral

$$\int_0^t f(s)dW(s) = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j)\Delta W_j$$

where $\tilde{s}_j \in [s_j\ s_{j+1}]$, $\Delta W_j = W(s_{j+1}) - W(s_j)$

The Riemann sum, $\lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j)\Delta W_j$, is a normal RV with mean = 0 and

$$\text{variance} = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j)^2\Delta s = \int_0^t f(s)^2\,ds$$

When $f(s)$ is piecewise continuous, the choice of $\tilde{s}_j \in [s_j, s_{j+1}]$ does not affect the limit. We can use any $\tilde{s}_j \in [s_j, s_{j+1}]$.

Another simple stochastic integral

$$\int_0^t f(s,W(s))ds = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j, W(\tilde{s}_j))\Delta s$$

When $f(s, w)$ is smooth, the choice of $\tilde{s}_j \in [s_j, s_{j+1}]$ does not affect the limit (homework problem).

A more complicated stochastic integral:

$$\int_0^t f(s,W(s))dW(s) = \lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j, W(\tilde{s}_j))\Delta W_j$$

where $\quad \tilde{s}_j \in [s_j\ s_{j+1}], \quad \Delta W_j = W(s_{j+1}) - W(s_j)$

Note that

- $f(s, W(s))$ is not a deterministic function of s.
- $f(\tilde{s}_j, W(\tilde{s}_j))$ is a random variable, potentially correlated with ΔW_j depending on the choice of $\tilde{s}_j \in [s_j, s_{j+1}]$.
- As a result, different choices of $\tilde{s}_j \in [s_j, s_{j+1}]$ lead to different results.
- Thus, integral $\int_0^t f(s,W(s))dW(s)$ is subject to different interpretations.

Appendix A The limit of $X(t)$ as radius $a \to 0$ while ρ_{mass} is fixed.

Recall that in the "simplified story", as $m \to 0$ while b and q are fixed, we have

$2D = O(1)$ and $(X(t) - X(0))$ converges to $\sqrt{2D}\,W(t)$

Now we consider the real story. As $a \to 0$ while ρ_{mass} is fixed, we have

$$m = O(a^3), \quad b = O(a), \quad q = \sqrt{2k_B T b} = O(\sqrt{a})$$

$$\beta = \frac{b}{m} = O(a^{-2}), \quad \gamma = \frac{q}{m} = O(a^{-2.5})$$

$$\frac{\gamma}{\beta} = O(a^{-0.5}), \quad D = \frac{1}{2}\left(\frac{\gamma}{\beta}\right)^2 = O(a^{-1}) \to \infty$$

The behavior of diffusion coefficient D suggests scaling the displacement by $\sqrt{a}$.

We show that $\sqrt{a}\left(X(t_1) - X(0)\right)$ converges to $cW(t)$ on any discrete time grid where coefficient $c \equiv \sqrt{a}\sqrt{2D} = O(1)$. Specifically, we show that for $t_2 > t_1 > 0$, as $a \to 0$,

- $\sqrt{a}\left(X(t_1) - X(0)\right) \to c\,N(0, t_1)$
- $\sqrt{a}\left(X(t_1 + t_2) - X(t_1)\right) \to c\,N(0, t_2)$
- $(X(t_1) - X(0))$ and $(X(t_1 + t_2) - X(t_1))$ are independent.

Using (E01), we write $\sqrt{a}\,(X(t_1) - X(0))$ as

$$\sqrt{a}\left(X(t_1) - X(0)\right) \sim (1 - e^{-\beta t_1})\frac{\sqrt{a}\,Y(0)}{\beta} + \underbrace{c\,N\left(0, \left(t_1 - \frac{2(1 - e^{-\beta t_1})}{\beta} + \frac{(1 - e^{-2\beta t_1})}{2\beta}\right)\right)}_{\text{containing } dW\text{'s in } [0, t_1]}$$

The Maxwell-Boltzmann distribution gives

$$Y(t) \sim N\left(0, \frac{\gamma^2}{\beta}\right) = O\left(\sqrt{\frac{\gamma^2}{\beta}}\right) = O\left(\sqrt{\frac{a^{-5}}{a^{-2}}}\right) = O(a^{-1.5})$$

$$\Longrightarrow \quad \frac{\sqrt{a}\,Y(t)}{\beta} = \frac{\sqrt{a}\,O(a^{-1.5})}{O(a^{-2})} = O(a) \to 0$$

Taking the limit as $a \to 0$ and using $\frac{1}{\beta}(1 - e^{-\beta t_1}) \to 0$, we obtain

- $\sqrt{a}\left(X(t_1) - X(0)\right) \xrightarrow{\text{as } a \to 0} \underbrace{c\,N(0, t_1)}_{\substack{\text{containing } dW\text{'s} \\ \text{in } [0, t_1]}}$

Similarly, we have

$$\sqrt{a}\left(X(t_1+t_2)-X(t_1)\right)\sim(1-e^{-\beta t_2})\frac{\sqrt{a}\,Y(t_1)}{\beta}+\underbrace{c\,N\left(0,\left(t_2-\frac{2(1-e^{-\beta t_2})}{\beta}+\frac{(1-e^{-2\beta t_2})}{2\beta}\right)\right)}_{\text{containing } dW\text{'s in } [t_1,t_1+t_2]}$$

- $\sqrt{a}\left(X(t_1+t_2)-X(t_1)\right)\xrightarrow{\text{as } a\to 0}\underbrace{c^2N(0,t_2)}_{\substack{\text{containing } dW\text{'s} \\ \text{in}[t_1,t_1+t_2]}}$

Again, $\sqrt{a}\left(X(t_1+t_2)-X(t_1)\right)-(1-e^{-\beta t_2})\dfrac{\sqrt{a}\,Y(t_1)}{\beta}$ contains dW's in $[t_1, t_1+t_2]$.

Since $(1-e^{-\beta t_2})\dfrac{\sqrt{a}\,Y(t_1)}{\beta}=O(a)\to 0$ as $a \to 0$, we arrive at

- $(X(t_1)-X(0))$ and $(X(t_1+t_2)-X(t_1))$ are independent in the limit of $a \to 0$.

Therefore, we conclude that $\sqrt{a}\left(X(t)-X(0)\right)$ converges to $cW(t)$ as $a \to 0$.

In other words, for a particle of small radius a, the displacement $(X(t) - X(0))$ is approximately $\dfrac{c}{\sqrt{a}}W(t)$ with the magnitude diverging to ∞ as $a \to 0$.

AM216 Stochastic Differential Equations

Lecture 09

List of topics in this lecture

- Convergence in probability, a sufficient condition for convergence in probability, a theorem for calculating the variance of sum of products
- Ito's interpretation and Stratonovich's interpretation of stochastic integrals, the relation between the two, proof of Ito's lemma
- Stochastic integrals based on axioms, the λ-chain rule

Recap

Going backward in time in equilibrium

Tool: Bayes theorem

Backward time evolution in equilibrium OU process

$$\left(Y(-t)\middle|Y(0)=y_2\right) \sim N\left(e^{-\beta t}y_2,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t})\right) \qquad \text{for } t>0$$

Forward time evolution in equilibrium OU process

$$\left(Y(t)\middle|Y(0)=y_0\right) \sim N\left(e^{-\beta t}y_0,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t})\right) \qquad \text{for } t>0$$

Time reversibility of equilibrium

Different interpretations of stochastic integrals

$$\int_0^t f(s,W(s))dW(s)=\lim_{N\to\infty}\sum_{j=0}^{N-1} f(\tilde{s}_j,W(\tilde{s}_j))\Delta W_j$$

where $\Delta s=\dfrac{t}{N}$, $s_j=j\Delta s$, $\Delta W_j=W_{j+1}-W_j$, $\tilde{s}_j\in[s_j\ s_{j+1}]$

Different choices of $\tilde{s}_j \in [s_j, s_{j+1}]$ lead to different results.

Ito's interpretation (Kiyosi Ito):

$$\tilde{s}_j = s_j$$

$$\int_0^t f(s,W(s))dW(s) = \lim_{N\to\infty}\sum_{j=0}^{N-1} f_j \Delta W_j\,, \qquad f_j \equiv f(s_j, W(s_j))$$

Stratonovich's interpretation (Ruslan Stratonovich):

$$\tilde{s}_j = \frac{1}{2}(s_j + s_{j+1})$$

$$\int_0^t f(s,W(s))dW(s) = \lim_{N\to\infty}\sum_{j=0}^{N-1} \frac{1}{2}(f_j + f_{j+1})\Delta W_j\,, \qquad f_j \equiv f(s_j, W(s_j))$$

Note: Stratonovich's interpretation is based on the trapezoidal rule; it is not exactly the Riemann sum with $\tilde{s}_j = (s_j + s_{j+1})/2$. The two are equivalent (see below).

Road map of the discussion:

1. We show that the Stratonovich's interpretation is equivalent to the Riemann sum with $\tilde{s}_j = s_{j+1/2} \equiv (s_j + s_{j+1})/2$.

$$\lim_{N\to\infty}\sum_{j=0}^{N-1}\left[\frac{1}{2}(f_j + f_{j+1}) - f_{j+1/2}\right]\Delta W_j = 0\,, \qquad f_{j+1/2} \equiv f(s_{j+1/2}, W(s_{j+1/2}))$$

2. We demonstrate the relation between the Ito interpretation and the Stratonovich interpretation in a simple example.
3. As a tool for connecting the Ito interpretation and the Stratonovich interpretation in the general case, we prove Ito's lemma.
4. We write out the relation between the Ito interpretation and the Stratonovich interpretation in the general case.

Preparation for discussion

Recall the convergence in probability.

Definition (convergence in probability)

Let $\{Q_N(\omega)\}$ be a sequence of random variables. We say that $\{Q_N(\omega)\}$ converges to q in probability as $N \to +\infty$, if for any $\varepsilon > 0$,

$$\lim_{N\to\infty} \Pr\left(|Q_N(\omega) - q| > \varepsilon\right) = 0$$

Theorem (a sufficient condition for convergence in probability)

Suppose $\lim\limits_{N\to\infty} E\big(Q_N(\omega)\big)=q$ and $\lim\limits_{N\to\infty} \mathrm{var}\big(Q_N(\omega)\big)=0$.

Then, $\{Q_N(\omega)\}$ converges to q in probability as $N \to +\infty$.

Proof: homework problem.

Theorem (a useful formula for calculating $\mathrm{var}\left(\sum_{j=0}^{N-1} Y_j X_j\right)$)

Suppose random variables $\{X_j, j = 0, 1, \ldots, N-1\}$ and $\{Y_k, k = 0, 1, \ldots, N-1\}$ satisfy

1. $E(X_j) = 0$ for all j,
2. X_j is independent of X_i for all $i \neq j$, and
3. X_j is independent of Y_k for all $k \leq j$.

Then we have $$\mathrm{var}\left(\sum_{j=0}^{N-1} Y_j X_j\right)=\sum_{j=0}^{N-1} E(Y_j^{\,2})E(X_j^{\,2}).$$

Remarks:

- Important note: the theorem does not require "Y_j is independent of Y_i for all $i \neq j$".
  Example: $X_j = (\Delta W_j)^2$, $Y_j = (W_j)^2$.
- We can write the conclusion as $\mathrm{var}\left(\sum_{j=0}^{N-1} Y_j X_j\right)=\mathrm{var}(Y_j X_j)$.

$$E(Y_j X_j)=E(Y_j)E(X_j)=0$$

$$\mathrm{var}(Y_j X_j)=E\big((Y_j X_j)^2\big)=E(Y_j^{\,2})E(X_j^{\,2})$$

Proof: homework problem.

Item 1 of the road map:

We state a general theorem that includes item 1 as a special case.

Theorem: (weighted average of two Riemann sums)

Let $f(s, w)$ be a smooth function of (s, w), and $\Delta s=\dfrac{t}{N}$, $s_j = j\Delta s$, $\Delta W_j = W_{j+1} - W_j$.

For any $0 \leq \beta \leq 1$, we have

$$\lim_{N\to\infty}\sum_{j=0}^{N-1}\Big[\big((1-\beta)f_j+\beta f_{j+1}\big)-f_{j+\beta}\Big]\Delta W_j=0$$

where $s_{j+\beta}\equiv s_j+\beta\Delta s$, $\quad f_{j+\beta}\equiv f(s_{j+\beta}, W(s_{j+\beta}))$

Proof: (Skip)

We look at the main steps in the proof for β = 1/2. Let

$$Q_N \equiv \sum_{j=0}^{N-1}\left(\frac{1}{2}(f_j+f_{j+1})-f_{j+1/2}\right)\Delta W_j$$

We only need to show $\lim_{N\to\infty} E\big(Q_N(\omega)\big)=q$ and $\lim_{N\to\infty} \mathrm{var}\big(Q_N(\omega)\big)=0$.

- Expand f_{j+1} and $f_{j+1/2}$ around s_j. Neglect O(Δs) terms in the expansions of f_{j+1} and $f_{j+1/2}$.

$$f_{j+1}=f_j+(f_w)_j(\Delta W_j)+O(\Delta s)$$

$$f_{j+1/2}=f_j+(f_w)_j(\Delta W_j^{(-)})+O(\Delta s), \qquad \Delta W_j^{(-)}\equiv W_{j+1/2}-W_j$$

$$\frac{1}{2}(f_j+f_{j+1})-f_{j+1/2}=(f_w)\frac{1}{2}\left(\Delta W_j^{(+)}-\Delta W_j^{(-)}\right)+O(\Delta s)$$

$$\Delta W_j^{(+)}\equiv W_{j+1}-W_{j+1/2}, \qquad \Delta W_j=\Delta W_j^{(-)}+\Delta W_j^{(+)}$$

- Multiply by $\Delta W_j=\Delta W_j^{(-)}+\Delta W_j^{(+)}$ and sum over j, we write Q_N as

$$Q_N=\frac{1}{2}\sum_{j=0}^{N-1}\underbrace{(f_w)_j}_{Y_j}\underbrace{\left((\Delta W_j^{(+)})^2-(\Delta W_j^{(-)})^2\right)}_{X_j}+\underbrace{\sum_{j=0}^{N-1}O(\Delta s)\Delta W_j}_{o(1)}$$

- Use the theorem to show $\lim_{N\to\infty} E\big(Q_N(\omega)\big)=q$ and $\lim_{N\to\infty} \mathrm{var}\big(Q_N(\omega)\big)=0$.

$$E(X_j)=E\left((\Delta W_j^{(+)})^2-(\Delta W_j^{(-)})^2\right)=0$$

$$E(X_j^2)=E\left((\Delta W_j^{(+)})^4-2(\Delta W_j^{(+)})^2(\Delta W_j^{(-)})^2+(\Delta W_j^{(-)})^4\right)=(\Delta s)^2$$

Here we used $E\left((\Delta W_j^{(-)})^2\right)=(\Delta s)/2$, $\quad E\left((\Delta W_j^{(-)})^4\right)=3(\Delta s)/2$.

Item 2 of the road map: (an example)

Key point: Ito interpretation and Stratonovich interpretation yield different values!

Example: $I=\int_0^t W(s)dW(s)$

Discretization: $\Delta s=\frac{t}{N}$, $s_j=j\Delta s$, $\Delta W_j=W_{j+1}-W_j$.

We first work out the Stratonovich interpretation:

$$I_{\text{Stratonovich}}=\lim_{N\to\infty}\sum_{j=0}^{N-1}\frac{1}{2}(W_j+W_{j+1})(\Delta W_j)=\lim_{N\to\infty}\sum_{j=0}^{N-1}\frac{1}{2}(W_j+W_{j+1})(W_{j+1}-W_j)$$

$$=\lim_{N\to\infty}\sum_{j=0}^{N-1}\frac{1}{2}\Big((W_{j+1})^2-(W_j)^2\Big)=\frac{1}{2}\Big((W_N)^2-(W_0)^2\Big)=\frac{1}{2}W(t)^2$$

Ito interpretation:

$$I_{\text{Ito}}=\lim_{N\to\infty}\sum_{j=0}^{N-1}W_j(\Delta W_j)=\lim_{N\to\infty}\sum_{j=0}^{N-1}\left(\frac{1}{2}(W_j+W_{j+1})-\frac{1}{2}(W_{j+1}-W_j)\right)\Delta W_j$$

$$=\underbrace{\lim_{N\to\infty}\sum_{j=0}^{N-1}\frac{1}{2}(W_j+W_{j+1})\Delta W_j}_{\text{Stratonovich}}-\lim_{N\to\infty}\sum_{j=0}^{N-1}\frac{1}{2}(\Delta W_j)^2$$

In Lecture 4, as a special case of Ito's lemma, we showed $\lim_{N\to\infty}\sum_{j=0}^{N-1}\Big((\Delta W_j)^2-\Delta s\Big)=0$, which leads to $\lim_{N\to\infty}\frac{1}{2}\sum_{j=0}^{N-1}(\Delta W_j)^2=\lim_{N\to\infty}\frac{1}{2}\sum_{j=0}^{N-1}\Delta s=\frac{1}{2}\int_0^t ds=\frac{1}{2}t$. We arrive at

$$I_{\text{Ito}}=\frac{1}{2}W(t)^2-\frac{1}{2}t=I_{\text{Stratonovich}}-\frac{1}{2}t \quad \text{for} \quad \int_0^t W(s)dW(s)$$

Item 3 of the road map:

Discretization: $\Delta s=\dfrac{t}{N}$, $s_j=j\Delta s$, $\Delta W_j=W_{j+1}-W_j$

Theorem (Ito's lemma)

$$\lim_{N\to\infty}\left(\sum_{j=0}^{N-1}g(s_j,W_j)(\Delta W_j)^2-\sum_{j=0}^{N-1}g(s_j,W_j)\Delta s\right)=0$$

Remark: The two Riemann sums and the corresponding integrals (after taking the limits) are still random variables.

Proof: Let

$$Q_N\equiv\sum_{j=0}^{N-1}\underbrace{g(s_j,W_j)}_{Y_j}\underbrace{\Big((\Delta W_j)^2-\Delta s\Big)}_{X_j}$$

$\{Q_N\}$ is a sequence of random variables. We only need to show

$$\lim_{N\to\infty}E\Big(Q_N(\omega)\Big)=0 \quad \text{and} \quad \lim_{N\to\infty}\text{var}\Big(Q_N(\omega)\Big)=0$$

We use the theorem for calculating $\text{var}\left(\sum_{j=0}^{N-1}Y_jX_j\right)$. We first check the 3 conditions.

1. $E(X_j)=0$ for all j,
2. X_j is independent of X_i for all $i\neq j$, and

3. X_j is independent of Y_k for all $k \le j$.

Remark: We don't have and we don't need "Y_j is independent of Y_i for all $i \ne j$".

It follows that $E(Q_N) = 0$ and

$$\mathrm{var}(Q_N) = \mathrm{var}\left(\sum_{j=0}^{N-1} Y_j X_j\right) = \sum_{j=0}^{N-1} E(Y_j^{\,2}) E(X_j^{\,2})$$

$$E(X_j^{\,2}) = E\Big(((\Delta W_j)^2 - \Delta s)^2\Big) = \mathrm{var}\Big((\Delta W_j)^2\Big) = 2(\Delta s)^2 \qquad \text{(see homework)}$$

$$E(Y_j^{\,2}) = E\Big(g(s_j, W_j)^2\Big) = O(1)$$

$$\mathrm{var}(Q_N) = \sum_{j=0}^{N-1} E(Y_j^{\,2}) E(X_j^{\,2}) = 2(\Delta s)^2 \sum_{j=0}^{N-1} O(1)$$

$$= 2(\Delta s)^2 O(N) = O(\Delta s) \;\to\; 0 \text{ as } N \to +\infty$$

End of proof of Ito's lemma

Item 4 of the road map: (Relation between Ito and Stratonovich interpretations)

We look at the general case: $\int_0^t f(s, W(s))\,dW(s)$.

We start with the Stratonovich interpretation:

$$I_{\text{Stratonovich}} = \lim_{N\to\infty} \sum_{j=0}^{N-1} \frac{1}{2}(f_j + f_{j+1})\Delta W_j\,, \qquad f_j \equiv f(s_j, W(s_j))$$

$$= \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[f_j \Delta W_j + \frac{1}{2}(f_{j+1} - f_j)\Delta W_j \right]$$

Expand $(f_{j+1} - f_j)$ around s_j. Neglect $O(\Delta s)$ terms in the expansion of $(f_{j+1} - f_j)$.

$$f_{j+1} - f_j = (f_w)_j \Delta W_j + O(\Delta s)$$

Multiply by ΔW_j and sum over j, we get

$$I_{\text{Stratonovich}} = \underbrace{\lim_{N\to\infty} \sum_{j=0}^{N-1} f_j \Delta W_j}_{\text{Ito}} + \lim_{N\to\infty} \sum_{j=0}^{N-1} \frac{1}{2}(f_w)_j (\Delta W_j)^2 + \underbrace{\lim_{N\to\infty} \sum_{j=0}^{N-1} O(\Delta s)\Delta W_j}_{=0}$$

Ito's lemma gives

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} \frac{1}{2}(f_w)_j (\Delta W_j)^2 = \lim_{N\to\infty} \frac{1}{2} \sum_{j=0}^{N-1} (f_w)_j \Delta s = \frac{1}{2} \int_0^t f_w(s, W(s))\,ds$$

Combining these results, we obtain the main theorem connecting the Stratonovich interpretation and the Ito interpretation in the general case.

Theorem:

For integral $\int_0^t f(s,W(s))dW(s)$, we have

$$I_{\text{Stratonovich}} = I_{\text{Ito}} + \frac{1}{2}\int_0^t f_w(s,W(s))ds$$

Stochastic integrals based on axioms

In the above, we interpreted stochastic integrals as limits of Riemann sums. Different choices of Riemann sums lead to different interpretations. Alternatively, we can calculate stochastic integrals based on a set of axioms.

Two axioms:

1) Fundamental theorem of calculus (FTC)

$$\int_a^b dH(t,W(t)) = H(t,W(t))\Big|_a^b$$

2) λ-chain rule

$$dH(t,W(t)) = H_t dt + H_w dW(t) + \left(\frac{1}{2} - \lambda\right) H_{ww} dt$$

Ito's interpretation: $\lambda = 0$

Stratonovich's interpretation: $\lambda = 0.5$

Meaning of the λ-chain rule

We compare the behaviors of increment $\Delta H \equiv H(t+\Delta t, u(t+\Delta t)) - \Delta H(t, u(t))$ respectively for $u(t)$ = smooth deterministic and for $u(t) = W(t)$.

Case 1: $u(t)$ = deterministic

- Expand around t to rewrite $\Delta H(t, u(t))$. Neglect $o(\Delta t)$ terms in the expansion.

$$\Delta H = H_t(t,u(t))\Delta t + H_u(t,u(t))\underbrace{\Delta u}_{O(\Delta t)} + o(\Delta t) \qquad \text{(E01A)}$$

Short notation: $\Delta H = H_t\big|_t \Delta t + H_u\big|_t \Delta u + o(\Delta t)\,, \qquad H_u\big|_t \equiv H_u(t,u(t))$

- Expand around $(t + \Delta t)$ to rewrite $\Delta H(t, u(t))$. Neglect $o(\Delta t)$ terms in the expansion.

$$H\big|_t = H\big|_{t+\Delta t} + H_t\big|_{t+\Delta t}(-\Delta t) + H_u\big|_{t+\Delta t}(-\Delta u) + o(\Delta t)\,, \qquad H_u\big|_{t+\Delta t} \equiv H_u(t+\Delta t, u(t+\Delta t))$$

$$\Delta H = H_t\big|_{t+\Delta t}\Delta t + H_u\big|_{t+\Delta t}\Delta u + o(\Delta t) \qquad \text{(E01B)}$$

(E01A) using $H_u|_t$ and (E01B) using $H_u|_{t+\Delta t}$ have the same form. We write it as a regular differential without specifying if H_u is $H_u|_t$ or $H_u|_{t+\Delta t}$.

$$dH = H_t dt + H_u du$$

Observation: For a smooth deterministic function, the differential has the same form whether we interpret H_u as $H_u|_t$ or as $H_u|_{t+\Delta t}$.

That is the beauty of deterministic calculus.

Case 2: $u(t) = W(t)$, the Wiener process

- Ito's interpretation:

  Expand around t to rewrite $\Delta H \equiv H(t+\Delta t, W(t+\Delta t)) - \Delta H(t, W(t))$.

  Neglect o(Δt) terms in the expansion.

$$\Delta H = H_t\big|_t \Delta t + H_w\big|_t \Delta W + \frac{1}{2}H_{ww}\big|_t(\Delta W)^2 + o(\Delta t), \qquad H_w\big|_t \equiv H_w(t, W(t)) \qquad \text{(E02A)}$$

  We replace $(\Delta W)^2$ by Δt to write out the differential. The differential has the form

$$dH = H_t dt + H_w dW + \frac{1}{2}H_{ww}dt$$

  with the understanding $H_w = H_w\big|_t$.

  This is the λ-chain rule with $\lambda = 0$.

- Stratonovich's interpretation:

  Expand around $(t+\Delta t)$ to rewrite ΔH. Neglect o(Δt) terms in the expansion.

$$H|_t = H|_{t+\Delta t} + H_t\big|_{t+\Delta t}(-\Delta t) + H_w\big|_{t+\Delta t}(-\Delta W) + \frac{1}{2}H_{ww}\big|_{t+\Delta t}(-\Delta W)^2 + o(\Delta t)$$

$$\text{where} \quad H_w\big|_{t+\Delta t} \equiv H_w(t+\Delta t, W(t+\Delta t))$$

$$\begin{aligned}\Delta H &= H_t\big|_{t+\Delta t}\Delta t + H_w\big|_{t+\Delta t}\Delta W - \frac{1}{2}H_{ww}\big|_{t+\Delta t}(\Delta W)^2 + o(\Delta t) \\ &= H_t\big|_t \Delta t + H_w\big|_{t+\Delta t}\Delta W - \frac{1}{2}H_{ww}\big|_t(\Delta W)^2 + o(\Delta t)\end{aligned} \qquad \text{(E02B)}$$

  Use the average of (E02A) and (E02B) to write ΔH as

$$\Delta H = H_t\big|_t \Delta t + \frac{1}{2}\left(H_w\big|_t + H_w\big|_{t+\Delta t}\right)\Delta W + o(\Delta t) \qquad \text{(E02C)}$$

  The differential has a diferent form

$$dH = H_t dt + H_w dW(t)$$

with the understanding $H_w = \frac{1}{2}\left(H_w\big|_t + H_w\big|_{t+\Delta t}\right)$.

This is the λ-chain rule with λ = 0.5.

Remarks:

- If we use the expansion around $(t + \Delta t)$, the differential has another form

$$dH = H_t dt + H_w dW - \frac{1}{2} H_{ww} dt$$

with the understanding $H_w = H_w\big|_{t+\Delta t}$.

This is the λ-chain rule with λ = 1.

- The λ-chain rule implicitly distinguishes the choices of H_w in the differential.
  The differential $dH \equiv H(t+dt, W(t+dt)) - H(t, W(t))$ is always the same.
  dH takes different forms depending on whether we interpret H_w as

$$H_w|_t \quad \text{or} \quad (H_w|_t + H_w|_{t+\Delta t})/2 \quad \text{or} \quad H_w|_{t+\Delta t}$$

$$dH = H_t dt + H_w dW + \frac{1}{2} H_{ww} dt \qquad H_w = H_w|_t$$

$$dH = H_t dt + H_w dW \qquad H_w = (H_w|_t + H_w|_{t+\Delta t})/2$$

$$dH = H_t dt + H_w dW - \frac{1}{2} H_{ww} dt \qquad H_w = H_w|_{t+\Delta t}$$

- To integrate $H_w dW$, we write $H_w dW = dH - (H_t + (0.5-\lambda)H_{ww})dt$.
  Different interpretations of $\int H_w dW$ are reflected in different values of λ (see below).

Use the axioms to calculate $\int_a^b f(t, W(t))dW(t)$

Strategy:

- Write the λ-chain rule as $H_w dW = dH - (H_t + (0.5-\lambda)H_{ww})dt$.
- Solve $H_w = f(t, w)$ to find $H(t, w)$.
- Calculate the integrate as

$$\int_a^b f(t, W(t))dW(t) = \underbrace{\int_a^b dH}_{\text{Fundamental theorem of calculus}} - \underbrace{\int_a^b \left(H_t + (\tfrac{1}{2} - \lambda)H_{ww}\right)dt}_{\text{See comments below}}$$

Comments:

The integrand ( ) varies with the value of λ. (reflecting different interpretations).

Given the integrand ( ), the integral is not affected by different interpretations.

Integral $\int_a^b g(t,W(t))dt$ is not affected by different interpretations.

$$\int_a^b g(t,W(t))dt = \lim_{N\to\infty}\sum_{j=0}^{N-1} g(\tilde{t}_j, W(\tilde{t}_j))\Delta t\,, \qquad \tilde{t}_j \in [t_j\, t_{j+1}]$$

The choice of $\tilde{t}_j \in [t_j\, t_{j+1}]$ does not matter (homework problem).

Procedure for calculating $\int_a^b f(t,W(t))dW(t)$

Step 1: Solve $H_w(t, w) = f(t, w)$ with $H(t, 0) = 0$ to define

$$H(t,w) = \int_0^w f(t,u)du \qquad \text{(this is a regular integral!)}$$

==> $f(t,W(t))dW(t) = H_w(t,W(t))dW(t)$

Step 2: Use the λ-chain rule to write $f(t, W(t))dW = dH - (\ )dt$.

λ-chain rule: $dH = H_t dt + H_w dW + \left(\tfrac{1}{2}-\lambda\right)H_{ww}dt$

==> $H_w dW = dH - \left(H_t + \left(\tfrac{1}{2}-\lambda\right)H_{ww}\right)dt$

==> $f(t,W(t))dW(t) = dH - \left(H_t + \left(\tfrac{1}{2}-\lambda\right)H_{ww}\right)dt$

Step 3: Differentiate $H(t, w)$ to calculate $H_t(t, w)$ and $H_{ww}(t, w)$.

Both are regular derivatives and regular functions.

Step 4: Use the fundamental theorem of calculus to calculate the integral

$$\int_a^b f(t,W(t))dW(t) = H(t,W(t))\Big|_a^b - \int_a^b \left(H_t + \left(\tfrac{1}{2}-\lambda\right)H_{ww}\right)dt$$

Example:

$$\int_a^b t W(t)^2 dW(t)$$

0. Identify function $f(t, w)$ in the integral

$f(t, w) = t\, w^2$

1. Solve $H_w(t, w) = f(t, w)$ with $H(t, 0) = 0$ to define

$$H(t,w) = \int_0^w t u^2 du = t\frac{w^3}{3}$$

==> $\quad tW(t)^2 dW(t) = H_w(t, W(t))dW(t)$

2. Use the λ-chain rule to write $f(t, W(t))dW = dH - (\ )dt$.

$$dH = H_t dt + H_w dW + \left(\tfrac{1}{2} - \lambda\right) H_{ww} dt$$

==> $\quad H_w dW = dH - \left(H_t + \left(\tfrac{1}{2} - \lambda\right) H_{ww}\right) dt$

==> $\quad tW(t)^2 dW(t) = dH - \left(H_t + \left(\tfrac{1}{2} - \lambda\right) H_{ww}\right) dt$

3. Differentiate to calculate $H_t(t, w)$ and $H_{ww}(t, w)$.

$$H_t(t, w) = \frac{w^3}{3}, \qquad H_{ww}(t, w) = 2tw$$

4. Use the fundamental theorem of calculus

$$\int_a^b tW(t)^2 dW(t) = H(t, W(t))\Big|_a^b - \int_a^b \left(H_t + \left(\tfrac{1}{2} - \lambda\right) H_{ww}\right) dt$$

$$= t\frac{W(t)^3}{3}\Bigg|_a^b - \int_a^b \left(\frac{W(t)^3}{3} + \left(\tfrac{1}{2} - \lambda\right) 2tW(t)\right) dt$$

Ito: $\quad \lambda = 0$;

Stratonovich: $\lambda = 0.5$

All terms in the result are random variables.

AM216 Stochastic Differential Equations

Lecture 11
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- Transition probability density $q(x, t \mid z, s)$ is NOT a probability density in z.
- Derivation of Kolmogorov forward equation, method of test function
- Meaning of backward equation: solution at (z, t) = the average reward at the end time T given starting at position z at time $(T-t)$
- Meaning of forward equation: solution at (x, t) = mass density at time t

---

Recap

Different interpretations of SDE

The Stratonovich interpretation of $dX = b(X,t)dt + \sqrt{a(X,t)}\,dW$ is equivalent to the Ito interpretation of $dX = \left(b(X,t) + \frac{1}{4}a_x(X,t)\right)dt + \sqrt{a(X,t)}\,dW$.

The "correct" interpretation is selected in the modeling process.

Transition probability density (a 4-variable function)

$$q(x, \underbrace{t}_{\text{end time}} \mid z, \underbrace{s}_{\text{starting time}}) \equiv \frac{1}{dx}\Pr\left(x < X(t) \le x + dx \,\middle|\, X(s) = z\right), \qquad t > s$$

Backward view (the law of total probability)

We fix (x, t) and view q as a function of (z, s):

$$\underbrace{q(x,t\,|\,z,s)}_{\substack{q(\bullet,s)\\ [s\to t]}} = \int \underbrace{q(x,t\,|\,z+y,s+ds)}_{\substack{q(\bullet,s+ds)\\ [s+ds\to t]}}\ \underbrace{q(z+y,s+ds\,|\,z,s)}_{\substack{\text{density of } dX\\ [s\to s+ds]}}\,dy$$

$$q(\,\cdot\,, s+ds) \longrightarrow q(\,\cdot\,, s)$$

We move the starting time backward from $(s+ds)$ to s.

Forward view (the law of total probability)

We fix (z, s) and view q as a function of (x, t):

$$\underbrace{q(x,t+dt\,|\,z,s)}_{\substack{q(\bullet,t+dt)\\ [s\to t+dt]}} = \int \underbrace{q(x,t+dt\,|\,y,t)}_{\substack{\text{density of } X(t+dt)|X(t)=y\\ [t\to t+dt]}}\ \underbrace{q(y,t\,|\,z,s)}_{\substack{q(\bullet,t)\\ [s\to t]}}\, dy$$

$$q(\cdot\,,t) \longrightarrow q(\cdot\,,t+dt)$$

We move the end time forward from t to $(t+dt)$.

Backward equation, final value problem, converting to IVP.

Before we derive the forward equation, we clarify two issues.

Issue 1: the "correct" interpretation of an SDE is selected in the modeling process.

We compare the two interpretations in the SDE below.

$$dX = \alpha X dW$$

Ito interpretation:

$$X(t+dt) - X(t) = \alpha X(t)dW(t)$$

$$E(X(t+dt) \mid X(t) = x) = x$$

Example 1: Consider a fair game between you and a casino.

Let $X(t)$ = your cash at time t.

- Suppose in each dt, you bet a small random percent of your current cash.

  In each dt, you are equally likely to win or lose that small random percent.

  $$E(X(t+dt) \mid X(t) = x) = x$$

  $\underbrace{X(t+dt)-X(t)=\alpha X(t)dW(t)}_{\text{Ito}}$ is appropriate for this situation.

- Suppose in each dt, you bet a small fixed percent of your current cash (for example, you bet $c\sqrt{dt}\,X(t)$ in each dt).

  $$E(X(t+\Delta t) \mid X(t) = x) = x$$

  $\underbrace{X(t+\Delta t)-X(t)=\alpha X(t)dW(t)}_{\text{Ito}}$ is appropriate for Δt = many steps of dt.

Stratonovich interpretation

$$X(t+dt)-X(t)=\alpha\frac{X(t)+X(t+dt)}{2}dW(t)$$

Let $Y(t)\equiv\log\big(X(t)\big)$, $dY\equiv Y(t+dt)-Y(t)$.

We examine the evolution of $Y(t)$. We write

$$X(t)=e^{Y(t)}, \quad X(t+dt)=e^{Y(t+dt)}=e^{Y(t)+dY}$$

Substitute into the SDE for X, divide by $e^{Y(t)+dY/2}$ and expand in dY

$$e^{Y(t)+dY}-e^{Y(t)}=\frac{\alpha}{2}\left(e^{Y(t)}+e^{Y(t)+dY}\right)dW(t)$$

$$e^{dY/2}-e^{-dY/2}=\frac{\alpha}{2}\left(e^{dY/2}+e^{-dY/2}\right)dW(t)$$

$$dY+O\left((dY)^3\right)=\alpha\left[1+O\left((dY)^2\right)\right]dW(t)$$

Remark: dividing by $e^{Y(t)+dY/2}$ made it simple!

Use $dY=\log(X(t)+dX)-\log X(t)\approx\dfrac{dX}{X(t)}\approx dW=O(\sqrt{dt})$ and neglect o(dt) terms.

$$dY=\alpha\, dW(t)$$

$$\log\left(X(t+dt)\right)-\log\left(X(t)\right)=\alpha dW(t) \quad\Rightarrow\quad X(t+dt)=X(t)e^{\alpha dW(t)}$$

$$E\left(\log X(t+dt)\middle|X(t)=x\right)=\log x$$

$$E\left(X(t+dt)\middle|X(t)=x\right)=E\left(x\,e^{\alpha dW(t)}\right)=x\,E\left(1+\alpha dW+\tfrac{1}{2}\alpha^2(dW)^2+o(dt)\right)$$
$$=x\left(1+\tfrac{1}{2}\alpha^2 dt+o(dt)\right)>x$$

Example 2:

Consider a game between you and a "casino" (the stock market).

Let $X(t)$ = your net worth at time t.

Suppose in each dt, you net worth is equally likely to be multiplied or divided by a factor close to 1 (for example, a factor of $\left(1+c\sqrt{dt}\right)$)

$$E\left(\log X(t+dt)\middle|X(t)=x\right)=\log x$$

$$\underbrace{X(t+dt)-X(t)=\alpha\frac{X(t)+X(t+dt)}{2}dW(t)}_{\text{Stratonovich}}$$ is appropriate for this situation.

$$E\left(X(t+dt)\middle|X(t)=x\right)=x\frac{1}{2}\left(1+c\sqrt{dt}+\frac{1}{1+c\sqrt{dt}}\right)=x\left(1+\frac{1}{2}c^2dt+o(dt)\right)>x$$

$$\underbrace{X(t+dt)-X(t)=\alpha X(t)dW(t)}_{\text{Ito}}$$ is not appropriate for this situation.

The two interpretations are related to each other.

Stratonovich of $dX = \mu X dt + \alpha X dW(t)$

is equivalent to Ito of the modified SDE, $dX = \left(\mu X + \frac{1}{2}\alpha^2 X\right)dt + \alpha X dW(t)$.

Now back to the discussion of transition probability density.

Issue 2: In general, $q(x, t \mid z, s)$ is NOT a probability density in z.

Example: Ornstein-Uhlenbeck process

$$dY = -\beta Y dt + \sqrt{\gamma^2}\, dW$$

Recall that previously we derived

$$\left(Y(t)\middle| Y(0) = y_0\right) \sim N\left(e^{-\beta t} y_0,\ \frac{\gamma^2}{2\beta}(1 - e^{-2\beta t})\right)$$

For simplicity, we set $\beta = 1$ and $\gamma^2/(2\beta) = 1$. Apply it to $[s, t]$ with $t > s$.

$$\left(Y(t)\middle| Y(s) = z\right) \sim N\left(e^{-(t-s)} z,\ (1 - e^{-2(t-s)})\right)$$

The transition probability density of Y is

$$q(x,t \mid z,s) = \frac{1}{\sqrt{2\pi(1 - e^{-2(t-s)})}} \exp\left(\frac{-(x - e^{-(t-s)} z)^2}{2(1 - e^{-2(t-s)})}\right)$$

As a function of x, it is a probability density.

We examine it as a function of z. For simplicity, we set $s = 0$, $t = 1$ and $x = 0$.

$$q(0,1 \mid z,0) = \frac{1}{\sqrt{2\pi(1 - e^{-2})}} \exp\left(\frac{-(0 - e^{-1} z)^2}{2(1 - e^{-2})}\right) = \frac{e^1}{\sqrt{2\pi(e^2 - 1)}} \exp\left(\frac{-z^2}{2(e^2 - 1)}\right)$$

$$= e^1 \cdot \rho_{N\left(0,\ (e^2 - 1)\right)}(z) \qquad \text{not a probability density}$$

Key observations:

- We should not expect $\int q(x, t \mid z, s) dz = 1$
- We should not expect $\int q(x, t \mid z, s) dz$ to be conserved with respect to s.

Derivation of the forward equation for SDE $dX = b(X,t)dt + \sqrt{a(X,t)}\, dW$

We fix (z, s) and view q as a function of (x, t): $\quad q(x, t) \equiv q(x, t \mid z, s)$

Forward view:

$$\underbrace{q(x,t+dt\,|\,z,s)}_{q(\bullet,t+dt)} = \int \underbrace{q(x,t+dt\,|\,y,t)}_{\substack{\text{density of}\\ X(t+dt)|X(t)=y}}\ \underbrace{q(y,t\,|\,z,s)}_{q(\bullet,t)}\ dy$$

Key for the derivation:

As $dt \to 0$, $q(x,t+dt\,|\,y,t)$ is significant only for small $|y-x|$. As a result, the integral is dominated by contribution from small $|y-x|$.

The old approach of expanding $q(y, t)$ around $y = x$ won't work!

$$q(y,t) = q(x+(y-x),t) = \cdots + q_x(x,t\,|\,z,s)(y-x) + \cdots$$

$$\int q(x,t+dt\,|\,y,t)q(y,t)dy = \cdots + \int \underbrace{q\big(x,t+dt\,|\,y,t\big)}_{\substack{\text{This is NOT}\\ \text{a density in } y!}}\ \underbrace{q_x(x,t)}_{\substack{\text{This is fine. It is}\\ \text{independent of } y}}\ (y-x)dy + \cdots$$

- Integrating a density leads to moments.

$$\int (x-y)\underbrace{q(x,t+dt\,|\,y,t)}_{\text{This is a density in } x.}\ dx = E(dX)$$

- Integrating a non-density leads to nowhere.

$$\int (y-x)\underbrace{q(x,t+dt\,|\,y,t)}_{\substack{\text{This is NOT}\\ \text{a density in } y!}}\ dy = \text{unknown}$$

Strategy:

We multiply it by a test function $h(x)$ and then integrate with respect to x.

Implementation

Let $h(x)$ be a smooth function with compact support.

Definition:

We say function $h(x)$ has compact support if there exists M such that

$$h(x) = 0 \quad \text{for} \quad |x| > M$$

We multiply both sides of the master equation by $h(x)$ and integrate with respect to x.

$$\text{LHS} = \int q(x,t+dt)h(x)dx = \int \Big[q(x,t) + q_t dt + o(dt)\Big]h(x)dx$$

$$\text{RHS} = \int \Big[\int q(y,t)q(x,t+dt\,|\,y,t)dy\Big]h(x)dx$$

Changing the order of integration leads to

$$\text{RHS} = \int q(y,t)\left[\int \underbrace{q(x,t+dt\,|\,y,t)}_{\text{This is a density in } x.}\ h(x)dx\right]dy \qquad \text{(E01)}$$

The inner integral is dominated by contribution from small $|x-y|$.

We expand $h(x)$ around $x = y$.

$$h(x)=h(y+(x-y))=h(y)+h_y(y)(x-y)+\frac{h_{yy}(y)}{2}(x-y)^2+O\left((x-y)^3\right)$$

In the inner integral, these expansion terms lead to moments of dX.

Recall the moments of dX for the SDE $dX=b(X,t)dt+\sqrt{a(X,t)}\,dW$ (Ito).

$$E\left((dX)^0\right)=\int q(x,t+dt\,|\,y,t)dx=1$$

$$E\left((dX)^1\right)=\int q(x,t+dt\,|\,y,t)(x-y)dx=b(y,t)dt+o(dt)$$

$$E\left((dX)^2\right)=\int q(x,t+dt\,|\,y,t)(x-y)^2dx=a(y,t)dt+o(dt)$$

$$E\left((dX)^n\right)=\int q(x,t+dt\,|\,y,t)(x-y)^n dx=o(dt)\,,\qquad n\geq 3$$

The inner integral becomes

$$\int q(x,t+dt\,|\,y,t)h(x)dx=h(y)+h_y(y)b(y,t)dt+\frac{h_{yy}(y)}{2}a(y,t)dt+o\left(dt\right)$$

- Substituting this result into the outer integral in (E01),
- integrating by parts, and using the compactness of $h(y)$, we get

$$\text{RHS}=\int\left[q-\left(b(y,t)q\right)_y dt+\frac{1}{2}\left(a(y,t)q\right)_{yy}dt\right]h(y)dy+o(dt),\qquad q\equiv q(y,t)$$

- Renaming variable y as x, we write it as

$$\text{RHS}=\int\left[q-\left(b(x,t)q\right)_x dt+\frac{1}{2}\left(a(x,t)q\right)_{xx}dt\right]h(x)dx+o(dt),\qquad q\equiv q(x,t)$$

- Subtracting $\int q(x,t)h(x)dx$ from both LHS and RHS,
- dividing by dt, and taking the limit as $dt\to 0$, we arrive at

$$\text{LHS}=\int q_t h(x)dx$$

$$\text{RSH}=\int\left[-\left(b(x,t)q\right)_x+\frac{1}{2}\left(a(x,t)q\right)_{xx}\right]h(x)dx,\qquad q\equiv q(x,t)$$

Since LHS = RHS for all test function $h(x)$, we conclude

$$q_t=-\left(b(x,t)q\right)_x+\frac{1}{2}\left(a(x,t)q\right)_{xx}$$

This is called the Fokker-Planck equation or the Kolmogorov forward equation.

Conservation form:

The forward equation has the conservation form

$$q_t = -\frac{\partial}{\partial x} J(x,t)$$

where $J(x,t) \equiv b(x,t)q - \frac{1}{2}\big(a(x,t)q\big)_x$ is the probability flux

Terminology: flux ≡ flow per unit time

Remarks:

- Solution of $q_t = -\frac{\partial}{\partial x} J(x,t)$ is conserved:

$$\int_a^b q(x,t_2)dx - \int_a^b q(x,t_1)dx = \underbrace{\int_{t_1}^{t_2} J(a,t)dt}_{\text{In-flow}} - \underbrace{\int_{t_1}^{t_2} J(b,t)dt}_{\text{Out-flow}}$$

  Change in $\int_a^b q(x,t)dx$ is attributed to in-flow at $x = a$ and out-flow at $x = b$.
- In contrast, the backward equation is not in the conservation form.

$$q_s = -b(z,s)q_z - \frac{1}{2}a(z,s)q_{zz}$$

  In general, solution of the backward equation is not conserved.

The initial value problem (IVP) for $q(x, t) \equiv q(x, t \mid z, 0)$

$$\begin{cases} q_t = -\big(b(x,t)q\big)_x + \frac{1}{2}\big(a(x,t)q\big)_{xx} \\ q(x,t \mid z,0)\big|_{t=0} = \delta(x-z) \end{cases}$$

We solve it forward from $t = 0$ to $t = T$.

Autonomous SDEs:

$$dX = b(X)dt + \sqrt{a(X)}\,dW, \qquad b(x,t) = b(x), \quad a(x,t) = a(x)$$

- There is no explicit dependence on time.
- If we shift in time, the evolution remains the same.

The IVP of backward equation has a simple form in the autonomous case.

Backward equation in the autonomous case:

We shift in time by $(T-\tau)$ to write

$\phi(z, \tau) \equiv q(x, T \mid z, T-\tau) = q(x, \tau \mid z, 0)$ where x and T are fixed.

The IVP for $\phi(z,\tau)$ is

$$\begin{cases} \phi_\tau = \beta(z,\tau)\phi_z + \dfrac{1}{2}\alpha(z,\tau)\phi_{zz}\,, \qquad \beta(z,\tau)\equiv b(z,T-\tau)\,,\quad \alpha(z,\tau)\equiv a(z,T-\tau) \\ \phi(z,\tau)\Big|_{t=0} = \delta(z-x) \end{cases}$$

In the autonomous case, $\beta(z,\tau) = b(z)$, $\alpha(z,\tau) = a(z)$. For simplicity, we change back to (deceptively) simple notations:

$$\tau \to t, \qquad \phi(z,\tau) \to q(z,t).$$

The IVP for $q(z,t) \equiv q(x, T \mid z, T-t) = q(x, t \mid z, 0)$ is

$$\begin{cases} q_t = b(z)q_z + \dfrac{1}{2}a(z)q_{zz} \\ q(z,t)\Big|_{t=0} = \delta(z-x) \end{cases} \qquad \text{where } x \text{ is a parameter.}$$

Remark: In applications, end time T is fixed and t in $q(z,t)$ refers to the time until the end time, corresponding to real time $(T-t)$.

Forward equation in the autonomous case:

The IVP for $q(x,t) \equiv q(x, t \mid z, 0)$ is

$$\begin{cases} q_t = -\big(b(x)q\big)_x + \dfrac{1}{2}\big(a(x)q\big)_{xx} \\ q(x,t)\Big|_{t=0} = \delta(x-z) \end{cases} \qquad \text{where } z \text{ is a parameter.}$$

Meaning of the backward equation with a *general initial condition*

We consider the autonomous SDE $dX = b(X)dt + \sqrt{a(X)}\,dW$.

$$\begin{cases} u_t = b(z)u_z + \dfrac{1}{2}a(z)u_{zz} \\ u(z,t)\Big|_{t=0} = u_0(z) \end{cases} \qquad \text{(BE\_IVP1)}$$

Recall two examples we studied.

- For the transition PD $q(z,t) \equiv q(x, T \mid z, T-t) = q(x, t \mid z, 0)$

$$q(z,t)\Big|_{t=0} = \delta(z-x)$$

- For the probability of winning bet $X(T) \ge x_C$ given $X(T-t) = z$

$$u(z,t) = \Pr\Big(X(T) \ge x_c \,\Big|\, X(T-t) = z\Big)$$

$$u(z,t)\Big|_{t=0} = \begin{cases} 1, & z \ge x_C \\ 0, & z < x_C \end{cases}$$

Here we look at a general initial condition: $u(z, t)|_{t=0} = u_0(z)$.

It is straightforward to verify that the solution of (BE_IVP1) is

$$u(z,t) = \int q(x,T \,|\, z,T-t)\, u_0(x)\, dx \qquad \text{(B01)}$$

<u>Observations</u>:

- $q(x, T \mid z, T-t)$ is the transition probability density.
- Real time T is a future time, for example, the expiration date of an option.
- Variable t in the backward equation is the time until the end time. Variable t corresponds to *real time* $(T-t)$.

<u>Meaning of solution $u(z, t)$</u>

Suppose the reward is determined <u>at *real time T*</u>, based on $X(T)$.

Let $u_0(x)$ be the reward function, which maps $X(T)$ to reward:

the amount of reward = $u_0(X(T))$

<u>Example</u>:

Let $X(t)$ = the market price of a stock at time t.

Consider a "call" option to buy the stock at price x_C at time T.

<u>Terminology:</u>

A call option = the right (not obligation) to buy a certain number of shares of the stock at a specified price at a preset time (expiration date).

A put option = the right (not obligation) to sell a certain number of shares of the stock at a specified price at a preset time (expiration date).

The amount of reward for owning the call option depends on $X(T)$. It is

$$u_0\big(X(T)\big) = \begin{cases} X(T) - x_C, & X(T) > x_C \\ 0, & X(T) \le x_C \end{cases}$$

Suppose X starts at position z at <u>*real time* $(T-t)$</u>. The conditional distribution of $X(T)$ given $X(T-t) = z$ is described by the transition PD

$q(x, T \mid z, T-t)$

The conditional expected amount of reward given $X(T-t) = z$ is

$$E\Big(u_0\big(X(T)\big)\Big| X(T-t) = z\Big) = \int \underbrace{q(x,T \,|\, z,T-t)}_{\substack{\text{transition density of} \\ X(T)=x|X(T-t)=z}} \underbrace{u_0(x)}_{\substack{\text{reward} \\ \text{for } X(T)=x}} dx \qquad \text{(B02)}$$

This is exactly the same as the solution $u(z, t)$ given in (B01).

Summary (meaning of the backward equation)

Suppose the reward is determined at *real time T*, as $u_0(X(T))$.

The solution of the backward equation, $u(z, t)$, is the expected amount of reward given $X(T-t) = z$.

$$\underbrace{u(z,t)}_{\substack{\text{solution of the}\\ \text{backward equation}}} = \underbrace{E\Big(u_0\big(X(T)\big)\Big|X(T-t)=z\Big)}_{\substack{\text{expected amount of reward}\\ \text{given } X(T-t)=z}}$$

The backward equation describes the backward time evolution of the expected amount of reward. The end time is fixed at T. In the backward time evolution, the start time is gradually moved backward from T to $(T-t)$.

In general, the expected amount of reward $q(z, t)$ is not conserved.

$$\int q(z,t_1)dz \neq \int q(z,t_2)dz$$

This is related to that the backward equation is NOT in the conservation form.

Meaning of the forward equation with a *general initial condition*

We consider the autonomous SDE $dX = b(X)dt + \sqrt{a(X)}\,dW$

$$\begin{cases} p_t = -\big(b(x)p\big)_x + \dfrac{1}{2}\big(a(x)p\big)_{xx} \\ p(x,t)\big|_{t=0} = p_0(x) \end{cases} \qquad \text{(FE\_IVP1)}$$

It is straightforward to verify that the solution of (FE_IVP1) is

$$p(x,t) = \int \underbrace{q(x,t\,|\,z,0)}_{\text{start time is 0}} p_0(z)dz \qquad \text{(F01)}$$

Observations:

- $q(x, t \mid z, 0)$ is the transition probability density.
- Variable t in the forward equation is the time elapsed since the start time.

Meaning of solution $p(x, t)$

Consider a set of X. For example, a set of 7 particles, $\{X_j(t), j = 1, 2, ..., 7\}$.

All averages are based on an ensemble of the set.

Ensemble = $\{ \{X_j(t, \omega), j = 1, 2, ..., 7\},\ \omega \in \Omega \}$

a collection of an infinite number of *independent* copies of the set.

Mass density at position x at time t is

$$\rho(x,t)=\frac{1}{dx}E_\omega\Big(\#\text{ of X}_j(t,\omega)\text{'s in }[x,x+dx]\Big)=\sum_{j=1}\rho(X_j(t)=x)$$

Notation: $\rho(X=x)\equiv\rho_X(x)$.

Let $p_0(x)$ be the mass density at time 0.

The mass density at time t is given by the law of total probability.

$$\sum_{j=1}\rho(X_j(t)=x)=\sum_{j=1}\int\underbrace{\rho(X_j(t)=x\,|\,X_j(0)=z)}_{q(x,t|z,0),\text{ independent of }j}\rho(X_j(0)=z)dz$$

$$=\int q(x,t\,|\,z,0)\underbrace{\sum_{j=1}\rho(X_j(0)=z)}_{p_0(z)}dz=\int q(x,t\,|\,z,0)p_0(z)dz \qquad \text{(F02)}$$

This is exactly the same as the solution $p(x, t)$ given in (F01).

Summary:

We consider a set of X because the mass density is more general.

Suppose $p_0(x)$ is the mass density of X at time 0.

The solution of the forward equation, $p(x, t)$ = the mass density of X at time t.

The forward equation describes the forward time evolution of mass density.

The mass density $p(x, t)$ is conserved.

$$\int_a^b p(x,t_2)dx-\int_a^b p(x,t_1)dx=\underbrace{\int_{t_1}^{t_2}J(a,t)dt}_{\text{In-flow}}-\underbrace{\int_{t_1}^{t_2}J(b,t)dt}_{\text{Out-flow}}$$

This is related to that the forward equation is in the conservation form.

A tricky issue: When the number of particle in the set is very large, one copy is enough for describing the behavior of the ensemble and thus is often called an ensemble. In that case, mass density is also called ensemble density.

Example: an ensemble of 3×10^{16} air molecules in a volume of 1 $(\text{mm})^3$.

AM216 Stochastic Differential Equations

Lecture 12
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- Forward/backward equations in terms of differential operators, inner product, adjoint operator
- An alternative derivation of the forward equation, average reward for a population
- Boundary conditions for the forward/backward equations
- Exit problem, probability of exit by time *t*, average exit time

---

Recap

Autonomous SDE: $dX = b(X)dt + \sqrt{a(X)}\,dW$ (Ito)

IVP of the backward equation with a general initial condition

$$\begin{cases} u_t = b(z)u_z + \frac{1}{2}a(z)u_{zz} \\ u(z,t)\big|_{t=0} = u_0(z) \end{cases}$$

Meaning:

$u_0(z)$: reward function; reward is determined at the end time T as $u_0(X(T))$.

$u(z, t)$ = The average reward given starting at position z at time $(T-t)$.

Variable t in the backward equation corresponds to *real time* $(T-t)$.

$u(z, t)$ is (in general) not conserved; the backward equation is not conservative.

IVP of the forward equation with a general initial value

$$\begin{cases} p_t = -\left(b(x)p\right)_x + \frac{1}{2}\left(a(x)p\right)_{xx} \\ p(x,t)\big|_{t=0} = p_0(x) \end{cases}$$

Meaning:

$p_0(z)$: mass density at time 0.

$p(x, t)$ = mass density at time t.

Variable t in the forward equation = *real time t*.

$p(z, t)$ is conserved; the forward equation is conservative.

Forward equation and backward equation in terms of differential operators

We consider the autonomous SDE:

$$dX = b(X)dt + \sqrt{a(X)}\,dW\,, \qquad \text{(Ito)}$$

We introduce linear differential operator L_Z.

$$L_z = b(z)\frac{\partial \bullet}{\partial z} + \frac{1}{2}a(z)\frac{\partial^2 \bullet}{\partial z^2}$$

which means $L_z[u] = b(z)\dfrac{\partial u}{\partial z} + \dfrac{1}{2}a(z)\dfrac{\partial^2 u}{\partial z^2}$

Short story:

1. Backward equation in terms of L_Z:

$$u_t = L_z\left[u\right]$$

2. Forward equation in terms of L_Z:

$$p_t = L_z^*\left[p\right]$$

where L_Z^* is the adjoint operator of L_Z, which we will introduce and discuss.

3. An alternative derivation of forward equation that is more intuitive and conceptually simpler than the method of test function.
4. Comments on ensemble average and boundary effect

Long story:

1. **Backward equation** can be written as

$$u_t = L_z\left[u\right].$$

This follows directly from the definition of operator L_Z.

To write out the forward equation in terms of L_Z, we need to define a few things.

Definition (inner product)

The inner product of two functions is defined as

$$\langle u_1, u_2 \rangle \equiv \int u_1(z) u_2(z) dz$$

Definition (adjoint operator)

The adjoint operator of L is denoted by L^*, and is defined by the condition

$\langle u, L^*[v] \rangle = \langle L[u], v \rangle$ for all functions $u(z)$ and $v(z)$ of compact support.

Example:

Let A be an $n\times n$ matrix. We view matrix A as a linear operator: $\mathbb{R}^n \to \mathbb{R}^n$

$$u \longrightarrow Au$$

For vectors, the inner product is

$$\langle u, v \rangle = \sum u_i v_i = u^T v$$

We use the definition of adjoint operator to find A^*.

$$\underbrace{\langle u, A^* v \rangle = \langle Au, v \rangle}_{\text{definition of } A^*} \to (Au)^T v \to u^T A^T v \to u^T (A^T v) \to \langle u, A^T v \rangle$$

==> $\langle u, A^* v \rangle = \langle u, A^T v \rangle$ for all vectors u and v.

==> $A^* = A^{\mathrm{T}}$

For matrix A, the adjoint operator of A is A^{T}.

Example:

Consider differential operator

$$D_x = a(x) \frac{\partial^2 \bullet}{\partial x^2}$$

We use the definition of adjoint operator to find D_x^*.

$$\underbrace{\langle u, D_x^*[v] \rangle = \langle D_x[u], v \rangle}_{\text{definition of } D_x^*} \to \int a(x) \frac{\partial^2 u}{\partial x^2} v(x) dx \to \int \frac{\partial^2 u}{\partial x^2} \big(a(x) v(x) \big) dx$$

Integrating by parts twice, we write the RHS as

$$\text{RHS} = -\int \frac{\partial u}{\partial x} \cdot \frac{\partial}{\partial x} \big(a(x) v(x) \big) dx = \int u(x) \cdot \frac{\partial^2}{\partial x^2} \big(a(x) v(x) \big) dx = \left\langle u, \frac{\partial^2}{\partial x^2} \big(a(x) v \big) \right\rangle$$

==> $\langle u, D_x^*[v] \rangle = \left\langle u, \frac{\partial^2}{\partial x^2} \big(a(x) v \big) \right\rangle$ for all functions $u(x)$ and $v(x)$.

==> $$D_x^* = \frac{\partial^2}{\partial x^2}\left(a(x)\bullet\right)$$

Example:

Consider differential operator

$$L_z = b(z)\frac{\partial\bullet}{\partial z} + \frac{1}{2}a(z)\frac{\partial^2\bullet}{\partial z^2}$$

The adjoint operator of L_z is

$$L_z^* = -\frac{\partial}{\partial z}\left(b(z)\bullet\right) + \frac{1}{2}\frac{\partial^2}{\partial z^2}\left(a(z)\bullet\right) \qquad \text{(homework problem)}$$

2. **Forward equation in terms of** L_Z:

Comparing the expression of forward equation we derived previously and the expression of L_Z^* in the example above, we write the forward equation as

$$p_t = L_z^*\left[p\right] \qquad \text{where } L_Z^* \text{ is the adjoint operator of } L_Z.$$

Note: We have changed variable x to variable z.

3. **A more intuitive derivation of the forward equation**

Logically,

- We discard the forward equation derived using the method of test function.
- Instead we derive the governing equation for the mass density.

First, we look at the IVP of the backward equation

$$\begin{cases} u_t = A[u] \\ u(x,0) = u_0(x) \end{cases} \qquad \text{where } A = L_x.$$

Note: We have changed variable z to variable x.

Function $u(x, t)$ has the meaning:

- The amount of reward is determined at the end time T as $u_0(X(T))$.
- $u(x, t)$ = average reward given starting at position x at time $(T-t)$.

$$u(x,t) = E\left(u_0(X(T)) \,\middle|\, X(T-t) = x\right)$$

- Variable t in the backward equation corresponds to *real time* $(T-t)$.

Next, we look at the IVP of the forward equation

$$\begin{cases} p_t = B[p] \\ p(x, T-t_0) = p_0(x) \end{cases}$$

We are going to show $B = A^*$.

Function $p(x, t)$ has the meaning:

- $p(x, T-t_0) = p_0(x)$ is the mass density at *real time* $(T-t_0)$.
- $p(x, t)$ = the mass density at *real time* $t > T-t_0$

Key observation:

Consider a population with mass density $p(x, T-t_0)$ at *real time* $(T-t_0)$.

We combine $u(\ )$ and $p(\ )$ to calculate the average reward for the population

$u(x, t_0)$, the conditional average reward given $X(T-t_0) = x$, and

$p(x, T-t_0)$ mass density of $X(T-t_0)$.

The law of total expectation gives

$$\text{Average reward} = \int u(x,t_0)p(x,T-t_0)dx \qquad \text{(Expression 1)}$$

On the other hand, by definition, the reward occurs at the end time T. The average reward for the population is determined by the reward function $u_0(x) \equiv u(x, 0)$ and the mass density $p(x, T)$ at time T.

$$\text{Average reward} = \int u_0(x)p(x,T)dx \qquad \text{(Expression 2)}$$

Equating the two expressions of average reward, we obtain

$$\int \underbrace{u(x,t_0)}_{\substack{\text{solved from the} \\ \text{backward Eq}}} \underbrace{p(x,T-t_0)}_{\substack{\text{mass density} \\ \text{at time } (T-t_0)}} dx = \int \underbrace{u_0(X)}_{\substack{\text{reward} \\ \text{function}}} \underbrace{p(x,T)}_{\substack{\text{solved from the} \\ \text{forward Eq}}} dx \quad \text{for all } t_0 > 0 \qquad \text{(E01)}$$

Writing out operator B

Given the mass density $p(x, T-t_0)$ at time $(T-t_0)$, we have two ways to calculate the average reward for the ensemble

- Solve the backward equation with the given $u_0(x)$ to calculate $u(x, t_0)$ …
- Solve the forward equation with the given $p(x, T-t_0)$ to calculate $p(x, T)$ …

This is how the backward equation and the forward equation are related in (E01). We use (E01) to write out operator B in terms of operator A.

Since (E01) is valid for all $t_0 > 0$, we set $t_0 = \Delta t$.

$$\int u(x,\Delta t)p(x,T-\Delta t)dx = \int u_0(x)p(x,T)dx \qquad \text{(E02)}$$

We set $p(x, T-\Delta t) = v(x)$.

We write $u(x, \Delta t)$ and $p(x, T)$ in terms of $u_0(x)$ and $v(x)$ as follows.

Backward equation: $u_t = A[u]$

$$u(x,\Delta t) = u(x,0) + \Delta t\, u_t(x,0) + o(\Delta t)$$
$$= u_0(x) + \Delta t\, A[u_0(x)] + o(\Delta t)$$

Forward equation: $p_t = B[p]$

$$p(x,T) = p(x,T-\Delta t) + \Delta t\, p_t(x,T-\Delta t) + o(\Delta t)$$
$$= v(x) + \Delta t\, B[v(x)] + o(\Delta t)$$

Substituting into (E02) leads to

$$\text{LHS} = \int \big(u_0(x) + \Delta t\, A[u_0(x)] + o(\Delta t)\big) v(x)\,dx$$
$$= \int u_0(x)v(x)\,dx + \Delta t \int A[u_0(x)]v(x)\,dx + o(\Delta t)$$

$$\text{RHS} = \int u_0(x)\big(v(x) + \Delta t\, B[v(x)] + o(\Delta t)\big)\,dx$$
$$= \int u_0(x)v(x)\,dx + \Delta t \int u_0(x)B[v(x)]\,dx + o(\Delta t)$$

Subtracting $\int u_0(x)v(x)dx$ from both LHS and RHS, dividing by Δt and taking the limit as $\Delta t \to 0$, we arrive at

$$\int A[u_0(x)]v(x)\,dx = \int u_0(x)B[v(x)]\,dx \qquad \text{for all } u_0(x) \text{ and } v(x)$$

In terms of inner product, it becomes

$$\langle u_0, B[v]\rangle = \langle A[u_0], v\rangle \qquad \text{for all } u_0(x) \text{ and } v(x)$$

which implies $B = A^*$.

The end of intuitive derivation of the forward equation

4. Comments on ensemble average and boundary effect

- Ensemble average

  On the RHS of (E01), the reward is averaged over all independent copies in the ensemble starting with mass density $p(x, T-t_0)$ at time $(T-t_0)$.

  $$\int u_0(X)p(x,T)\,dx\,, \qquad p(x,T) = \text{density based on the intact ensemble}$$

  If the ensemble is modified in $[(T-t_0), T]$, then (E01) is no longer valid.

  Example: Consider a call option of a stock.

  A call option is the right (not obligation) to buy a certain number of shares of the stock at a specified price at a preset time (expiration date). Let

  T = the expiration time

x_C = specified price in the call option

$X(t)$ = the stock price at time t.

t_0 = time until expiration; corresponding to real time $(T-t_0)$

x_0 = the starting value of $X(T-t_0)$

The reward for the call option holder is realized at expiration and is determined by the market price of the stock at expiration. The reward function is

$$u_0\big(X(T)\big) = \begin{cases} X(T)-x_C\,, & X(T) > x_C \\ 0\,, & X(T) \le x_C \end{cases}$$

The expected reward given $X(T-t_0)$ is

$$E\Big(u_0(X(T))\Big|X(T-t_0)=x_0\Big)$$

This is the average over all independent copies in the ensemble starting with $X(T-t_0) = x_0$. The price of the call option reflects the expected reward. We will discuss the option pricing later (Black Scholes model).

Let $(T-t_2) > (T-t_0)$ be a later time. $t_2 < t_0$ means a shorter time until expiration. When the stock price $X(T-t_2) = x_2$ becomes known, the expected reward is updated to

$$E\Big(u_0(X(T))\Big|X(T-t_2)=x_2\Big)$$

This is the average over a modified ensemble, consisting of those copies in the original ensemble satisfying $X(T-t_2) = x_2$. When the original ensemble is modified, the expected reward is also modified.

$$E\Big(u_0(X(T))\Big|X(T-t_2)=x_2\Big) \ne E\Big(u_0(X(T))\Big|X(T-t_0)=x_0\Big)$$

<u>End of example</u>

Boundary effect is another way of modifying the original ensemble.

<u>Question:</u> In the derivation of forward equation above, if a boundary is present, how should we deal with the boundary effect?

- <u>Boundary effect in the derivation of forward equation</u>

Recall that in the derivation of backward equation, we start with $X(t) = x$ away from boundary; we select dt small enough such that

Pr(X hitting boundary in $[t, t+dt]$) = negligible …

In the derivation of forward equation above, we select $u_0(x)$ and $v(x)$ with supports not touching the boundary, and we select $t_0 = \Delta t$ small enough such that starting within the supports, the probability of touching boundary in $[t, t+dt]$ is negligible.

<u>Summary:</u>

Evolution equation itself is not affected by boundary effect.

The evolution of a given initial condition, however, is affected by boundary effect.

End of long story

Boundary conditions

SDE: $dX = b(X)dt + \sqrt{a(X)}\,dW$ (Ito)

Absorbing boundary at $x = L$

- When a particle gets to $x = L$, it is removed from the set of particles.
- When a game reaches $x = L$, it is ended (removed from the set of ongoing games).

For the forward equation, the absorbing boundary is described by

$$p(x,t)\Big|_{x=L} = 0$$

That is, the mass density at $x = L$ is zero.

For the backward equation,

$u(x, t)$ = average reward given starting at position x at time $(T-t)$

The absorbing boundary is described by

$$u(x,t)\Big|_{x=L} = 0$$

That is, starting at $x = L$, it is removed immediately and thus cannot get any reward.

Reflecting boundary at $x = L$

- When a particle tries to go through $x = L$, it is not allowed to pass through; it is not removed; instead it is "turned back".

For the forward equation, the reflecting boundary is described by

$$J(x,t)\Big|_{x=L} = 0$$

where $J(x,t) \equiv b(x)p - \frac{1}{2}\big(a(x)p\big)_x$ is the flux.

That is, the flux through $x = L$ is zero.

For the backward equation, the reflecting boundary is described by

$$\frac{\partial u(x,t)}{\partial x}\Bigg|_{x=L} = 0$$

Derivation:

At $x = L$, we set the increment $(dX \mid X(T-t) = L)$ as follows.

$$dX = -\left|b(L)dt + \sqrt{a(L)}dW\right| = -\sqrt{a(L)}\,|dW| + O(dt)$$

==> $E(dX) = -\sqrt{a(L)}\,E(|dW|) + O(dt) = O(\sqrt{dt})$ (This is the key)

$$u(L,t) = E\left(u(L+dX, t-dt)\right) = E\left(u(L,t) + u_x(L,t)dX + O(dt)\right)$$

==> $u_x(L,t)E(dX) = O(dt)$

==> $u_x(L,t) = \dfrac{O(dt)}{E(dX)} \to 0$

Next we look at applications of the forward and the backward equations.

Exit problem:

Suppose $X(t)$ is governed by the SDE

$$dX = b(X)dt + \sqrt{a(X)}\,dW \qquad \text{(Ito)}$$

Consider the problem of exiting (i.e., escaping from) a prescribed region.

We study the time until escape, also called the exit time or the escape time.

<u>Probability of exit by time *t*</u>

Let Y = the exit time (a random variable).

Let $u(x, t)$ = probability of exiting the region by time t given starting at x at time 0.

$$u(x,t) \equiv \Pr\left(Y \le t \,\middle|\, X(0) = x\right)$$

<u>Governing equation for $u(x, t)$</u>

For x inside the region, when dt is small enough, we have

$$u(x,t) = E\left(u(x+dX, t-dt)\right) + o(dt)$$

Taylor expansion + moments of dX leads to the backward equation

$$u_t = b(x)u_x + \frac{1}{2}a(x)u_{xx}$$

<u>Average exit time</u>

Let $T(x)$ be the average exit time given that $X(0) = x$.

$$T(x) \equiv E\left(Y \,\middle|\, X(0) = x\right)$$

<u>Governing equation for $T(x)$</u>

For x inside the region, when dt is small enough, we have

$$T(x) = E\left(T(x+dX)\right) + dt + o(dt)$$

Taylor expansion + moments of dX leads to an ODE for $T(x)$.

$$\frac{1}{2}a(x)T_{xx}+b(x)T_{x}=-1$$

We look at a few examples before discussing "escape of a Brownian particle".

Example: The particle undergoes pure diffusion with no net drift:

$a(x) = 1$, $b(x) = 0$

The region is $[0, L]$. Exit can occur at either $x = 0$ or $x = L$.

We have seen this example in the Gambler's ruin problem (fair game) where

x: your initial cash; L: total cash of casino + you

$T(x)$ is the time until breaking the bank or bankrupt.

The boundary value problem (BVP) for $T(x)$ is

$$\begin{cases} T_{xx}=-2 \\ T(0)=0, \quad T(L)=0 \end{cases}$$

The solution is

$T(x) = x\,(L-x)$

In particular, we have

$T(L/2) = L^2/4$

Example: The particle undergoes diffusion with a net drift:

$a(x) = 1$, $b(x) = b$

The region is $[L_1, L_2]$. Exit can occur at either L_1 or L_2.

This example is similar to the Gambler's ruin problem (biased game).

The boundary value problem (BVP) for $T(x)$ is

$$\begin{cases} T_{xx}+2bT_{x}=-2 \\ T(L_1)=0, \quad T(L_2)=0 \end{cases}$$

The solution is

$$T(x)=\frac{1}{b}(L_2-x)-\frac{1}{b}(L_2-L_1)\cdot\frac{\exp(2b(L_2-x))-1}{\exp(2b(L_2-L_1))-1}$$

(homework problem)

For $b > 0$ (a net drift in the positive direction), we look at the limit as $L_1 \to -\infty$ (the lower boundary disappears).

$$T(x) \to \frac{1}{b}(L_2 - x) \qquad \text{as } L_1 \to -\infty .$$

This is consistent with the picture of a deterministic escape.

<u>A model for the casino:</u>

We model the situation in a casino as a two-player game:

- the casino is one player with $b > 0$ (positive net drift) and with initial cash x;
- all other gamblers are collectively viewed as the other player

Here the casino is the player in focus.

Suppose the casino starts with no cash ($x = 0$) but has a line of credit. It is solvent as long as the balance is above L_1 ($L_1 < 0$).

The end of game is defined as either the casino's balance dropping below L_1 (which is very unlikely) or the casino winning L_2 amount from the other player (getting L_2 amount of revenue). The average time until the end of game is

$$T(0) \to \frac{1}{b} L_2 \qquad \text{as } L_1 \to -\infty$$

This is the average time for the casino to get L_2 amount of revenue.

For the casino, the end of one game is also the start of a new game (transferring L_2 amount to the revenue account and resetting cash balance to 0).

AM216 Stochastic Differential Equations

Lecture 13

List of topics in this lecture

- Exit problem: reflecting boundary condition for the average exit time $T(x)$
- Escape of a Brownian particle from a potential well: Langevin equation, Smoluchowski-Kramers approximation, over-damped Langevin equation
- Non-dimensionalization, exact integral solution of $T(x)$
- Escape from a deep potential well, Kramers' approximate solution of $T(x)$

Recap

Stochastic differential equation:

$$dX = b(X)dt + \sqrt{a(X)}\, dW$$

The associated backward equation:

$$u_t = L_x[u], \qquad L_x \equiv b(x)\frac{\partial \bullet}{\partial x} + \frac{1}{2}a(x)\frac{\partial^2 \bullet}{\partial x^2}$$

Meaning of $u(x, t)$

$u(x, t)$ = average reward at end time T given $X(T-t) = x$

Absorbing boundary at $x = L$: $\quad u(x,t)\big|_{x=L} = 0$

Reflecting boundary at $x = L$: $\quad \left.\dfrac{\partial u(x,t)}{\partial x}\right|_{x=L} = 0$

The associated forward equation:

$$p_t = L_x^*[p], \qquad L_x^* \equiv \text{adjoint of } L_x = -\frac{\partial\big(b(x)\bullet\big)}{\partial x} + \frac{1}{2}\frac{\partial^2\big(a(x)\bullet\big)}{\partial x^2}$$

Meaning of $p(x, t)$

$p(x, t)$ = mass density at time t.

Absorbing boundary at $x = L$: $\quad p(x,t)\big|_{x=L} = 0$

Reflecting boundary at $x = L$: $\quad J(x,t)\big|_{x=L} = 0, \quad J(x,t) \equiv b(x)p - \tfrac{1}{2}\big(a(x)p\big)_x$

Exit problem of $dX = b(X)dt + \sqrt{a(X)}\, dW$

$T(x) = E(\text{time until exit} \mid X(0) = x)$

Governing equation: $\frac{1}{2}a(x)T_{xx} + b(x)T_x = -1$

Exit problem: boundary conditions for $T(x)$

Absorbing boundary at $x = 0$

By definition, we have $T(0) = 0$

Reflecting boundary at $x = 0$

We look at what happens when X starts with $X(0) = 0$.

$$E\left(dX \middle| X(0)=0\right) = E\left(\left|b(0)dt + \sqrt{a(0)}\, dW\right|\right)$$

$$= E\left(\sqrt{a(0)}\,|dW| + O(dt)\right) = O(\sqrt{dt})$$

We write $T(0)$ as

$$T(0) = E(T(dX)) + dt = E\left(T(0) + T'(0)dX + O(dt)\right) + dt$$

$$= T(0) + T'(0)E(dX) + O(dt)$$

==> $\quad 0 = T'(0)E(dX) + O(dt)$

==> $\quad T'(0) = \dfrac{O(dt)}{E(dX)} = \dfrac{O(dt)}{O(\sqrt{dt})} = O(\sqrt{dt}) \to 0 \quad \text{as } dt \to 0$

Conclusion: at reflecting boundary $x = 0$, $T(x)$ satisfies

$$T'(0) = 0$$

In comparison, $u(x, t)$, the probability of exiting by time t, satisfies

$$\left.\frac{\partial u(x,t)}{\partial x}\right|_{x=0} = 0$$

Example:

Suppose X is governed by $dX = b(X)dt + \sqrt{a(X)}\, dW$ with

$a(x) = 1, \quad b(x) = b$

Note: When $a(x) \equiv \text{const}$, Ito = Stratonovich

We consider the escape from $[L_1, L_2]$ where

L_1 is a reflecting boundary and

L_2 is an absorbing boundary.

Draw a slope over $[L_1, L_2]$, tilted either downward ($b > 0$) or upward ($b < 0$), with an exit at L_2 and a dead end at L_1.

The boundary value problem (BVP) for $T(x)$ is

$$\begin{cases} T_{xx} + 2bT_x = -2 \\ T'(L_1) = 0\,, \quad T(L_2) = 0 \end{cases}$$

We follow the procedure below to solve the BVP.

*) find a particular solution of the nonhomogeneous equation;

*) find a general solution of the homogeneous equation;

*) superpose the two and enforce boundary conditions; ...

The solution for a reflecting boundary at L_1

$$T(x) = \frac{1}{b}(L_2 - x) - \frac{1}{2b^2}\cdot\frac{\exp(2b(L_2 - x)) - 1}{\exp(2b(L_2 - L_1))} \qquad \text{(homework problem)}$$

We discuss 3 cases.

Case 1: $b > 0$; L_1 = negative and large.

Taking the limit as $L_1 \to -\infty$, we have

$$T(x) \to \frac{1}{b}(L_2 - x) \quad \text{as } L_1 \to -\infty$$

Remark: when the bias is driving X toward the exit and the other end of the region is far away, $T(x)$ is affected by i) how far x is from the exit and ii) the bias. $T(x)$ is not affected by the size of the region $(L_2 - L_1)$.

Case 2: $b = 0$; $[L_1, L_2]$ stays finite.

We can solve the BVP directly or we can take the limit as $b \to 0$.

We first expand $e^{bz}-1$ and e^{-bw} as $b \to 0$

$$e^{bz} - 1 = 1 + bz + \frac{1}{2}b^2z^2 + O(b^3) - 1 = bz\left(1 + \frac{1}{2}bz + O(b^2)\right)$$

$$e^{-bw} = \left(1 - bw + O(b^2)\right)$$

$$\frac{e^{bz}-1}{e^{bw}}=(e^{bz}-1)e^{-bw}=bz\left(1+\frac{1}{2}bz+O(b^2)\right)\left(1-bw+O(b^2)\right)$$

$$=bz\left(1+b(\frac{1}{2}z-w)+O(b^2)\right)$$

Apply the expansion to the analytical solution, we obtain

$$T(x)=\frac{1}{b}(L_2-x)-\frac{1}{2b^2}\cdot\frac{\exp(2b(L_2-x))-1}{\exp(2b(L_2-L_1))}$$

$$=\frac{1}{b}(L_2-x)-\frac{1}{2b^2}2b(L_2-x)\left(1+b\left((L_2-x)-2(L_2-L_1)\right)+O(b^2)\right)$$

$$=(L_2-x)\left(2(L_2-L_1)-(L_2-x)+O(b)\right)$$

For b = 0, the solution is

$$T(x)=(L_2-x)\left(2(L_2-L_1)-(L_2-x)\right)$$

Remark: when the bias is zero, $T(x)$ is affected by i) how far x is from the exit and ii) the size of the region (L_2–L_1).

Case 3: $b = -k < 0$ (where $k > 0$).

We rewrite $T(x)$ in terms of parameter $k > 0$.

$$T(x)=\frac{1}{b}(L_2-x)-\frac{1}{2b^2}\left(\exp(2b(L_2-x))-1\right)\exp(-2b(L_2-L_1))$$

$$=\frac{-1}{k}(L_2-x)+\frac{1}{2k^2}\left(1-\exp(-2k(L_2-x))\right)\exp(2k(L_2-L_1))$$

(pulling out the dominant factor)

$$=\frac{1}{2k^2}\exp(2k(L_2-L_1))\left[1-\exp(-2k(L_2-x))-2k(L_2-x)\exp(-2k(L_2-L_1))\right]$$

When $2k(L_2-x)$ is moderately large, for example, $2k(L_2-x) \geq 5$, we have

$$\exp(-2k(L_2-x))<<1$$
$$2k(L_2-x)\exp(-2k(L_2-L_1))<<1$$

It follows that $T(x)$ is approximately (in the sense of small relative error)

$$T(x)\approx\frac{1}{2k^2}\exp(2k(L_2-L_1))=\underbrace{(L_2-L_1)^2}_{\text{width}}\frac{1}{2\left(k(L_2-L_1)\right)^2}\exp(\underbrace{2k(L_2-L_1)}_{\text{depth}})$$

Remark: when the bias $b = -k$ is driving X away from exit and when the depth of slope, $k(L_2-L_1)$, is moderately large, $T(x)$ has two properties

- $T(x)$ is independent of x as long as x is not too close to the exit.
- $T(x)$ is exponentially large, depending on the depth and the width of slope.

In this example, we derived the two properties based on the analytical solution. We will see that these two properties are generally valid when the bias is against the exit.

Escape of a Brownian particle from a potential well

Model equations

Consider a particle undergoes Brownian motion in a potential well $V(x)$. The potential exerts a position-dependent conservative force $-V'(x)$ on the particle.

We consider the problem of a Brownian particle escaping from a potential well. This problem serves as a model for a wide spectrum of application problems, for example, breaking of a molecular bond, activation in a chemical reaction, …

The stochastic motion of the particle is governed by Newton's second law.

$$dX = Y\,dt$$

$$m\,dY = -\underbrace{bY\,dt}_{\text{Viscous drag}} - \underbrace{V'(X)dt}_{\text{Force from potential}} + \underbrace{\sqrt{2k_B T b}\,dW}_{\text{Brownian force}}$$

X: position

Y: velocity

m: mass

b: drag coefficient

This equation is called Langevin equation (named after Paul Langevin).

In the limit of small particle (i.e., particle size converging to zero), we have

$$0 = -bY\,dt - V'(X)dt + \sqrt{2k_B T b}\,dW$$

(The derivation is more complicated than setting $m = 0$!)

The small particle limit is called the Smoluchowski-Kramers approximation (named after Marian Smoluchowski and Hans Kramers), which we will discuss separately.

Writing (Ydt) as dX, we obtain an equation for X.

$$0 = -b\,dX - V'(X)dt + \sqrt{2k_B T b}\,dW$$

$$\Longrightarrow \quad dX = -\frac{1}{b}V'(X)dt + \sqrt{2\frac{k_B T}{b}}\,dW$$

We write it in terms of the diffusion coefficient, $D = k_B T/b$.

$$dX = -D\frac{V'(X)}{k_B T}dt + \sqrt{2D}\,dW$$

This equation is called the over-damped Langevin equation.

The physical (dimensional) exit problem

Suppose a particle is governed by the over-damped Langevin equation. We consider the problem of the particle escaping from $[0, L]$ where

$x = 0$ is a reflecting boundary and

$x = L$ is an absorbing boundary.

Draw a potential over $[0, L]$.

Scales for non-dimensionalization

At room temperature (~295K),

$k_B T \approx 4.1\ pN\cdot\text{nm} = 4.1\times10^{-21}\ N\cdot m$ (Joule)

- $k_B T$ serves as the energy scale for normalizing potential $V(x)$.

  $[\, k_B T\,]$ = Energy

- L (the width of the region) serves as the length scale for normalizing X.

  $[\, L\,]$ = Length

- Diffusion coefficient D has the dimension

  $$[\,D\,] = \frac{(\text{Length})^2}{\text{Time}}$$

- We construct a time scale from L and D.

  $$\left[\frac{L^2}{D}\right] = \text{Time}$$

Non-dimensional variables

We define

$$X_{new} = \frac{X_{old}}{L} \qquad ==> \qquad X_{old} = L X_{new}$$

$$t_{new} = \frac{D}{L^2}t_{old} \qquad ==> \qquad t_{old} = \frac{L^2}{D}t_{new}$$

$$V_{new}(X_{new}) = \frac{1}{k_B T} V_{old}(X_{old}) \qquad ==> \qquad V_{old}(X_{old}) = k_B T\, V_{new}(X_{new})$$

Non-dimensional SDE:

We start with the physical SDE:

$$dX = -D \cdot \frac{V'(X)}{k_B T} dt + \sqrt{2D}\, dW \, .$$

We write all old variables in terms of new variables.

$$dX_{old} = L dX_{new} \, , \qquad dt_{old} = \frac{L^2}{D} dt_{new}$$

$$\frac{1}{k_B T} V_{old}{}'(X_{old}) = \frac{dV_{new}}{dX_{old}} = \frac{dV_{new}}{dX_{new}} \cdot \frac{dX_{new}}{dX_{old}} = V_{new}{}'(X_{new}) \frac{1}{L}$$

$$dW(t_{old}) = \sqrt{dt_{old}} \underbrace{\frac{dW(t_{old})}{\sqrt{dt_{old}}}}_{\sim N(0,1)} = \sqrt{\frac{L^2}{D} dt_{new}} \frac{dW(t_{new})}{\sqrt{dt_{new}}} = \sqrt{\frac{L^2}{D}}\, dW(t_{new})$$

Substituting these terms into the equation, we obtain

$$\underbrace{L dX_{new}}_{dX_{old}} = -D \cdot \underbrace{V_{new}{}'(X_{new}) \frac{1}{L}}_{\frac{1}{k_B T} V_{old}{}'(X_{old})} \cdot \underbrace{\frac{L^2}{D} dt_{new}}_{dt_{old}} + \sqrt{2D} \underbrace{\sqrt{\frac{L^2}{D}}\, dW(t_{new})}_{dW(t_{old})}$$

$$==> \qquad dX_{new} = -V_{new}{}'(X_{new}) dt_{new} + \sqrt{2}\, dW(t_{new})$$

For conciseness, we recycle the simple notion and write the equation as

$$dX = -V'(X) dt + \sqrt{2}\, dW$$

Now all quantities in the equation are dimensionless.

Exact solution of the dimensionless average escape time

Let $T(x)$ be the dimensionless average escape time.

Recall that for $dX = b(X)dt + \sqrt{a(X)}\, dW$, the governing equation for $T(x)$ is

$$\frac{1}{2} a(x) T_{xx} + b(x) T_x = -1$$

Substituting $a(x) = 2$ and $b(x) = -V'(x)$, we write out the BVP for $T(x)$

$$\begin{cases} T_{xx} - V'(x) T_x = -1 \\ T'(0) = 0\,, \quad T(1) = 0 \end{cases} \qquad \text{(T\_BVP1)}$$

Theorem:

The solution of (T_BVP1) is

$$\boxed{T(x) = \int_x^1 dy \exp(V(y)) \int_0^y ds \exp(-V(s))} \qquad \text{(T\_SOL1)}$$

Derivation:

We use the method of integrating factor.

Multiplying $T_{xx} - V'(x)\, T_x = -1$ by $\exp(-V(x))$, we write the ODE as

$$\exp(-V(x)) T_{xx} - \exp(-V(x)) V'(x) T_x = -\exp(-V(x))$$

$$\Longrightarrow \quad \left(\exp(-V(x)) T_x\right)_x = -\exp(-V(x))$$

Integrating from 0 to y, and using $T_x(0) = 0$, we obtain

$$\exp(-V(y)) T_x(y) = -\int_0^y ds \exp(-V(s))$$

$$\Longrightarrow \quad T_x(y) = -\exp(V(y)) \int_0^y ds \exp(-V(s))$$

Integrating from x to 1, and using $T(1) = 0$, we arrive at

$$T(x) = \int_x^1 dy \exp(V(y)) \int_0^y ds \exp(-V(s))$$

End of derivation

(T_SOL1) is the exact solution of (T_BVP1). It does not have any approximation error.

Next, we find an approximation to (T_SOL1) when the potential well is deep.

Escape from a deep potential well: an approximate solution

We consider the potential shown. Specifically,

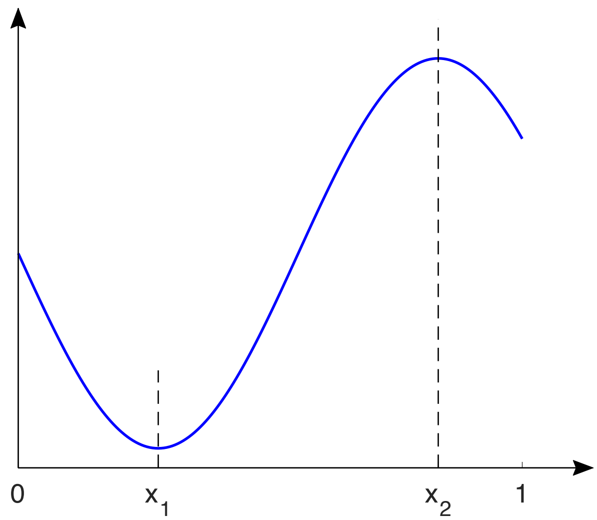
- $V(x)$ decreases monotonically in $(0, x_1)$;
- $V(x)$ attains the only minimum at $x_1 > 0$;
- $V(x)$ increases monotonically in (x_1, x_2);
- $V(x)$ attains the only maximum at $x_2 > x_1$; and
- $V(x)$ decreases monotonically in $(x_2, 1)$.

The depth of the potential well is defined as the height from bottom to top.

$\Delta G \equiv V(x_2) - V(x_1)$

We consider the case of moderately large ΔG, for example, $\Delta G \geq 10$.

Goal: Find an approximate solution for $T(x)$.



Writing the potential in terms of its depth

Since only $V'(x)$ appears in the stochastic differential equation, we shift $V(x)$ by a constant to make $V(x_1) = 0$. We write potential $V(x)$ as

$$V(x) = \Delta G \cdot \phi(x)$$

where $\phi(x)$ satisfies i) $\phi(x) \geq 0$ for $x \in (0, 1)$, ii) $\phi(x_1) = 0$ and iii) $\phi(x_2) = 1$.

Case 1: $\phi''(x) \neq 0$ at both of the two extrema.

Writing the exact solution in terms of ΔG and $\phi(x)$, we have

$$T(x) = \int_x^1 dy \exp(\Delta G \cdot \phi(y)) \int_0^y ds \exp(-\Delta G \cdot \phi(s))$$

In the inner integral $\int_0^y ds$, the dominant contribution comes from near $s = x_1$. When s gets away from x_1, $\phi(s)$ is positive and the integrand $\exp(-\Delta G\, \phi(s))$ is exponentially small. As a result, we only need to capture the integrand approximately near $s = x_1$.

We expand $\phi(s)$ near $s = x_1$.

$$\phi(s) = \underbrace{\phi(x_1)}_{=0} + \underbrace{\phi'(x_1)}_{=0}(s - x_1) + \frac{1}{2}\underbrace{\phi''(x_1)}_{>0}(s - x_1)^2 + \cdots$$

For $y > x_1$, the inner integral is

$$\int_0^y ds \exp(-\Delta G \cdot \phi(s)) \approx \int_0^y ds \exp\left(-\Delta G \cdot \frac{1}{2}\phi''(x_1)(s-x_1)^2\right)$$

$$\approx \int_{-\infty}^{+\infty} ds \exp\left(\frac{-1}{2}\Delta G \cdot \phi''(x_1)(s-x_1)^2\right) = \underbrace{\sqrt{\frac{2\pi}{\Delta G \cdot \phi''(x_1)}}}_{\text{independent of } y} \qquad \text{for } y > x_1$$

Here we used the integration formula

$$\int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}\alpha u^2\right) du = \sqrt{\frac{2\pi}{\alpha}} \int_{-\infty}^{+\infty} \underbrace{\frac{1}{\sqrt{2\pi\alpha^{-1}}} \exp\left(\frac{-u^2}{2\alpha^{-1}}\right)}_{\text{Normal distribution}} du = \sqrt{\frac{2\pi}{\alpha}}$$

For $y < x_1$, the inner integral is negligible relative to its value for $y > x_1$.

Summary of the inner integral:

For $y > x_1$, $\int_0^y ds \exp(-\Delta G \cdot \phi(s)) \approx$ high constant, independent of y.

For $y < x_1$, $\int_0^y ds \exp(-\Delta G \cdot \phi(s)) \approx 0$.

In the outer integral $\int_x^1 dy$, the factor $\exp(\Delta G \cdot \phi(y))$ attains its maximum at $y = x_2$ where $\phi(x_2) = 1$. When y gets away from x_2, $\exp(\Delta G \cdot \phi(y))$ decreases rapidly relative to its maximum at $y = x_2$. Also for y near x_2, the inner integral takes its high constant value. Thus, in the outer integral, the dominant contribution comes from near $y = x_2$, we only need to capture the integrand approximately near $y = x_2$. We expand $\phi(y)$ near $y = x_2$.

$$\phi(y) = \underbrace{\phi(x_2)}_{=1} + \underbrace{\phi'(x_2)}_{=0}(y-x_2) + \frac{1}{2}\underbrace{\phi''(x_2)}_{<0}(y-x_2)^2 + \cdots$$

For $x < x_2$, (i.e., the starting point is inside the potential well), $T(x)$ is

$$T(x) \approx \int_x^1 dy \underbrace{\exp(\Delta G \cdot \phi(y))}_{\text{focus on near } y = x_2} \underbrace{\int_0^y ds \exp(-\Delta G \cdot \phi(s))}_{= \text{high constant value}}$$

$$\approx \int_x^1 dy \exp\left(\Delta G + \Delta G \cdot \frac{1}{2}\phi''(x_2)(y-x_2)^2\right) \sqrt{\frac{2\pi}{\Delta G \cdot \phi''(x_1)}}$$

$$\approx \exp(\Delta G)\cdot\sqrt{\frac{2\pi}{\Delta G\cdot\phi''(x_1)}}\int_{-\infty}^{+\infty} dy \exp\left(\frac{-1}{2}\Delta G\underbrace{(-\phi''(x_2))}_{>0}(y-x_2)^2\right)$$

$$=\exp(\Delta G)\cdot\underbrace{\sqrt{\frac{2\pi}{\Delta G\cdot\phi''(x_1)}}\sqrt{\frac{2\pi}{\Delta G\cdot(-\phi''(x_2))}}}_{\text{independent of } x} \qquad \text{for } x<x_2$$

Kramers' approximate solution for $T(x)$:

When the potential height ΔG is moderately large and the starting point x is inside the potential well, $T(x)$ is approximately (in the sense of small relative error)

$$T(x)\approx \exp(\Delta G)\cdot\frac{1}{\Delta G}\underbrace{\sqrt{\frac{(2\pi)^2}{\phi''(x_1)\cdot(-\phi''(x_2))}}}_{\text{independent of } x} \qquad \text{for } x<x_2$$

This is part of Kramers' theory of reaction kinetics.

Remarks:

1. Q: Why do we want an approximate solution?
   A: It gives us a clear picture on the behaviors of $T(x)$.
2. When ΔG is moderately large, $T(x)$ has two properties:
   - $T(x)$ is independent of x as long x is inside the potential well.
   - $T(x)$ is exponentially large.

   We will look at the dependence on the width and the height of potential when we go back to the physical average exit time T_{phy}.

Case 2: $\phi''(x_1) > 0$ at x_1; $\phi''(x_2) = 0$ and $\phi^{(4)}(x_2) < 0$ at x_2 (skip)

We expand $\phi(y)$ near $y = x_2$.

$$\phi(y)=\frac{1}{4!}\phi^{(4)}(x_2)(y-x_1)^4+\cdots$$

For $x < x_2$, we have

$$T(x)\approx \exp(\Delta G)\cdot\sqrt{\frac{2\pi}{\Delta G\cdot\phi''(x_1)}}\int_{-\infty}^{+\infty} dy \exp\left(\frac{-1}{4!}\Delta G(-\phi^{(4)}(x_2))(y-x_2)^4\right)$$

$$= \exp(\Delta G) \cdot \sqrt{\frac{2\pi}{\Delta G \cdot \phi''(x_1)}} \frac{(3/2)^{1/4}\Gamma(1/4)}{\left(\Delta G(-\phi^{(4)}(x_2))\right)^{1/4}} \qquad \text{for } x < x_2$$

Here we used the integral formula

$$\int_{-\infty}^{+\infty} \exp(-bu^4)du = 2\int_0^{+\infty} \exp(-bu^4)du$$

(change of variables $bu^4 = w$)

$$= \frac{1}{2b^{1/4}} \int_0^{+\infty} \exp(-w) w^{-3/4} dw = \frac{1}{2b^{1/4}} \Gamma(1/4), \qquad \Gamma(1/4) \approx 3.6256$$

<u>Kramers' approximate solution</u> for $T(x)$:

When the potential height ΔG is moderately large and the starting point x is inside the potential well, $T(x)$ is approximately (in the sense of small relative error)

$$T(x) \approx \exp(\Delta G) \cdot \frac{1}{(\Delta G)^{3/4}} \underbrace{\sqrt{\frac{2\pi}{\phi''(x_1)}} \cdot \frac{(3/2)^{1/4}\Gamma(1/4)}{(-\phi^{(4)}(x_2))^{1/4}}}_{\text{independent of } x} \qquad \text{for } x < x_2$$

AM216 Stochastic Differential Equations

Lecture 14

List of topics in this lecture

- Kramers' theory of reaction kinetics, physical exit time, effects of energy barrier, diffusion coefficient, and potential width
- Memoryless property of exit time, exponential distribution, escape rate
- Application of Kramers' theory: a simple model of ignition
- Feynman-Kac formula, fatality/growth rate, path integral $u(x, t, T)$, interpretation of path integral as relative population, governing equation of $u(x, t, T)$

Recap

Escape of a Brownian particle from a potential well

Smoluchowski-Kramers approximation in the limit of small particle

$$\underbrace{m dY = -bY dt - V'(X) dt + \sqrt{2k_B T b}\, dW}_{\text{Langevin equation}} \quad ==> \quad \underbrace{dX = -\frac{D}{k_B T} V'(X) dt + \sqrt{2D}\, dW}_{\text{over-damped Langevin equation}}$$

Dimensionless SDE

$$dX = -V'(X) dt + \sqrt{2}\, dW$$

Exact integral solution of the average exit time

$$T(x) = \int_x^1 dy \exp(V(y)) \int_0^y ds \exp(-V(s))$$

Deep potential well

$V(x) = \Delta G\, \phi(x)$, $\quad$ $\min \phi(x) = \phi(x_1) = 0$, $\quad$ $\max \phi(x) = \phi(x_2) = 1$,

ΔG is moderately large.

Kramers' approximate solution of $T(x)$

$$T(x) \approx \exp(\Delta G) \cdot \frac{1}{\Delta G} \sqrt{\frac{(2\pi)^2}{\phi''(x_1) \cdot (-\phi''(x_2))}} \quad \text{independent of } x \text{ for } x < x_2$$

$T(x)$ is independent of the starting position x when x is inside the potential well.

Kramers' theory of reaction kinetics

Physical escape time in terms of physical quantities

Recall the non-dimensionalization.

$$t = \frac{D}{L^2} t_{\text{phy}}, \qquad T(x) = \frac{D}{L^2} T_{\text{phy}}(x_{\text{phy}}), \qquad \Delta G = \frac{1}{(k_B T)} \Delta G_{\text{phy}}$$

Substituting these into the expression of $T(x)$, we get

$$T(x) = \exp(\Delta G) \cdot \frac{1}{\Delta G} \sqrt{\frac{(2\pi)^2}{\phi''(x_1) \cdot (-\phi''(x_2))}}$$

$$\Longrightarrow \qquad \frac{D}{L^2} T_{phy}(x_{phy}) = \exp\left(\frac{\Delta G_{phy}}{k_B T}\right) \cdot \frac{k_B T}{\Delta G_{phy}} \sqrt{\frac{(2\pi)^2}{\phi''(x_1) \cdot (-\phi''(x_2))}}$$

Caution on the notation:

- T in $(k_B T)$ is the temperature.
- $T(x)$ is the average exit time.

The physical escape time has the expression

$$T_{phy}(x_{phy}) = \underbrace{\frac{L^2}{D}}_{\substack{\text{Effect of}\\ \text{mobility}}} \cdot \underbrace{\exp\left(\frac{\Delta G_{phy}}{k_B T}\right) \frac{k_B T}{\Delta G_{phy}}}_{\substack{\text{Effect of}\\ \text{energy barrier}}} \underbrace{\sqrt{\frac{(2\pi)^2}{\phi''(x_1) \cdot (-\phi''(x_2))}}}_{\substack{\text{Effect of}\\ \text{relative geometry}}}$$

We can see how the physical escape time scales with other physical quantities.

- When the width of potential L is doubled, T_{phy} is increased by a factor of 4.
  It is more difficulty to escape from a wide potential well.
- When the diffusion coefficient D is doubled, T_{phy} is halved.
  It is easier for a smaller particle to escape.
- T_{phy} increases exponentially with the energy barrier ΔG_{phy}. When ΔG_{phy} is increased by $2.3 k_B T$, T_{phy} is increased by a factor of 10.
  By far, the energy barrier ΔG_{phy} has the dominant influence on T_{phy}.

An example:

Consider the escape of a 1-nm (diameter) particle from a potential well of width 0.5nm.

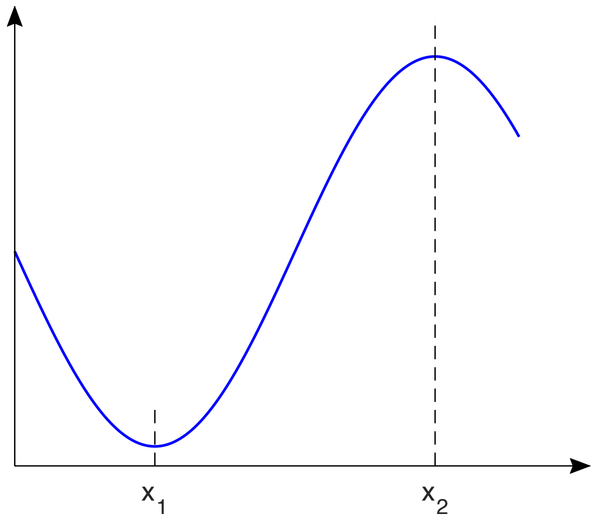
Particle radius: a = 0.5nm; viscosity of water: $\eta = 0.01\ \text{g(cm)}^{-1}\text{s}^{-1}$.

Diffusion coefficient: $D = \dfrac{k_B T}{6\pi\eta a} = 4.350 \times 10^8\ \text{nm}^2\text{s}^{-1}$.

Potential: $V(x) = \Delta G\, \phi(x)$, $\phi(x) = \frac{1}{2} + \frac{1}{2}\sin\left(\pi(1.8x - 1.0)\right)$.

$$x_1 = \arg\min \phi(x) = \frac{5}{18}, \qquad x_2 = \arg\max \phi(x) = \frac{15}{18}$$

$$\phi''(x_1) = \frac{1}{2}(1.8\pi)^2, \qquad \phi''(x_2) = \frac{-1}{2}(1.8\pi)^2$$



Substituting these quantities into the expression of T_{phy}, we obtain

$$T_{phy}(x_{phy}) = \exp\left(\frac{\Delta G_{phy}}{k_B T}\right)\frac{k_B T}{\Delta G_{phy}}\left(2.258\times 10^{-10}\,\text{s}\right)$$

- $\Delta G_{\text{phy}} = 10\, k_B T$ ==> $T_{\text{phy}} = 4.974\times 10^{-7}$ s = 0.497 μs
- $\Delta G_{\text{phy}} = 20\, k_B T$ ==> $T_{\text{phy}} = 5.478\times 10^{-3}$ s = 5.48 ms
- $\Delta G_{\text{phy}} = 40\, k_B T$ ==> $T_{\text{phy}} = 1.329\times 10^{+6}$ s = 15.38 days

Distribution of the random exit time

Let $Y(\omega)$ denote the random exit time. In the above, we studied

$T(x) \equiv E(Y(\omega)|X(0) = x)$

Question: What can we say about the distribution of $Y(\omega)$?

Answer: For a deep potential well, the escape process is memoryless.

Specifically, the solution of $T(x)$ tells us

$$T(x) \approx \exp(\Delta G)\cdot \frac{1}{\Delta G}\sqrt{\frac{(2\pi)^2}{\phi''(x_1)\cdot(-\phi''(x_2))}} \quad \text{independent of } x$$

That is, the average exit time is memoryless. Mathematically it gives us

$$E(Y-t_0 \mid Y>t_0)=E(Y) \qquad \text{independent of } t_0 \qquad \text{(E01)}$$

Let $\rho(t)$ be the probability density of Y.

Previously (in Lecture 2) we derived $\rho(t)$ based on the memoryless property of Y. It turns out that the memoryless property of $E(Y)$ is sufficient for deriving $\rho(t)$.

We write (E01) in terms of $\rho(t)$.

$$\frac{1}{\int_{t_0}^{\infty}\rho(t)dt}\int_{t_0}^{\infty}(t-t_0)\rho(t)dt=E(Y) \qquad \text{independent of } t_0 \qquad \text{(E01B)}$$

Let $G(t)\equiv\int_t^{\infty}\rho(s)ds$. We have $\rho(t)=-G'(t)$.

Carrying out integration by parts in the numerator and identify the denominator as $G(t_0)$, we write (E01B) as

$$\int_{t_0}^{\infty}G(t)dt=E(Y)G(t_0) \qquad \text{(E01C)}$$

Differentiating with respect to t_0, we arrive at

$$\frac{-1}{E(Y)}G(t_0)=G'(t_0) \qquad \text{(the same ODE as we obtained previously.)}$$

We conclude that $Y(\omega)$ has the exponential distribution:

$$\rho(t)=r\exp(-rt), \qquad r\equiv\frac{1}{E(Y)}=\frac{1}{T(x)}$$

The <u>escape rate</u>, r, describes the <u>conditional probability of escaping per time</u>:

$$r=\frac{1}{\Delta t}\Pr\left(\text{escaping in } (t_0, t_0+\Delta t) \,\middle|\, \text{having not escaped by } t_0\right)$$

<u>The physical escape rate</u>

$$r_{phy}=\frac{1}{T_{phy}(x_{phy})}=\underbrace{\frac{D}{L^2}}_{\text{Effect of mobility}}\cdot\underbrace{\exp\left(\frac{-\Delta G_{phy}}{k_B T}\right)\frac{\Delta G_{phy}}{k_B T}}_{\text{Effect of energy barrier}}\underbrace{\sqrt{\frac{\phi''(x_1)\cdot(-\phi''(x_2))}{(2\pi)^2}}}_{\text{Effect of relative geometry}}$$

This is the <u>Kramers' theory of reaction kinetics</u> (named after Hans Kramers).

<u>Remarks:</u>

- The chemical reaction between molecules A and B requires activation, which means crossing over an energy barrier. The energy barrier represents the situation where molecule A has to fluctuate to an energetically unfavorable configuration before reacting with molecule B.
- Crossing over an energy barrier is mathematically an escape process.

- When the energy barrier is large, the escape process is memoryless and is described by a reaction rate, which has a strong dependence on the temperature.

$$r_{phy} \sim \exp\left(\frac{-\Delta G_{phy}}{k_B T}\right)$$

- Another aspect of the chemical reaction is the probability of encounter between molecules A and B, which is affected by their concentrations.

A simple model of ignition

Let T_0 = the ambient temperature.

$T(t)$ = the spot temperature at time t at an interface of gasoline and air

(where locally there is a mix of gasoline and air)

Governing equation for $T(x)$

$T(t)$ is governed by Newton's law of cooling

$$\frac{dT(t)}{dt} = \underbrace{-\mu(T(t)-T_0)}_{\text{cooling}} + \underbrace{\alpha \exp\left(\frac{-\Delta G}{T(t)}\right)}_{\substack{\text{heat generated}\\ \text{by reaction}}}$$

Note: For simplicity, the Boltzmann coefficient k_B has been absorbed into ΔG.

Let $y(t) \equiv (T(t)-T_0)/T_0$, the normalized temperature increase.

We expand the non-linear term in the ODE for small y.

$T(t) = T_0(1+y(t))$

$$\frac{-\Delta G}{T(t)} = \frac{-\Delta G}{T_0(1+y(t))} = \frac{-\Delta G}{T_0}(1-y(t)+\cdots) = \frac{-\Delta G}{T_0} + \frac{\Delta G}{T_0} y(t) + \cdots$$

$$\exp\left(\frac{-\Delta G}{T(t)}\right) = \exp\left(\frac{-\Delta G}{T_0}\right)\exp\left(\frac{\Delta G}{T_0} y(t)+\cdots\right) = \exp\left(\frac{-\Delta G}{T_0}\right)\left(1+\frac{\Delta G}{T_0} y(t)+\cdots\right)$$

Substituting the expansion in the ODE yields

$$T_0 \frac{dy}{dt} = \underbrace{-\mu T_0 y}_{\text{cooling}} + \underbrace{\alpha \exp\left(\frac{-\Delta G}{T_0}\right)\left(1+\frac{\Delta G}{T_0} y+\cdots\right)}_{\substack{\text{heat generated}\\ \text{by reaction}}}$$

Dividing by T_0 and neglecting higher order terms, we obtain

Linearized ODE for $y(t)$

$$\frac{dy}{dt} = \underbrace{\left(\frac{\alpha \Delta G}{T_0^{\,2}} \exp\left(\frac{-\Delta G}{T_0}\right) - \mu\right)}_{\equiv \lambda(T_0)} y + \underbrace{\frac{\alpha}{T_0} \exp\left(\frac{-\Delta G}{T_0}\right)}_{\equiv q} \equiv \lambda(T_0) y + q$$

We study the behavior of IVP for $\lambda > 0$ and for $\lambda < 0$.

$$\begin{cases} y' = \lambda y + q \\ y(0) = 0 \end{cases}$$

Exact solution of $y(t)$:

$$y(t) = \left(e^{\lambda t} - 1\right) \frac{q}{\lambda}$$

$\lambda < 0$: $y(t) \to q/(-\lambda)$ as $t \to +\infty$

The temperature stabilizes at a finite value. No combustion.

$\lambda > 0$: $y(t) \to +\infty$ as $t \to +\infty$

The temperature increases unbounded. Combustion.

The ignition temperature T_0^* is the solution of $\lambda(T_0^*) = 0$.

$\lambda(T_0)$ is an increasing function of T_0 for $\Delta G/T_0 > 2$.

$$\lambda(T_0) \equiv \frac{\alpha \Delta G}{T_0^{\,2}} \exp\left(\frac{-\Delta G}{T_0}\right) - \mu$$

$$\Longrightarrow \quad \frac{d\lambda}{dT_0} = \frac{\alpha \Delta G}{T_0^{\,3}} \exp\left(\frac{-\Delta G}{T_0}\right)\left(\frac{\Delta G}{T_0} - 2\right) > 0$$

Remark: The case of $\Delta G/T_0 < 2$ is irrelevant since Kramers' theory of reaction kinetics is valid only for large $\Delta G/T_0$.

Feynman-Kac formula for the backward equation

We are back to the time-dependent (non-autonomous) SDE

$$dX = b(X,t)dt + \sqrt{a(X,t)}\,dW \qquad \text{(Ito interpretation)}$$

The moments of $(dX \mid X(t) = x)$ are

$$E\left(dX \middle| X(t) = x\right) = b(x,t)dt + o(dt)$$

$$E\left((dX)^2 \middle| X(t) = x\right) = a(x,t)dt + o(dt)$$

$$E\left((dX)^n \middle| X(t) = x\right) = o(dt) \quad \text{for } n \ge 3$$

<u>Definition of</u> $u(x, t, T)$

$$u(x,t,T) \equiv E\left(\exp\left(-\int_t^T \psi(X(s),s)ds \right) \middle| X(t) = x \right)$$

<u>Meaning of</u> $u(x, t, T)$

<u>Case1: $\psi(z, s)>0$:</u>

We view $\psi(z, s)$ as the fatality rate of a "cell" at time s with $X(s) = z$.

Pr(fatality in $[s, s+\Delta s]$ | having survived to s with $X(s) = z$)

$= \psi(z, s) \times \Delta s$

Let us follow one <u>particular path $x(s)$</u> from t to T.

We discretize the path on a time grid

$$\Delta s = \frac{T-t}{N}, \qquad s_j = t + j\Delta s, \quad s_0 = t, \quad s_N = T$$

Along the given path $x(s)$, the probability of surviving from s_j to s_{j+1} is

Pr(surviving to s_{j+1} | having survived to s_j)

$= 1 - $ Pr(fatality in $[s_j, s_{j+1}]$ | having survived to s_j)

$$= 1 - \psi(x(s_j), s_j)\Delta s \approx \exp\left(-\psi(x(s_j), s_j)\Delta s\right)$$

Along the given path $x(s)$, the probability of surviving from t to T is

Pr(surviving to T | having survived to t)

$$= \prod_{j=0}^{N-1} \exp\left(-\psi(x(s_j), s_j)\Delta s\right) = \exp\left(-\sum_{j=0}^{N-1} \psi(x(s_j), s_j)\Delta s \right)$$

$$\longrightarrow \exp\left(-\int_t^T \psi(x(s), s)ds \right) \qquad \text{as } N \to \infty$$

We average the surviving probability <u>over all paths</u> starting at $X(t) = x$.

$$u(x,t,T) \equiv E\left(\exp\left(-\int_t^T \psi(X(s),s)ds \right) \middle| X(t) = x \right)$$

= probability of surviving from t to T | $X(t) = x$

<u>Case2: $\psi(z, s)<0$:</u>

We interpret $[-\psi(z, s)]>0$ as the growth rate of a cell at time s with $X(s) = z$.

Pr(split into two in $[s, s+\Delta s]$ | having survived to s with $X(s) = z$)

$= (-\psi(z, s)) \times \Delta s$

$u(x, t, T)$ = expected population at time T <u>relative to that at time t</u> | $X(t) = x$.

The general case

We interpret $\psi(z, s)$ as the fatality/growth rate of a cell at time s with $X(s) = z$.

(Outcome in $[s, s+\Delta s]$ | having survived to s with $X(s) = z$)

$$=\begin{cases} \text{fatality with prob } =\psi(z,s)\Delta s & \text{if } \psi(z,s)>0 \\ \text{split into two with prob } =\big(-\psi(z,s)\big)\Delta s & \text{if } \psi(z,s)<0 \end{cases}$$

After splitting into two, both new cells are associated with the same $X(s)$ path. In this way, the evolution of $X(s)$ is solely governed by the SDE, not affected by $\psi(z, s)$.

Some sample path of $X(s)$ may have no cell; some may have many cells.

$u(x, t, T)$ = expected population at time T relative to that at time t | $X(t) = x$.

Examples:

1. $X(s)$ = temperature of a site at time s

   $u(x, t, T)$ = expected bacteria population at T relative to that at t | $X(t) = x$.
2. $X(s)$ = collective population of all predators in a region at time s

   $u(x, t, T)$ = expected population of a prey at T relative to that at t | $X(t) = x$.
3. $X(s)$ = oil price at time s.

   $u(x, t, T)$ = expected stock price of an oil company at T relative to that at t | $X(t) = x$.

Governing equation for $u(x, t, T)$

We apply the backward view on

$$u(x,t,T)=E\left(\exp\left(-\int_t^T \psi(X(s),s)ds\right)\middle| X(t)=x\right)$$

$[\,t \to T\,]$ is divided into $[t \to t+\Delta t\,]$ and $[\,t+\Delta t \to T\,]$.

$$\exp\left(-\int_t^T \psi(X(s),s)ds\right)=\exp\left(-\int_t^{t+dt} \psi(X(s),s)ds\right)\exp\left(-\int_{t+dt}^T \psi(X(s),s)ds\right)$$

$$=\underbrace{(1-\psi(x,t)dt)}_{\text{independent of path}}\exp\left(-\int_{t+dt}^T \psi(X(s),s)ds\right)+o(dt)$$

Averaging over all paths starting at $X(t) = x$, we get

$$u(x,t,T)=(1-\psi(x,t)dt)E\left(\exp\left(-\int_{t+dt}^T \psi(X(s),s)ds\right)\middle| X(t)=x\right)+o(dt) \qquad \text{(E02)}$$

On the RHS, the average is over $\{X(s),\ t \le s \le T\}$. We use the law of total expectation to rewrite is as averaging over $\{X(s) \mid X(t+\Delta t),\ t+\Delta t \le s \le T\}$ and then over $dX(t)$.

For any quantity Q, we have

$$E_{\{X(s),t\le s\le T\}}(Q|\,X(t)=x)=E_{dX}\left(E_{\{X(s),t+dt\le s\le T\}}(Q|X(t+dt)=x+dX)\right)$$

Apply this result to the expectation in (E02)

$$E\left(\exp\left(-\int_{t+dt}^{T}\psi(X(s),s)ds\right)\middle|X(t)=x\right)$$

$$=E_{dX}\left(E_{\{X(s),t+dt\le s\le T\}}\left(\exp\left(-\int_{t+dt}^{T}\psi(X(s),s)ds\right)\middle|X(t+dt)=x+dX\right)\right)$$

Definition of $u(x, t, T)$

$$=E_{dX}\left(u(x+dX,t+dt,T)\right)$$

Taylor expansion

$$=E_{dX}\left(u(x,t,T)+u_t dt+u_x dX+\tfrac{1}{2}u_{xx}(dX)^2\right)+o(dt)$$

Using moments of dX

$$=u(x,t,T)+u_t dt+u_x b(x,t)dt+\tfrac{1}{2}u_{xx}a(x,t)dt+o(dt)$$

Substituting this result into the RHS of (E02), we write (E02) as

$$u\left(x,t,T\right)=(1-\psi(x,t)dt)\left(u(x,t,T)+u_t dt+u_x b(x,t)dt+\tfrac{1}{2}u_{xx}a(x,t)dt\right)+o(dt)$$

$$=u(x,t,T)+u_t dt+u_x b(x,t)dt+\tfrac{1}{2}u_{xx}a(x,t)dt-\psi(x,t)u\,dt+o(dt)$$

Dividing by dt and taking the limit at $dt \to 0$, we obtain the governing equation

$$0=u_t+b(x,t)u_x+\frac{1}{2}a(x,t)u_{xx}-\psi(x,t)u$$

It is the backward equation with a fatality/growth term.

<u>The final condition</u>: $u(x, t, T)|_{t=T} = 1$

<u>The final value problem (FVP)</u>

$$\begin{cases} 0=u_t+b(x,t)u_x+\dfrac{1}{2}a(x,t)u_{xx}-\psi(x,t)u \\ u(x,t,T)\Big|_{t=T}=1 \end{cases}$$

The solution of the FVP is given by

$$u(x,t,T)=E\left(\exp\left(-\int_{t}^{T}\psi(X(s),s)ds\right)\middle|X(t)=x\right)$$

This is called the Feynman-Kac path integral formula for the backward equation (named after Richard Feynman and Mark Kac).

A more general case of Feynman-Kac formula

Definition of $u(x, t, T)$

$$u(x,t,T) \equiv E\left(\exp\left(-\int_t^T \psi(X(s),s)ds \right) f(X(T)) \middle| X(t)=x \right)$$

Meaning of $u(x, t, T)$

$\psi(z, s)$ is the fatality/growth rate of a cell at time s with $X(s) = z$.

$f(z)$ is the reward for a cell surviving to time T with $X(T) = z$.

$u(x, t, T)$ = expected reward at final time T per unit population at time t | $X(t) = x$.

Each cell of the population gets its own reward. The growth increases the population size and increases the reward for the population.

Governing equation for $u(x, t, T)$

The governing equation is not affected by function $f(z)$.

$$0 = u_t + b(x,t)u_x + \frac{1}{2}a(x,t)u_{xx} - \psi(x,t)u$$

The final condition: $u(x, t, T)|_{t=T} = f(x)$

Note: the effect of $f(z)$ is contained in the final condition.

The final value problem (FVP)

$$\begin{cases} 0 = u_t + b(x,t)u_x + \frac{1}{2}a(x,t)u_{xx} - \psi(x,t)u \\ u(x,t,T)\Big|_{t=T} = f(x) \end{cases}$$

The solution of the FVP is given by the Feynman-Kac path integral formula

$$u(x,t,T) = E\left(\exp\left(-\int_t^T \psi(X(s),s)ds \right) f(X(T)) \middle| X(t)=x \right)$$

Feynman-Kac formula for the forward equation

Definition of $u(x, t)$

$$u(x,t) \equiv E\left(\delta\big(X(t)-x\big)\exp\left(-\int_0^t \psi(X(s),s)ds \right) \right)$$

where $\psi(x, t)$ is the fatality/growth rate of a cell at time t with $X(t) = x$.

Items of the discussion:

1) We need to explain the δ function in the average.
2) We need to derive the governing equation for $u(x, t)$.
3) We need to explain the meaning of $u(x, t)$ and discuss the distribution of $X(0)$.

Item #1: We first explain the δ function in the average.

View #1: Approximate δ( ) using a boxcar function.

Let $I_{[x,x+\Delta x]}(z)$ be the indicator function defined as

$$I_{[x,x+\Delta x]}(z)=\begin{cases} 1, & x\le z\le x+\Delta x \\ 0, & \text{otherwise}\end{cases}$$

$u(x, t)$ can be viewed as

$$u(x,t)=\lim_{\Delta x\to 0}\frac{1}{\Delta x}E\left(I_{[x,x+\Delta x]}(X(t))\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

View #2: Use the method of test function

We integrate the product $h(x)u(x, t)$.

$$\int h(x)u(x,t)dx=E\left(\left(\int h(x)\delta(X(t)-x)dx\right)\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

which leads to

$$\boxed{\int h(x)u(x,t)dx=E\left(h(X(t))\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)}$$

Here $h(x)$ is any smooth function that decays to zero rapidly as $|x|\to\infty$.

The two views are equivalent to each other. We are going to use view #2 in the derivation of the governing equation for $u(x, t)$.

AM216 Stochastic Differential Equations

Lecture 15
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- Feynman-Kac formula for the forward equation, path integral $u(x, t)$, interpretation of path integral as mass density of the surviving cell population at time t, governing equation of $u(x, t)$
- An application of Feynman-Kac formula: reconstructing potential $V(x)$ from a set of sample paths of particle position; exploring non-equilibrium with an applied force; modeling the effect of applied force as a fatality/growth rate

---

Recap

Feynman-Kac formula for the backward equation

We are back to the time-dependent SDE

$$dX = b(X,t)dt + \sqrt{a(X,t)}\,dW$$

$u(x, t, T)$ is defined by the Feynman-Kac path integral formula

$$u(x,t,T) = E\left(\exp\left(-\int_t^T \psi(X(s),s)ds \right) f(X(T)) \middle| X(t) = x \right)$$

Meaning of $u(x, t, T)$

$\psi(x, s)$ = fatality/growth rate of a cell at time s with $X(s) = x$.

$f(x)$ = reward for a cell surviving to time T with $X(T) = x$.

$u(x, t, T)$ = expected reward at final time T per unit population at time t

Each cell of the population gets its own reward. The growth increases the population size and increases the reward for the population.

Governing equation for $u(x, t, T)$ and the FVP

$$\begin{cases} 0 = u_t + b(x,t)u_x + \dfrac{1}{2}a(x,t)u_{xx} - \psi(x,t)u \\ u(x,t,T)\big|_{t=T} = f_0(x) \end{cases}$$

The solution of the FVP is given by the Feynman-Kac path integral formula.

Feynman-Kac formula for the forward equation (continued)

$$u(x,t)=E\left(\delta\big(X(t)-x\big)\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

Items of the discussion:

1) We need to explain the δ function in the average.
2) We need to derive the governing equation for $u(x, t)$.
3) We need to explain the meaning of $u(x, t)$ and discuss the distribution of $X(0)$.

Item #1: Explaining the δ function in the average

View 1: $$u(x,t)=\lim_{\Delta x\to 0}\frac{1}{\Delta x}E\left(I_{[x,x+\Delta x]}(X(t))\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

View 2: $$\int h(x)u(x,t)dx=E\left(h(X(t))\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right) \qquad \text{for any } h(x)$$

(We use this view to derive the equation for $u(x, t)$).

Item #2: Derivation of the governing equation for $u(x, t)$

We start with View #2 at $(t+\Delta t)$. We apply the forward view on $[0 \to t+\Delta t]$.

$[0 \to t+\Delta t]$ is divided into $[0 \to t]$ and $[t \to t+\Delta t]$.

$$\underbrace{\int h(x)u(x,t+dt)dx}_{\text{LHS}}=\underbrace{E\left(h(X(t)+dX)\exp\left(-\int_0^{t+dt} \psi(X(s),s)ds\right)\right)}_{\text{RHS}}$$

$$=E\left(h(X(t)+dX)\exp\left(-\int_t^{t+dt} \psi(X(s),s)ds\right)\times\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

In the RHS, we expand $h(X(t)+dX)$ and $\exp\left(-\int_t^{t+dt} \psi(X(s),s)ds\right)$

$$\text{RHS}=E\left(\left[h(X(t))+h'(X(t))dX+\frac{1}{2}h''(X(t))(dX)^2\right]\big(1-\psi(X(t),t)dt\big)\right.$$
$$\left.\times\exp\left(-\int_0^t \psi\big(X(s),s\big)ds\right)\right)$$

$$=E\left(\left[h(X(t))+h'(X(t))dX+\frac{1}{2}h''(X(t))(dX)^2-h(X(t))\psi(X(t),t)dt\right]\right.$$
$$\left.\times\underbrace{\exp\left(-\int_0^t \psi\big(X(s),s\big)ds\right)}_{\text{independent of } dX}\right)$$

In the RHS, the average is over $\{X(s),\ 0 \le s \le t+\Delta t\}$. We use the law of total expectation to rewrite is as averaging over $(dX(t) \mid X(t))$ and then over $\{X(s),\ 0 \le s \le t\}$.

$$E_{\{X(s),0\le s\le t+dt\}}(Q) = E_{\{X(s),0\le s\le t\}}\left(E_{dX(t)}(Q\,\big|\,X(t))\right)$$

We use the moments of $(dX(t) \mid X(t))$:

$$E_{dX(t)}\left(h'(X(t))dX\,\big|\,X(t)\right) = h'(X(t))b(X(t),t)dt$$

$$E_{dX(t)}\left(h''(X(t))(dX)^2\,\big|\,X(t)\right) = h''(X(t))a(X(t),t)dt$$

$$\text{RHS} = E\left(\left[h(X(t)) + h'(X(t))b(X(t),t)dt + \frac{1}{2}h''(X(t))a(X(t),t)dt - h(X(t))\psi(X(t),t)dt\right]\times\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

View #2 (method of test function) of $u(x, t)$ gives

$$E\left(g(X(t))\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right) = \int g(x)u(x,t)dx \qquad \text{for any } g(x).$$

In the RHS, we set respectively $g(x) = h(x)$ for term 1, $g(x) = h'(x)b(x, t)$ for term 2, …,

$$\text{RHS} = \int\left(h(x) + h'(x)b(x,t)dt + \frac{1}{2}h''(x)a(x,t)dt - h(x)\psi(x,t)dt\right)u(x,t)dx$$

In the RHS, we integrate by parts; in the LHS, we expand $u(x, t+dt)$.

$$\text{RHS} = \int h(x)\left(u(x,t) - \left(b(x,t)u\right)_x dt + \frac{1}{2}\left(a(x,t)u\right)_{xx} dt - \psi(x,t)u\,dt\right)dx$$

$$\text{LHS} = \int h(x)u(x,t+dt)dx = \int h(x)\left(u(x,t) + u_t dt\right)dx$$

Subtracting $\int h(x)\, u(x, t)dx$, dividing by dt, and taking the limit as $dt \to 0$, we obtain

$$\underbrace{\int h(x)u_t\,dx}_{\text{LHS}} = \underbrace{\int h(x)\left(-\left(b(x,t)u\right)_x + \frac{1}{2}\left(a(x,t)u\right)_{xx} - \psi(x,t)u\right)dx}_{\text{RHS}}$$

Since LHS = RHS for arbitrary test function $h(x)$, we arrive at

$$u_t = -\left(b(x,t)u\right)_x + \frac{1}{2}\left(a(x,t)u\right)_{xx} - \psi(x,t)u$$

This is the governing PDE for $u(x, t)$.

It is the forward equation with a fatality/growth term.

Item #3: Meaning of $u(x, t)$

View #1 of $u(x, t)$ gives

$$u(x,t)=\lim_{\Delta x\to 0}\frac{1}{\Delta x}E_{\omega}\left(I_{[x,x+\Delta x]}(X(t,\omega))\exp\left(-\int_0^t\psi(X(s,\omega),s)ds\right)\right)$$

The average is based on an ensemble $\{X(s, \omega), \omega \in \Omega\}$. It is worthwhile to emphasize that the evolution of $X(s)$ is governed solely by the SDE

$$dX=b(X,t)dt+\sqrt{a(X,t)}\,dW, \qquad \text{independent of } \psi(x, s)$$

In particular, $X(s)$ does not die or split. In $u(x, t)$, the fatality/growth effect of $\psi(x, s)$ is reflected in the factor $\exp\left(-\int_0^t\psi(X_j(s),s)ds\right)$.

To interpret $u(x, t)$, we consider the "cell" population associated with $X(s)$. The cell population includes the fatality/growth effect of $\psi(x, s)$, in the form of termination/addition of cells.

For a cell at time s with $X(s) = z$, the outcome of the cell in $[s, s+\Delta s]$ is

(Outcome in $[s, s+\Delta s]$ | having survived to s with $X(s) = z$)

$$=\begin{cases}\text{fatality with prob } =\psi(z,s)\Delta s & \text{if } \psi(z,s)>0\\ \text{split into two with prob } =\left(-\psi(z,s)\right)\Delta s & \text{if } \psi(z,s)<0\end{cases}$$

We write $u(x, t)$ in terms of the cell population and associated $X(s)$.

$$u(x,t)=\lim_{\Delta x\to 0}\frac{1}{\Delta x}E_{\omega}\left(\#\text{ of cells surviving to time } t \text{ with } X(t,\omega)\in[x,x+dx]\right)$$

= mass density in x of the surviving cell population at time t.

The size of surviving population is different from the size of starting population.

Initial condition for $u(x, t)$

$u(x, 0) = f_0(x)$ = mass density in x of the starting cell population.

Remarks on the Feynman-Kac formula for the forward equation

- The evolution of $X(s)$ is governed solely by the SDE, independent of $\psi(x, s)$. $X(s)$ does not die or split.
- In the Feynman-Kac formula, the average is over an ensemble of $X(s)$. The fatality/growth effect of $\psi(x, s)$ is reflected in the factor $\exp\left(-\int_0^t\psi(X_j(s),s)ds\right)$.

  The Feynman-Kac formula is good for calculating $u(x, t)$ from an ensemble of $X(s)$.
- Given a LARGE set of sample paths of $\{X_j(s), j = 1, 2, \ldots, N\}$, we have

$$u(x,t) \approx \frac{1}{N \cdot \Delta x} \sum_{X_j(t) \in [x, x+\Delta x]} \exp\left(-\int_0^t \psi(X_j(s), s) ds \right)$$

The calculation does not directly involve the cell population.

- To interpret $u(x, t)$ defined in the Feynman-Kac formula, we follow the surviving cell population associated with $X(s)$, which includes the effect of fatality/growth.
- Although we interpret $\psi(x, s)$ as the fatality/growth rate, the Feynman-Kac formula is well defined for any $\psi(x, s)$, not associated with any physical fatality/growth.
- When $\psi(x, s)$ is not associated with any physical fatality/growth, the cell population exists only in our mathematical imagination (see the application below).

An application of Feynman-Kac formula

The big picture: When $b(x, t)$ in the SDE is unknown but a LARGE set of sample paths is available, we can use the Feynman-Kac formula to calculate the unknown $b(x, t)$.

Consider a particle diffusing in a potential well.

$X(t)$: position of particle at time t

$V(x)$: static potential well

The stochastic motion is governed by the over-damped Langevin equation.

$$dX = -\frac{D}{k_B T} V'(X) dt + \sqrt{2D}\, dW$$

After non-dimensionalization, we have

$$dX = -V'(X) dt + \sqrt{2}\, dW$$

What we can measure (data)

A large set of sample paths $\{X_j(t),\ j = 1, 2, \ldots, N\}$

Goal:

To reconstruct potential $V(x)$ from the data

Method:

We use Feymann-Kac formula to reconstruct potential $V(x)$.

Items of the discussion

A. Forward equation and equilibrium of probability density

B. Estimating potential from equilibrium measurements

C. A practical issue with equilibrium data

D. Exploring non-equilibrium with an applied force

E. Steps in constructing potential $V(x)$

Item A:

Forward equation (Fokker-Planck equation) of $dX = -V'(X)dt + \sqrt{2}\,dW$

Let $\rho(x, t)$ = probability density of X at time t.

Note: probability density = mass density of a population of one path.

The time evolution of $\rho(x, t)$ is governed by the forward equation

$$\rho_t = -\big(b(x)\rho\big)_x + \frac{1}{2}\big(a(x)\rho\big)_{xx}\,, \qquad a(x) = 2, \quad b(x) = -V'(x)$$

$$\Longrightarrow \quad \rho_t = \big(V'(x)\rho\big)_x + \rho_{xx}$$

We write the forward equation in conservation form

$$\rho_t = -\frac{\partial}{\partial x}J(x,t)\,, \qquad J(x,t) \equiv -\big(V'(x)\rho + \rho_x\big)$$

where $J(x, t)$ is the probability flux.

Equilibrium distribution

At equilibrium, the probability flux must be identically zero.

$$J(x) = 0, \qquad \text{for all } x$$

$$\Longrightarrow \quad V'(x)\rho + \rho_x = 0$$

$$\Longrightarrow \quad \big(\exp(V(x))\rho(x)\big)' = 0$$

$$\Longrightarrow \quad \exp(V(x))\rho(x) = \text{const}$$

$$\Longrightarrow \quad \rho^{(\text{eq})}(x) \propto \exp(-V(x))$$

As expected, the equilibrium is the Maxwell-Boltzmann distribution.

Caution: A steady state is different from equilibrium.

When $J(x)$ = const ≠ 0 for all x, we still have $\rho = \rho(x)$, independent of t.

It is called a steady state, which is different from equilibrium.

At equilibrium, we must have $J(x) = 0$ for all x.

In the terminology of deterministic dynamical systems, “steady state” and “equilibrium” are usually not distinguished.

Item B:

Estimating potential from equilibrium measurements

Suppose the system is at equilibrium and we measure a large set of sample paths $\{X_j(t), j = 1, 2, \ldots, N\}$.

The equilibrium density $\rho^{(eq)}(x)$ can be calculated as

$$\rho^{(eq)}(x) \approx \frac{1}{N \cdot \Delta x} \sum_{X_j(t_k) \in [x, x+\Delta x]} 1 \qquad \text{at a particular time level } t_k$$

where N is the number of sample paths.

To fully utilize the data set, we average $\rho^{(eq)}(x)$ over all t_k's

$$\rho^{(eq)}(x) \approx \frac{1}{K_T} \sum_{k=1}^{K_T} \left[\rho^{(eq)}(x) \text{ estimated at } t_k \right]$$

where K_T is the number of time levels.

Recall that $\rho^{(eq)}(x) \propto \exp(-V(x))$. We write potential $V(x)$ as

$$V(x) = -\log \rho^{(eq)}(x) + \underbrace{C}_{\text{additive constant}}$$

Item C:

A practical issue with equilibrium data

Unfortunately, the approach of using only equilibrium measurements does not work well. It requires an impractically large amount of data.

At equilibrium, a region of high $V(x)$ value is visited only very infrequently.

$$\rho^{(eq)}(x) \propto \exp(-V(x))$$

Consider a set of discrete sites (intervals of x). For a site with probability 10^{-8}, we need to sample 10^9 times to get 10 visits to that particular site. It is practically impossible to accurately estimate $\rho^{(eq)}(x)$ in a region of high $V(x)$ value.

Remedy: We need to perturb the system to non-equilibrium.

Item D:

Exploring non-equilibrium with an applied force

Let $F(t)$ be the applied force (non-dimensionalized).

In experiments, $F(t)$ is controlled. In AFM (Atomic Force Microscopy) experiments, the force is controlled by moving an actuator to stretch an elastic link.

$$\mathrm{F}^{(AFM)}(t) = k\int_0^t u(s)ds$$

where k = spring constant; $u(s)$ velocity of actuator at time s.

Stochastic differential equation in the presence of an applied force

The applied force tilts the static potential. At time t, the tilted potential is

$$H(x,t) = V(x) - \underbrace{F(t)\cdot x}_{\substack{\text{Effect of}\\ \text{applied force}}}$$

Replacing $V'(x)$ with $H_x(x, t)$, we get the new SDE.

$$dX = -H_x(X,t)dt + \sqrt{2}\,dW$$

In the presence of an applied force, potential $H(x, t)$ changes with time. As a result, the system is not at equilibrium and the Boltzmann distribution does not apply.

Nevertheless we consider a "hypothetical" density, $\rho^{(F)}(x, t)$, defined as

$$\rho^{(F)}(x,t) \equiv \frac{1}{Z}\exp\big(-H(x,t)\big) = \frac{1}{Z}\exp\big(-V(x) + F(t)\cdot x\big)$$

$$\text{where} \quad Z = \int \exp\big(-V(x)\big)dx$$

Remark:

- We call $\rho^{(F)}(x, t)$ "hypothetical" density because it is not the mass density of any physical population. In particular, $\rho^{(\mathrm{F})}(x, t)$ is NOT the probability density of particle position. In the discussion below, we interpret $\rho^{(F)}(x, t)$ as the mass density of a "hypothetical" cell population, whose existence/fatality/growth is only in our mathematical imagination.

Advantage of working with $H(x, t)$ and $\rho^{(\mathrm{F})}(x, t)$

With a properly designed force schedule $F(t)$, a region of relatively high $V(x)$ value becomes a region of relatively low $H(x, t)$ value in the tilted potential.

In this way, different regions of $V(x)$ can be very well explored/sampled at different time t with a time-dependent force schedule $F(t)$.

Item E:

Steps in constructing potential $V(x)$

1. Find the governing PDE for $\rho^{(\mathrm{F})}(x, t)$
2. Identify the "fatality/growth" term $\psi(x, s)$ in the PDE
3. Express $\rho^{(\mathrm{F})}(x, t)$ in the Feynman-Kac formula

4. Use the Feynman-Kac formula to calculate $\rho^{(F)}(x, t)$ from a set of sample paths.
5. Determine potential $V(x)$ from $\rho^{(F)}(x, t)$.

Step 1: Find the governing PDE for $\rho^{(F)}(x, t)$

Let $\rho(x, t)$ be the density of particle position in the presence of applied force $F(t)$.

$\rho(x, t)$ is NOT the same as $\rho^{(F)}(x, t)$.

$$\rho(x, t) \neq \rho^{(F)}(x, t)$$

Stochastic motion of particle is governed by

$$dX = -H_x(X,t)dt + \sqrt{2}\,dW\,, \qquad H(x,t) \equiv V(x) - F(t)\cdot x$$

The forward equation (Fokker-Planck equation) for $\rho(x, t)$ is

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left(H_x(x,t)\rho + \frac{\partial}{\partial x}\rho \right)$$

We write the Fokker-Planck equation in terms of a differential operator.

$$\rho_t = L_{\{H\}}[\rho] \qquad \text{(FP\_forced)}$$

where $\quad L_{\{H\}}[\bullet] = \dfrac{\partial}{\partial x}\left(H_x(x,t)\bullet + \dfrac{\partial}{\partial x}\bullet \right)$

We find the governing equation for the "hypothetical" density $\rho^{(F)}(x, t)$.

$$\rho^{(F)}(x,t) \equiv \frac{1}{Z}\exp\big(-H(x,t)\big) = \frac{1}{Z}\exp\big(-V(x) + F(t)\cdot x\big), \qquad Z = \int \exp\big(-V(x)\big)dx$$

Substitute $\rho^{(F)}(x, t)$ into operator $L_{\{H\}}[\,\cdot\,]$ and into operator $\partial/\partial t$

$$L_{\{H\}}\left[\rho^{(F)}(x,t)\right] = \frac{\partial}{\partial x}\left(H_x(x,t)\rho^{(F)}(x,t) + \frac{\partial}{\partial x}\rho^{(F)}(x,t) \right) = 0$$

$$\rho_t^{(F)}(x,t) = F'(t)x\cdot\rho^{(F)}(x,t)$$

It follows that $\rho^{(F)}(x, t)$ satisfies

$$\rho_t^{(F)} = L_{\{H\}}\left[\rho^{(F)}\right] + \underbrace{F'(t)x\cdot\rho^{(F)}}_{\text{fatality/growth}}$$

Step 2: Identify the "fatality/growth" term $\psi(x, s)$ in the PDE

$\rho^{(F)}(x, t)$ satisfies the forward equation with a fatality/growth term

$$\rho_t^{(F)} = L_{\{H\}}\left[\rho^{(F)}\right] - \psi(x,t)\cdot\rho^{(F)}\,, \qquad \psi(x,t) = -F'(t)x$$

Step 3: Express $\rho^{(F)}(x, t)$ using the Feynman-Kac formula

$$\rho^{(F)}(x,t) = E\left(\delta(X(t)-x)\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right), \quad \psi(x,s) = -F'(s)x$$

$$\Longrightarrow \quad \rho^{(F)}(x,t) = E\left(\delta(X(t)-x)\exp\left(\int_0^t F'(s)X(s)ds\right)\right)$$

<u>Remark:</u> The ensemble of $X(s)$ is good for calculating $\rho^{(F)}(x, t)$. For that calculation, we don't need the hypothetical cell population that includes the stochastic addition/termination of cells.

<u>Step 4:</u> Calculate $\rho^{(F)}(x, t)$ from a set of sample paths.

Suppose we apply force $F(t)$ and measure a set of sample paths $\{X_j(s), j = 1, 2, \ldots, N\}$.

At each time level t_k, $\rho^{(F)}(x, t_k)$ can be calculated using the Feynman-Kac formula.

$$\rho^{(F)}(x,t_k) \approx \frac{1}{N\cdot\Delta x}\sum_{X_j(t_k)\in[x,\le x+\Delta x]} \exp\left(\int_0^{t_k} F'(s)X_j(s)ds\right) \qquad \text{at each time level } t_k$$

where N is the number of sample paths.

<u>Step 5:</u> Determine potential $V(x)$ from $\rho^{(F)}(x, t)$

Note that $\rho^{(F)}(x, t)$ and $\rho^{(eq)}(x)$ are related by

$$\rho^{(eq)}(x) = \frac{1}{Z}\exp\big(-V(x)\big)$$

$$\rho^{(F)}(x,t) = \frac{1}{Z}\exp\big(-V(x)+F(t)x\big)$$

$$\Longrightarrow \quad \rho^{(eq)}(x) = \rho^{(F)}(x,t)\exp\big(-F(t)x\big)$$

Once $\rho^{(F)}(x, t_k)$ is obtained at a time level t_k, we use it to calculate a sample version of equilibrium density $\rho^{(eq)}(x)$.

$$\rho^{(eq)}(x) = \rho^{(F)}(x,t_k)\exp\big(-F(t_k)x\big) \qquad \text{at each time level } t_k$$

Then we average the sample versions of $\rho^{(eq)}(x)$ over all t_k's.

$$\rho^{(eq)}(x) \approx \frac{1}{K_T}\sum_{k=1}^{K_T}\rho^{(F)}(x,t_k)\exp\big(-F(t_k)x\big)$$

where K_T is the number of time levels.

Once $\rho^{(eq)}(x)$ is accurately estimated, we write potential $V(x)$ as

$$V(x) = -\log\rho^{(eq)}(x) + \underbrace{C}_{\text{additive constant}}$$

<u>Remarks</u> on constructing potential $V(x)$

1. The "hypothetical" density $\rho^{(F)}(x, t)$ contains potential $V(x)$ and applied force $F(t)$, which allows us to extract potential $V(x)$ from $\rho^{(F)}(x, t)$ once $\rho^{(F)}(x, t)$ is obtained.
2. $\rho^{(F)}(x, t)$ satisfied the forward equation with a fatality/growth term $\psi(x, t) = -F'(t)x$. This hypothetical fatality/growth term exists only in our mathematical imagination. It does not correspond to the fatality/growth of any physical process.
3. The path integral expression (Feynman-Kac formula) of $\rho^{(F)}(x, t)$ allows us to calculate $\rho^{(F)}(x, t)$ from sample paths $\{X_j(s), j = 1, 2, \ldots, N\}$ and applied force $F(t)$.
4. Once $\rho^{(F)}(x, t)$ is obtained, we extract potential $V(x)$ from $\rho^{(F)}(x, t)$.

AM216 Stochastic Differential Equations

Lecture 16
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- Black-Scholes option pricing model: stochastic evolution of the stock price, call option, put option, strike price, expiration date, the option price function
- Delta hedging portfolio: 1 unit of delta hedging of time t, continuously revised delta hedging, mathematical view of delta hedging, net gain from maintaining a prescribed delta hedging portfolio
- Governing equation for the option price function, the final value problem

---

Recap

Feynman-Kac formula for the forward equation

Stochastic differential equation (SDE)

$$dX = b(X,t)dt + \sqrt{a(X,t)}\,dW$$

$u(x, t)$ is defined using path integral

$$u(x,t) = E\left(\delta\big(X(t)-x\big)\exp\left(-\int_0^t \psi(X(s),s)ds\right)\right)$$

$u(x, t)$ satisfies the forward equation with a fatality/growth term

$$u_t = -\big(b(x,t)u\big)_x + \frac{1}{2}\big(a(x,t)u\big)_{xx} - \psi(x,t)u$$

Meaning of $u(x, t)$

= mass density in x of the surviving cell population at time t.

Evolution of $X(s)$ is governed solely by the SDE, independent of $\psi(z, s)$.

Cell population is affected by the fatality/growth effect of $\psi(z, s)$.

When the SDE is unknown but a large set of sample paths $\{X_j(s)$, j = 1, 2, …, $N\}$ is available, we can estimate $u(x, t)$ from the data.

$$u(x,t) \approx \frac{1}{N\cdot\Delta x}\sum_{x<X_j(t)\le x+\Delta x} \exp\left(-\int_0^t \psi(X_j(s),s)ds\right)$$

Black-Scholes option pricing model

The underlying stock:

Consider the stock of ABC company:

One share of stock = a fraction of ownership of ABC company.

Example:

There are 1,805 million shares of Walt Disney Company stock.

Each share = $1/(1.8\times10^9)$ fraction of Walt Disney Company.

Roadmap of the discussion

1. Evolution of the stock price
2. Options associated with the stock
3. Mathematical formulation of the option price function
4. Delta hedging portfolio
5. Total gain in $[0, T]$ for maintaining a prescribed delta hedging
6. Governing PDE for the option price function

1. Evolution of the stock price

Let $S(t)$ = stock price of ABC company at time t

(stock price = per share price in the open financial market).

Assumption:

$S(t)$ is a geometric Brownian motion with a geometric drift.

Specifically, stock price $S(t)$ is governed by the SDE

$$dS = \mu S\,dt + \sigma S\,dW \quad \text{(Ito interpretation)}$$

Parameters:

Volatility: σ

Drift: μ

Comment:

Here we model the stock price using the Ito interpretation of the SDE

$$dS = \mu S\,dt + \sigma S\,dW \qquad \text{(S01A)}$$

In some literature, the stock price is model as the SDE

$$d\left(\log S\right) = \mu_B dt + \sigma\,dW \qquad \text{(S01B)}$$

(S01A) is subject to different interpretations; (S01B) is not.

Previously, we derived that

(S01B) is equivalent to the Stratonovich interpretation of (S01A) with $\mu = \mu_B$,

which is equivalent to the Ito interpretation of the modified SDE

$$dS = \left(\mu_B + \frac{1}{2}\sigma^2\right) S\,dt + \sigma S\,dW$$

We will use (S01A) with the Ito interpretation. When comparing the results of (S01A) and (S01B), keep in mind $\mu = (\mu_B + \sigma^2/2)$.

2. Options associated with the stock

An option is a contract that gives the owner (the holder) of the option the right, but not the obligation,

to buy (sell)

a specified number of shares of the underlying stock,

at a specified price

prior to or on a specified date

Terminology and notations:

- right ≠ obligation

  (the option owner has no obligation to buy or sell the stock).
- call option = the right to buy
- put option = the right to sell
- 1 contract of call (put) option = the right to buy (sell) 100 shares
- strike price, K = the specified purchase (sell) price for a call (put) option
- expiry (expiration date), T = the specified data
- exercising a call (put) option: exercising the right to buy (sell)

Examples of options

Price of DIS stocks on May 4th 2022 = \$116.19/share

(DIS 220603 120 call)

= the right to buy, strike price = \$120/share, expiry = 06/03/2022

(DIS 220603 110 put)

= the right to sell, strike price = \$110/share, expiry = 06/03/2022

Caution: In the open market, option prices are listed for 1/100 contract

= price for the right to buy (sell) one share.

Price of (DIS 220603 120 call) option on May 4th 2022 = \$3.55

The cost of buying 1 contract of (DIS 220603 120 call) = 100×\$3.55 = \$355.

Option underwriter, option buyers/sellers

Option underwriter: every option contract is underwritten by someone who sells it in the open market for a financial gain.

The option underwriter is the original option seller, and is obligated to honor the current owner of the option's right to buy (sell).

The option underwriter bets on that the future evolution of the underlying stock will make the option worth less than it is originally sold for.

Option buyer: a person who buys an option contract in the open market by paying a negotiated amount (the market price) to an option seller.

An option buyer bets on that the future evolution of the underlying stock will make the option worth more than it is bought for.

An option buyer can subsequently become an option seller by selling the acquired option in the open market.

The option underwriter can subsequently become an option buyer by buying back the option in the open market, which terminates the obligation.

Comments:

- Options have no assets of their own. Their values are solely derived from the underlying stock. For this reason, options are called financial derivatives.
- Options are not issued by the company of the underlying stock. Options are underwritten by financial gamblers. Option buyers pay money to own the right. Option underwriters receive money from buyers in exchange for granting option holders the right to buy or sell.
- Options are much more volatile than the underlying stock because the risk of options is much higher than that of the stock. The option buyer may win big if the underlying stock moves significantly in one direction; may lose all money invested in buying the options if the stock moves the other direction. For the option underwriter, the risk is even worse; the loss may be unlimited.
- Options are not limited to stocks. Options can be written on anything that has a market (and a market price). For example, oil, natural gas, wheat, corn, soy bean, gold, silver, copper, steel, lumber, …
- In particular, options can be written on other options. They are the derivatives of derivatives. They are even more volatile (more risky!).

3. Mathematical formulation of the option price function

Mathematical unit of an option

For mathematical convenience, we introduce 1 "unit" of option.

1 unit of call (put) option

= 1/100 contract of call (put) option

= the right to buy (sell) 1 share of the stock at price K at time T.

Note: This is the European style option.

The American style option = the right to … at any time $t \le T$.

We study the European style option because it is mathematically simpler and the two are essentially equivalent.

The option price function

Consider an option of a specified type (call or put), a specified strike price and a specified expiry.

Key assumption:

The option price at the current time t is a deterministic function of the current stock price $S(t)$ and the current time t.

We focus on the call option. The put option can be discussed in a similar way.

Let $C(s, t)$ denote the price per unit of the call option at time t when $S(t) = s$.

$C(s, t)$ is a deterministic function of two variables (s, t).

Both the stock price and the option price are stochastic. But the two are related by a deterministic function: $C(S(t), t)$. The randomness in the option price is solely caused by the randomness in the stock price $S(t)$.

List of variables and parameters:

$S(t)$: stock price at time t

$C(S(t), t)$: option price at time t

σ: volatility

Volatility (σ) predicts the likelihood of the stock making big moves.

μ: geometric drift in the SDE of $S(t)$

As we will see, drift (μ) does not affect the option price.

r: interest rate

Interest rate (r) affects investors' appetite for risk.

K: the strike price

T: the expiry (expiration date)

List of assumptions

- The stock does not pay a dividend. This is for mathematical simplicity.
- The underlying stock price $S(t)$ is governed by

  $dS = \mu S\,dt + \sigma S\,dW$ (Ito interpretation)
- Volatility (σ) and interest rate (r) are known.
- The stochastic option price and the stochastic stock price are related by $C(S(t), t)$ where $C(s, t)$ is a deterministic function of 2 variables. Function $C(s, t)$ also depends on parameters (σ, r, K, T).
- We can buy/sell any amount of option/stock, including short selling.

  Short selling = selling something we don't own.
- There is no bid-ask spread. At any time, the price we can buy is the same as the price we can sell. Again, this is for mathematical simplicity.
- There is a single interest rate. We can borrow/lend any amount at this known rate.

  Again, this is for mathematical simplicity.
- There is no transaction fee associated with buying/selling stock/option, no transaction fee associated with borrowing/lending.

  For large financial firms, this is not completely unrealistic.

  Again, this is for mathematical simplicity.

Key question for shaping the mathematical formulation

Suppose I am a market maker and I am required to set and publish $C(s, t)$, the deterministic function connecting the stock price and the option price.

How should I set function $C(s, t)$ to avoid guaranteed loss?

Could someone design a scheme of trading stocks/options based on the published function $C(s, t)$ to make a guaranteed gain?

Answer:

We study the delta hedging portfolio.

4. Delta hedging portfolio

We study the delta hedging involving a call option.

The delta hedging involving a put option can be discussed in a similar way.

Composition of the delta hedging

Consider the call option with strike price K and expiry T.

1 unit of delta hedging of time t

= being short one unit of call option and long $C_S(S(t), t)$ shares of stock

= owning (−1) unit of call option and $C_S(S(t), t)$ shares of stock.

Note: $C_S(s, t)$ is the derivative of $C(s, t)$.

$$C_s\big(S(t),t\big)=\frac{\partial C(s,t)}{\partial s}\bigg|_{s=S(t)}$$

Since we can buy/sell any amount (positive or negative, small or large), we can have (−1) unit of delta hedging, or any positive or negative fraction of delta hedging.

(−1) unit of delta hedging of time t

= owning (+1) unit of call option and $-C_S(S(t), t)$ shares of stock.

Mathematical view of the delta hedging:

- The composition of 1 unit delta hedging of time t is determined by the stock price at time t and the published function $C(s, t)$.
- The composition of 1 unit delta hedging of time t varies with t. To maintain 1 unit delta hedging of the current time, buying/selling stocks is needed.

  To update 1 unit delta hedging of time t to that of time $(t+dt)$, we need to

  buy $[C_S(S(t+dt), t+dt) - C_S(S(t), t)]$ shares of stock at time $(t+dt)$.

  This is called continuously revised delta hedging.
- In the mathematical view, at time $(t+dt)$ when updating 1 unit delta hedging of time t to that of time $(t+dt)$, there are two transactions:

  selling 1 unit delta hedging of time t at time $(t+dt)$, and

  buying 1 unit delta hedging of time $(t+dt)$ at time $(t+dt)$.

  "of time t" refers to the composition of 1 unit delta hedging.

  "at time $(t+dt)$" refers to the time of buying/selling.
- We maintain a portfolio of $F(S(t), t)$ units of delta hedging of time t, at time t, over a time period. For that purpose, we update the portfolio as follows.

  selling $F(S(t), t)$ units of delta hedging of time t at time $(t+dt)$, and

  buying $F(S(t+dt), t+dt)$ units of delta hedging of time $(t+dt)$ at time $(t+dt)$.

  These are the two transactions at each grid point over a time period.
- We view the selling as associated with time interval $[t, t+dt]$ and view the buying as associated with time interval $[t+dt, t+2dt]$. In this way, we start a time interval with no position and end the time interval with no position.

  The actions over time interval $[t, t+dt]$ are

  start with no position at time t,

at time t, buy $F(S(t), t)$ units of delta hedging of time t;

hold it during $[t, t+dt]$;

at time $(t+dt)$, sell $F(S(t), t)$ units of delta hedging of time t;

end with no position at time $(t+dt)$.

Note: This mathematical view involves a lot of "hypothetical" trading transactions. They are for the purpose of facilitating the calculation of net gain in $[t, t+dt]$. Many of these "hypothetical" trading transactions cancel each other. In real operation, the actual trading transactions needed are a lot less.

5. **Total gain in [0, *T*]** for maintaining a prescribed delta hedging

Suppose that over time period $[0, T]$, we maintain a portfolio of $F(S(t), t)$ units of delta hedging of time t, at time t. We do so by carrying out the actions described above in each time interval $[t, t+dt]$. Here $F(s, t)$ is a function to be specified.

We first calculate the net gain in each time interval $[t, t+dt]$.

Net gain in $[t, t+dt]$ for maintaining a prescribed delta hedging

- At time t, we start with 0 cash.
- At time t, after buying $F(S(t), t)$ units of delta hedging of time t,

  the cash balance at time t is

$$B(t)=\underbrace{F\big(S(t),t\big)}_{\substack{\text{\# of units of}\\ \text{delta hedging}}}\Big[\underbrace{1}_{\substack{\text{\# of units}\\ \text{of option}}}\times\underbrace{C\big(S(t),t\big)}_{\substack{\text{option price}\\ \text{at time } t}}-\underbrace{C_s\big(S(t),t\big)}_{\substack{\text{\# of shares}\\ \text{of stock}}}\times\underbrace{S(t)}_{\substack{\text{stock price}\\ \text{at time } t}}\Big]$$

Note: The cash balance may be positive or negative.

- The interest earned in $[t, t+dt]$ from the cash balance is

  $I = B(t) \times r\, dt$

Note:

The interest earned may be positive or negative, depending on $B(t)$.

Negative earning = the interest cost of borrowing money.

- At time $(t+dt)$, after selling $F(S(t), t)$ units of delta hedging of time t,

  the change in cash balance at time $(t+dt)$ is

$$\Delta B=\underbrace{F\big(S(t),t\big)}_{\substack{\text{\# of units of}\\ \text{delta hedging}}}\Big[\underbrace{-1}_{\substack{\text{\# of units}\\ \text{of option}}}\times\underbrace{C\big(S(t+dt),t+dt\big)}_{\substack{\text{option price}\\ \text{at time } t+\Delta t}}+\underbrace{C_s\big(S(t),t\big)}_{\substack{\text{\# of shares}\\ \text{of stock}}}\times\underbrace{S(t+dt)}_{\substack{\text{stock price}\\ \text{at time } t+\Delta t}}\Big]$$

Note:

The composition of delta hedging of time t is unchanged in $[t, t+dt]$.

The prices of stock and option in the delta hedging do fluctuate in $[t, t+dt]$.

To facilitate the calculation of ΔB, we introduce short notations:

$$S \equiv S(t), \qquad dS \equiv S(t+dt)-S(t)$$

==> $S(t+dt) = S+dS$

We write ΔB as

$$\Delta B = F(S,t)\Big[-C(S+dS,t+dt)+C_s(S,t)(S+dS)\Big]$$

Expanding the RHS in terms of dS and dt, we get

$$\Delta B = F(S,t)\begin{bmatrix} -C(S,t)-\underbrace{C_s(S,t)dS}_{dS\text{ term}}-C_t(S,t)dt-\frac{1}{2}C_{ss}(S,t)(dS)^2 \\ +C_s(S,t)S+\underbrace{C_s(S,t)dS}_{dS\text{ term}}+o(dt) \end{bmatrix}$$

By the special design of delta hedging, the two dS terms cancel each other.

$$\Delta B = F(S,t)\left[-C(S,t)+C_s(S,t)S-C_t(S,t)dt-\frac{1}{2}C_{ss}(S,t)(dS)^2+o(dt)\right]$$

- Net gain in time interval $[t, t+dt]$ is

  $$\text{net gain} = B(t)+I+\Delta B = B(t)(1+r\,dt)+\Delta B$$

Let $g(t)$ denote the rate of net gain in $[t, t+dt]$ (net gain per time). We have

$$g(t)dt = B(t)(1+r\,dt)+\Delta B$$

$$= \underbrace{F(S,t)\Big[C(S,t)-C_s(S,t)S\Big]}_{B(t)}(1+r\,dt)$$

$$+\underbrace{F(S,t)\left[-C(S,t)+C_s(S,t)S-C_t(S,t)dt-\frac{1}{2}C_{ss}(S,t)(dS)^2\right]}_{\Delta B}$$

$$= F(S,t)\left[\Big(C(S,t)-C_s(S,t)S\Big)r\,dt-C_t(S,t)dt-\frac{1}{2}C_{ss}(S,t)(dS)^2\right]$$

We use the SDE to express $(dS)^2$ in terms of dW.

$$dS = \mu S\,dt+\sigma S\,dW$$

==> $(dS)^2 = \sigma^2 S^2 (dW)^2 + o(dt)$

We write the net gain in $[t, t+dt]$ as

$$g(t)dt = F(s,t)\left[\Big(C(s,t)-C_s(s,t)s\Big)r\,dt - C_t(s,t)dt - \frac{1}{2}C_{ss}(s,t)\sigma^2 s^2 (dW)^2\right]\Bigg|_{s=S(t)}$$

Total gain in [0, *T*] for maintaining a prescribed delta hedging

$$G_{\text{Total}} = \int_0^T g(t)dt$$

$$= \int_0^T F(s,t)\left[\Big(C(s,t)-C_s(s,t)s\Big)r\,dt - C_t(s,t)dt - \frac{1}{2}C_{ss}(s,t)\sigma^2 s^2 (dW)^2\right]\Bigg|_{s=S(t)}$$

Recall Ito's lemma: we can replace $(dW)^2$ with dt. We obtain

$$G_{\text{Total}} = \int_0^T F(s,t)\left[\Big(C(s,t)-C_s(s,t)s\Big)r - C_t(s,t) - \frac{1}{2}C_{ss}(s,t)\sigma^2 s^2\right]\Bigg|_{s=S(t)} dt$$

6. **Governing PDE** for $C(s, t)$

Select the schedule of delta hedging, function $F(s, t)$

Recall that we maintain $F(S(t), t)$ units of delta hedging of time t, at time t.

We select function $F(s, t)$ to make the total gain positive.

Based on the published function $C(s, t)$, we select $F(s, t)$ as

$$F(s,t) \equiv \Big(C(s,t)-C_s(s,t)s\Big)r - C_t(s,t) - \frac{1}{2}C_{ss}(s,t)\sigma^2 s^2$$

With the selected $F(s, t)$, the total gain in $[0, T]$ becomes

$$G_{\text{Total}} = \int_0^T \left(\Big(C(s,t)-C_s(s,t)s\Big)r - C_t(s,t) - \frac{1}{2}C_{ss}(s,t)\sigma^2 s^2\right)^2\Bigg|_{s=S(t)} dt \qquad \text{(G01)}$$

Observations:

- The integrand in (G01) is a stochastic process (a random variable at each t). After integration, the total gain is still a random variable, a function of $S(t, \omega)$.
- Since the integrand is always non-negative, the total gain is always non-negative. That is, over any realized path $S(t)$, we will never lose money. The delta hedging trading scheme is completely risk-free!
- If the integrand is non-zero in a region of (s, t), then the trading scheme gives a positive gain whenever the realized path $S(t)$ goes through the region. For the market maker who sets $C(s, t)$, it is a guaranteed loss with no possibility of gain. When setting $C(s, t)$, the market maker must get rid of this risk.

Therefore, the integrand in the total gain (G01) must be identically zero.

$$\big(C(s,t)-C_s(s,t)s\big)r-C_t(s,t)-\frac{1}{2}C_{ss}(s,t)\sigma^2 s^2=0 \qquad \text{for all } (s,t)$$

Governing equation for $C(s, t)$

It follows that $C(s, t)$ satisfies the PDE

$$C_t(s,t)+\frac{1}{2}\sigma^2 s^2 C_{ss}(s,t)=r\big(C(s,t)-sC_s(s,t)\big)$$

At end time T (expiry), the price of the call option is simply the amount of money the call option holder can make when

- exercising the option to buy the stock (if the strike price is lower), and
- then immediately selling the stock in the open market.

$$C(s,T)\Big|_{s=S(T)}=\begin{cases} S(T)-K\,, & \text{if } S(T)>K \\ 0\,, & \text{otherwise} \end{cases}$$

The final condition for $C(s, t)$ is

$$C(s,t)\Big|_{t=T}=\max(s-K,0)$$

The FVP for $C(s, t)$ is

$$\begin{cases} C_t(s,t)+\dfrac{1}{2}\sigma^2 s^2 C_{ss}(s,t)=r\big(C(s,t)-sC_s(s,t)\big) \\ C(s,t)\Big|_{t=T}=\max(s-K,0) \end{cases}$$

Notice that the drift term μ is not in the FVP.

AM216 Stochastic Differential Equations

Lecture 17

List of topics in this lecture

- Analytical expression of the option price function $C(s, t)$
- Expected reward at time T for buying the option at time t_0
- Nominal value at time T of the amount $C(s_0, t_0)$ at time t_0
- Effect of interest rate: option price increase with interest rate
- Effect of volatility: option price increase with volatility

---

Recap

Black-Scholes option pricing model

Evolution of the underlying stock price

$$dS = \mu S\,dt + \sigma S\,dW \qquad \text{(Ito interpretation)}$$

Options associated with a stock

1 unit of call option = the right to buy 1 share of the stock at price K at time T.

1 unit of put option = the right to sell 1 share of the stock at price K at time T.

Mathematical formulation of the option price function

The option price at the current time t is a deterministic function of the current stock price $S(t)$ and the current time t:

Option price = $C(S(t), t)$ where $C(s, t)$ is a deterministic function.

The key question:

Suppose I am a market maker and I am required to set and publish $C(s, t)$.

How should I set function $C(s, t)$ to avoid a guaranteed loss?

Delta hedging portfolio

1 unit of delta hedging of time t

= owning (-1) unit of call option and $C_S(S(t), t)$ shares of stock.

Caution: the composition of delta hedging varies with t.

Net gain in $[0, T]$ for maintaining a prescribed delta hedging

Suppose that over time period [0, T], we maintain a portfolio of $F(S(t), t)$ units of delta hedging of time t, at time t.

$$G_{\text{Total}} = \int F(s,t)\left(\left(C(s,t) - C_s(s,t)s\right)r - C_t(s,t) - \frac{1}{2}C_{ss}(s,t)\sigma^2 s^2\right)\bigg|_{s=S(t)} dt$$

We set $F(s, t)$ to make $G_{\text{Total}} = \int (\quad)^2 dt$.

To avoid a guaranteed loss for the market maker, ( ) must be identically zero.

Governing FVP of $C(s, t)$

$$\begin{cases} C_t(s,t) + \frac{1}{2}\sigma^2 s^2 C_{ss}(s,t) = r\left(C(s,t) - sC_s(s,t)\right) \\ C(s,t)\big|_{t=T} = \max(s-K, 0) \end{cases}$$

Analytical expression of $C(s, t)$

We use change of variables to rewrite the FVP as an IVP.

New time variable

$\tau = T - t$ time to expiration

==> $t = T - \tau$

New spatial (price) variable

$$x = \log\frac{s}{K} + \left(r - \frac{1}{2}\sigma^2\right)(T-t)$$

$$\Longrightarrow \quad s = K\exp\left(x - \left(r - \frac{1}{2}\sigma^2\right)\tau\right)$$

New price function

$$u(x,\tau) = e^{r(T-t)}C(s,t)$$

$$\Longrightarrow \quad C(s,t) = e^{-r\tau}u(x,\tau)$$

Derivatives of $C(s, t)$

We start with the derivatives of (τ, x) with respect to (t, s).

$$\frac{\partial\tau}{\partial t} = -1, \qquad \frac{\partial\tau}{\partial s} = 0$$

$$\frac{\partial x}{\partial t} = -\left(r - \frac{1}{2}\sigma^2\right), \qquad \frac{\partial x}{\partial s} = \frac{1}{s}$$

We express derivatives of $C(s, t)$ in terms of those of $u(x, \tau)$ using the chain rule.

$$\frac{\partial}{\partial t}C(s,t)=\frac{\partial}{\partial \tau}\left[e^{-r\tau}u(x,\tau)\right]\cdot\frac{\partial \tau}{\partial t}+\frac{\partial}{\partial x}\left[e^{-r\tau}u(x,\tau)\right]\cdot\frac{\partial x}{\partial t}$$

$$=re^{-r\tau}u(x,\tau)-e^{-r\tau}\frac{\partial}{\partial \tau}u(x,\tau)-\left(r-\frac{1}{2}\sigma^2\right)e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)$$

$$\frac{\partial}{\partial s}C(s,t)=\frac{\partial}{\partial x}\left[e^{-r\tau}u(x,\tau)\right]\cdot\frac{\partial x}{\partial s}=e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\cdot\frac{1}{s}$$

$$\frac{\partial^2}{\partial s^2}C(s,t)=\frac{\partial}{\partial s}\left[e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\cdot\frac{1}{s}\right]$$

$$=\frac{\partial}{\partial s}\left[e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\right]\cdot\frac{1}{s}-e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\cdot\frac{1}{s^2}$$

$$=e^{-r\tau}\frac{\partial^2}{\partial x^2}u(x,\tau)\cdot\frac{1}{s^2}-e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\cdot\frac{1}{s^2}$$

PDE for $u(x, \tau)$

Substituting these derivatives into the PDE for $C(s, t)$, we obtain the PDE for $u(x, \tau)$.

$$\underbrace{re^{-r\tau}u(x,\tau)-e^{-r\tau}\frac{\partial}{\partial \tau}u(x,\tau)-\left(r-\frac{1}{2}\sigma^2\right)e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)}_{C_t(s,t)\equiv T_1+T_2+T_3}$$

$$\underbrace{+\frac{1}{2}\sigma^2\left[e^{-r\tau}\frac{\partial^2}{\partial x^2}u(x,\tau)-e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\right]}_{\frac{1}{2}\sigma^2 s^2 C_{ss}(s,t)\equiv T_4+T_5}=\underbrace{r\left(e^{-r\tau}u(x,\tau)-e^{-r\tau}\frac{\partial}{\partial x}u(x,\tau)\right)}_{r\left(C(s,t)-sC_s(s,t)\right)\equiv T_6+T_7}$$

Combining T_1 with T_6, first part of T_3 with T_7, second part of T_3 with T_5, we obtain

$$-e^{-r\tau}\frac{\partial}{\partial \tau}u(x,\tau)+\frac{1}{2}\sigma^2 e^{-r\tau}\frac{\partial^2}{\partial x^2}u(x,\tau)=0$$

which leads to a simple PDE for $u(x, \tau)$

$$\boxed{u_\tau(x,\tau)=\frac{1}{2}\sigma^2 u_{xx}(x,\tau)}$$

where $u(x, \tau)$ is related to $C(s, t)$ by

$$C(s,t)=e^{-r\tau}u(x,\tau), \qquad x=\log\frac{s}{K}+\left(r-\frac{1}{2}\sigma^2\right)\tau, \quad \tau=T-t\,.$$

Initial condition for $u(x, \tau)$

We use $s(x,\tau)\Big|_{\tau=0} = K\exp(x-(r-\frac{1}{2}\sigma^2)\tau)\Big|_{\tau=0} = Ke^x$.

$$u(x,\tau)\Big|_{\tau=0} = C(s,t)\Big|_{t=T} = \max\Big(Ke^x - K, 0\Big) = K\max\Big((e^x-1),0\Big)$$

The initial value problem (IVP) for $u(x, \tau)$

$$\begin{cases} u_\tau(x,\tau) = \dfrac{1}{2}\sigma^2 u_{xx}(x,\tau) \\ u(x,\tau)\Big|_{\tau=0} = K\begin{cases} (e^x-1), & x>0 \\ 0, & x<0 \end{cases} \end{cases}$$

Solution of the IVP

We use the fundamental solution of the heat equation to write out $u(x, \tau)$.

$$u(x,\tau) = \int_{-\infty}^{\infty} u(y,0)\frac{1}{\sqrt{2\pi\sigma^2\tau}}\exp\left(\frac{-(y-x)^2}{2\sigma^2\tau}\right)dy$$

$$= \frac{K}{\sqrt{2\pi\sigma^2\tau}}\int_0^{\infty}(e^y-1)\exp\left(\frac{-(y-x)^2}{2\sigma^2\tau}\right)dy \equiv K(I_2 - I_1)$$

We express I_1 and I_2 in terms of the error function. Recall that

$$F_{N(0,\sigma^2\tau)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}\exp\left(\frac{-\xi^2}{2\sigma^2\tau}\right)d\xi = \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x}{\sqrt{2\sigma^2\tau}}\right)\right)$$

We write out integral I_1 in terms of the error function.

$$I_1 = \frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_0^{\infty}\exp\left(\frac{-(y-x)^2}{2\sigma^2\tau}\right)dy$$

change of variables $\xi = x - y$

$$= \frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_{-\infty}^{x}\exp\left(\frac{-\xi^2}{2\sigma^2\tau}\right)d\xi = \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x}{\sqrt{2\sigma^2\tau}}\right)\right)$$

For integral I_2, we first complete the square in the exponent.

$$I_2 = \frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_0^{\infty}\exp\left(\frac{-(y-x)^2}{2\sigma^2\tau}+y\right)dy$$

$$= \frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_0^{\infty}\exp\left(\frac{-\left[y^2-2(x+\sigma^2\tau)y+(x+\sigma^2\tau)^2\right]}{2\sigma^2\tau}+x+\frac{\sigma^2\tau}{2}\right)dy$$

$$= \exp\left(x + \frac{\sigma^2\tau}{2}\right)\frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_0^{\infty}\exp\left(\frac{-(y-x-\sigma^2\tau)^2}{2\sigma^2\tau}\right)dy$$

We then use change of variables $\xi = x + \sigma^2\tau - y$ to write I_2 as

$$I_2 = \exp\left(x + \frac{\sigma^2\tau}{2}\right)\frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_{-\infty}^{(x+\sigma^2\tau)}\exp\left(\frac{-\xi^2}{2\sigma^2\tau}\right)dy$$

$$= \exp\left(x + \frac{\sigma^2\tau}{2}\right)\frac{1}{2}\left(1 + \text{erf}\left(\frac{x+\sigma^2\tau}{\sqrt{2\sigma^2\tau}}\right)\right)$$

Combining I_1 and I_2, we arrive at the expression of $u(x, \tau)$

$$\boxed{u(x,\tau) = \frac{K}{2}\left\{\exp\left(x + \frac{\sigma^2\tau}{2}\right)\left(1 + \text{erf}\left(\frac{x+\sigma^2\tau}{\sqrt{2\sigma^2\tau}}\right)\right) - \left(1 + \text{erf}\left(\frac{x}{\sqrt{2\sigma^2\tau}}\right)\right)\right\}}$$

Solution of $C(s,t) = e^{-r\tau}u(x,\tau)$

$$\boxed{\begin{aligned} &C(s,t) = \frac{e^{-r\tau}K}{2}\left\{\exp\left(x + \frac{\sigma^2\tau}{2}\right)\left(1 + \text{erf}\left(\frac{x+\sigma^2\tau}{\sqrt{2\sigma^2\tau}}\right)\right) - \left(1 + \text{erf}\left(\frac{x}{\sqrt{2\sigma^2\tau}}\right)\right)\right\} \\ &\qquad x = \log\frac{s}{K} + \left(r - \frac{1}{2}\sigma^2\right)\tau, \quad \tau = T - t \end{aligned}}$$

Function $\phi(\eta, \omega)$

For analyzing various trends, we introduce

$$\eta \equiv x + \frac{1}{2}\sigma^2\tau = \log\frac{s}{K} + r\tau, \qquad \omega \equiv \frac{1}{2}\sigma^2\tau$$

==> $\quad x + \sigma^2\tau = \eta + \omega, \qquad 2\sigma^2\tau = 4\omega$

We write $C(s, t)$ as a function of (η, ω).

$$\boxed{C(s,t) = \frac{e^{-r\tau}K}{2}\phi(\eta,\omega), \qquad \eta = \log\frac{s}{K} + r\tau, \quad \omega = \frac{1}{2}\sigma^2\tau} \qquad \text{(C-1)}$$

where function $\phi(\eta, \omega)$ is defined as

$$\phi(\eta,\omega) = e^{\eta}\left[1 + \text{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right] - \left[1 + \text{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right] \qquad \text{(F-1)}$$

Derivatives of $\phi(\eta, \omega)$

$$\frac{d}{dz}\text{erf}(z)=\frac{d}{dz}\int_0^z \frac{2}{\sqrt{\pi}}\exp(-s^2)ds=\frac{2}{\sqrt{\pi}}\exp(-z^2)$$

We use this result to calculate $\phi_\eta(\eta, \omega)$

$$e^\eta \frac{\partial}{\partial \eta}\text{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)=\frac{2}{\sqrt{\pi}}e^\eta \exp\left(-\frac{(\eta^2+2\eta\omega+\omega^2)}{4\omega}\right)\frac{1}{\sqrt{4\omega}}$$

$$\frac{\partial}{\partial \eta}\text{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)=\frac{2}{\sqrt{\pi}}\exp\left(-\frac{(\eta^2-2\eta\omega+\omega^2)}{4\omega}\right)\frac{1}{\sqrt{4\omega}}$$

Noticing that the two exponentials are the same.

$$e^\eta \exp\left(-\frac{(\eta^2+2\eta\omega+\omega^2)}{4\omega}\right)=\exp\left(-\frac{(\eta^2-2\eta\omega+\omega^2)}{4\omega}\right)$$

We obtain

$$\boxed{\frac{\partial \phi(\eta,\omega)}{\partial \eta}=e^\eta\left(1+\text{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right)>0} \qquad \text{(DF-1)}$$

$\phi(\eta, \omega)$ is an increasing function of η.

We calculate $\phi_\omega(\eta, \omega)$ in a similar way.

$$\frac{\partial \phi(\eta,\omega)}{\partial \omega}=e^\eta \exp\left(\frac{-(\eta^2+2\eta\omega+\omega^2)}{4\omega}\right)\left(-\frac{\eta}{4\omega^{3/2}}+\frac{1}{4\omega^{1/2}}\right)$$
$$-\exp\left(\frac{-(\eta^2-2\eta\omega+\omega^2)}{4\omega}\right)\left(-\frac{\eta}{4\omega^{3/2}}-\frac{1}{4\omega^{1/2}}\right)$$

Again, the two exponentials are the same. We obtain

$$\boxed{\frac{\partial \phi(\eta,\omega)}{\partial \omega}=\exp\left(\frac{-(\eta^2-2\eta\omega+\omega^2)}{4\omega}\right)\frac{1}{2\omega^{1/2}}>0} \qquad \text{(DF-2)}$$

$\phi(\eta, \omega)$ is an increasing function of ω.

We combine (F-1) and (DF-1) to calculate $(e^{-\eta}\phi(\eta, \omega))_\eta$.

$$\boxed{\frac{\partial\left(e^{-\eta}\phi(\eta,\omega)\right)}{\partial \eta}=e^{-\eta}\left(1+\text{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right) > 0} \qquad \text{(DF-1B)}$$

$e^{-\eta}\phi(\eta, \omega)$ is an increasing function of η.

Summary:

- Both $\phi(\eta, \omega)$ and $e^{-\eta}\phi(\eta, \omega)$ are increasing functions of η.

- $\phi(\eta, \omega)$ is an increasing function of ω.

<u>Expression of</u> $C_s(s, t)$

We use (DF-1) to calculate $C_s(s, t)$, which is needed in the delta hedging.

$$C(s,t)=\frac{e^{-r\tau}K}{2}\phi(\eta,\omega)\,,\qquad \eta=\log\frac{s}{K}+r\tau\,,\quad \omega=\frac{1}{2}\sigma^2\tau$$

$$\frac{\partial C(s,t)}{\partial s}=\frac{e^{-r\tau}K}{2}\frac{\partial\phi(\eta,\omega)}{\partial\eta}\cdot\frac{d\eta}{ds}=\frac{e^{-r\tau}K}{2}\mathrm{e}^{\eta}\left(1+\mathrm{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right)\cdot\frac{1}{s}\,,\qquad \mathrm{e}^{\eta}=\frac{s}{K}e^{r\tau}$$

We arrive at

$$\boxed{C_s(s,t)=\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right),\qquad \eta=\log\frac{s}{K}+r\tau\,,\quad \omega=\frac{1}{2}\sigma^2\tau}$$

Expected reward for buying the option at time t_0

We compare the rewards of buying the option vs not buying.

Let $s_0 \equiv S(t_0)$ be the stock price at time t_0.

$C(s_0, t_0)$ is the amount needed to buy the option at time t_0.

<u>Nominal value at time T</u> of the amount $C(s_0, t_0)$ at time t_0

$$\left[\text{NV at } T \text{ of } C(s_0,t_0) \text{ at } t_0\right]=\mathrm{e}^{r(T-t_0)}C(s_0,t_0)=\mathrm{e}^{r\tau_0}C(s_0,t_0)\,,\qquad \tau_0=T-t_0$$

This is the "reward" at time T for putting $C(s_0, t_0)$ in savings at time t_0.

We write it in terms of (η, ω), using equation (C-1)

$$\mathrm{e}^{r\tau_0}C(s_0,t_0)=\frac{K}{2}\phi(\eta,\omega)\,,\qquad \eta=\log\frac{s_0}{K}+r\tau_0\,,\quad \omega=\frac{1}{2}\sigma^2\tau_0\,,\quad \tau_0=T-t_0$$

Next we calculate the expected reward at time T for buying the option at time t_0.

<u>Evolution of</u> $Y = \log(S)$

$$dS=\mu S\,dt+\sigma S\,dW \quad (\text{Ito})\,,\quad \text{starting at } S(t_0)=s_0$$

This SDE (Ito) for S corresponds to the SDE below for $Y \equiv \log(S)$

$$dY=\left(\mu-\tfrac{1}{2}\sigma^2\right)dt+\sigma\,dW\,,\qquad Y=\log(S) \quad \text{starting at } Y(t_0)=\log(s_0)$$

$$\Longrightarrow\qquad Y(T)=Y(t_0)+\left(\mu-\tfrac{1}{2}\sigma^2\right)(T-t_0)+\sigma\left(W(T)-W(t_0)\right)$$

$$\Longrightarrow\qquad Y(T)=\log(s_0)+\left(\mu-\tfrac{1}{2}\sigma^2\right)\tau_0+N\left(0,\sigma^2\tau_0\right),\qquad \tau_0=T-t_0$$

<u>The probability density of $Y(T)$.</u>

$$\rho_Y(y)=\rho_{N\left(\log(s_0)+\mu\tau_0-\frac{1}{2}\sigma^2,\,\sigma^2\tau_0\right)}(y)=\frac{1}{\sqrt{2\pi\sigma^2\tau_0}}\exp\left(\frac{-\left(y-\log(s_0)-(\mu-\frac{1}{2}\sigma^2)\tau_0\right)^2}{2\sigma^2\tau_0}\right)$$

<u>Expected reward at time T for owning the option</u> is

$$E\left(\max(S(T)-K,0)\middle|S(t_0)=s_0\right)=E\left(\max(e^{Y(T)}-K,0)\right)$$

$$=\int_{-\infty}^{\infty}\max\left(\mathrm{e}^y-K,0\right)\rho_Y(y)dy=\int_{\log(K)}^{\infty}(\mathrm{e}^y-K)\rho_Y(y)dy\equiv J_2-J_1$$

We work out J_1 and J_2 similar to what did previously on I_1 and I_2.

In J_1, we use change of variables: $y=-\xi$.

$$J_1=K\int_{\log(K)}^{\infty}\rho_Y(y)dy=\frac{K}{\sqrt{2\pi\sigma^2\tau_0}}\int_{-\infty}^{-\log(K)}\exp\left(\frac{-\left(\xi+\log(s_0)+\mu\tau_0-\frac{1}{2}\sigma^2\tau_0\right)^2}{2\sigma^2\tau_0}\right)d\xi$$

$$=\frac{K}{2}\left(1+\mathrm{erf}\left(\frac{\eta_\mu-\omega}{\sqrt{4\omega}}\right)\right),\qquad \eta_\mu=\log\frac{s_0}{K}+\mu\tau_0\,,\quad \omega=\frac{1}{2}\sigma^2\tau_0$$

In J_2, we complete square and then use change of variables: $y=-\xi$.

$$J_2=\int_{\log(K)}^{\infty}e^y\rho_Y(y)dy=\frac{1}{\sqrt{2\pi\sigma^2\tau_0}}\int_{\log(K)}^{\infty}\exp\left(\frac{2\sigma^2\tau_0 y-\left(y-\log(s_0)-\mu\tau_0+\frac{1}{2}\sigma^2\tau_0\right)^2}{2\sigma^2\tau_0}\right)dy$$

$$=\frac{1}{\sqrt{2\pi\sigma^2\tau_0}}\exp\left(\log(s_0)+\mu\tau_0\right)\int_{\log(K)}^{\infty}\exp\left(\frac{-\left(y-\log(s_0)-\mu\tau_0-\frac{1}{2}\sigma^2\tau_0\right)^2}{2\sigma^2\tau_0}\right)dy$$

$$=\frac{K}{\sqrt{2\pi\sigma^2\tau_0}}\exp\left(\log\frac{s_0}{K}+\mu\tau_0\right)\int_{-\infty}^{-\log(K)}\exp\left(\frac{-\left(\xi+\log(s_0)+\mu\tau_0+\frac{1}{2}\sigma^2\tau_0\right)^2}{2\sigma^2\tau_0}\right)d\xi$$

$$=\frac{K}{2}\mathrm{e}^{\eta_\mu}\left(1+\mathrm{erf}\left(\frac{\eta_\mu+\omega}{\sqrt{4\omega}}\right)\right),\qquad \eta_\mu=\log\frac{s_0}{K}+\mu\tau_0\,,\quad \omega=\frac{1}{2}\sigma^2\tau_0$$

Combing J_1 and J_2, we write the expected reward at time T as

$$E\left(\max(S(T)-K,0)\middle|S(t_0)=s_0\right)=\frac{K}{2}\left\{\mathrm{e}^{\eta_\mu}\left(1+\mathrm{erf}\left(\frac{\eta_\mu+\omega}{\sqrt{4\omega}}\right)\right)-\left(1+\mathrm{erf}\left(\frac{\eta_\mu-\omega}{\sqrt{4\omega}}\right)\right)\right\}$$

$$= \frac{K}{2}\phi(\eta_\mu, \omega), \qquad \eta_\mu = \log\frac{s_0}{K} + \mu\tau_0, \quad \omega = \frac{1}{2}\sigma^2\tau_0$$

We compare the two rewards.

$$\underbrace{e^{r\tau_0}C(s_0, t_0)}_{\text{reward for putting it in savings}} = \frac{K}{2}\phi(\eta_r, \omega), \qquad \eta_r = \log\frac{s_0}{K} + r\tau_0, \quad \omega = \frac{1}{2}\sigma^2\tau_0$$

$$\underbrace{E\Big(\max(S(T) - K, 0)\Big| S(t_0) = s_0\Big)}_{\text{expected reward for buying the option}} = \frac{K}{2}\phi(\eta_\mu, \omega), \qquad \eta_\mu = \log\frac{s_0}{K} + \mu\tau_0, \quad \omega = \frac{1}{2}\sigma^2\tau_0$$

<u>Conclusions of the comparison:</u>

- Both rewards are given by $\phi(\eta, \omega)$ with respectively η_r and η_μ.

  In (DF-1), we derived $\phi_\eta(\eta, \omega) > 0$. It follows that $\phi(\eta, \omega)$ is an increasing function of η, which, in turn, is an increasing function of r.

- The reward at time T of owning the option is a random variable. The reward may be less than the cost $C(s_0, t_0)$. In fact, it has a large probability of 0 reward.

- The principle of risk and reward:

  A higher risk demands a greater <u>expected</u> reward.

  The principle of risk and reward tells us

  $$\phi(\eta_\mu, \omega) > \phi(\eta_r, \omega)$$

  ==> $\eta_\mu > \eta_r$

  ==> $\mu > r$

- In Appendix A, we do a three-way comparison by including the expected reward at time T of buying $C(s_0, t_0)$ amount of stock at time t_0.

- The principle of risk and reward is valid in the broader sense when we use the perceived risk and the perceived expected rewards, in all forms, received from all sources.

<u>Example:</u>

The reasoning for buying a lottery ticket.

I may assign a significant value to the excitement of a possible winning (the expected reward in all forms received from all sources is higher).

Or I may (falsely) believe that my number selection scheme will significantly increase my chance of winning (the perceived expected reward is higher than the actual value; the perceived risk is lower than the actual value).

The effect of interest rate *r* on $C(s, t)$

We rewrite $C(s, t)$ given in (C-1) to transfer the dependence on r into η.

$$C(s,t)=\frac{K}{2}e^{-r\tau}\phi(\eta,\omega)\,, \qquad \eta=\log\frac{s}{K}+r\tau\,, \quad \omega=\frac{1}{2}\sigma^2\tau$$

$$=\frac{s}{2}e^{-\eta}\phi(\eta,\omega)\,, \qquad e^{-\eta}=\frac{K}{s}e^{-r\tau}$$

In (DF-1B) we derived $(e^{-\eta}\phi_\eta(\eta, \omega))_\eta > 0$. It follows that

$$\frac{\partial C(s,t)}{\partial r}=\frac{s}{2}\frac{\partial\left(e^{-\eta}\phi(\eta,\omega)\right)}{\partial\eta}\cdot\frac{d\eta}{dr}=\frac{s}{2}\frac{\partial\left(e^{-\eta}\phi(\eta,\omega)\right)}{\partial\eta}\cdot\tau > 0$$

Conclusion:

Option price $C(s, t)$ increases with interest rate r.

The effect of volatility σ

The effect of volatility σ is contained in variable ω.

$$C(s,t)=\frac{K}{2}e^{-r\tau}\phi(\eta,\omega)\,, \qquad \eta=\log\frac{s}{K}+r\tau\,, \quad \omega=\frac{1}{2}\sigma^2\tau$$

In (DF-2) we derived $\phi_\omega(\eta, \omega) > 0$. It follows that

$$\frac{\partial C(s,t)}{\partial\sigma}=\frac{K}{2}e^{-r\tau}\frac{\partial\phi(\eta,\omega)}{\partial\omega}\cdot\frac{d\omega}{d\sigma}=\frac{K}{2}e^{-r\tau}\frac{\partial\phi(\eta,\omega)}{\partial\omega}\cdot\sigma\tau > 0$$

Conclusion:

Option price $C(s, t)$ increases with volatility σ.

The case of unknown σ

We can estimate σ from historic stock price data of the underlying stock and then use the estimated σ to predict the option price $C(s, t)$.

Conversely, we can use the current market price $C(s, t)$ of the option to estimate investors' perceived future volatility of the underlying stock.

- $C(s, t)$ increases with σ monotonically.
- For each realized sample of market price $C(s, t)$, there is a corresponding estimated value of perceived future volatility σ.

Appendix A: expected reward of buying option vs buying stock vs savings

We already calculated the two rewards.

$$\underbrace{e^{r\tau_0}C(s_0,t_0)}_{\text{reward for putting it in savings}} = \frac{K}{2}\phi(\eta_r,\omega), \quad \eta_r = \log\frac{s_0}{K} + r\tau_0, \quad \omega = \frac{1}{2}\sigma^2\tau_0$$

$$\underbrace{E\Big(\max(S(T)-K,0)\Big|S(t_0)=s_0\Big)}_{\text{expected reward for buying the option}} = \frac{K}{2}\phi(\eta_\mu,\omega), \quad \eta_\mu = \log\frac{s_0}{K} + \mu\tau_0, \quad \omega = \frac{1}{2}\sigma^2\tau_0$$

We calculate the expected reward at time T of buying $C(s_0, t_0)$ amount of stock at time t_0.

The stock price is governed by

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(t_0) = s_0$$

Note that in the Ito interpretation, $S(t)$ is independent of $dW(t) \equiv W(t+dt) - W(t)$.

Taking the average of both sides, we get

$$dE\big(S(t)\big) = \mu E\big(S(t)\big)dt, \quad E\big(S(t_0)\big) = s_0$$

$$\Longrightarrow \quad \frac{dE\big(S(t)\big)}{dt} = \mu E\big(S(t)\big)$$

$$\Longrightarrow \quad E\big(S(t)\big) = e^{\mu(t-t_0)}s_0$$

$$E\big(S(T)\big) = e^{\mu\tau_0}s_0, \quad \tau_0 = T - t_0$$

The expected reward at time T of buying $C(s_0, t_0)$ amount of stock at time t_0 is

$$\underbrace{e^{\mu\tau_0}C(s_0,t_0)}_{\text{expected reward for buying the stock}} = e^{\mu\tau_0 - r\tau_0}\underbrace{e^{r\tau_0}C(s_0,t_0)}_{\text{reward for putting it in savings}}$$

Using the result of $e^{r\tau_0}C(s_0,t_0)$, we write $e^{\mu\tau_0}C(s_0,t_0)$ in terms of η_r, ω and η_μ.

$$\underbrace{e^{\mu\tau_0}C(s_0,t_0)}_{\text{expected reward for buying the stock}} = \frac{K}{2}e^{\eta_\mu-\eta_r}\phi(\eta_r,\omega), \quad \eta_r = \log\frac{s_0}{K} + r\tau_0, \quad \eta_\mu = \log\frac{s_0}{K} + \mu\tau_0$$

Recall i) $\eta_\mu > \eta_r$ and ii) $e^{-\eta}\phi(\eta, \omega)$ is an increasing function of η. We obtain

$$\phi(\eta_r,\omega) < e^{\eta_\mu-\eta_r}\phi(\eta_r,\omega) < \phi(\eta_\mu,\omega)$$

Therefore we arrive at

$$\underbrace{\mathrm{e}^{r\tau_0}C(s_0,t_0)}_{\text{reward for putting it in savings}} < \underbrace{e^{\mu\tau_0}C(s_0,t_0)}_{\substack{\text{expected reward for}\\ \text{buying the stock}}} < \underbrace{E\Big(\max(S(T)-K,0)\Big|S(t_0)=s_0\Big)}_{\text{expected reward for buying the option}}$$

AM216 Stochastic Differential Equations

Lecture 18
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- The case of time-dependent drift
- Estimating volatility from market prices of options
- Scaling laws of option price
- In-the-money vs out-of-the-money options, short-term vs long-term
- Properties of option price, variations in option price

Recap

Black-Scholes option pricing model

Analytical expression of the option price function $C(s, t)$

$$C(s,t)=\frac{e^{-r\tau}K}{2}\phi(\eta,\omega)=\frac{s}{2}e^{-\eta}\phi(\eta,\omega)\,,\qquad \eta=\log\frac{s}{K}+r\tau\,,\quad \omega=\frac{1}{2}\sigma^2\tau$$

$$\phi(\eta,\omega)=\mathrm{e}^{\eta}\left[1+\mathrm{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right]-\left[1+\mathrm{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right]$$

- Both $\phi(\eta, \omega)$ and $\mathrm{e}^{-\eta}\phi(\eta, \omega)$ are increasing functions of η.
- $\phi(\eta, \omega)$ is an increasing function of ω.

Two expected rewards at time T

$$\underbrace{\mathrm{e}^{r\tau_0}C(s_0,t_0)}_{\text{reward for putting it in savings}}=\frac{K}{2}\phi(\eta_r,\omega)\,,\qquad \eta_r=\log\frac{s_0}{K}+r\tau_0\,,\quad \omega=\frac{1}{2}\sigma^2\tau_0$$

$$\underbrace{E\Big(\max(S(T)-K,0)\Big|S(t_0)=s_0\Big)}_{\text{expected reward for buying the option}}=\frac{K}{2}\phi(\eta_\mu,\omega)\,,\qquad \eta_\mu=\log\frac{s_0}{K}+\mu\tau_0\,,\quad \omega=\frac{1}{2}\sigma^2\tau_0$$

Principle of risk and reward:

$$\phi(\eta_\mu,\omega)>\phi(\eta_r,\omega)\qquad ==> \qquad \eta_\mu > \eta_r \qquad ==> \qquad \mu > r$$

Effect of interest rate r :

$C(s, t)$ increases with r

Effect of volatility σ :

$C(s, t)$ increases with σ

The case of time-dependent drift $\mu(t)$

$$dS = \mu(t)Sdt + \sigma S dW$$

We re-visit the process of deriving the PDE for $C(s, t)$.

…

At time t, buy $F(S(t), t)$ unit of delta hedging of time t .

At time $(t+dt)$, sell $F(S(t), t)$ unit of delta hedging of time t purchased at time t.

Change in cash balance at time $(t+dt)$ after selling

$$\Delta B = \underbrace{F\big(S(t),t\big)}_{\substack{\text{\# of units of}\\ \text{delta hedge}}}\left[\underbrace{-1}_{\substack{\text{\# of units}\\ \text{of option}}} \times \underbrace{C\big(S(t+dt),t+dt\big)}_{\substack{\text{option price}\\ \text{at time } t+\Delta t}} + \underbrace{C_s\big(S(t),t\big)}_{\substack{\text{\# of shares}\\ \text{of stock}}} \times \underbrace{S(t+dt)}_{\substack{\text{stock price}\\ \text{at time } t+\Delta t}}\right]$$

written in short notation

$$= F(S,t)\Big[-C(S+dS,t+dt)+C_s(S,t)(S+dS)\Big]$$

Taylor expansion in terms of dS and dt

$$= F(S,t)\left[\begin{array}{l} -C(S,t)-\underbrace{C_s(S,t)dS}_{dS\text{ term}}-C_t(S,t)dt-\dfrac{1}{2}C_{ss}(S,t)(dS)^2 \\ +C_s(S,t)S+\underbrace{C_s(S,t)dS}_{dS\text{ term}}+o(dt) \end{array}\right]$$

By the special design of delta hedging portfolio, the two dS terms cancel each other.

$$(dS)^2 = \big(\mu(t)Sdt + \sigma S dW\big)^2 = \sigma^2 S^2 (dW)^2 + o(dt)$$

The effect of $\mu(t)$ is buried in the $o(dt)$ term, which disappears as $dt \to 0$.

…

Conclusion: $\mu(t)$ does not affect $C(s, t)$.

In reality, $\mu(t)$ is unknown and unpredictable.

Estimating volatility from market prices of options

Recall that $C(s, t)$ is an increasing function of σ. Based on the observed market stock price $S(t)$ and market option price $C(S(t), t)$, we can estimate the perceived volatility.

Example:

Current time: t = the closing of May 26th, 2020.

Current stock price of Walt Disney Co: $s \equiv S(t)$ = \$120.95.

Interest rate: r = 2%/year; μ is not needed.

Table 1: Market prices of call options expiring June 5th, and the predicted volatility.

Expiry: T = June 5th 2020, τ = 8 trading days.

K, strike price	$C(s, t)$, market price of the option,	Predicted volatility σ^2 [1/year]
\$121	\$2.91	0.1140
\$122	\$2.47	0.1158
\$123	\$1.98	0.1092
\$124	\$1.63	0.1096
\$125	\$1.35	0.1118
\$126	\$1.05	0.1081

$E(\sigma^2)$ = 0.1114/year ==> $\sqrt{E(\sigma^2)} = 0.334/\sqrt{\text{year}}$

==> volatility = 33.4% fluctuation in a year

Table 2: Market prices of call options expiring June 19th, and the predicted volatility.

Expiry: T = June 19th 2020, τ = 18 trading days.

K, strike price	$C(s, t)$, market price of the option,	Predicted volatility σ^2 [1/year]
\$121	\$4.40	0.1138
\$122	\$3.90	0.1121
\$123	\$3.45	0.1112
\$124	\$3.05	0.1109
\$125	\$2.61	0.1067
\$126	\$2.31	0.1081

$E(\sigma^2)$ = 0.1105/year ==> $\sqrt{E(\sigma^2)} = 0.332/\sqrt{\text{year}}$

==> volatility = 33.2% fluctuation in a year

Scaling laws

Interest-rate-adjusted strike price at time t

To exercise the option at time T (buying the stock at strike price K), we need to allocate cash $e^{-r\tau}K$ at time t, which will grow to K at time T. Here $\tau = T - t$.

(Price K at time T) <==> (price $e^{-r\tau}K$ at time t)

We define the interest-rate-adjusted-strike price as $K_c \equiv e^{-r\tau}K$.

We write $C(s, t)$ and η in terms of K_c.

$$C(s,t) = \frac{e^{-r\tau}K}{2}\phi(\eta,\omega) = \frac{K_c}{2}\phi(\eta,\omega), \qquad \eta = \log\frac{s}{K_c}, \quad \omega = \frac{1}{2}\sigma^2\tau$$

Non-dimensional ratios

We normalize the rate-adjusted strike price and the option price by the stock price.

$$q \equiv \frac{K_c}{s}, \qquad Q \equiv \frac{C(s,t)}{s}$$

Q is a function of q and $\omega = \sigma^2\tau/2$; Q has no explicit dependence on s.

$$Q(q,\omega) = \frac{q}{2}\phi(\eta,\omega) = \frac{e^{-\eta}}{2}\phi(\eta,\omega), \qquad \eta = \log\frac{1}{q}, \quad \omega = \frac{1}{2}\sigma^2\tau$$

Scaling property 1: at certain conditions, $C(S(t), t) \propto S(t)$

Given t, given $K/S(t)$ and given σ, we have

$$q = \frac{e^{-r(T-t)}K}{S(t)} = \text{fixed}, \qquad \omega = \frac{1}{2}\sigma^2(T-t) = \text{fixed} \quad \Rightarrow \quad Q(q,\omega) = \text{fixed}$$

==> $C(S(t),t) = S(t)\cdot Q(q,\omega) \propto S(t)$

Conclusion:

Given t, given $K/S(t)$ and given σ, the option price is proportional to the stock price.

Example

Stock 1 is at $S_1(t) = 45$ and the call option on stock 1 with $K = 50$ is at $C_1 = 2$.

Suppose stock 2 has the same volatility and is at $S_2(t) = 90$.

The call option on stock 2 with $K = 100$ and the same T should be at $C_2 = 4$.

Scaling property 2: effect of volatility σ is only in the combination $\omega = \sigma^2\tau/2$.

Given K_c and given $S(t)$, we have $q = \dfrac{K_c}{S(t)} = \text{fixed}$.

$$C(S(t),t) = S(t)\cdot Q\big(q,\omega\big), \qquad \omega \equiv \frac{\sigma^2\tau}{2}$$

Conclusion:

Given K_c and given $S(t)$, when $\omega \equiv \sigma^2\tau/2$ is unchanged, the option price is unchanged.

Example:

Suppose stock 1 has volatility σ_1 and stock 2 has volatility $\sigma_2 = 2\sigma_1$. Suppose both stocks are currently at the same price $S_1(t) = S_2(t) = S(t)$.

We consider two options with times to expiry (τ_1, τ_2) and strike prices (K_1, K_2).

- the call option on stock 1 with τ_1 and $K_1 = \exp(r\tau_1)K_c$
- the call option on stock 2 with $\tau_2 = \tau_1/4$ and $K_2 = \exp(r\tau_2)K_c$

These two options should have the same option price.

Scaling property 3: effect of interest rate r is only in the combination $K_c \equiv e^{-r\tau}K$.

Given $S(t)$ and given $\omega \equiv \sigma^2\tau/2$, we have

$$C(S(t),t) = S(t)\cdot Q(q,\omega), \qquad q \equiv \frac{K_c}{S(t)} = \frac{e^{-r\tau}K}{S(t)}$$

Conclusion:

Given $S(t)$ and given $\omega \equiv \sigma^2\tau/2$, when $K_c \equiv e^{-r\tau}K$ is unchanged, the option price is unchanged even when both r and K are increased.

In-the-money vs out-of-the-money options, short-term vs long-term

In the terminology of option market, in-the-money and out-of-the-money call options are defined as follows.

Conventional definitions of in-the-money and out-of-the-money

In-the-money call option:

$K < S(t)$ (strike price < current stock price)

If the option is exercised immediately, there is a reward (not counting the premium already paid for the option)

Out-of-the-money call option:

$K > S(t)$ (strike price > current stock price)

If the option is exercised immediately, there is no reward.

Mathematically, we define in-the-money and out-of-the-money slightly differently.

Mathematical definitions of in-the-money and out-of-the-money

In-the-money call option (mathematical definition)

$K_c \equiv e^{-r\tau} K < S(t)$ (rate-adjusted strike price < current stock price)

There is a reward if the option holder does the followings

i) sells the stock at time t,

ii) set aside the proceeds (which will grow to $S(t)e^{r\tau}$ at time T), and

iii) at time T, use cash K to exercises the option to buy back the stock, which yields a cash balance of $S(t)e^{r\tau} - K > 0$.

Out-of-the-money call option (mathematical definition)

$K_c \equiv e^{-r\tau} K > s$ (rate-adjusted strike price > current stock price)

There is no reward if the option holder does the steps described above.

Remark:

The two sets of definitions are essentially the same for small $r\tau$.

Short-term in-the-money option:

$$C(s,t) = \frac{K_c}{2}\phi(\eta,\omega), \qquad \eta = \log\frac{s}{K_c}, \quad \omega = \frac{1}{2}\sigma^2\tau$$

$$\phi(\eta,\omega) = e^{\eta}\left[1+\text{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right] - \left[1+\text{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right]$$

Mathematically, short-term means small ω and in-the-money means $\eta > 0$.

We expand $\text{erf}(z)$ for $z > 0$ and large.

$$\text{erf}(z) \approx 1 - \frac{\exp(-z^2)}{z\sqrt{\pi}} \approx 1 \qquad \text{for } z > 0 \text{ and large}$$

We expand $\phi(\eta, \omega)$ and $C(s, t)$ for $\eta > 0$ and small ω.

$$\frac{\eta+\omega}{\sqrt{4\omega}} > 0 \text{ and large}, \qquad \frac{\eta-\omega}{\sqrt{4\omega}} > 0 \text{ and large}$$

$$\phi(\eta,\omega) \approx e^{\eta} - 1$$

$$C(s,t) \approx \frac{K_c}{2}(e^{\eta}-1) = s - K_c, \qquad \eta = \log\frac{s}{K_c}, \quad \omega = \frac{1}{2}\sigma^2\tau$$

(A more careful expansion: $C(s,t) \approx s - K_c + s\dfrac{\sigma^3\tau^{3/2}}{\eta^2\sqrt{2\pi}}\exp\left(\dfrac{-\eta^2}{2\sigma^2\tau}\right) \approx s - K_c$).

Conclusion:

For a very short-term option that is significantly in the money, the option price is essentially $(S(t) - K_c)$.

Short-term out-of-the-money option:

Mathematically, short-term means small ω and out-the-money means $\eta < 0$.

We expand $\mathrm{erf}(z)$ for $z < 0$ and large.

$$\mathrm{erf}(z) \approx -1 + \frac{\exp(-z^2)}{z\sqrt{\pi}} \approx -1 \qquad \text{for } z < 0 \text{ and large}$$

We expand $\phi(\eta, \omega)$ and $C(s, t)$ for $\eta < 0$ and small ω.

$$\frac{(\eta+\omega)}{\sqrt{4\omega}} < 0 \text{ and large}, \qquad \frac{(\eta-\omega)}{\sqrt{4\omega}} < 0 \text{ and large}$$

$$\phi(\eta,\omega) \approx 0$$

$$C(s,t) \approx 0$$

(A more careful expansion: $C(s,t) \approx s\dfrac{\sigma^3\tau^{3/2}}{\eta^2\sqrt{2\pi}}\exp\left(\dfrac{-\eta^2}{2\sigma^2\tau}\right) \approx 0$)

Conclusion:

For a very short-term option that is significantly out-of-the-money, the option price is essentially 0.

Long-term options:

Mathematically, long-term means $\omega \equiv \sigma^2\tau/2 \gg 1$, which almost does not exist in the real world of option market because $\omega \gg 1$ corresponds to a really long time.

Example:

$\sigma^2 = 0.25/\text{year}$ (50% fluctuation in a year, a significant fluctuation)

$\omega \equiv \sigma^2\tau/2 \geq 10$ requires $\tau \geq 80$ years

Even at $\omega = 10$, we have

$$\frac{\omega}{\sqrt{4\omega}} = \frac{\sigma^2\tau/2}{\sqrt{2\sigma^2\tau}} = \frac{10}{\sqrt{40}} \approx 1.58, \quad \text{not a large number.}$$

In the real world of option market, "long-term" refers to the case of $\omega \equiv \sigma^2\tau/2 \sim O(1)$, which does not have a simple asymptotic expression.

Properties of $C(s, t)$ (call options)

Option price is always positive

$C(S(t), t) \le 0$ is absolutely impossible.

Otherwise, we can make a risk-free gain by "buying" the option and doing nothing.

It makes no sense to exercise before expiry T

Exercising at time t requires cash K at time t.

Exercising at time T requires cash $e^{-r\tau}K < K$ at time t.

American style options and European style options are essentially the same (for stocks that do not pay a dividend).

Option price is always above $S(t)-e^{-r\tau}K$

$C(S(t), t) < S(t)-e^{-r\tau}K$ is absolutely impossible.

If $C(S(t), t) < S(t)-e^{-r\tau}K$, we can make a risk-free gain by doing the steps below.

- Sell the stock at $S(t)$ at time t.
- Buy the option at $C(S(t), t)$ at time t.

  The two actions above yield a cash balance of $S(t)- C(S(t), t) > e^{-r\tau}K$, which will grow to $e^{r\tau}(S(t)-C(S(t), t)) > K$ at time T.
- At time T, use cash K to exercise the option to close the position.

  We end with a positive cash balance: $e^{r\tau}(S(t)-C(S(t), t)) - K > 0$.

Option price is always below the stock price

$C(S(t), t) > S(t)$ is absolutely impossible.

If $C(S(t), t) > S(t)$, we can make a risk-free gain by doing the steps below.

- Sell the option at $C(S(t), t)$ at time t.
- Buy the stock at $S(t)$ at time t.

  The two actions above yield a cash balance of $C(S(t), t)-S(t) > 0$.
- At time T, the stock may or may not be called away.

We end with a positive cash balance, $e^{r\tau}(C(S(t), t) - S(t)) > 0$ plus the stock (if the option is not exercised) or additional amount K (if the stock is called away).

Variations in option price

Change in the option price vs change in the stock price

We calculate $\partial C/\partial s$ from the expression of $C(s, t)$.

$$C(s,t)=\frac{e^{-r\tau}K}{2}\phi(\eta,\omega)=\frac{s}{2}e^{-\eta}\phi(\eta,\omega), \qquad \eta=\log\frac{s}{K}+r\tau, \quad \omega=\frac{1}{2}\sigma^2\tau$$

$$\phi(\eta,\omega)=e^{\eta}\left[1+\mathrm{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right]-\left[1+\mathrm{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right]$$

Recall that $(e^{-\eta}\phi(\eta, \omega))_\eta > 0$. We obtain

$$\frac{\partial C(s,t)}{\partial s}=\frac{1}{2}e^{-\eta}\phi(\eta,\omega)+\frac{s}{2}\frac{\partial\left(e^{-\eta}\phi(\eta,\omega)\right)}{\partial\eta}>\frac{1}{2}e^{-\eta}\phi(\eta,\omega)=\frac{C(s,t)}{s}$$

$$\Longrightarrow \quad \frac{C(s+\Delta s,t)-C(s,t)}{\Delta s}>\frac{C(s,t)}{s}$$

$$\Longrightarrow \quad \left|\frac{C(s+\Delta s,t)-C(s,t)}{C(s,t)}\right| > \left|\frac{\Delta s}{s}\right|$$

Percentage-wise, the option price is more volatile than the stock price.

Change in the option price vs change in the strike price

We calculate $\partial C/\partial K$ from the expression of $C(s, t)$.

$$C(s,t)=\frac{s}{2}e^{-\eta}\phi(\eta,\omega), \qquad \eta=\log\frac{s}{K}+r\tau, \quad \omega=\frac{1}{2}\sigma^2\tau$$

Note that the effect of K is contained in η.

$$\frac{\partial C(s,t)}{\partial K}=\frac{s}{2}\frac{\partial\left(e^{-\eta}\phi(\eta,\omega)\right)}{\partial\eta}\cdot\frac{\partial\eta}{\partial K}=\frac{-s}{2K}\frac{\partial\left(e^{-\eta}\phi(\eta,\omega)\right)}{\partial\eta}<0$$

Recall that previously we derived

$$\frac{\partial}{\partial\eta}\left(e^{-\eta}\phi(\eta,\omega)\right)=e^{-\eta}\left(1+\mathrm{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right) < 2e^{-\eta}$$

$$\Longrightarrow \quad 0 < -\frac{\partial C}{\partial K} < \frac{s}{K}e^{-\eta}=e^{-r\tau}$$

$$\Longrightarrow \quad 0 < \frac{C\big|_K - C\big|_{K+\Delta K}}{\Delta K} < e^{-r\tau}$$

$$\Longrightarrow \quad \left| C\big|_K - C\big|_{K+\Delta K} \right| < e^{-r\tau} |\Delta K|$$

Conclusion:

- The change in option price is less than the change in strike price.
- The effect of strike price decreases as the time to expiry increases.

Price of actively traded options

Price of the at-the-money option

$$C(s,t) = K_c \frac{1}{2} \phi(\eta,\omega), \qquad \eta = \log\frac{s}{K_c}, \quad \omega = \frac{1}{2}\sigma^2\tau$$

At-the-money means $s = K_c$ and thus $\eta = 0$.

$$\phi(\eta,\omega)\Big|_{\eta=0} = \left[1 + \mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)\right] - \left[1 - \mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)\right] = 2\,\mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)$$

When the stock price is at $s = K_c$, the option price is

$$C(s,t) = e^{-r\tau} K \cdot \mathrm{erf}\left(\frac{\sqrt{\sigma^2\tau}}{2\sqrt{2}}\right) \quad \text{at } s = e^{-r\tau}K$$

Price of a near-the-money option

$$C(s,t) = K_c \frac{1}{2} \phi(\eta,\omega), \qquad \eta = \log\frac{s}{K_c}, \quad \omega = \frac{1}{2}\sigma^2\tau$$

Near-the-money means $s \approx K_c$ and thus $\eta \approx 0$.

$$\eta = \log\frac{s}{K_c} = \log\left(1 + \frac{s}{K_c} - 1\right) \approx \frac{s}{K_c} - 1$$

We expand $C(s, t)$ around $\eta = 0$.

$$\phi(\eta,\omega)\Big|_{\eta=0} = \left[e^{\eta}\left(1 + \mathrm{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right) - \left(1 + \mathrm{erf}\left(\frac{\eta-\omega}{\sqrt{4\omega}}\right)\right)\right]\Bigg|_{\eta=0} = 2\,\mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)$$

$$\frac{\partial}{\partial \eta}\phi(\eta,\omega)\Big|_{\eta=0} = e^{\eta}\left(1+\mathrm{erf}\left(\frac{\eta+\omega}{\sqrt{4\omega}}\right)\right)\Bigg|_{\eta=0} = 1+\mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)$$

$$\phi(\eta,\omega) \approx \phi(\eta,\omega)\big|_{\eta=0} + \eta\frac{\partial}{\partial \eta}\phi(\eta,\omega)\Big|_{\eta=0}$$

We use the expansion of $\phi(\eta, \omega)$ to approximate $C(s, t)$.

$$C(s,t) \approx \frac{K_c}{2}\left[\phi(\eta,\omega)\big|_{\eta=0} + \eta\frac{\partial}{\partial \eta}\phi(\eta,\omega)\Big|_{\eta=0}\right]$$

$$\approx \frac{K_c}{2}\left[2\mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right) + \left(\frac{s}{K_c}-1\right)\left(1+\mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)\right)\right]$$

$$\approx K_c \mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right) + (s-K_c)\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\sqrt{\omega}}{2}\right)\right)$$

When the stock price is near $s = K_c$, the option price is

$$C(s,t) \approx e^{-r\tau}K\cdot \mathrm{erf}\left(\frac{\sqrt{\sigma^2\tau}}{2\sqrt{2}}\right) + (s-e^{-r\tau}K)\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\sqrt{\sigma^2\tau}}{2\sqrt{2}}\right)\right) \quad \text{for } s \approx e^{-r\tau}K$$

AM216 Stochastic Differential Equations

Lecture 19
Copyright by Hongyun Wang, UCSC

List of topics in this lecture

- Smoluchowski-Kramers approximation, an intuitive derivation based on ODE
- Time scale of inertia, time scale of thermal excitation
- Equipartition of energy, root-mean-square velocity of a particle
- Characters of molecular motors vs macroscopic motors
- Time scale of Smoluchowski-Kramers approximation

Smoluchowski-Kramers approximation

Consider the stochastic motion of a small particle in water

It is governed by the Langevin equation (Newton's second law)

$$dX = Y\,dt$$

$$m dY = -bY\,dt + F(X,t)dt + q\,dW \qquad \text{(S01)}$$

where

X: position

Y: velocity

$m = (4\pi/3)a^3$: mass of the particle

a: radius of the particle

$b = 6\pi\eta a$: drag coefficient of the particle

$F(X, t)$: external force

$q = \sqrt{2k_B T\,b}$: magnitude of thermal excitation

Claim:

As $a \to 0$, the stochastic motion is approximately governed by

$$dX = \frac{F(X,t)}{b}dt + \sqrt{2D}\,dW, \qquad D = \frac{k_B T}{b}$$

This equation is called the over-damped Langevin equation.

This process is called the Smoluchowski-Kramers approximation.

We are going to "derive" the Smoluchowski-Kramers approximation in several ways.

An intuitive derivation based on the result of a deterministic ODE

Consider a deterministic ODE

$$\begin{cases} y' = -\lambda(y - g(t)) \\ y(0) = y_0 \end{cases} \qquad \text{(D01)}$$

where λ is positive and large.

Theorem

The solution of (D01) satisfies

$$\lim_{\lambda \to +\infty} y(t;\lambda) = g(t) \qquad \text{for } t > 0$$

Proof:

We solve (D01) analytically. First, we rewrite it as

$$y' + \lambda y = \lambda g(t)$$

Multiplying by the integrating factor, we have

$$e^{\lambda t} y' + \lambda e^{\lambda t} y = \lambda g(t) e^{\lambda t}$$

$$\Longrightarrow \quad \left(e^{\lambda t} y\right)' = \lambda g(t) e^{\lambda t}$$

Integrating from 0 to t, we get

$$e^{\lambda t} y(t) - y_0 = \lambda \int_0^t g(s) e^{\lambda s}\, ds$$

$$\Longrightarrow \quad y(t) = e^{-\lambda t} y_0 + \lambda \int_0^t g(s) e^{\lambda(s-t)}\, ds$$

Applying change of variables $u = t - s$, we write $y(t)$ as

$$y(t) = e^{-\lambda t} y_0 + \lambda \int_0^t g(t-u) e^{-\lambda u}\, du$$

For λ positive and large, the dominant contribution of the integral comes from the region near $u = 0$. We expand function g near $u = 0$.

$$y(t) = e^{-\lambda t} y_0 + \lambda \int_0^t \left[g(t) - g'(t) u + \tfrac{1}{2} g''(t) u^2 + \cdots \right] e^{-\lambda u}\, du$$

An integration formula:

$$\int_0^t u^k e^{-\lambda u}\,du = \frac{1}{\lambda^{k+1}}\int_0^{(\lambda t)} w^k e^{-w}\,dw = \frac{1}{\lambda^{k+1}}\big(\underbrace{\int_0^\infty w^k e^{-w}\,dw}_{=k!} + \text{T.S.T.}\big)$$

T.S.T. = Transcendentally small term with respect to (λt)

$$\boxed{\int_0^t u^k e^{-\lambda u}\,du = \frac{1}{\lambda^{k+1}}\left(k! + \text{T.S.T.}\right)}$$

$k = 0$: $\quad \int_0^t e^{-\lambda u}\,du = \frac{1}{\lambda}\left(1 + \text{T.S.T.}\right)$

$k = 1$: $\quad \int_0^t u e^{-\lambda u}\,du = \frac{1}{\lambda^2}\left(1 + \text{T.S.T.}\right)$

Using the integration formula, we obtain

$$y(t) = \underbrace{e^{-\lambda t} y_0}_{\text{T.S.T.}} + \lambda\left[g(t)\frac{1}{\lambda} - g'(t)\frac{1}{\lambda^2} + g''(t)\frac{1}{\lambda^3} + \cdots\right] + \text{T.S.T.}$$

Neglecting the TST, we arrive at

$$y(t) = g(t) - g'(t)\frac{1}{\lambda} + g''(t)\frac{1}{\lambda^2} + \cdots$$

$$\boxed{y(t) = g(t) + O(1/\lambda)}$$

Remarks:

- When (λt) is moderately large, the influence of initial condition $y(0)$ is TST.
- When λ is LARGE λ and t is not too small, (λt) is moderately large and we have $y(t) \approx g(t)$.
- When $(\lambda t) < 1$, $y(t)$ is highly affected by $y(0)$ and $y(t) \approx g(t)$ is invalid.

  In particular, for $(\lambda t) \ll 1$, solution $y(t)$ has the expansion

$$y(t) = e^{-\lambda t} y_0 + \lambda\int_0^t g(t-u)e^{-\lambda u}\,du = y_0 + O(\lambda t)$$

Applying the theorem "formally" to the SDE

We ignore the fact that (S01) is a stochastic differential equation. We treat it "formally" as a deterministic ODE and write it in the form $y' = -\lambda\,[y - g(t)]$.

$$m\,dY = -bV\,dt + F(X,t)\,dt + q\,dW$$

==> $$\frac{dY}{dt} = -\frac{b}{m}\left[Y - \left(\frac{F(X,t)}{b} + \frac{q}{b}\frac{dW}{dt}\right)\right] \qquad \text{(S01B)}$$

It has the form $y' = -\lambda\,[y - g(t)]$ where

$$\lambda \equiv \frac{b}{m}, \qquad g(t) \equiv \left(\frac{F(X,t)}{b} + \frac{q}{b}\frac{dW}{dt}\right)$$

As $a \to 0$, we have

$b = O(a)$, $\qquad m = O(a^3)$

==> $$\lambda = \frac{b}{m} = O(a^{-2}) \to \infty \qquad \text{as } a \to 0$$

We "formally" apply the theorem above to (S01B) to obtain

$$Y(t) = \left(\frac{F(X,t)}{b} + \frac{q}{b}\frac{dW}{dt}\right), \qquad q = \sqrt{2k_B T b}, \qquad \frac{q}{b} = \sqrt{\frac{2k_B T}{b}} = \sqrt{2D}$$

Multiplying by dt and using $Ydt = dX$, we arrive at

$$dX = \frac{F(X,t)}{b}dt + \sqrt{2D}\,dW$$

This is the <u>over-damped Langevin equation</u>.

A more rigorous derivation

<u>What we learned in the ODE</u> $y' = -\lambda\,[y - g(t)]$

- The time scale of the influence of $y(0)$ is $O(1/\lambda)$.
- For $t \in [0, O(1/\lambda)]$, we don't have $y(t) = g(t)$.
- For LARGE λ and $t \gg O(1/\lambda)$, the influence of $y(0)$ disappears and we have

  $y(t) \approx g(t)$.

We consider the case of $F(x, t) \equiv F_0$. We discuss

- Time scale of inertia
- Time scale of thermal excitation.
- Equipartition of energy, root-mean-square velocity of a particle
- Time scale of the Smoluchowski-Kramers approximation

<u>Time scale of inertia</u>

$$m\,dY = -bY\,dt + F_0 dt + q\,dW$$

$$\text{==>} \quad m\,d\big(Y - F_0/b\big) = -b\big(Y - F_0/b\big)dt + q\,dW$$

Let $V(t) \equiv \big(Y(t) - F_0/b\big)$.

$$m\,dV = -bV\,dt + q\,dW \qquad \text{an Ornstein-Uhlenbeck process}$$

Previously, for an OU process, we derived

$$V(t) = V(0)e^{-\beta t} + N\left(0,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t})\right), \qquad \beta = \frac{b}{m}, \qquad \gamma = \frac{q}{m}$$

$$\text{==>} \quad Y(t) - \frac{F_0}{b} = \left(Y(0) - \frac{F_0}{b}\right)e^{-2\beta t} + N\left(0,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t})\right)$$

We write $Y(t)$ in terms of $t_0 \equiv 1/\beta = m/b$.

$$Y(t) = Y(0)e^{-t/t_0} + \frac{F_0}{b}\left(1-e^{-t/t_0}\right) + N\left(0,\ \frac{\gamma^2 t_0}{2}(1-e^{-2t/t_0})\right), \qquad t_0 \equiv \frac{1}{\beta} = \frac{m}{b}$$

$t_0 \equiv m/b$ has the dimension of time and is called the time scale of inertia.

Observation: the effect of inertia is a matter of time scale.

- The regime of $t << t_0$

$$Y(t) = Y(0) + O(t/t_0) + N\big(0,\ O(t/t_0)\big) \approx Y(0)$$

  In this regime, the inertia is dominant. The velocity at time t is almost entirely determined by the initial velocity.

- The regime of $t = O(t_0)$

$$Y(t) = Y(0)e^{-t/t_0} + \frac{F_0}{b}\left(1-e^{-t/t_0}\right) + N\left(0,\ \frac{\gamma^2 t_0}{2}(1-e^{-2t/t_0})\right)$$

  In this regime, the remaining effect of inertia is still significant while other terms are no longer negligible.

- The regime of $t >> t_0$

$$Y(t) = \frac{F_0}{b} + N\left(0,\ \frac{\gamma^2 t_0}{2}\right) + \text{T.S.T.}$$

  In this regime, the effect of inertia is negligible. The distribution of velocity at time t is independent of the initial velocity.

Examples (time scale of inertia)

The time scale of inertia for a bead of radius a in water.

$b = 6\pi\eta a$ (drag coefficient)

$$m = \rho\frac{4\pi a^3}{3}$$

==> $$t_0 = \frac{m}{b} = \frac{2\rho a^2}{9\eta} \propto a^2$$

- For a bead of 1 μm diameter in water, we have

  a = 0.5 μm = 0.5 × 10^−4^ cm (radius)

  $\rho = 1$ g (cm)$^{-3}$ (density of bead material)

  η = 0.01 poise = 0.01 g (cm)$^{-1}$ s^{-1} (viscosity of water)

==> $$t_0 = \frac{2\rho a^2}{9\eta} \approx 5.6\times 10^{-8} s = 56\,\text{ns} \qquad (\text{ns} = 10^{-9}\text{ s})$$

- For a bead of 10 nm diameter in water, we have

$$t_0 = \frac{2\rho a^2}{9\eta} \approx 5.6\times 10^{-12}\,\text{s} = 5.6\,\text{ps}$$

For molecular motors, we are concerned with reactions and motions in time scale of ms (ms = 10^{-3} s). So the effect of inertia can be neglected. If we want to know their detailed dynamics in time scale of ps, the inertia plays the dominant role.

Time scale of thermal excitation

$$Y(t) = Y(0)e^{-t/t_0} + \frac{F_0}{b}\left(1-e^{-t/t_0}\right) + N\left(0,\ \frac{\gamma^2 t_0}{2}(1-e^{-2t/t_0})\right)$$

==> $$\text{var}\left(Y(t)\middle|Y(0)\right) = (\gamma^2 t_0/2)(1-e^{-2t/t_0})$$

Given $Y(0)$, the variance var$(Y(t)|Y(0))$ comes from thermal excitation.

Observation: the time scale of thermal excitation is also t_0.

- For $t << t_0$

$$\text{var}\left(Y(t)\middle|Y(0)\right) = \frac{\gamma^2 t_0}{2}(1-e^{-2t/t_0}) \approx \gamma^2 t$$ which grows linearly with t.

- For $t \gg t_0$

$$\text{var}\left(Y(t)\middle|Y(0)\right) = \frac{\gamma^2 t_0}{2}(1-e^{-2t/t_0}) \approx \frac{\gamma^2 t_0}{2}$$ which has reached its saturation level.

Equipartition of energy

In the absence of an external driving force ($F_0 = 0$), after reaching equilibrium ($t \gg t_0$),

$$E\big(Y(t)\big) = 0$$

$$E\big(Y(t)^2\big) = \frac{\gamma^2 t_0}{2} = \frac{q^2}{2m^2}\cdot\frac{m}{b} = \frac{2k_B T b}{2mb} = \frac{k_B T}{m}, \qquad q = \sqrt{2k_B T b}$$

$$\Longrightarrow \qquad \frac{1}{2} m E\big(Y(t)^2\big) = \frac{1}{2} k_B T$$

This is called equipartition of energy:

At equilibrium, the thermal energy associated with each degree of freedom is $k_B T/2$, independent of particle size and independent of mass and density.

Root-mean-square velocity of a particle

The root-mean-square (RMS) velocity gives us the typical magnitude of the stochastic instantaneous velocity.

$$\text{RMS velocity} \equiv \sqrt{E(Y^2)} = \sqrt{\frac{k_B T}{m}} = \sqrt{\frac{3k_B T}{4\pi\rho a^3}} \propto a^{-3/2}$$

Examples (RMS velocity)

- RMS velocity of a 1 μm bead (diameter) in water:

  $k_B T = 4.1$ pN·nm $= 4.1\times10^{-14}$ g (cm)^2 s^{-2}

  $\rho = 1$ g (cm)$^{-3}$

  $a = 0.5$ μm $= 0.5 \times 10^{-4}$ cm

  $$\sqrt{E(Y^2)} = \sqrt{\frac{3k_B T}{4\pi\rho a^3}} = 0.28 \text{ cm/s} = 2.8\times10^3 \text{ μm/s} = 2800 \text{ body-size/s}$$

- RMS velocity of a 10 nm bead (diameter) in water:

  $$\sqrt{E(Y^2)} = 280 \text{ cm}/s = 2.8 m/s = 2.8\times10^8 \text{ body-size/s.}$$

Example (magnitude of thermal excitation)

Consider a bottle of water. Suppose the thermal energy is used to drive all water molecules to move in the same direction with the same velocity and with no relative motion with respect to each other. This uniform velocity would be > 500 m/s.

With a velocity > 500m/s, a bottle of water is lethal.

Characters of molecular motors

- Time scale of inertia is short ~ ns
- Average velocity is ~ 1 μm, small in the absolute scale, large relative to the body size of molecular motors.
- Velocity fluctuations ≫ average velocity
- Velocity fluctuations ≫ average velocity.

Characters of macroscopic motors (e.g., vehicles)

- Time scale of inertia is long ~ s (or longer)
- Average velocity is ~ 10m/s (20 miles/h or higher).
- Velocity fluctuations ≪ average velocity.

Time scale of Smoluchowski-Kramers approximation

Recall that $V(t) \equiv \left(Y(t) - F_0 / b\right)$ is an Ornstein-Uhlenbeck process

Previously, for an Ornstein-Uhlenbeck process, we derived

$$V(t) = V(0)e^{-\beta t} + N\left(0,\ \frac{\gamma^2}{2\beta}(1-e^{-2\beta t})\right), \qquad \beta = \frac{b}{m}, \qquad \gamma = \frac{q}{m}$$

$$\int_0^t V(s)ds = \frac{(1-e^{-\beta t})}{\beta}V(0) + \underbrace{N\left(0,\ \left(\frac{\gamma}{\beta}\right)^2\left(t - \frac{2(1-e^{-\beta t})}{\beta} + \frac{(1-e^{-2\beta t})}{2\beta}\right)\right)}_{\text{containing } dW\text{'s in } [0,t]}$$

We write the particle position in terms of $t_0 \equiv 1/\beta$.

$$X(t) - X(0) = \int_0^t Y(s)ds = \frac{F_0}{b}t + \int_0^t V(s)ds\,, \qquad V(s) \equiv Y(s) - \frac{F_0}{b}$$

$$= \frac{F_0}{b}t + (1-e^{-t/t_0})t_0\left(Y(0) - \frac{F_0}{b}\right) + N\left(0,\ (\gamma t_0)^2 t_0\left(\frac{t}{t_0} - 2(1-e^{-t/t_0}) + \frac{(1-e^{-2t/t_0})}{2}\right)\right)$$

We examine the magnitudes of various terms as particle radius $a \to 0$.

$$m = \rho\frac{4\pi a^3}{3} = O(a^3)$$

$b = 6\pi\,\eta\,a = O(a)$

$$q = \sqrt{2k_B T b} = O(a^{1/2})$$

$$\beta = \frac{b}{m} = O(a^{-2}), \qquad t_0 = \frac{1}{\beta} = O(a^2)$$

$$\gamma = \frac{q}{m} = O(a^{-5/2})$$

$$\gamma^2 t_0 = O(a^{-3})$$ we need this quantity in the equilibrium of V.

$$Y(0) - \frac{F_0}{b} = N\left(0,\ \frac{\gamma^2 t_0}{2}\right) = O(\sqrt{\gamma^2 t_0}) = O(a^{-3/2})$$ based on equilibrium of $V(0)$

$$t_0\left(Y(0) - \frac{F_0}{b}\right) = O(a^{1/2})$$

$$(\gamma t_0)^2 = O(a^{-1}), \qquad (\gamma t_0)^2 t_0 = O(a)$$

For $t/t_0 \gg 1$, we have

$$X(t) - X(0) = \underbrace{\frac{F_0}{b}t}_{\text{Term III}} + \underbrace{t_0\left(Y(0) - \frac{F_0}{b}\right)}_{\text{Term I}} + \underbrace{N\left(0,\ (\gamma t_0)^2 t_0\left(\frac{t}{t_0}\right)\right)}_{\text{Term II}} + \text{T.S.T.} \qquad \text{(SK0)}$$

We compare Term I and Term II.

$$\text{Term I} = O(a^{1/2})$$
$$\text{Term II} = \sqrt{\frac{t}{t_0}}\sqrt{(\gamma t_0)^2 t_0} = \sqrt{\frac{t}{t_0}}\,O(a^{1/2}) >> O(a^{1/2})$$

For $t/t_0 \gg 1$, we neglect Term I, and keep Term II and Term III.

$$X(t) - X(0) = \frac{F_0}{b}t + N\left(0,\ (\gamma t_0)^2 t\right), \qquad (\gamma t_0)^2 = \left(\frac{q}{m}\cdot\frac{m}{b}\right)^2 = \frac{2k_B T}{b} = 2D$$

For $dt/t_0 \gg 1$ (i.e., on a "coarse" grid), $X(t)$ is governed by

$$dX = \frac{F_0}{b}dt + \sqrt{2D}\,dW \qquad \text{(SK1)}$$

This is the Smoluchowski-Kramers approximation (S-K approximation).

Several remarks on the S-K approximation

<u>The S-K approximation</u>

= neglecting Term I and keeping Term II and Term III.

Term I

Term I is the displacement attributed to the initial velocity.

$$(1-e^{-t/t_0})t_0\left(Y(0)-F_0/b\right) \longrightarrow t_0V(0) \qquad \text{for } t>>t_0$$

Here we study $(Y(0)–F_0/b)$, the deviation from the constant driving/drifting.

Based on $V(0) \sim N(0, k_BT/m)$, the RMS displacement due to inertia is

$$\text{RMS inertia displacement} = t_0\sqrt{E\left(V(0)^2\right)} = \frac{m}{b}\sqrt{\frac{k_BT}{m}} = \frac{\sqrt{mk_BT}}{b}$$

$$= \frac{\sqrt{\rho\frac{4}{3}\pi a^3 k_BT}}{6\pi\eta a} = \sqrt{a}\cdot\sqrt{\frac{\rho k_BT}{27\pi\eta^2}} = \sqrt{\frac{a}{[\text{nm}]}}\times 7\times 10^{-3}\,[\text{nm}] \propto a^{1/2}$$

Examples:

RMS inertia displacement = 0.16 nm for a 1μm bead

RMS inertia displacement = 0.016 nm for a 10nm bead

Term II vs Term I

Term II is the displacement due to diffusion.

For $t = O(t_0)$, the RMS displacement due to diffusion (Term II) is comparable to the RMS displacement due to inertia (Term I).

$$\text{RMS diffusion displacement } = \sqrt{2Dt_0} = \sqrt{2\frac{k_BT}{b}\cdot\frac{m}{b}}$$

$$= \sqrt{2}\frac{\sqrt{mk_BT}}{b} = \sqrt{2}\times(\text{RMS inertia displacement})$$

Time scale of the coarse grid in the S-K approximation

For $t \gg t_0$, Term II dominates Term I. Neglecting Term I is valid on a coarse grid of time scale $t \gg t_0$. Neglecting Term I is not valid on any fine grid of time scale $t \le t_0$.

The assumption of $F(x, t) \equiv F_0$

This assumption is valid when

$$\frac{\Delta F}{F} \equiv \frac{F\left(x+\Delta x,t+\Delta t\right)-F(x,t)}{F(x,t)} \approx \frac{\partial F}{\partial x}\cdot\frac{\Delta x}{F}+\frac{\partial F}{\partial t}\cdot\frac{\Delta t}{F} << 1 \qquad \text{for } t_{\text{scale}} >> \Delta t >> t_0$$

where $\Delta x \equiv x(t+\Delta t) - x(t)$, Δt is the time scale of the coarse grid in the S-K approximation, and t_{scale} is the time scale of physical evolution.

Term III vs Term II

Term III is the displacement due to external force.

For $t = O(t_0)$, Term III is much smaller than Term II (diffusion displacement) unless F_0 is extraordinarily large. We calculate how large F_0 needs to be, in order to make Term III comparable to Term II for $t = O(t_0)$.

$$\frac{F_0}{b}t_0 = \sqrt{2Dt_0} \qquad \Longrightarrow \qquad F_0 = \sqrt{\frac{2Db^2}{t_0}} = \sqrt{\frac{2k_B T\, b}{t_0}}$$

It is sensible to examine the force per mass.

$$\frac{F_0}{m} = \sqrt{\frac{2k_B T\, b}{m^2 t_0}} = \sqrt{\frac{2k_B T\, b^2}{m^3}} = \sqrt{\frac{2k_B T\,(6\pi\eta a)^2}{(\rho\pi a^3 4/3)^3}}\,, \qquad t_0 = \frac{m}{b}$$

For a small particle, this quantity is huge.

$$\frac{F_0}{m} = \begin{cases} = 7.2\times 10^{10}\,\dfrac{\text{Gravity}}{\text{Mass}} & \text{for a 10nm bead} \\ = 7.2\times 10^{3}\,\dfrac{\text{Gravity}}{\text{Mass}} & \text{for a 1}\mu\text{m bead} \end{cases}$$

Why do we neglect Term I but not Term III?

In real applications, Term III is much smaller than Term II for $t = O(t_0)$. However, Term III is proportional to the time while Term II is proportional to the square root of time. For $t \gg t_0$, eventually, Term III will catch up with Term II.

For $t \gg t_0$, Term I does not grow; it converges to a constant $O(a^{1/2})$ that is negligible in comparison with Term II.

Diffusion Models

1 The forward diffusion

1 SDE of the forward diffusion

Let $X(0)$ be a random sample from an underlying distribution that, in applications, is represented by a data set. Mathematically, we view $X(0)$ as a random variable (RV) with a continuous distribution. We evolve $X(t)$ stochastically in time by i) making it gradually forget the current position and ii) adding gradually independent noise to it. This process is called the forward diffusion and is achieved by the stochastic differential equation (SDE) below.

$$dX(t) = \underbrace{-\beta X(t)dt}_{\text{drift to 0}} + \underbrace{\sigma dW(t)}_{\text{diffusion}}, \tag{1}$$

$$dX(t) \equiv X(t+dt) - X(t), \quad dW(t) \sim N(0, dt)$$

2 Solution of the forward diffusion

(1) is an Ornstein-Uhlenbeck process. Previously obtained its solution

$$X(t) = X(0)e^{-\beta t} + \sqrt{\sigma^2 \frac{(1-e^{-2\beta t})}{2\beta}}\, Z, \qquad Z \sim N(0,1) \tag{2}$$

As $t \to \infty$, the original information of $X(0)$ is lost and $X(t)$ converges to a pure noise.

$$X(\infty) \equiv \lim_{t\to\infty} X(t) \sim N(0, \sigma_\infty^2), \qquad \sigma_\infty^2 \equiv \frac{\sigma^2}{2\beta} \tag{3}$$

3 Fokker-Planck equation of the forward diffusion

We consider a general SDE. Let $p(x,t)$ be the probability density of $X(t)$ at time t.

$$dX(t) = \underbrace{f(X(t))dt}_{\text{drift}} + \underbrace{\sigma dW(t)}_{\text{diffusion}}, \qquad dW(t) \sim N(0, dt) \tag{4}$$

For SDE (4), the probability density $p(x,t)$ is governed by the Fokker-Planck equation.

$$\begin{aligned}\frac{\partial p(x,t)}{\partial t} &= \underbrace{\nabla\cdot\Big(-f(x)p(x,t)\Big)}_{\text{drift}} + \underbrace{\frac{1}{2}\sigma^2\nabla^2 p(x,t)}_{\text{diffusion}} \\ &= \frac{\partial}{\partial x}\Big(-f(x)p(x,t)\Big) + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}p(x,t) \end{aligned} \tag{5}$$

In applications, $x \equiv \mathbf{x} \in \mathbb{R}^n$ is in a very high dimensional space (i.e., a 1024×1024 image has $2^{20}\approx 10^6$ pixels). In (5), we used the notations for the general case of vector $\mathbf{x}$.

$$\begin{aligned}&\nabla\cdot\Big(-f(x)p(x,t)\Big) \equiv \nabla\cdot\Big(-f(\mathbf{x})p(\mathbf{x},t)\Big),\\ &\nabla^2 p(x,t) \equiv \nabla\cdot\Big(\nabla p(\mathbf{x},t)\Big)\end{aligned}$$

The first line of (5) is for the general case of vector $\mathbf{x}$; the second line gives the equation for the case of one-dimensional x. Notice in particular the relation between the drift term in the SDE and the that in the Fokker-Planck equation.

$$dX(t) = f(X(t))dt + \sigma dW(t) \quad\Longleftrightarrow\quad \frac{\partial p(x,t)}{\partial t} = \nabla\cdot\Big(-f(x)p(x,t)\Big) + \frac{\sigma^2}{2}\nabla^2 p(x,t)$$

The relation allows us i) to write out the Fokker-Planck equation corresponding to a given SDE or ii) to write out the SDE corresponding to a given Fokker-Planck equation. We will use ii).

In SDE (1), $f(x) = -\beta x$. The Fokker-Planck equation and equilibrium $p_\infty(x)$ are

$$\frac{\partial p(x,t)}{\partial t} = \underbrace{\nabla\cdot\Big((\beta x)p(x,t)\Big)}_{\text{drift}} + \underbrace{\frac{1}{2}\sigma^2\nabla^2 p(x,t)}_{\text{diffusion}} \tag{6}$$

$$\begin{aligned}&0 = \nabla\cdot\Big((\beta x)p_\infty(x) + \frac{1}{2}\sigma^2\nabla p_\infty(x)\Big)\\ &\xrightarrow{\text{eq: zero flux everywhere}} \quad (\frac{2\beta}{\sigma^2}x)p_\infty(x) + \nabla p_\infty(x) = 0\\ &\xrightarrow{\text{integrating factor}} \quad \nabla\Big(\exp(\frac{\beta}{\sigma^2}\|x\|^2)p_\infty(x)\Big) = 0\\ &\longrightarrow \quad p_\infty(x) \propto \exp(\frac{-\|x\|^2}{2\sigma_\infty^2}), \qquad \sigma_\infty^2 \equiv \frac{\sigma^2}{2\beta}\end{aligned} \tag{7}$$

which simply confirms the equilibrium (3) we found in SDE (1).

2 The reverse diffusion

The objective of diffusion models is to sample the underlying distribution of $X(0)$. The basic idea of diffusion models is the following.

1. At a large time T, start with a random sample of pure noise: $X(T) \sim N(0,\sigma_\infty^2)$.

2. Evolve $X(t)$ backward in time to reverse the process of the forward diffusion. To make the time reversal well-posed, we the score function, which is learned from data.

3. When we reach time 0, we get a sample of the distribution of $X(0)$.

1 Time reversal of an SDE

We consider SDE (4): $dX(t) = f(X(t))dt + \sigma dW(t)$. Based on the SDE, for small dt, the forward time evolution from $X(t)$ to $X(t+dt)$ is approximately a Gaussian.

$$\Big(X(t+dt)\Big|X(t)=x_0\Big) \sim N\big(x_0 + f(x_0)dt,\ \sigma^2 dt\big) + o(dt)$$

Let $p(x,t) \equiv \rho_{X(t)}(x_0)$ be the probability density of X at time t. The transition density of the reverse time evolution from $X(t+dt) = x_1$ to $X(t)$ is given by Bayes theorem.

$$\begin{aligned}
&\rho_{(X(t)\big|X(t+dt)=x_1)}(x_0) \propto \rho_{(X(t+dt)\big|X(t)=x_0)}(x_1)\rho_{X(t)}(x_0) \\
&\quad\propto \exp\Big(\frac{-\big(x_1 - x_0 - f(x_0)dt\big)^2}{2\sigma^2 dt}\Big)p(x_0,t) \\
&\quad\propto \exp\Big(\frac{-(x_0-x_1)^2}{2\sigma^2 dt} - \frac{2(x_0-x_1)f(x_0) + f^2(x_0)dt}{2\sigma^2} + \ln p(x_0,t)\Big)
\end{aligned} \tag{8}$$

For small dt, the RHS of (8) is dominated by the factor $\exp\big(\frac{-(x_0-x_1)^2}{2\sigma^2 dt}\big)$, which is significant only in the region of $(x_0 - x_1) = O(\sqrt{dt})$. Note that density $\rho_{(X(t)\big|X(t+dt)=x_1)}(x_0)$ is a function of x_0 with x_1 as a parameter. In (8), we expand $f(x_0)$ and $\ln p(x_0,t)$ about x_1.

$$\begin{aligned}
&f(x_0) = f(x_1) + f'(x_1)(x_0 - x_1) + \cdots \\
&\ln p(x_0,t) = \ln p(x_1,t) + \frac{\partial \ln p(x,t)}{\partial x}\Big|_{x_1}(x_0-x_1) + \frac{1}{2}\frac{\partial^2 \ln p(x,t)}{\partial x^2}\Big|_{x_1}(x_0-x_1)^2 + \cdots
\end{aligned}$$

Using these expansions, we write the exponent in (8) as

$$\begin{aligned}
&\text{Exponent in (8)} = \Big(\frac{-(x_0-x_1)^2}{2\sigma^2 dt} - \frac{2(x_0-x_1)f(x_0) + f^2(x_0)dt}{2\sigma^2} + \ln p(x_0,t)\Big) \\
&= \frac{-1}{2\sigma^2 dt}\Big((x_0-x_1)^2 + 2f(x_1)dt(x_0-x_1) + 2f'(x_1)dt(x_0-x_1)^2 \\
&\quad - 2\sigma^2 dt\frac{\partial \ln p(x,t)}{\partial x}\Big|_{x_1}(x_0-x_1) - \sigma^2 dt\frac{\partial^2 \ln p(x,t)}{\partial x^2}\Big|_{x_1}(x_0-x_1)^2 + \cdots\Big) + const \\
&= \frac{-1}{2\sigma^2 dt}\Big((x_0-x_1)^2 + 2c_1 dt(x_0-x_1) + c_2 dt(x_0-x_1)^2 + \cdots\Big) + const
\end{aligned} \tag{9}$$

where coefficients c_1 and c_2 are

$$c_1 \equiv f(x_1) - \sigma^2\frac{\partial \ln p(x,t)}{\partial x}\Big|_{x_1}, \qquad c_2 \equiv 2f'(x_1) - \sigma^2\frac{\partial^2 \ln p(x,t)}{\partial x^2}\Big|_{x_1} \tag{10}$$

Note that there is some "inconsistency" between x and t in coefficients c_1 and c_2: $X(t+dt) = x_1$ is the position of X at time $(t+dt)$. We will address this issue shortly.

In the above, we neglected terms $(dt)^r(x_0 - x_1)^k$ for $r \geqslant 2$ or $k \geqslant 3$ with $r = 1$. We complete the square to write the quadratic form in (9) as (homework)

$$\text{Exponent in (8)} = \frac{-(1+c_2 dt)}{2\sigma^2 dt}\Big(x_0 - x_1 + \frac{c_1 dt}{1+c_2 dt}\Big)^2 + \cdots \tag{11}$$

(11) implies that the reverse time step $\big(X(t)\big|X(t+dt) = x_1\big)$ is also approximately a Gaussian.

$$\Big(X(t)\Big|X(t+dt) = x_1\Big) \sim N\Big(\frac{-c_1 dt}{1+c_2 dt}, \frac{\sigma^2 dt}{1+c_2 dt}\Big)$$

Using (10) and neglecting $o(dt)$, we write the drift and diffusion terms as

$$\frac{-c_1 dt}{1+c_2 dt} = -c_1 dt + o(dt) = \big(-f(x) + \sigma^2 \nabla \ln p(x, t+dt)\big)\Big|_{x_1} dt + o(dt)$$

$$\frac{\sigma^2 dt}{1+c_2 dt} = \sigma^2 dt + o(dt)$$

Note that in the expressions of drift and diffusion terms above, we have replaced $\ln p(x,t)$ with $\ln p(x, t+dt)$ to make the position and the time both at time $(t+dt)$. The starting point of the time reversal is $X(t+dt) = x_1$. The reverse time evolution from $X(t+dt)$ to $X(t)$ is

$$X(t) = X(t+dt) + \big(-f(x) + \sigma^2 \nabla \ln p(x, t+dt)\big)\Big|_{X(t+dt)} dt + \sigma\sqrt{dt}\, N(0,1)$$

We shift the time and write out the reverse time evolution from $X(t)$ to $X(t-dt)$.

$$X(t-dt) = X(t) + \big(-f(x) + \sigma^2 \nabla \ln p(x,t)\big)\Big|_{X(t)} dt + \sigma\sqrt{dt}\, N(0,1)$$

In SDE (1), $f(x) = -\beta x$. The reverse time evolution of SDE (1) is

$$X(t-dt) = X(t) + \big((\beta x) + \sigma^2 \nabla \ln p(x,t)\big)\Big|_{X(t)} dt + \sigma\sqrt{dt}\, N(0,1) \tag{12}$$

Here $\nabla \ln p(x,t)$ is called the score function. Let $\tau \equiv (T-t)$, $t = (T-\tau)$, and

$$Y(\tau) \equiv X(t) = X(T-\tau), \qquad q(x,\tau) \equiv p(x,t) = p(x, T-\tau)$$

The evolution from $\underbrace{Y(\tau) \text{ to } Y(\tau+d\tau)}_{\text{forward in } \tau}$ corresponds to that from $\underbrace{X(T-\tau) \text{ to } X(T-\tau-d\tau)}_{\text{backward in } t}$.

$$\boxed{dY(\tau) = \big((\beta x) + \sigma^2 \nabla \ln q(x,\tau)\big)\Big|_{Y(\tau)} d\tau + \sigma dW(\tau), \quad Y(T) \sim N(0, \sigma_\infty^2)} \tag{13}$$

2 Fokker-Planck equation for reversing the density

SDE (13) describes the reverse time evolution of stochastic process $X(t)$, which means paths $\{Y(\tau)\colon 0 \leqslant \tau \leqslant T\}$ of SDE (13) are statistically indistinguishable from the flipped forward diffusion paths $\{X(T-t)\colon 0 \leqslant t \leqslant T\}$ of SDE (1).

Recall the objective of diffusion models: drawing samples of the underlying density of $X(0)$. This objective can be achieved by SDE (13), which reverses the pure noise density of $X(T)$ at large T to the density of $X(0)$. However, for the goal of mapping pure noise to a desired density, it does not require the microscopic reverse time evolution of stochastic process $X(t)$. For example, to map $N(0,1)$ at $t=0$ to $N(1,e^{-1})$ at $t=1$, we can do it in many ways.

$$dX = -\beta(X-b)dt + \sqrt{a^2}dW, \quad b = \frac{1}{1-e^{-\beta}}, \quad a^2 = 2\beta\frac{(e^{-1}-e^{-2\beta})}{1-e^{-2\beta}}, \quad \text{for any } \beta > 1.$$

$$dX = \frac{-1}{2}\big(X - \frac{1}{1-e^{-1/2}}\big)dt, \quad \text{a deterministic ODE.}$$

$$dX = \frac{-1}{2}\big(X - t - 2\big)dt, \quad \text{a deterministic ODE.}$$

Fokker-Planck equation (6) evolves the density of $X(0)$ at $t=0$ forward in time to pure noise at large T. We explore reversing (6) in time to evolve the pure noise density at large T to the target density of $X(0)$ at $t=0$. The straightforward time reversal of a diffusion equation is ill-posed. We need to find a well-posed way to reverse (6) in time.

Let $\tau \equiv (T-t)$, $t=(T-\tau)$, and $q(x,\tau) \equiv p(x,t) = p(x,T-\tau)$. Note that evolving backward in t from T to 0 corresponds to evolving forward in τ from 0 to T. For $q(x,\tau)$, (6) becomes

$$\frac{\partial q(x,\tau)}{\partial \tau} = \nabla \cdot \Big(-(\beta x)q(x,\tau)\Big) - \frac{1}{2}\sigma^2\nabla^2 q(x,\tau) \tag{14}$$

This PDE is ill-posed for evolving $q(x,\tau)$ forward in τ because the coefficient of the diffusion term is negative. To make the PDE well-posed, we write the negative diffusion coefficient as the sum of a positive part and a larger negative part.

$$\begin{aligned}-\frac{1}{2}\sigma^2\nabla^2 q(x,\tau) &= \frac{\gamma^2}{2}\sigma^2\nabla^2 q(x,\tau) - \frac{(1+\gamma^2)}{2}\sigma^2\nabla\cdot\Big(\nabla q(x,\tau)\Big)\\ &= \frac{\gamma^2}{2}\sigma^2\nabla^2 q(x,\tau) - \frac{(1+\gamma^2)}{2}\sigma^2\nabla\cdot\Big((\nabla \ln q(x,\tau))q(x,\tau)\Big)\end{aligned}$$

We keep the positive part as the diffusion and move the negative part into the drift term with the help of a key component. We rewrite Fokker-Planck equation (14) as

$$\begin{aligned}\frac{\partial q(x,\tau)}{\partial \tau} &= \nabla\cdot\Big(-(\beta x)q(x,\tau) - \frac{(1+\gamma^2)}{2}\sigma^2(\nabla \ln q(x,\tau))q(x,\tau)\Big) + \frac{\gamma^2}{2}\sigma^2\nabla^2 q(x,\tau)\\ &= \nabla\cdot\Big(\big(-(\beta x) - \frac{(1+\gamma^2)}{2}\sigma^2\nabla\ln q(x,\tau)\big)q(x,\tau)\Big) + \frac{\gamma^2}{2}\sigma^2\nabla^2 q(x,\tau)\end{aligned} \tag{15}$$

(15) is well-posed for evolving $q(x,\tau)$ forward in τ (i.e., evolving $p(x,t)$ backward in t). The key component needed is the score function $\nabla \ln q(x,\tau)$. (15) describes a collection of Fokker-Planck equations, one for each value of $\gamma \geqslant 0$.

3 SDE for reversing the density

For each $\gamma \geqslant 0$, the SDE corresponding to Fokker-Planck equation (15) is

$$\boxed{dY(\tau) = \Big((\beta x) + \frac{(1+\gamma^2)}{2}\sigma^2 \nabla \ln q(x,\tau)\Big)\Big|_{Y(\tau)} d\tau + \gamma\sigma dW(\tau)} \tag{16}$$

For each $\gamma \geqslant 0$, SDE (16) evolves the pure noise density at large T to the target density of $X(0)$ at $t=0$. We make several comments on SDE (16).

- When $\gamma = 1$, SDE (16) becomes SDE (13), which describes the reverse time evolution of SDE (1). That is, paths $\{Y(\tau)\colon 0 \leqslant \tau \leqslant T\}$ of SDE (13) are statistically indistinguishable from the flipped forward diffusion paths $\{X(T-t)\colon 0 \leqslant t \leqslant T\}$ of SDE (1).
- For $\gamma \neq 1$, paths $\{Y(\tau)\colon 0 \leqslant \tau \leqslant T\}$ of SDE (16) are statistically different from the flipped forward diffusion paths $\{X(T-t)\colon 0 \leqslant t \leqslant T\}$ of SDE (1)
- If at $\tau = 0$ we start $Y(0)$ with an ensemble of samples of pure noise, then at $\tau = T$ the evolved ensemble by SDE (16) contains samples of the target distribution.
- The capability of SDE (16) to sample the target distribution is valid for every $\gamma \geqslant 0$, including $\gamma = 0$, which gives a deterministic ODE instead of an SDE.
- The capability of SDE (16) to sample the target distribution depends on that we know the score function $\nabla \ln q(x,\tau) \equiv \nabla \ln p(x, T-\tau)$.

3 Learning the score function $\nabla \ln p(x,t)$ from data

The objective of sampling from the target distribution of $X(0)$ is achieved by using SDE (16) to evolve the pure noise density to the target density. In SDE (16), we need the score function $\nabla \ln q(x,\tau) = \nabla \ln p(x, T-\tau)$, which is not immediately available even if the distribution $X(0)$ is given. In applications, the underlying distribution of $X(0)$ is represented in a set of data points of $X(0)$. We need to construct (learn) the score function from the given data set.

1 Conceptual framework for learning $\nabla \ln p(x,t)$ at time t

Recall that SDE (12) is the microscopic time reversal of SDE (1). Sample paths obtained $\{X(t)\}$ from the forward diffusion in SDE (1) are also sample paths of the reverse diffusion in SDE (12).

In principle, the score function $\nabla \ln p(x,t)$ can be estimated from these sample paths.

$$\begin{aligned}
\text{SDE (12):}\quad & X(t-dt) = X(t) + \big((\beta x) + \sigma^2 \nabla \ln p(x,t)\big)\Big|_{X(t)} dt + \sigma\sqrt{dt}\, N(0,1) \\
\Longrightarrow\quad & E\Big(X(t-dt)\Big|X(t)=x\Big) = x + x\beta dt + \sigma^2 \nabla \ln p(x,t) dt + o(dt) \\
\Longrightarrow\quad & \nabla \ln p(x,t) = E\Big(\frac{X(t-dt) - X(t)(1+\beta dt)}{\sigma^2 dt}\Big|X(t)=x\Big) + o(1) \qquad (17)
\end{aligned}$$

Although this will work in principle, there are two practical issues.

1) To estimate $\nabla \ln p(x,t)$ at x, we need a large number of sample paths $\{X(t)\}$ arriving at x at time t. These sample paths are needed in $E\big(\bullet\,\big|X(t)=x\big)$. While it is conceptually possible to obtain these sample paths, practically it is very difficult.

2) $\nabla \ln p(x,t)$ is a function of (x,t). How do we represent it in a computational form? especially when we work with a real data set with a finite number of points.

3) SDE (12) is for the limit of infinitesimal dt. At any finite dt, the finite difference version has a discretization error of $o(dt)$, leading to an error of $o(1)$ in the estimated $\nabla \ln p(x,t)$.

We first address issues 1) and 2) above. We rewrite (17) in a simpler form. Let $s(x)$ denote the unknown function we wish to determine. Let $(X(\omega), Y(\omega))$ be a pair of random variables satisfying i) the support of $X(\omega)$ is $(-\infty, +\infty)$, covering the support of the unknown function $s(x)$, and ii) the condition average of $Y(\omega)$ given $X(\omega)=x$ is $s(x)$.

$$s(x) = E\Big(Y\Big|X{=}x\Big) \qquad (18)$$

Note that for any random variable U, its mean μ_U satisfies

$$\begin{aligned}
& E\Big((U-\lambda)^2\Big) = E\Big((U-\mu_U)^2\Big) + (\lambda-\mu_U)^2 \geqslant E\Big((U-\mu_U)^2\Big) \\
& \mu_U = \arg\min_{\lambda} E\Big((U-\lambda)^2\Big)
\end{aligned}$$

At any x, applying this result to the conditional average in (18) yields

$$E\Big((Y-\lambda(X))^2\Big|X{=}x\Big) \geqslant E\Big((Y-s(X))^2\Big|X{=}x\Big) \quad \text{for any function } \lambda(x)$$

Using the law of total expectation, we obtain

$$\begin{aligned}
E\Big((Y-\lambda(X))^2\Big) &= E\Big(E\big((Y-\lambda(X))^2\big|X\big)\Big) \\
&\geqslant E\Big(E\big((Y-s(X))^2\big|X\big)\Big) = E\Big((Y-s(X))^2\Big) \quad \text{for any function } \lambda(x)
\end{aligned}$$

It follows that

$$\boxed{s(x) = \arg\min_{\{\lambda(x)\}} E\Big((Y-\lambda(X))^2\Big)} \qquad (19)$$

(19) is the key component in extracting function $s(x)$ from data of (X, Y) satisfying (18). It tells us that a function can be determined in a functional least squares problem.

In applications, random vector (X, Y) is represented by a data set of finite size:

$$\text{Data set of } (X, Y): \quad D = \left\{(X^{(j)}, Y^{(j)})\right\} \tag{20}$$

Accordingly, function $\lambda(x)$ in (19) is represented by a neural network

$$\text{Neural network representation: } \lambda(x) = \lambda(x;\, \theta)$$

Functional least squares problem (19) becomes

$$\boxed{\begin{cases} \theta^{(opt)} = \arg\min\limits_{\theta} \sum\limits_{j} \left(Y^{(j)} - \lambda(X^{(j)};\, \theta)\right)^2 \\ s(x) = \lambda(x;\, \theta^{(opt)}) \end{cases}} \tag{21}$$

2 Precise formulation for learning $\nabla \ln p(x, t)$ at time t

We now address issue 3): the discretization error for finite dt. Recall that SDE (1), as an Ornstein Uhlenbeck process, has an analytical solution. As a result, sample paths $\{X(t)\}$ in the forward diffusion can be generated exactly with no discretization error. With starting point $X(0)$, the forward diffusion is given in (2). We rewrite is as

$$X(t) = X(0)e^{-\beta t} + \sigma_t Z, \quad Z \sim N(0, 1), \quad \sigma_t^2 \equiv \sigma^2 \frac{(1 - e^{-2\beta t})}{2\beta} \tag{22}$$

We use this analytical solution to derive an exact formulation for extracting the score function, which has no discretization error. In the forward diffusion, $X(0)$ is from the underlying distribution and noise Z is independent of $X(0)$. Together $X(0)$ and Z determine $X(t)$. To facilitate the discussion, We introduce two conditional probability densities. Let

- $p(x_t|x_0) \equiv p\big((x_t, t)|(x_0, 0)\big)$ be the conditional density of $X(t)$ given $X(0) = x_0$;
- $p(x_0|x_t) \equiv p\big((x_0, 0)|(x_t, t)\big)$ be the conditional density of $X(0)$ given $X(t) = x_t$.

The forward diffusion solution (22) gives

$$\begin{aligned} &\big(X(t)|X(0) = x_0\big) \sim N(x_0 e^{-\beta t}, \sigma_t^2), \qquad p(x_t|x_0) \propto \exp\Big(\frac{-(x_t - x_0 e^{-\beta t})^2}{2\sigma_t^2}\Big) \\ &\nabla_{x_t} \ln p(x_t|x_0) = \frac{\nabla_{x_t} p(x_t|x_0)}{p(x_t|x_0)} = \frac{-1}{\sigma_t^2}(x_t - x_0 e^{-\beta t}) \end{aligned} \tag{23}$$

The essence of reverse diffusion is denoising. Conditional on $X(t) = x_t$, the added Z that produces $X(t)$ in (22) is no longer Gaussian. We derive the conditional average of Z given $X(t) = x_t$.

$$E\Big(\sigma_t Z \Big| X_t = x_t\Big) = E\Big((X(t) - X(0)e^{-\beta t})\Big| X(t) = x_t\Big) = \int (x_t - x_0 e^{-\beta t}) p(x_0|x_t) dx_0 \tag{24}$$

We use the expression of $\nabla_{x_t} p(x_t|x_0)$ in (23) to rewrite it as

$$E\Big(\sigma_t Z \Big| X_t = x_t\Big) = -\sigma_t^2 \int \nabla_{x_t} p(x_t|x_0) \frac{p(x_0|x_t)}{p(x_t|x_0)} dx_0 \tag{25}$$

We apply Bayes' theorem.

$$\underbrace{\frac{p(x_0|x_t)}{p(x_t|x_0)} = \frac{p(x_0)}{p(x_t)}}_{\text{Bayes' theorem}}, \qquad E\Big(\sigma_t Z \Big| X_t = x_t\Big) = -\sigma_t^2 \int \nabla_{x_t} p(x_t|x_0) \frac{p(x_0)}{p(x_t)} dx_0 \tag{26}$$

where $p(x_0)$ and $p(x_t)$ are short notations for $p(x_0) \equiv p(x_0, 0)$ and $p(x_t) \equiv p(x_t, t)$. Note that in (26), the integration variable is x_0. We move $p(x_t)$ and ∇_{x_t} out of the integral.

$$E\Big(\sigma_t Z \Big| X_t = x_t\Big) = -\sigma_t^2 \frac{1}{p(x_t)} \nabla_{x_t} \Big(\underbrace{\int p(x_t|x_0) p(x_0) dx_0}_{=p(x_t)} \Big)$$

$$= -\sigma_t^2 \frac{\nabla_{x_t} p(x_t)}{p(x_t)} = -\sigma_t^2 \nabla_{x_t} \ln p(x_t) = \boxed{-\sigma_t^2 \nabla_{x_t} \ln p(x_t, t)} \tag{27}$$

We use $X(t) = X(0)e^{-\beta t} + \sigma_t Z$, and we denote x_t simply as x.

$$\boxed{\begin{aligned} &\nabla \ln p(x,t) = E\Big(\frac{-Z}{\sigma_t} \Big| X(0)e^{-\beta t} + \sigma_t Z = x\Big), \\ &\sigma_t^2 \equiv \sigma^2 \frac{(1-e^{-2\beta t})}{2\beta}, \qquad Z \sim N(0,1) \end{aligned}} \tag{28}$$

(28) is an improvement over (17) in two aspects: i) (28) is exact with no discretization error, and ii) (28) is valid exactly for any $t > 0$.

We follow the approach outlined in (19) and (21) to formulate the task of extracting $\nabla \ln p(x,t)$ as that of training a neural network. To apply the result of (19), we identify

$$X = X(0)e^{-\beta t} + \sigma_t Z, \qquad Y = \frac{-Z}{\sigma_t}$$

Based on (19), the functional least squares problem for $\nabla \ln p(x,t)$ is

$$\boxed{\nabla \ln p(x,t) = \underset{\{\lambda(x)\}}{\arg\min}\, E\Big(\Big[\frac{Z}{\sigma_t} + \lambda\big(X(0)e^{-\beta t} + \sigma_t Z\big)\Big]^2\Big), \quad Z \sim N(0,1)} \tag{29}$$

Random vector $\big(X(0), Z\big)$ is represented by the data set below.

$$\text{Data set:} \quad D = \big\{(X_0^{(i)}, Z^{(i,j)})\big\}, \quad Z^{(i,j)} \overset{\text{i.i.d.}}{\sim} N(0,1) \tag{30}$$

In the data set, for each sample $X_0^{(i)}$ of $X(0)$, there are many independent realizations $Z^{(i,j)}$ of Z, corresponding to independent realizations of $X(t)$: $X_t^{(i,j)} = X_0^{(i)} e^{-\beta t} + \sigma_t Z^{(i,j)}$.

$\nabla \ln p(x,t)$ at time t is represented by a trained neural network.

$$\boxed{\begin{cases} \theta^{(opt)} = \arg\min\limits_{\theta} \sum\limits_{i,j} \Big[\dfrac{Z^{(i,j)}}{\sigma_t} + \lambda\big(X_0^{(i)} e^{-\beta t} + \sigma_t Z^{(i,j)};\, \theta\big) \Big]^2, \\ \qquad\qquad Z^{(i,j)} \overset{\text{i.i.d.}}{\sim} N(0,1), \quad \text{neural network: } \lambda(x;\, \theta) \\ \nabla \ln p(x,t) = \lambda(x;\, \theta^{(opt)}) \end{cases}} \tag{31}$$

3 Formulation for learning $\nabla \ln p(x,t)$ as a function of (x,t)

Let $\{X_0^{(i)}\}$ be a set of data points from the underlying distribution of $X(0)$. In applications, the underlying distribution is represented by data set $\{X_0^{(i)}\}$. The objective of diffusion models is to draw new samples from the underlying distribution.

In the time direction, we use a sequence of time instances $\{t_k\}$ to cover $[0,T]$. Random vector $\big(X(0), Z\big)$ used in $X(t) = X(0)e^{-\beta t} + \sigma_t Z,\ t \in [0,T]$ is represented by the data set below.

$$\text{Data set:} \quad D = \big\{(X_0^{(i)}, Z^{(i,j,t_k)})\big\}, \quad Z^{(i,j,t_k)} \overset{\text{i.i.d.}}{\sim} N(0,1) \tag{32}$$

In the data set, for each sample $X_0^{(i)}$ of $X(0)$, there are many independent realizations $Z^{(i,j,t_k)}$ of Z, corresponding to many independent realizations of $X(t_k)$ at many t_k.

$$X_{t_k}^{(i,j)} = X_0^{(i)} e^{-\beta t_k} + \sigma_{t_k} Z^{(i,j,t_k)}$$

To model functions of (x,t), we adopt a neural network of the form $\lambda(x,t;\, \theta)$. The score function $\nabla \ln p(x,t)$ as a function of (x,t) is represented by a trained neural network.

$$\boxed{\begin{cases} \theta^{(opt)} = \arg\min\limits_{\theta} \sum\limits_{i,j,t_k} \Big[\dfrac{Z^{(i,j,t_k)}}{\sigma_{t_k}} + \lambda\big(X_0^{(i)} e^{-\beta t_k} + \sigma_{t_k} Z^{(i,j,t_k)}, t_k;\, \theta\big) \Big]^2, \\ \qquad\qquad Z^{(i,j,t_k)} \overset{\text{i.i.d.}}{\sim} N(0,1), \quad \text{neural network: } \lambda(x,t;\, \theta) \\ \nabla \ln p(x,t) = \lambda(x,t;\, \theta^{(opt)}) \end{cases}} \tag{33}$$

AM216 Stochastic Differential Equations

Lecture 20

List of topics in this lecture

- A mathematical derivation of Smoluchowski-Kramers approximation
- Scaling of X, Y and $F(X, t)$
- The method of solvability condition

Recap

Smoluchowski-Kramers approximation

Langevin equation for a small particle in water

$$dX = Y\,dt$$

$$m dY = -bY\,dt + F(X,t)dt + \sqrt{2k_B T b}\,dW$$

where X : position; Y : velocity;

a : radius of the particle;

$m = 4\pi/3\ a^3 = O(a^3)$: mass of the particle

$b = 6\pi\ \eta\ a = O(a)$: drag coefficient of the particle

Time scale of inertia and thermal excitation: $t_0 = m/b = O(a^2)$

S-K approximation:

For $t_{\text{scale}} \gg t \gg t_0$, $X(t)$ satisfies

$$X(t) - X(0) = \underbrace{\frac{F_0}{b}t}_{\text{Term III}} + \underbrace{t_0\left(Y(0) - \frac{F_0}{b}\right)}_{\text{Term I}} + \underbrace{N\left(0,\ (\gamma t_0)^2 t_0\left(\frac{t}{t_0}\right)\right)}_{\text{Term II}} + \text{T.S.T.} \qquad \text{(SK0)}$$

where t_{scale} is the time scale of physical evolution. We neglect Term I, and keep Term II and Term III. On a "coarse" grid, $X(t)$ is governed by

$$dX = \frac{F(X,t)}{b}dt + \sqrt{2D}\,dW$$

A mathematical derivation of Smoluchowski-Kramers approximation

Let $\varepsilon \equiv \sqrt{t_0 / t_{\text{scale}}} = O(a/R)$ be a small parameter.

Here t_{scale} is the time scale of physical evolution and R is a size scale whose time scale of inertia is t_{scale}. We assume that the time is measured in units of t_{scale}, and the particle position/size is measured in units of R. We write $O(a/R)$ simply as $O(a)$.

Scaling of position X

The over-damped Langevin equation for X is

$$dX = \frac{F(X,t)}{b}dt + \sqrt{2D}\,dW\,, \qquad D = O(a^{-1}) = O(\varepsilon^{-1})$$

This equation is not free of ε. When $F = 0$ and $t = O(1)$, we have

$$X(1) - X(0) \sim \sqrt{2D} = O(\varepsilon^{-1/2})$$

We select a scaling for X to make it $O(1)$.

$$\hat{X} = \sqrt{\varepsilon} \qquad \Rightarrow \qquad \hat{X}(1) - \hat{X}(0) = O(1)$$

The over-damped Langevin equation after scaling X

We consider the case where for $t = O(1)$, the effect of external driving force is comparable to that of the diffusion

$$\frac{F(X,t)}{b} = O(\sqrt{D}) = O(\varepsilon^{-1/2})\,, \qquad b = O(a) = O(\varepsilon)$$

$$==> \qquad F(X,t) = O(\sqrt{\varepsilon})$$

After scaling X, we have

$$\sqrt{\varepsilon}\,dX = \sqrt{\varepsilon}\frac{F(X,t)}{b}dt + \sqrt{\varepsilon}\sqrt{2D}\,dW$$

$$==> \qquad \boxed{d\hat{X} = f(\hat{X},t)dt + \sqrt{2D_0}\,dW}$$

where

$$D_0 \equiv \varepsilon D = O(1)\,, \qquad f(\hat{X},t) \equiv \sqrt{\varepsilon}\frac{F(X,t)}{b} = O(1)$$

Note: after scaling X, the over-damped Langevin equation is free of ε.

Scaling of velocity Y

We scale Y separately from the scaling of X. Here Y is the instantaneous velocity, dominated by thermal excitations, much larger than $X(1)$–$X(0)$.

When $F = 0$, the equipartition of energy gives us

$$\sqrt{E(Y^2)} = \sqrt{\frac{k_B T}{m}} = O(\varepsilon^{-3/2}), \qquad m = O(a^3) = O(\varepsilon^3)$$

We select a scaling for Y to make it O(1)

$$\hat{Y} = \varepsilon^{3/2} Y \qquad \Rightarrow \qquad \sqrt{E(\hat{Y}^2)} = O(1)$$

<u>The FULL Langevin equation after scaling</u> of Y

$$m dY = -bY\,dt + F(X,t)dt + b\sqrt{2D}\,dW$$

==> $$\frac{m}{b} dY = -Y\,dt + \frac{F(X,t)}{b} dt + \sqrt{2D}\,dW$$

==> $$\varepsilon^2(\varepsilon^{-3/2} d\hat{Y}) = -(\varepsilon^{-3/2}\hat{Y})dt + \varepsilon^{-1/2} f(\hat{X},t)dt + \sqrt{2(\varepsilon^{-1} D_0)}\,dW$$

==> $$\varepsilon^{1/2} d\hat{Y} = -\varepsilon^{-3/2}\hat{Y}\,dt + \varepsilon^{-1/2} f(\hat{X},t)dt + \varepsilon^{-1/2}\sqrt{2D_0}\,dW$$

==> $$\boxed{d\hat{Y} = \frac{-1}{\varepsilon^2}\hat{Y}\,dt + \frac{1}{\varepsilon} f(\hat{X},t)dt + \frac{1}{\varepsilon}\sqrt{2D_0}\,dW}$$

Since X and Y are scaled <u>separately</u>, after scaling they are related by

$$dX = Y\,dt \qquad \Rightarrow \qquad d(\varepsilon^{1/2} X) = \varepsilon^{-1}(\varepsilon^{3/2} Y)dt \qquad \Rightarrow \qquad d\hat{X} = \varepsilon^{-1}\hat{Y}\,dt$$

<u>Notation:</u>

<u>After scaling</u>, we recycle the simple notation of (X, Y, t) to write out the starting SDE and the end SDE of the S-K approximation.

<u>The starting SDE</u> of S-K approximation

$$\begin{aligned} dX &= \frac{1}{\varepsilon} Y\,dt \\ dY &= \frac{-1}{\varepsilon^2} Y\,dt + \frac{1}{\varepsilon} f(X,t)dt + \frac{1}{\varepsilon}\sqrt{2D_0}\,dW \end{aligned} \qquad \text{(S-1)}$$

<u>The end SDE</u> of S-K approximation

$$dX = f(X,t)dt + \sqrt{2D_0}\,dW \qquad \text{(S-2)}$$

<u>Backward equation</u> for the starting SDE

For the starting SDE (S-1), we consider function

$$u(x,y,t) \equiv \Pr\left\{ X \text{ having crossed } x_B \text{ by time } t \,\middle|\, X(0) = x, Y(0) = y \right\}$$

We derive the governing equation of $u(x, y, t)$. For dt small enough, we have

$$u(x,y,t)=E\Big(u(x+dX,y+dY,t-dt)\Big)+o(dt) \qquad \text{(E-1)}$$

Note that in the SDE, ε = fixed. As $dt\to 0$, we have $dX=O(dt)$ and $dY=O(\sqrt{dt})$.

Specifically we have the moments of dX and dY from the SDE.

$$\Big(dX\Big|X(0)=x,Y(0)=y\Big)=\frac{y}{\varepsilon}dt \qquad \text{not a random variable}$$

$$E\Big(dY\Big|X(0)=x,Y(0)=y\Big)=\left(\frac{-1}{\varepsilon^2}y+\frac{1}{\varepsilon}f(x,t)\right)dt$$

$$E\Big((dY)^2\Big|X(0)=x,Y(0)=y\Big)=\frac{1}{\varepsilon^2}2D_0dt+o(dt)$$

$$E\Big((dY)^n\Big|X(0)=x,Y(0)=y\Big)=o(dt) \quad \text{for } n\ge 3$$

We expand (E-1)

$$u(x,y,t)=E\Big(u(x+dX,y+dY,t-dt)\Big)+o(dt)$$

$$=E\left(u(x,y,t)+u_x dX+u_y dY+\frac{1}{2}u_{yy}(dY)^2-u_t dt\right)+o(dt)$$

$$=u+u_x\frac{y}{\varepsilon}dt+u_y\left(\frac{-1}{\varepsilon^2}y+\frac{1}{\varepsilon}f(x,t)\right)dt+\frac{1}{2}u_{yy}\frac{1}{\varepsilon^2}2D_0dt-u_t dt+o(dt)$$

…

The governing equation for $u(x, y, t)$ is the backward equation of (S-1).

$$\varepsilon^2 u_t=-yu_y+D_0u_{yy}+\varepsilon\Big(yu_x+f(x,t)u_y\Big) \qquad \text{(BE-1)}$$

Backward equation for the end SDE

For the end SDE (S-2), we consider function

$$w(x,t)\equiv \Pr\Big\{X \text{ having crossed } x_B \text{ by time } t\Big|X(0)=x\Big\}$$

The governing equation for $w(x, t)$ is the backward equation of (S-2).

$$\frac{\partial w}{\partial t}=D_0\frac{\partial^2 w}{\partial x}+f(x,t)\frac{\partial w}{\partial x} \qquad \text{(BE-2)}$$

Convergence of $u(x, y, t)$ to $w(x, t)$

For backward equation (BE-1), we seek solutions of the form

$$u(x,y,t)=u_0(x,y,t)+\varepsilon u_1(x,y,t)+\varepsilon^2 u_2(x,y,t)+\cdots$$

We substitute the expansion into (BE-1) and keep terms up to O(ε^2).

$$\varepsilon^2\frac{\partial u_0}{\partial t}=-y\left(\frac{\partial u_0}{\partial y}+\varepsilon\frac{\partial u_1}{\partial y}+\varepsilon^2\frac{\partial u_2}{\partial y}\right)+D_0\left(\frac{\partial^2 u_0}{\partial y^2}+\varepsilon\frac{\partial^2 u_1}{\partial y^2}+\varepsilon^2\frac{\partial^2 u_2}{\partial y^2}\right)$$
$$+\varepsilon\left(y\left(\frac{\partial u_0}{\partial x}+\varepsilon\frac{\partial u_1}{\partial x}\right)+f(x,t)\left(\frac{\partial u_0}{\partial y}+\varepsilon\frac{\partial u_1}{\partial y}\right)\right)+o(\varepsilon^2)$$

$$\Longrightarrow\quad 0=\underbrace{\left(-y\frac{\partial u_0}{\partial y}+D_0\frac{\partial^2 u_0}{\partial y^2}\right)}_{O(1)\text{ terms}}+\varepsilon\underbrace{\left(-y\frac{\partial u_1}{\partial y}+D_0\frac{\partial^2 u_1}{\partial y^2}+y\frac{\partial u_0}{\partial x}+f(x,t)\frac{\partial u_0}{\partial y}\right)}_{O(\varepsilon)\text{ terms}}$$
$$+\underbrace{\varepsilon^2\left(-y\frac{\partial u_2}{\partial y}+D_0\frac{\partial^2 u_2}{\partial y^2}+y\frac{\partial u_1}{\partial x}+f(x,t)\frac{\partial u_1}{\partial y}-\frac{\partial u_0}{\partial t}\right)}_{O(\varepsilon^2)\text{ terms}}+o(\varepsilon^2)$$

Below, we show that

- $u_0(x, y, t) = u_0(x, t)$, independent of y.
- $u_0(x, t)$ satisfies backward equation (BE-2).

Step A: $u_0(x, y, t)$ is independent of y.

The balance of O(1) terms gives the equation

$$-y\frac{\partial u_0}{\partial y}+D_0\frac{\partial^2 u_0}{\partial y^2}=0$$

Multiplying by the integrating factor yields

$$\exp\left(\frac{-y^2}{2D_0}\right)\left(\frac{-y}{D_0}\right)\frac{\partial u_0}{\partial y}+\exp\left(\frac{-y^2}{2D_0}\right)\frac{\partial^2 u_0}{\partial y^2}=0$$

$$\Longrightarrow\quad \frac{\partial}{\partial y}\left[\exp\left(\frac{-y^2}{2D_0}\right)\frac{\partial u_0}{\partial y}\right]=0$$

$$\Longrightarrow\quad \exp\left(\frac{-y^2}{2D_0}\right)\frac{\partial u_0}{\partial y}=c_1(x,t)\quad\text{independent of } y$$

==> $\quad \dfrac{\partial u_0}{\partial y} = c_1(x,t)\exp\left(\dfrac{y^2}{2D_0}\right)$

==> $\quad u_0 = c_0(x,t) + c_1(x,t)\int_0^y \exp\left(\dfrac{z^2}{2D_0}\right)dz$

Recall that $u(x, y, t)$ = probability.

==> $\quad u(x, y, t)$ is bounded as $y \to \infty$.

==> $\quad u_0(x, y, t)$, as the leading term, is bounded as $y \to \infty$

==> $\quad c_1(x,t) = 0$

==> $\quad u_0 = c_0(x,t)$ independent of y

Step B: $\quad u_0(x, t)$ satisfies backward equation (BE-2)

Step B1: $\quad$ The balance of O(ε) terms gives

$$-y\frac{\partial u_1}{\partial y} + D_0\frac{\partial^2 u_1}{\partial y^2} + y\frac{\partial u_0}{\partial x} + f(x,t)\frac{\partial u_0}{\partial y} = 0$$

We write it as an equation for u_1 and multiply by the integrating factor

$$-y\frac{\partial u_1}{\partial y} + D_0\frac{\partial^2 u_1}{\partial y^2} = -y\frac{\partial u_0}{\partial x}$$

==> $\quad \exp\left(\dfrac{-y^2}{2D_0}\right)\left(\dfrac{-y}{D_0}\right)\dfrac{\partial u_1}{\partial y} + \exp\left(\dfrac{-y^2}{2D_0}\right)\dfrac{\partial^2 u_1}{\partial y^2} = \exp\left(\dfrac{-y^2}{2D_0}\right)\left(\dfrac{-y}{D_0}\right)\dfrac{\partial u_0}{\partial x}$

==> $\quad \dfrac{\partial}{\partial y}\left(\exp\left(\dfrac{-y^2}{2D_0}\right)\dfrac{\partial u_1}{\partial y}\right) = \dfrac{\partial}{\partial y}\left(\exp\left(\dfrac{-y^2}{2D_0}\right)\right)\dfrac{\partial u_0}{\partial x}$

Noticing that $\partial u_0/\partial x$ is independent of y and integrating in y, we get

$$\exp\left(\frac{-y^2}{2D_0}\right)\frac{\partial u_1}{\partial y} = \exp\left(\frac{-y^2}{2D_0}\right)\frac{\partial u_0}{\partial x} + c_3(x,t)$$

==> $\quad \dfrac{\partial u_1}{\partial y} = \dfrac{\partial u_0}{\partial x} + c_3(x,t)\exp\left(\dfrac{y^2}{2D_0}\right)$

==> $\quad u_1 = y\dfrac{\partial u_0}{\partial x} + c_2(x,t) + c_3(x,t)\int_0^y \exp\left(\dfrac{z^2}{2D_0}\right)dz$

u_1 is solvable without imposing any constraint on u_0.

To find a constraint on u_0, we need to examine $O(\varepsilon^2)$ terms.

Step B2: The balance of $O(\varepsilon^2)$ terms gives

$$-y\frac{\partial u_2}{\partial y}+D_0\frac{\partial^2 u_2}{\partial y^2}+y\frac{\partial u_1}{\partial x}+f(x,t)\frac{\partial u_1}{\partial y}-\frac{\partial u_0}{\partial t}=0$$

==> $$-y\frac{\partial u_2}{\partial y}+D_0\frac{\partial^2 u_2}{\partial y^2}=\frac{\partial u_0}{\partial t}-y\frac{\partial u_1}{\partial x}-f(x,t)\frac{\partial u_1}{\partial y}$$

==> $$L_1[u_2]=\frac{\partial u_0}{\partial t}+L_2[u_1]$$

where operators L_1 and L_2 are defined as

$$L_1[\cdot]\equiv -y\frac{\partial \cdot}{\partial y}+D_0\frac{\partial^2 \cdot}{\partial y^2}$$

$$L_2[\cdot]\equiv -y\frac{\partial \cdot}{\partial x}-f(x,t)\frac{\partial \cdot}{\partial y}$$

Theorem (solvability condition)

Equation $L[u] = g$ is solvable if and only if $\langle g, v\rangle = 0$ for all v satisfying $L^*[v] = 0$ (with $v(y) \to 0$ rapidly as $|y| \to \infty$)

A simple demonstration of the theorem:

Let A be an $m\times n$ matrix.

$A\,u = b$ is solvable if and only if $b \in \mathrm{Col}(A)$

if and only if $\langle b, v\rangle$ for all $v \in \mathrm{Col}(A)^{\perp} = \mathrm{Nul}(A^{\mathrm{T}})$.

if and only if $\langle b, v\rangle$ for all v satisfying $A^{\mathrm{T}}v = 0$.

Important note:

When operator L_1 is in variable y, we view $v(y)$ as a function of y and the inner-product $\langle g, v\rangle$ is an integral with respect to y.

Step B3: Solvability of $L_1[u_2]=\dfrac{\partial u_0}{\partial t}+L_2[u_1]$.

u_2 is solvable if and only if $\left\langle \dfrac{\partial u_0}{\partial t}+L_2[u_1], v\right\rangle=0$ for all v satisfying $L_1^*[v] = 0$.

Step B4: Solution of $L_1^*[v] = 0$

$$L_1[\cdot] = -y\frac{\partial \cdot}{\partial y} + D_0\frac{\partial^2 \cdot}{\partial y^2} \quad ==> \quad L_1^*[\cdot] = \frac{\partial (y\,\cdot)}{\partial y} + D_0\frac{\partial^2 \cdot}{\partial y^2}$$

$L_1^*[v] = 0$ yields

$$\frac{\partial (y\,v)}{\partial y} + D_0\frac{\partial^2 v}{\partial y^2} = 0$$

$$==> \quad \frac{y}{D_0}v + \frac{\partial v}{\partial y} = d_1$$

$$==> \quad \exp\left(\frac{y^2}{2D_0}\right)\frac{y}{D_0}v + \exp\left(\frac{y^2}{2D_0}\right)\frac{\partial v}{\partial y} = d_1\exp\left(\frac{y^2}{2D_0}\right)$$

$$==> \quad \frac{\partial}{\partial y}\left[\exp\left(\frac{y^2}{2D_0}\right)v\right] = d_1\exp\left(\frac{y^2}{2D_0}\right)$$

$$==> \quad \exp\left(\frac{y^2}{2D_0}\right)v(y) = d_1\int_0^y \exp\left(\frac{z^2}{2D_0}\right)dz + d_0$$

$$==> \quad v(y) = \underbrace{d_1\exp\left(\frac{-y^2}{2D_0}\right)\int_0^y \exp\left(\frac{z^2}{2D_0}\right)dz}_{\text{Not decaying to zero rapidly as } y\to\infty} + d_0\exp\left(\frac{-y^2}{2D_0}\right)$$

Anyway, we select $\quad v(y) = d_0\exp\left(\frac{-y^2}{2D_0}\right)$

Step B5: Back to solvability of $L_1[u_2] = \frac{\partial u_0}{\partial t} + L_2[u_1]$.

We set d_0 to make $v(y)$ a normal density.

$$v(y) = \frac{1}{\sqrt{2\pi D_0}}\exp\left(\frac{-y^2}{2D_0}\right) = \rho_{N(0,D_0)}(y)$$

u_2 is solvable if and only if $\int_{-\infty}^{+\infty}\left(\frac{\partial u_0}{\partial t} + L_2[u_1]\right)\rho_{N(0,D_0)}(y)dy = 0$.

Notice that $\partial u_0/\partial t$ is independent of y. The solvability condition gives us

$$\frac{\partial u_0}{\partial t}+\int_{-\infty}^{+\infty} L_2[u_1]\rho_{N(0,D_0)}(y)dy=0 \qquad \text{(Cond-1)}$$

<u>Step B6:</u> Calculation of $\int_{-\infty}^{+\infty} L_2[u_1]\rho_{N(0,D_0)}(y)dy$

Recall that $u_1=y\frac{\partial u_0}{\partial x}+c_2(x,t)+c_3(x,t)\int_0^y \exp\left(\frac{z^2}{2D_0}\right)dz$. We calculate $L_2[u_1]$.

$$L_2[u_1]=-y\frac{\partial u_1}{\partial x}-f(x,t)\frac{\partial u_1}{\partial y}=\underbrace{-y^2\frac{\partial^2 u_0}{\partial x^2}-f(x,t)\frac{\partial u_0}{\partial x}}_{\text{effect of } u_0}\underbrace{-y\frac{\partial c_2}{\partial x}}_{\text{effect of } c_2}$$

$$\underbrace{-y\frac{\partial c_3}{\partial x}\int_0^y \exp\left(\frac{z^2}{2D_0}\right)dz-f(x,t)c_3\exp\left(\frac{y^2}{2D_0}\right)}_{\text{effect of } c_3}$$

We examine the inner product of each part with $\rho_{N(0,D)}(y)$.

$u_0(x,t)$ is independent of y.

$$\int_{-\infty}^{+\infty}\underbrace{\left[-y^2\frac{\partial^2 u_0}{\partial x^2}-f(x,t)\frac{\partial u_0}{\partial x}\right]}_{\text{effect of } u_0}\rho_{N(0,D_0)}(y)dy=-D_0\frac{\partial^2 u_0}{\partial x^2}-f(x,t)\frac{\partial u_0}{\partial x}$$

$c_2(x,t)$ is independent of y.

$$\int_{-\infty}^{+\infty}\underbrace{-y\frac{\partial c_2}{\partial x}}_{\text{effect of } c_2}\rho_{N(0,D_0)}(y)dy=0$$

$c_3(x,t)$ is independent of y.

$$\int_{-\infty}^{+\infty}\underbrace{\left[-y\frac{\partial c_3}{\partial x}\int_0^y \exp\left(\frac{z^2}{2D_0}\right)dz-f(x,t)c_3\exp\left(\frac{y^2}{2D_0}\right)\right]}_{\text{effect of } c_3}\rho_{N(0,D_0)}(y)dy=\infty \quad \text{if } c_3\neq 0.$$

To make $\int_{-\infty}^{+\infty} L_2[u_1]\rho_{N(0,D_0)}(y)dy$ finite, we must have $c_3(x,t)\equiv 0$.

Thus, we obtain the expression

$$\int_{-\infty}^{+\infty} L_2[u_1]\rho_{N(0,D_0)}(y)dy=-D_0\frac{\partial^2 u_0}{\partial x^2}-f(x,t)\frac{\partial u_0}{\partial x} \qquad \text{(Res-1)}$$

Step B7: Governing equation for u_0

Substitute result (Res-1) into solvability condition (Cond-1), we conclude

$$\frac{\partial u_0}{\partial t} = D_0 \frac{\partial^2 u_0}{\partial x^2} + f(x,t)\frac{\partial u_0}{\partial x}$$

which is backward equation (BE-2).

Remark:

In the derivation above, the key is the method of solvability condition.

Below we illustrate the method of solvability condition in two simple examples.

Example: (the method of solvability condition)

Consider solving a 2×2 linear system $Ax = b$ where

$$A = \begin{pmatrix} 1 & 1+\varepsilon \\ 1+\varepsilon & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1-\varepsilon \\ 1+\varepsilon \end{pmatrix}$$

The exact solution is

$$x_{\text{exa}} = \begin{pmatrix} \dfrac{3+\varepsilon}{2+\varepsilon} \\ \dfrac{-(1+\varepsilon)}{2+\varepsilon} \end{pmatrix} \approx \begin{pmatrix} 1.5 \\ -0.5 \end{pmatrix}$$

Here we use asymptotic analysis to solve this example.

The goal is to see the application of solvability condition in a simple setting.

We seek solutions of the form

$$x = x^{(0)} + \varepsilon x^{(1)} + \cdots$$

We write matrix A and vector b as

$$A = A^{(0)} + \varepsilon A^{(1)}, \qquad A^{(0)} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad A^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$b = b^{(0)} + \varepsilon b^{(1)}, \qquad b^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad b^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

We expand linear system $Ax - b = 0$ into O(1) terms, O(ε) terms, …

$$\left(A^{(0)} + \varepsilon A^{(1)}\right)\left(x^{(0)} + \varepsilon x^{(1)} + \cdots\right) - \left(b^{(0)} + \varepsilon b^{(1)}\right) = 0$$

==> $\left(A^{(0)}x^{(0)}-b^{(0)}\right)+\varepsilon\left(A^{(0)}x^{(1)}+A^{(1)}x^{(0)}-b^{(1)}\right)+\cdots=0$

First we look at the O(1) terms.

$$A^{(0)}x^{(0)}=b^{(0)}$$

This linear system is underdetermined.

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \end{pmatrix}=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

==> $x_1^{(0)}+x_2^{(0)}=1$ (CNS-1)

From the O(1) terms, we obtain only one constraint for two unknowns.

Next we look at the O(ε) terms.

$$A^{(0)}x^{(1)}=-A^{(1)}x^{(0)}+b^{(1)} \qquad \text{(L-1)}$$

Matrix $A^{(0)}$ is symmetric. The null space of $A^{(0)}$ is

$$\text{Nul}\left(A^{(0)}\right)=\left\{v_0\right\}, \qquad v_0=\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Linear system (L-1) is solvable if and only if

$$\left\langle -A^{(1)}x^{(0)}+b^{(1)},\ v_0\right\rangle=0$$

==> $x_1^{(0)}-x_2^{(0)}-2=0$ (CNS-2)

Combining constraints (CNS-1) and (CNS-2), we conclude

$$x^{(0)}=\begin{pmatrix} 1.5 \\ -0.5 \end{pmatrix}$$

Example: (the method of solvability condition)

Consider the BVP of 2nd order linear ODE

$$\begin{cases} L[u]+\varepsilon x u=3\sin(2x), & L[u]\equiv u''+u \\ u(0)=0, \quad u(\pi)=0 \end{cases}$$

Here we use asymptotic analysis to solve this example.

We seek solutions of the form

$$u(x)=u^{(0)}(x)+\varepsilon u^{(1)}(x)+\cdots$$

We expand the ODE into O(1) terms, O(ε) terms, …

$$L\left[u^{(0)\prime\prime}(x)+\varepsilon u^{(1)\prime\prime}(x)\right]+\varepsilon x u^{(0)}(x)-3\sin(2x)=0$$

$$\Longrightarrow \quad \left(L\left[u^{(0)}(x)\right]-3\sin(2x)\right)+\varepsilon\left(L\left[u^{(1)}(x)\right]+x u^{(0)}(x)\right)+\cdots=0$$

First we look at the O(1) terms.

$$\begin{cases} L\left[u^{(0)}\right]=3\sin(2x) \\ u^{(0)}(0)=0\,, \qquad u^{(0)}(\pi)=0 \end{cases}$$

This BVP is underdetermined.

$$u^{(0)}=-\sin(2x)+c\sin(x) \quad \text{is a solution for any } c. \qquad \text{(CNS-3)}$$

Next we look at the O(ε) terms.

$$\begin{cases} L\left[u^{(1)}\right]=-x u^{(0)}(x) \\ u^{(1)}(0)=0\,, \qquad u^{(1)}(\pi)=0 \end{cases} \qquad \text{(BVP-1)}$$

Operator L with zero-BCs is self-adjoint (symmetric).

The null space of L with zero-BCs is

$$\text{Nul}\left(L\right)=\left\{v_0(x)\right\}, \qquad v_0(x)=\sin(x)$$

(BVP-1) is solvable if and only if

$$\left\langle -x u^{(0)}(x),\ v_0\right\rangle=0$$

$$\Longrightarrow \quad \int_0^{\pi}\left(x\sin(2x)-c\,x\sin(x)\right)\sin(x)dx=0$$

$$\Longrightarrow \quad -\frac{8}{9}-c\frac{\pi^2}{4}=0 \qquad \Longrightarrow \quad c=\frac{-32}{9\pi^2} \qquad \text{(CNS-4)}$$

Combining constraints (CNS-3) and (CNS-4), we conclude

$$u^{(0)}=-\sin(2x)-\frac{32}{9\pi^2}\sin(x)$$