AM216 Stochastic Differential Equations

Lecture 03
Copyright by Hongyun Wang, UCSC

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

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

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

- CF and the first moment

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

  Proof:

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

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

- 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 ξ = 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:

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

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

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

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

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>

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

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

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?

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

Version 1: 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 must open one of the two remaining boxes.
- The host must open an empty box to show it is empty.
- The host must 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.

Version 2: 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 if and only if 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.

Caution: 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.

Version 3: 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 repeated games.

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

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

<u>Version 1</u> 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.

<u>Key observation:</u>

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

Stochastic differential equation

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

Or in a more concise form

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

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

We need to introduce $W(t)$.

<u>Definition:</u>

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

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

<u>Remark:</u> 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 $t_2 \geq t_1 \geq 0$, $W(t_1)$ 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 $\phi_X(\delta) = 1 + E(X)\delta - \dfrac{E(X^2)}{2}\delta^2 + \cdots$

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

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

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

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

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