Topic 7: Multiple Continuous Random Variables

Lecture Outline

- Conditioning on an Event
- Joint Probability Density Functions
- Conditioning on a Random Variable
  - 條件機率 Conditional probability conditioned on a continuous RV
  - 條件期望值 Conditional expectation
  - 條件機率密度函數 Conditional PDF
- Continuous Bayes Rule
- Variants of Total Probability Theorem (total probability theorem 的四個變形)

Reading : Textbook 3.4- 3.6

Multiple Random Variables

Quite naturally, we definitely will have to deal with ***multiple RVs*** in a real-world problem:

- The total amount of time we need to wait in the line at the supermarket counter
- What information can the multiple antennas of a WiFi router provide to the receiver?

Again, we are particularly interested in the following:

- How different random variables are *related* to each other? 不同隨機變數間之關聯性 (正相關、負相關、零相關、獨立、條件機率、條件期望值、條件機率密度函數)

- How to learn the *behavior* of the sum of all RVs?
  Ex: 分析網路系統中資料流量進出某網路節點所耗費的時間延遲 (排隊理論, queueing)

- How to learn the *behavior* or the *true value* of one RV when it is buried in the sum with other random variables?
  Ex: 如何從 接收訊號(Y)=傳送訊號(S)+雜訊(N) 中擷取 S 的資訊。其中雜訊一般是 continuous RV，在**數位通訊**的技術下，傳送訊號 S 是discrete。接收訊號 Y 為連續。

Conditioning on an Event

Definition

Similar to the definition of unconditional PDF, the *conditional PDF* of a continuous random variable X, conditioned on a particular event E with $P(E) > 0$, is a function $f_{X|E}$ that satisfies

$$P(X \in B|E) = \int_B f_{X|E}(x)dx \qquad (1)$$

Special Case

Conditioning on the event E that X belongs to a subset A of the real line, we have

$$P(X \in B|X \in A) = \frac{P(X \in A, X \in B)}{P(X \in A)} = \frac{\int_{A \cap B} f_X(x)dx}{P(X \in A)} \qquad (2)$$

Therefore, comparing (2) with (1), we have

$$f_{X|\{X \in A\}} = \begin{cases} \frac{f_X(x)}{P(X \in A)}, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

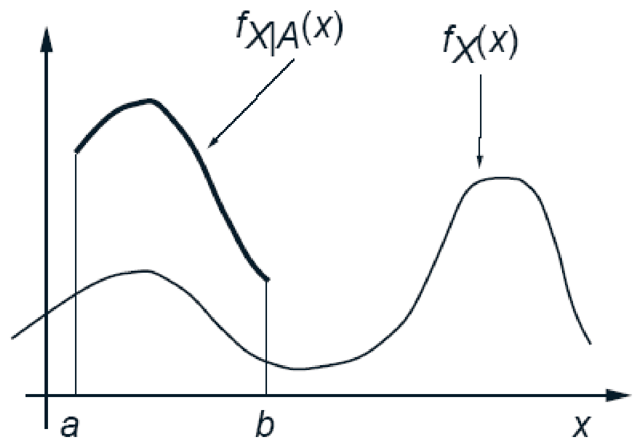
Illustration of Conditioning on an Event

Special Case

Conditioning on the set when X belongs to a subset A of the real line, we have

$$f_{X|\{X \in A\}} = \begin{cases} \frac{f_X(x)}{P(X \in A)}, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

We see it's just a *scaling* of $f_x(x)$. So the PDF within the conditioning set has the same shape as the unconditional PDF



The unconditional PDF f_X and the conditional PDF $f_{X|A}$, where A is the interval $[a, b]$. Note that within the conditioning event A, $f_{X|A}$ retains the same shape as f_X, except that it is scaled along the vertical axis.

Example: The Exponential Random Variable Is Memoryless

Example 3.13 (Memoryless Property of Exponential)

The time T until a new light bulb burns out can be modeled by an exponential random variable with parameter λ. Alice turns the light on, leaves the room, and when she returns, t time units later, finds that the light bulb is still on, which corresponds to the event $A=\{T>t\}$.

Let X be the additional time until the light bulb burns out. What is the conditional CDF of X, given the event A?

Total Probability Theorem Using Conditional PDF**

If $A_1, A_2, \ldots, A_n$ are disjoint events with $P(A_i) > 0$ for each i, that form a partition of the sample space, then

$$f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)$$

- 這是常用的技巧，可協助計算出複雜 X 的PDF。
  How to justify the equation?
- See Example 3.14 in the textbook
- Another example:

Let X be a standard normal random variable. A new random variable Y is defined as follows: We flip a coin. If the outcome is a head, then $Y = X$. And if the outcome of the coin flip is a tail, then $Y = -X$. Assume the coin flip is independent with X. Please find the PDF of Y.

Total Probability Theorem using Conditional PDF

(Example 3.14)

The metro train arrives at the station near your home every quarter hour starting at 6:00 a.m. You walk into the station every morning between 7:10 and 7:30 a.m. and your arrival time is a uniform random variable over this interval. What is the PDF of the time you have to wait for the first train to arrive?

Total Expectation Theorem using Conditional Expectation

Definition

The conditional expectation is defined by

$$E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

Total Expectation Theorem

If $A_1, A_2, \ldots, A_n$ are disjoint events with $P(A_i) > 0$ for each i, that form a partition of the sample space, then

$$E[X] = \sum_{i=1}^{n} P(A_i)E[X|A_i]$$

See Example 3.17 in textbook.

Total Expectation Theorem using Conditional Expectation

(Example 3.17)
Suppose that the random variable X has the piecewise constant PDF

$$f_X(x) = \begin{cases} 1/3, & \text{if } 0 \leq x \leq 1, \\ 2/3, & \text{if } 1 < x \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Find $E[X]$ and var(X).

Joint Probability Density Function

We say that two continuous random variables associated with a common experiment are jointly continuous and can be described in terms of a joint PDF $f_{X,Y}$, if $f_{X,Y}$ is a nonnegative function that satisfies

$$P((X,Y) \in B) = \int\int_{(x,y)\in B} f_{X,Y}(x,y)dxdy$$

for every subset B of the 2-dimensional plane.

- The probability that (X,Y) falls within a rectangle $B = \{(x,y)|a \leq x \leq b, c \leq y \leq d\}$ is

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$

Joint Probability Density Function

- PDF $f_{X,Y}(x,y)$ must satisfy *normalization* equation, i.e.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

- To interpret the PDF, we let δ be very small and consider the probability of a small rectangle. We have

$$P(a \leq X \leq a+\delta, c \leq Y \leq c+\delta) \approx f_{X,Y}(a,c)\cdot\delta^2.$$

we can view $f_{X,Y}(a, c)$ as the "probability per unit area" in the vicinity of (a, c)

Marginal PDF from Joint PDF

The marginal PDFs $f_X(x)$ and $f_Y(y)$ of *continuous* X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

- Recall that for *discrete* random RVs, we have

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$
$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

- 本質上，上述諸式均源自於 total probability theorem

Expectation

- If X and Y are jointly continuous random variables, and *g* is some function, then $Z = g(X, Y)$ is also a random variable. the **expected value rule** is

$$E[g(X,Y)] = \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dxdy$$

- Using the expected value rule, for any scalars a, b, we have

$$E[aX + bY] = aE[X] + bE[Y]$$

Conditioning on a Random Variable

Definition

Let X and Y be continuous random variables with joint PDF $f_{X,Y}$. For any fixed y with $f_Y(y) > 0$, the conditional PDF of X given that $Y = y$, is defined by

$$f_{X|Y}(x|y) \triangleq \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

1. Analogous to the formula in the discrete PMF case
2. $f_{X|Y}(x|y)$ is a legitimate PDF (check normalization)

See example 3.19 in the textbook.

Interpretation of Conditioning on a Random Variable

Interpretation

Fix some small positive numbers δ_1 and δ_2, and condition on the event $B = \{y \leq Y \leq y + \delta_2\}$. We have

$$\begin{aligned} P(x \leq X \leq x + \delta_1 | y \leq Y \leq y + \delta_2) &\approx \frac{f_{X,Y}(x,y)\delta_1\delta_2}{f_Y(y)\delta_2} \\ &= f_{X|Y}(x|y)\delta_1 \end{aligned}$$

We can think of the limiting case where δ_2 decreases to zero and write

$$P(x \leq X \leq x + \delta_1 | Y = y) \approx f_{X|Y}(x|y)\delta_1$$

More generally,

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y)dx$$

Conditional Expectation

Definition

The conditional expectation is defined by

$$E[X|Y=y] \triangleq \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

- The expected value rule applies to conditional expectation

$$E[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

- Total expectation theorem

$$E[X] = \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y)dy$$

Example

The joint PDF of X and Y is $f_{X,Y}(x,y) = \dfrac{e^{-x/y}e^{-y}}{y}, \quad 0 < x < \infty, 0 < y < \infty$

Find the conditional expectation $E[X|Y=y]$.

Total Expectation Theorem

Total expectation theorem for continuous random variable

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y) dy$$

- Consider any event A. Let X be the random variable that takes the value 1 if event A occurs and the value 0 otherwise (Such RV is called an indicator). In this case, we have $E[X]=P(A)$.

$$\begin{cases} X = 1, & \text{if } A \text{ occurs,} \\ X = 0, & \text{otherwise.} \end{cases}$$

Then, we have the following version of total probability theorem, when conditioned on continuous random variable Y (See page 30 of this topic)

$$\boxed{P(A) = \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y) dy}$$

Independence

Two continuous random variables X and Y are independent if their joint PDF for all x and y is the product of the marginal PDFs:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

Suppose X and Y are independent.

- Similar to the discrete case, the random variables $g(X)$ and $h(Y)$ are independent, for any functions g and h
- We have $E[XY] = E[X]E[Y]$

  or more generally,

  $$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

- IF **X and Y are independent,** we have

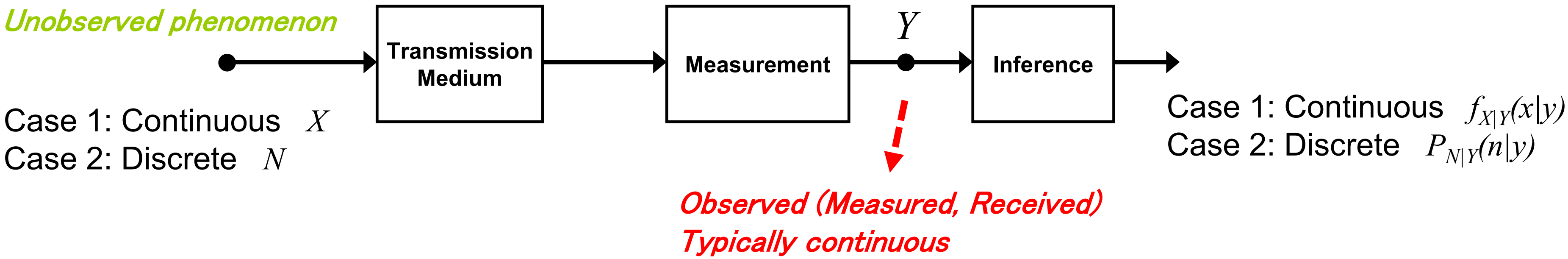
  $$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$$

Continuous Bayes' Rule

The conditional PDF $f_{X|Y}(x|y)$ can by obtained via

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int f_X(t) f_{Y|X}(y|t) dt}$$

Statistical Inference using Bayes' Rule
使用貝氏定理進行預測、推論



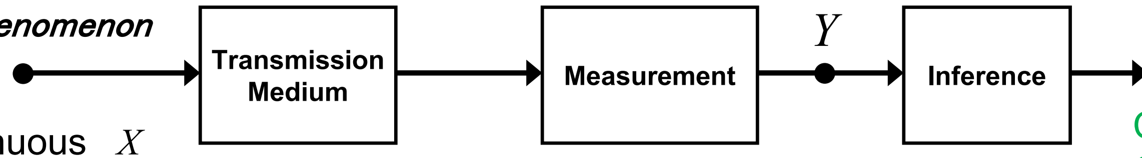
- ***Unobserved phenomenon*** (**cause**) describes something we want to know about. But we can only observe a certain result (**effect**) at the receiving end.

  For example,
  - Unknown transmitted bit (0 or 1) in a communication system
  - Unknown presence of an airplane in a radar system
  - Unknown disease of a patient

- **Objective:** (藉由觀察 Y 的值來推論 X 或 N 的值。此處 Y 是蒐集獲得之資料、訊號)
  Determine which cause among many candidates is the most likely, by checking the posterior probability $P_{N|Y}(n|y)$ *for all possible* n (or posterior density $f_{X|Y}(x|y)$)

Statistical Inference using Bayes' Rule

Unobserved phenomenon



Case 1: Continuous X
Case 2: Discrete N

Observed (Measured, Received)
Typically continuous, due to noise in the observation

Case 1: Continuous $f_{X|Y}(x|y)$
Case 2: Discrete $P_{N|Y}(n|y)$

1. Continuous case: The *unobserved* phenomenon is a continuous random variable *X, the posterior density of the unknown X given observed Y=y is*

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int f_X(t) f_{Y|X}(y|t) dt}$$

2. Discrete case: The *unobserved* phenomenon is a discrete random variable *N, the posterior probability of N given the observed $Y=y$ is*

$$P(N = n|Y = y) = \frac{P_N(n) f_{Y|N}(y|N = n)}{\sum_i P_N(i) f_{Y|N}(y|N = i)}$$

Example – Signal Detection in Communication Systems**

A binary signal S is transmitted, and we are given that $P(S = 1) = p$ and $P(S = -1) = 1-p$. The received signal is $Y = N+S$, where N is normal noise, with zero mean and variance σ^2, independent of S. What is the probability that $S = 1$, given that we have observed $Y=y$?

Joint CDF

If X and Y are two random variables associated with the same experiment, we define their joint CDF by

$$F_{X,Y}(x,y) \triangleq P(X \leq x, Y \leq y)$$

- The joint CDF $$\begin{aligned} F_{X,Y}(x,y) &\triangleq P(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(s,t) ds dt \end{aligned}$$

- The PDF can be recovered from the PDF by differentiating:

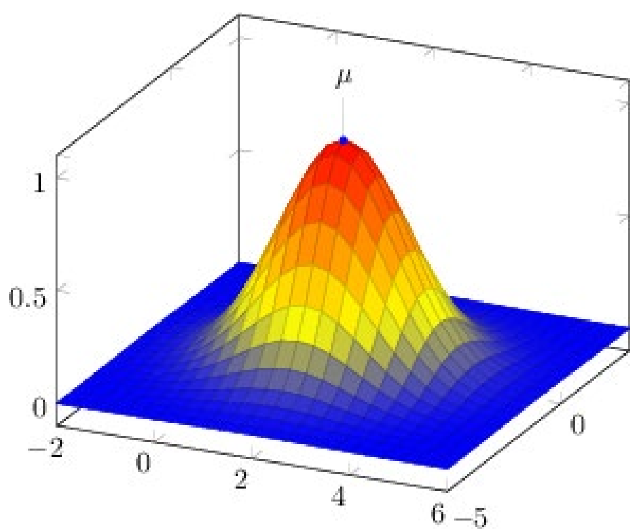
$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

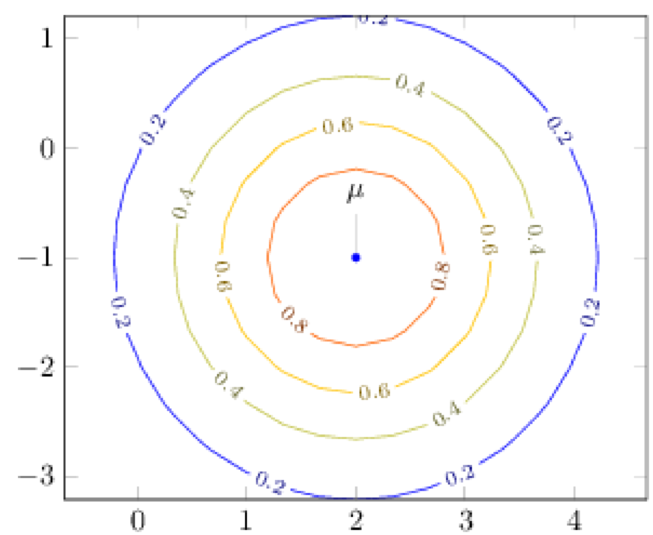
Joint PDF of Independent Gaussian RVs

Let X and Y be two independent Gaussian random variables with means μ_x ,μ_y, and variances σ_x^2, σ_y^2 respectively. The joint PDF is given by

- The joint PDF has a bell shape centered at (μ_x ,μ_y). Whether the PDF is a *tall-thin* bell or *wide-fat* bell is determined by the variances σ_x^2, σ_y^2

- Very often we are interested in knowing the contours (等高線圖、切面圖) of the joint PDF. The contours are sets of points at which the PDF takes a constant value.





Summary of Total Probability Theorem and Its Variants

- We have already learned several different versions of *total probability theorem*. The original version is given in the form of probability of events (Section 1.4, page 28)

$$P(G) = \sum_{i=1}^{K} P(G \mid F_i) P(F_i)$$

- According to the types of underlying RVs in forming G and F_i, the TPT can be further generalized to the following variants:
  - Variant 1: Discrete RV in G – Discrete RV in F_i (DD Type)
  - Variant 2: Continuous RV in G – Continuous RV in F_i (CC Type)
  - Variant 3: Continuous RV in G – Discrete RV in F_i (CD Type)
  - Variant 4: Discrete RV in G – Continuous RV in F_i (DC Type)

- 通則: *When continuous RV is involved, the sum and PMF in original TPT need to be modified to integral and PDF*

第一類變形: Discrete RV in G – Discrete RV in F_i

- Variant 1: Discrete RV in G – Discrete RV in F_i (DD Type)

$$\boxed{\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x,y) \\ &= \sum_y p_{X|Y}(x|y)p_Y(y) \end{aligned}}$$

  - ✓ This follows directly from original version by setting $G=\{X=x\}$ and $F_i=\{Y=y_i\}$
  - ✓ See Example 2.14 and 2.15.
  - ✓ Another possible forms:

$$p_X(x) = \sum_{i=1}^{n} p_{X|A_i}(x)P(A_i)$$
$$P(B) = \sum_y P(B|Y=y)p_Y(y)$$

  - ✓ Example:

    Let X, Y be discrete RVs. Compute the probability $P(X+Y=w)$.

第二類: Continuous RV in G – Continuous RV in F_i

- Variant 2: Continuous RV in G – Continuous RV in F_i (CCType)

$$\boxed{\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)dy \end{aligned}}$$

  - ✓ See page 172
  - ✓ Example 3.19, on page 179 (continuous Bayes rule)

第三類: Continuous RV in G – Discrete RV in F_i

- Variant 3: Continuous RV in G – Discrete RV in F_i (CDType)

$$\boxed{f_Y(y) = \sum_i f_{Y|N}(y|i) p_N(i)}$$

  - ✓ See page 180 and Example 3.20.
  - ✓ A more general form is (see page 167)

$$\boxed{f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)}$$

  - ✓ Example 3.14, on page 168.
  - ✓ See the example on page 6 of this topic.

第四類: Discrete RV in G – Continuous RV in F_i

- Variant 4: Discrete RV in G – Continuous RV in F_i (CDType)

$$\boxed{p_N(n) = \int_{-\infty}^{\infty} P(N=n|Y=y) f_Y(y) dy}$$

  - ✓ See page 181 and page 182.
  - ✓ A more general form is (page 181 of textbook, page 18 of this topic)

$$\boxed{P(A) = \int_{-\infty}^{\infty} P(A|Y=y) f_Y(y) dy}$$

  - ✓ You can justify this relation by taking the integral over y to the 1st equation on page 181 of the textbook

  - ✓ Example:

    Let X, Y be indepdent continuous RVs. Compute the PDF of $X+Y$.