Topic 5: *Multiple* Discrete Random Variables

Lecture Outline

- Joint PMF and Marginal PMF
- Linearity of Expectation
- Conditional PMF
  - Total Probability Theorem (TPT) in Discrete Random Variables **(第一個TPT變形)**

- Independent Random Variables
- Conditional Expectation

Reading : Textbook 2.5- 2.7

Quite naturally, we have to deal with ***multiple RVs*** in real-world problems:

- The money we earn in several consecutive plays of poker games
- The numbers of 3 pointers Steph Curry can make in 4 different quarters
- In a communication system, the receiver signal can often be modeled as the signal that contains the (unknown) transmitted signal plus (unknown) noise

We are particularly interested in knowing:

- 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?)
- How to learn the *behavior* of the sum of all RVs?
- 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

Joint PMF

We can fully describe two discrete RVs X and Y by the *joint PMF*.

Suppose two *discrete* RVs X and Y are defined over the same sample space, the *joint* PMF $p_{X,Y}(x,y)$ is defined by

$$p_{X,Y}(x,y) = \mathrm{P}(X=x \text{ and } Y=y)$$

$$= \mathrm{P}\big(\{\omega : X(\omega) = x \text{ and } Y(\omega) = y\} \big)$$

From the additivity and normalization axioms of probability, we must have

$$\Sigma_{x,y}\, p_{X,Y}(x,y) = 1$$

Example:

Rolling two 4-sided die. Let X be the outcome of the first roll and Y of the second, then

$$p_{X,Y}(x,y) = 1/16; \;\; for\ x=1,2,3,4;\ y=1,2,3,4$$

Joint PMF and Marginal PMF

Suppose X and Y are two random variables defined on a common experiment with a joint PMF $p_{X,Y}(x,y)$. The ***marginal*** PMFs $p_X(x)$ and $p_Y(y)$ can be obtained from joint PMF as follows

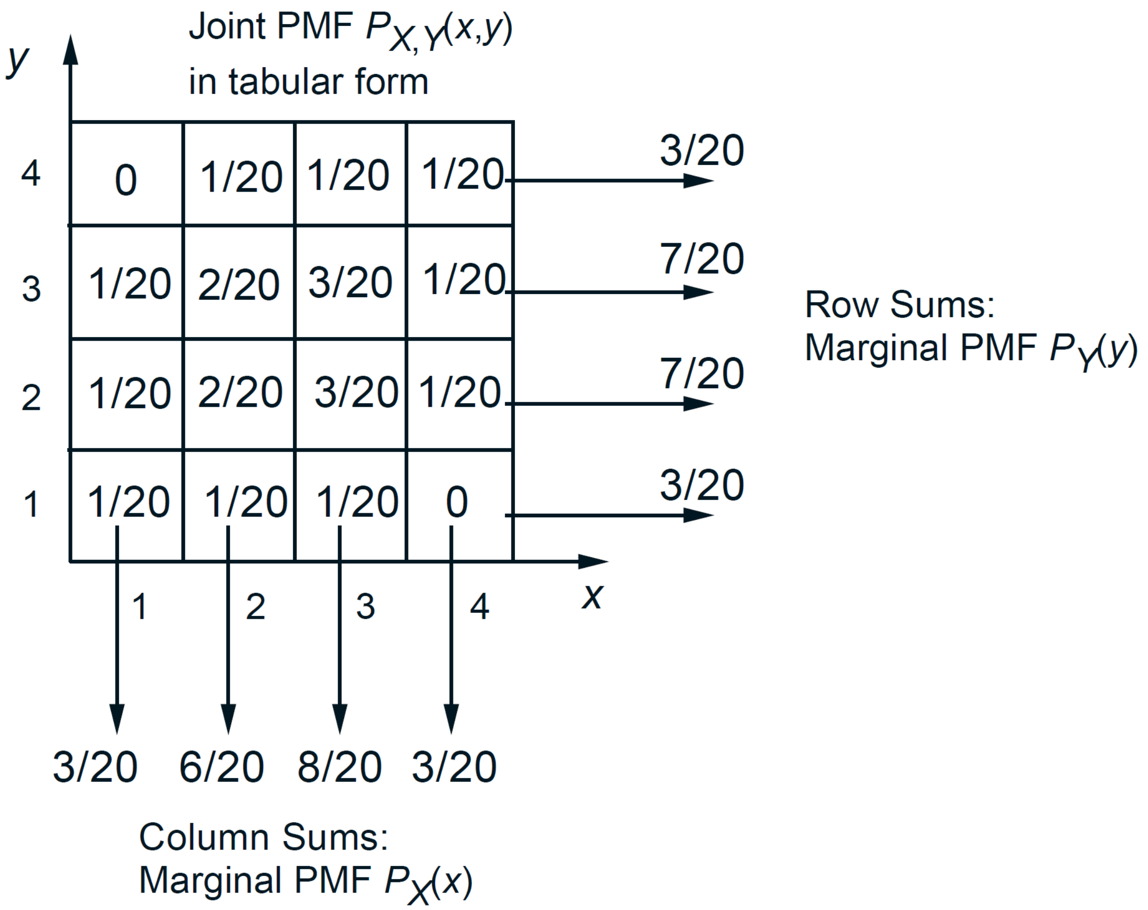
$$\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x,y) \\ p_Y(y) &= \sum_x p_{X,Y}(x,y) \end{aligned}$$

Remarks:

- The above relation can be explained by the *total probability theorem*
- The marginal PMF is just the "individual" PMF $p_X(x)$ or $p_Y(y)$
  （單獨RV的PMF、原本個別的PMF、邊際PMF）

Example: Calculating Marginal PMF from Joint PMF

Example 2.9:



Functions of Multiple Random Variables

A function $Z = g(X, Y)$ of the random variables X and Y defines another random variable. Its PMF can be calculated from the joint PMF $p_{X,Y}$ according to

$$p_Z(z) = \sum_{\{(x,y) | g(x,y) = z\}} p_{X,Y}(x, y)$$

The expected value rule for functions takes the following form

$$E[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$$

Example:

We are often interested in finding $E[XY]$, which measures the **similarity** between X and Y.

Linearity of Expectation

As an example, consider the linear function $g(X,Y)=aX+bY+c$ of random variables X and Y, where a, b, and c are scalars (constants). The expectation of $g(X,Y)$ is given by

$$\begin{aligned} E[aX+bY+c] &= \sum_{x,y}(ax+by+c)p_{X,Y}(x,y) \\ &= \sum_{x,y} axp_{X,Y}(x,y) + \sum_{x,y} byp_{X,Y}(x,y) + \sum_{x,y} cp_{X,Y}(x,y) \\ &= a\sum_{x} x \sum_{y} p_{X,Y}(x,y) + b\sum_{y} y \sum_{x} p_{X,Y}(x,y) + c \\ &= aE[X] + bE[Y] + c \end{aligned}$$

Expectation is linear

We see that the expectation of the sum of random variables is the sum of the expectations.

Do Example 2.9 and Example 2.11.

Consider 3 random variables X, Y, and Z.

Joint PMF

For all possible triplets of numerical values (*x*,*y*,*z*)

$$p_{X,Y,Z}(x,y,z) = P(X = x, Y = y, Z = z)$$

Marginal PMF from Joint PMF

$$p_{X,Y}(x,y) = \sum_{z} p_{X,Y,Z}(x,y,z)$$

$$p_X(x) = \sum_{y,z} p_{X,Y,Z}(x,y,z)$$

Linearity of Expectation

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E[X_i]$$

This is an extremely useful relation!

Example - Mean of Binomial R.V.

One way to define a Binomial random variable is the following:
Flip a coin (with bias *p*) *n* times independently and let X_i = 1 if the i-th toss is a head and 0 if it is a tail.

Define the random variable $Y = \sum_{i=1}^{n} X_i$. Then, Y is exactly a binomial RV.

$$\begin{aligned} E[Y] &= E\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} E[X_i] \\ &= np \end{aligned}$$

Conditional PMF

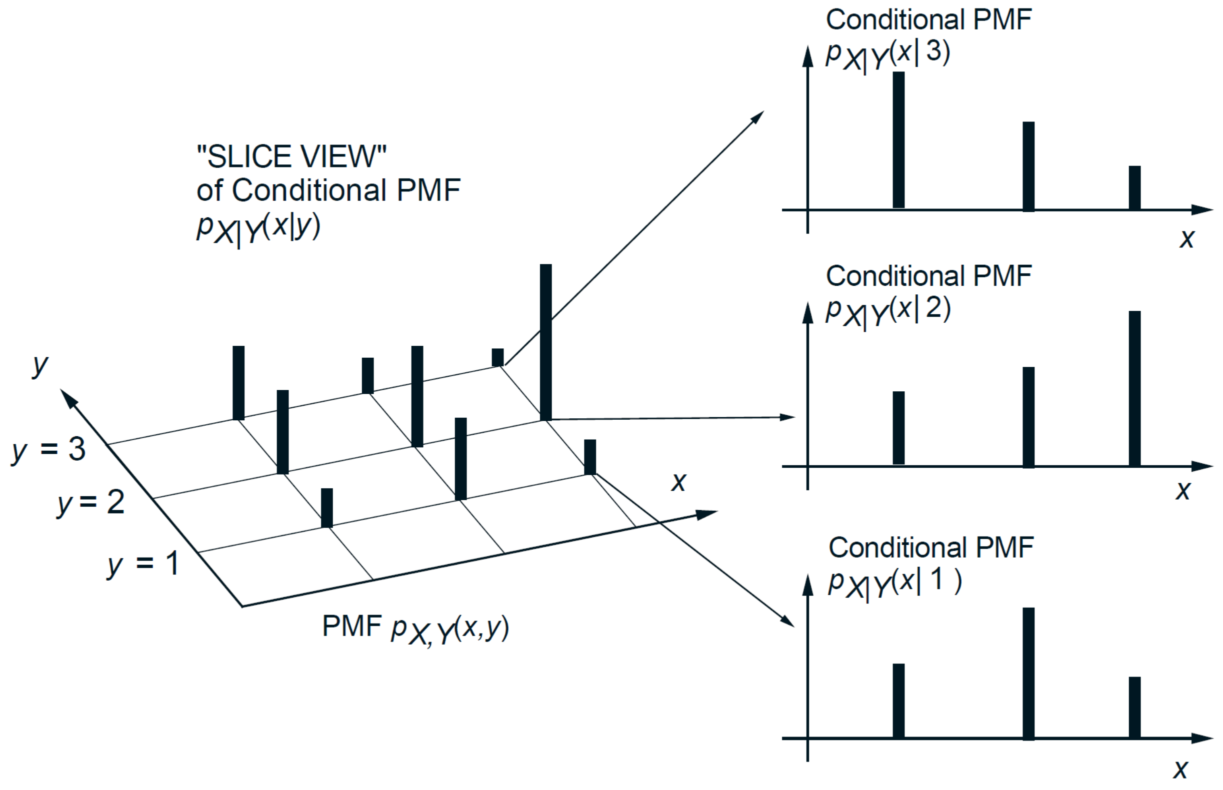
The **conditional PMF** of a random variable X, conditioned on a particular event A with **P**(A) > 0, is defined by

$$
\begin{aligned}
p_{X|A}(x) &\triangleq P(X = x|A) \\
&= \frac{P(\{X = x\} \cap A)}{P(A)}
\end{aligned}
$$

In the special case $A=\{Y=y\}$, we have a PMF conditioned on a random variable Y

$$
\begin{aligned}
p_{X|Y}(x|y) &\triangleq P(X = x|Y = y) \\
&= \frac{P(X = x, Y = y)}{P(Y = y)} \\
&= \frac{p_{X,Y}(x, y)}{p_Y(y)}
\end{aligned}
$$

Visualization of Conditional PMF



$$\sum_{x,y} p_{X,Y}(x,y) = 1$$

$$\sum_{x} p_{X|Y}(x|y) = 1$$

Properties of Conditional PMF

- Conditional PMF is just conditional probability in different notations
  - Satisfies the 3 probability axioms

- Normalization

$$\sum_x p_{X|Y}(x|y) = 1$$

- Chain rule

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)$$

- **Total probability Theorem**

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

  - This is the first variant (變形) of the total probability theorem. We will see three more later in Chapter 3.
  - Please do Example 2.14 and Example 2.15.

Example *(Problem 43 on page 134)*

Suppose that X and Y are *independent and identically distributed* (**i.i.d.**) geometric random variables with parameter p, then

$$P(X=i \mid X+Y=n)=\frac{1}{n-1}, \qquad i=1,\ldots,n-1.$$

Remarks:

- Physical interpretations (物理意義)?
- The acronym "**i.i.d.**" appears very often. So please memorize what that means.

Independence

Definition:

We say that two random variables X and Y are *independent* if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad \text{for all } x, y$$

- Independence means the experimental value of Y tells us nothing about the value of X
- Equivalent to saying that the events $\{ \omega : X(\omega) = x\}$ and $\{\omega : Y(\omega) = y\}$ are *independent* events for all possible outcomes x and y
- The above definition of independence is equivalent to

$$p_{X|Y}(x|y) = p_X(x) \quad \text{for all } y \text{ with } p_Y(y) > 0 \text{ and all } x$$

Important Facts from Independent RVs

- If X and Y are independent random variables, then we can prove that

$$E[XY] = E[X]E[Y]$$

  - 為什麼我們會對 $E[XY]$是否等於 $E[X]\ E[Y]$ 感興趣?
  - If X and Y are independent, then we have $E[XY]= E[X]\ E[Y]$. On the contrary, if $E[XY]= E[X]\ E[Y]$, can we say that X and Y are independent? The answer is **NO**.

  We will look into the above two issues more deeply in Chapter 4.

- If X and Y are independent random variables, then $g(X)$ and $h(Y)$, for *any functions* g and h, are also independent. (See the proof in Problem 44.)
  - If X and Y are independent random variables, then for *any functions* g and h

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

Important Facts from Independent RVs

- If X and Y are independent random variables, then

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

But *in general*, this relation is not always true. That is, if X and Y are NOT independent, then we do NOT know whether the above relation is true.

- 可以延伸到： 若 $X_1, X_2, \ldots, X_n$ 為獨立隨機變數(independent RVs), 則

$$\text{var}(X_1+X_2+\ldots+X_n)=\sum_{i=1}^{n}\text{var}(X_i)$$

Multiple Independent Random Variables

A collection of random variables $\{X_1, X_2, \ldots, X_n\}$ is said to be independent if the n-fold joint PMF is a product of marginal PMFs:

$$p_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

In the special case where the marginal probabilities $p_{X_i}(x_i)$ are all the same, we say that these n random variables are *independent and identically distributed* (***i.i.d.***)

Example - Variance of Binomial

One way to define a Binomial random variable is the following:

Flip a coin (with bias *p*) *n* times independently and let X_i = 1 if the i-th toss is a head and 0 if it is a tail. Define the random variable $Y = \sum_{i=1}^{n} X_i$. Then, Y is exactly a binomial RV.

$$\begin{aligned} E[Y^2] &= E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \\ &= E\left[\sum_{i=1}^{n}\sum_{m=1}^{n} X_i X_m\right] \\ &= \sum_{i=1}^{n}\sum_{m=1}^{n} E[X_i X_m] \\ &= np + (n^2 - n)p^2 \end{aligned}$$

$$\begin{aligned} \mathrm{var}(Y) &= E[Y^2] - (E[Y])^2 \\ &= np(1-p) \end{aligned}$$

Example – Sample Mean

Let $X_1, X_2, \ldots, X_n$ be *n* i.i.d. random variables with mean μ and variance σ^2. The sample mean is defined by

$$S_n \triangleq \frac{1}{n} \sum_{i=1}^{n} X_i$$

- We often use sample mean to estimate the mean value of X_i
- What are the mean and variance of S_n?

As *n* grows, the variance of the sample mean decreases

Definitions

Conditional expectation of X given an event A:

$$E[X|A] \triangleq \sum_x x p_{X|A}(x)$$

Conditional expectation of X given $Y=y$:

$$E[X|Y=y] \triangleq \sum_x x p_{X|Y}(x|y)$$

- Conditional expectation $E[X|Y=y]$ is a function of y, *it's a numerical value depending on y.* On the other hand, $E[X|Y]$ is a function of the random variable Y, and is therefore a random variable.
  請說服自己, $E[X|Y]$ 是***隨機變數***, 而 $E[X|Y=y]$ 是一個和 y ***有關的實數*** *(也就是 以 y 為變數的**函數值**)*

- Conditional probability (in particular the posterior probability) and conditional expectation play very critical roles in system designs under uncertainty.

Example

If X and Y are i.i.d. geometirc random variables with parameters p. Find the conditional expected value of X, given that X+Y=n.

Total Expectation Theorem**

Total expectation theorem states unconditional expectation can be calculated using conditional expectation, just like the counterpart of unconditional probability.

$$E[X] = \sum_y p_Y(y) E[X|Y = y]$$

Please do Example 2.16 and Example 2.17.

Remark:

Comparing with the expected value rule, we can re-write the total expectation theorem more compactly as

$$E[X] = E\big[E[X|Y]\big]$$

這是經常使用到的關係式！

Example – Sum of a Random Number of Random Variables

Consider the sum

$$Y = X_1 + X_2 + \cdots + X_N$$

where N is a *random variable* that takes positive integer values, and $X_1, X_2, \ldots, X_N$ are $i.i.d.$ random variables. We assume that N is independent with X_i for all i.

Then, we have

$$E[Y] = m \cdot E[N]$$

where $E[X_i]=m$ for all i.

Practical applications:

N : the number of customers entering a department store

X_i: amount of money spent by the ith customer

Y : total money spent by customers in the store

$E[Y]$: expected money spent in the store

Properties of Conditional Expectation

- Conditional expectation is simply expectation with respect to a conditional PMF. It inherits many properties from ordinary expectation
  - Linearity
  - Expected value of functions of random variables

$$E[g(X)|Y=y] = \sum_x g(x) p_{X|Y}(x|y)$$

- $E[Y|Y=y]$=?

- When conditioning on a random variable, that random variable plays the role of a constant

$$E[g(X,Y)|Y=y] = E[g(X,y)|Y=y]$$

觀念澄清

For independent random variables X and Y, which of the following statements is correct, for an appropriate function $g()$ and set A?

$$\text{(i)}\;\; P\Big[g(X,Y)\in A \;\Big|\; Y=y_0\Big] = P\Big[g(X,y_0)\in A\Big] \quad \text{(correct?)}$$

$$\text{(ii)}\;\; P\Big[g(X,Y)\in A\cap Y=y_0\Big] = P\Big[g(X,y_0)\in A\Big] \quad \text{(correct?)}$$

Example:

Toss a dice twice, and let the outcome for the 1st toss and the 2nd toss be X_1 and X_2, respectively.

What is the probability $P[\,\{X_1+X_2\leq 8\}\cap\{X_1=5\}\,]$=?

What is the probability $P[X_1+X_2\leq 8\mid X_1=5]$=?