MSBA7003 Decision Analytics

Tutorial 01

Guoxing He

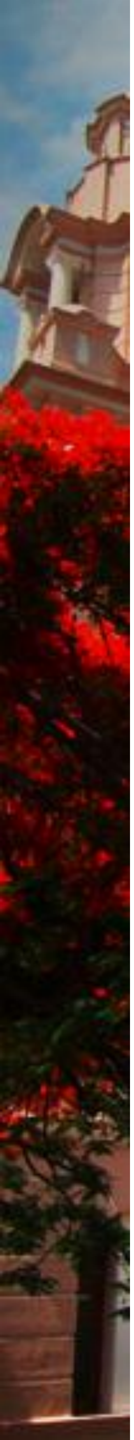
2024-2025

Email: gxhe@connect.hku.hk

Office: Room 119, K. K. Leung Building

香 港 大 學

THE UNIVERSITY OF HONG KONG



Class notes

- The template of answer sheet

1					1
	1				2
		1			3
			1	1	4
80	200	30	40		5
A	B	C	D	E	Number of the Question
Name:	Gan Ultraman				
University Numb	3030030300				

Please fill the name into cell "B8" with the surname name in front of the given name.
Please fill the UID into cell "B9".
Please fill in 1 in the corresponding cells according to chocies for each question.
In this example, questions 1 through 4 are multiple-choice, and question 5 is a fill-in-the-blank question. The answers for each question are as follows:
For question 1, "A" is chosen.
For question 2, "B" is chosen.
For question 3, "C" is chosen.
For question 4, "D" and "E" are chosen.
For question 5, fill in the blanks with 80, 200, 30, and 40, respectively. Leave cell 'E5' blank if there are only four questions requiring a answer.
For question 5, you also need to upload your detailed calculations in a separate file as the appendix.

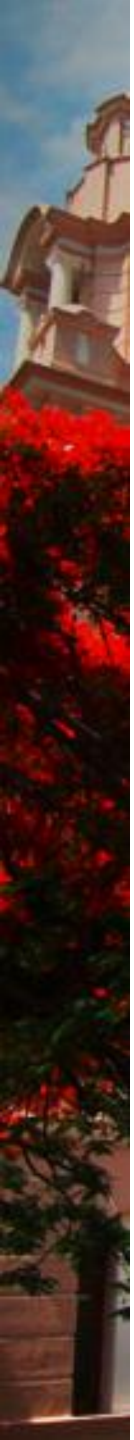
- Grading policy of assignment
  - The correct answer may consist of one or more options.
  - Full marks for the correct answer and zero otherwise.

- DO NOT
  - Need to submit a pdf file with your solutions for choice questions
  - Zip the xlsx file
  - Submit the xlsm file (empty file when opening).
  - Rely on ChatGPT

香 港 大 學
THE UNIVERSITY OF HONG KONG

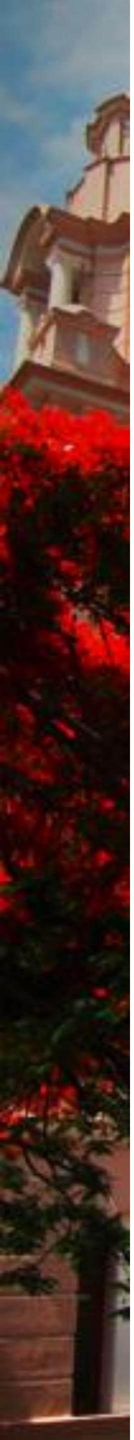
Tutorial

- Explanation of assignments and after-class exercise
- More exercises
  - Similar to in-class exercises and assignments
- Software
  - Python is most used
  - R, Excel (English Version)



Agenda

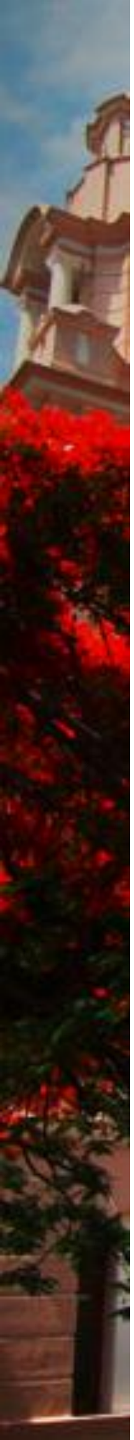
- Joint Distribution
  - Corpus Data
- Conditional Probability and Independence
  - Coin Tosses Examples
  - Conceptual Questions
- Bayesian Learning and Decision Making
  - The Monty Hall Problem
  - Café Du Donut
- Coding For Bayesian Update
  - Toss Coins In Python



Corpus Data

- Assume we have a corpus of a 100 words (a corpus is a collection of text). We tabulate the words, their frequencies and probabilities in the corpus are as follows:

Words (w)	Occurrences c(w)	Probabilities P(w)	Length (x)	Vowels (y)
the	30	0.30	3	1
to	16	0.16	2	1
some	15	0.15	4	2
grade	10	0.10	5	2
point	9	0.09	5	2
fail	8	0.08	4	2
pass	8	0.08	4	1
HK	4	0.04	2	0

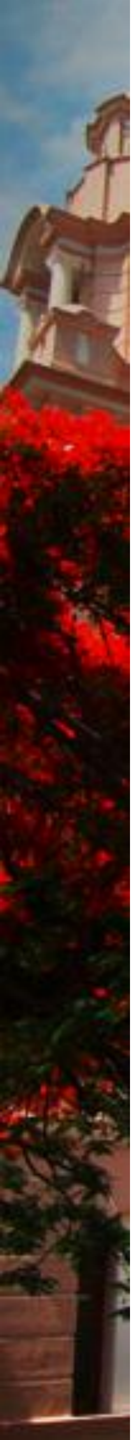


Corpus Data

- We can now define the following random variables:
  - X: the length of the word;
  - Y: number of vowels in the word.

- The probabilities of some events:
  - $P(2 \leq X \leq 3) = P(\text{to}) + P(\text{HK}) + P(\text{the}) = 0.16 + 0.3 + 0.04 = 0.5$
  - $P(2 \leq Y) = P(\text{some}) + P(\text{grade}) + P(\text{point}) + P(\text{fail}) = 0.42$

Words (w)	Occurrences c(w)	Probabilities P(w)	Length (x)	Vowels (y)
the	30	0.30	3	1
to	16	0.16	2	1
some	15	0.15	4	2
grade	10	0.10	5	2
point	9	0.09	5	2
fail	8	0.08	4	2
pass	8	0.08	4	1
HK	4	0.04	2	0



Joint Distribution

- We can describe the joint distribution between word length (X) and number of vowels (Y):
  - Let $f(x, y) = P(X = x, Y = y)$.
  - Examples:
    - $f(4,2) = P(\text{fail}) + P(\text{some}) = 0.15 + 0.08 = 0.23;$
    - $f(3,1) = P(\text{the}) = 0.3;$
    - $f(5,0) = 0.$

Words (w)	Occurrences c(w)	Probabilities P(w)	Length (x)	Vowels (y)
the	30	0.30	3	1
to	16	0.16	2	1
some	15	0.15	4	2
grade	10	0.10	5	2
point	9	0.09	5	2
fail	8	0.08	4	2
pass	8	0.08	4	1
HK	4	0.04	2	0

Joint Distribution		y		
		0	1	2
x	2	0.04	0.16	0
	3	0	0.30	0
	4	0	0.08	0.23
	5	0	0	0.19

香 港 大 學
THE UNIVERSITY OF HONG KONG

Corpus Data

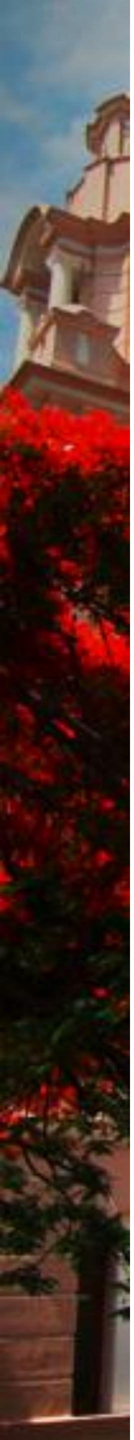
- According to the joint distribution $f(x, y)$, we can calculate the marginal distribution (f_X and f_Y) and the conditional distribution.

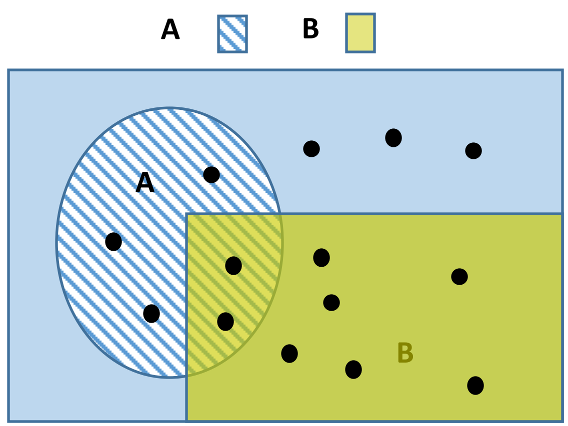
		y			f_X
		0	1	2	
	2	0.04	0.16	0	
x	3	0	0.30	0	
	4	0	0.08	0.23	
	5	0	0	0.19	

Conditional Probability

The idea of conditioning

Assume 14 equally likely outcomes

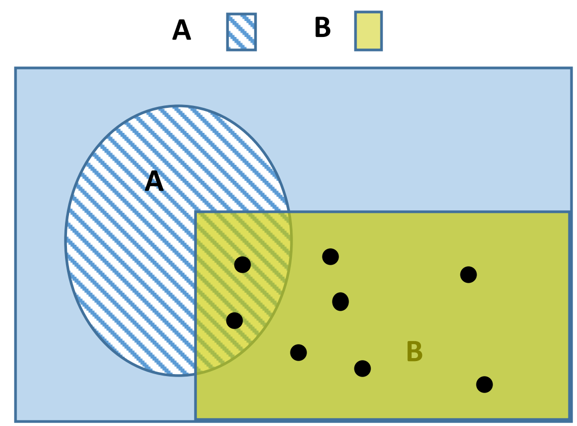




$$Pr(A) = \frac{5}{14} \quad Pr(B) = \frac{8}{14}$$

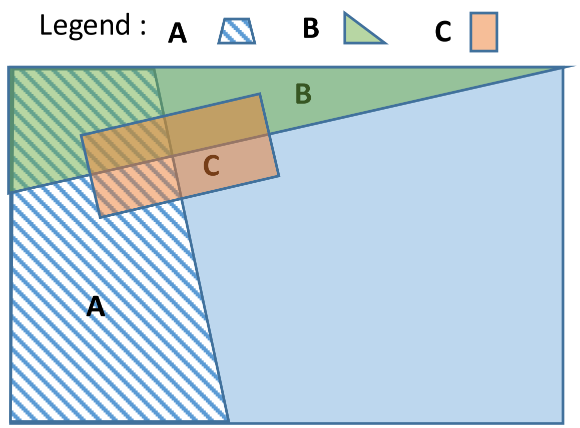
Use new information to revise a model

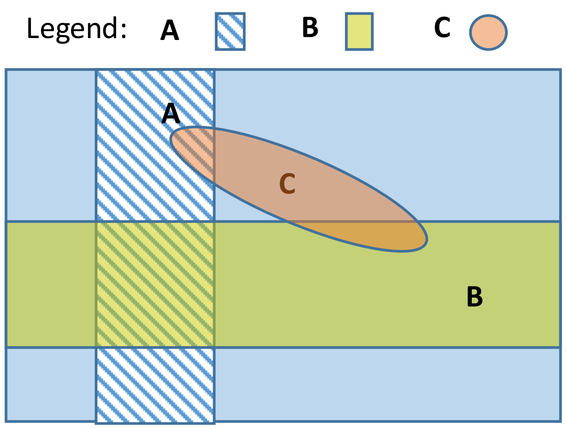
If told B occurred:



$$Pr(A|B) = \frac{2}{8} \quad Pr(B|B) = 1$$

Conditional independence



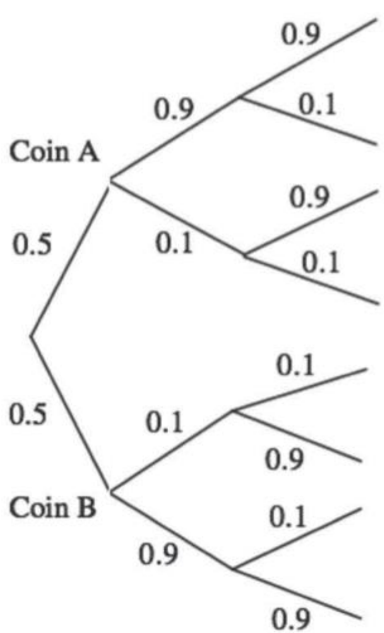


Conditional independent

Assume A and B are independent ($A \perp B$), A and B are not independent if we are given C occurred.

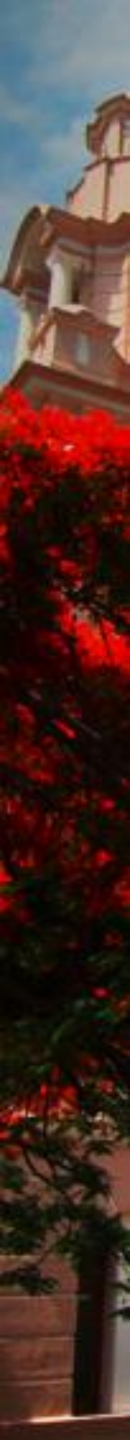
Conditioning may affect independence

- Two unfair coins, A and B:
  - Pr (H|coin A) = 0.9, Pr (H|coin B) = 0.1
- Choose either coin with equal probability



Compare :

- Pr(toss 11 = H)
- Pr (toss 11 = H|first 10 tosses are heads)



Independence vs. pairwise independence

- Two independent fair coin tosses
  - H_1: First toss is H
  - H_2 : Second toss is H
- C: the two tosses had the same result

HH	HT
TH	TT

H_1 = {HH, HT} , H_2 = {HH, TH}, C = {HH,TT}

Pr(H_1) = 1/2, Pr(H_2) = 1/2 , Pr(C)=1/2

Pr(H1∩C) = P(H1 ∩ H2) =Pr(H2 ∩C) = 1/4

Pr(H1 ∩ C ∩ H2) =Pr(HH) = 1/4

H_1, H_2, and C are pairwise independent, but not independent

Exercise

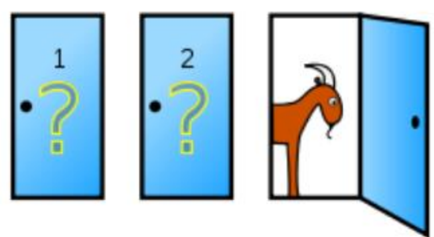
- Suppose $P(A) = 0.4$ and $P(B) = 0.6$. Are A and B mutually exclusive?

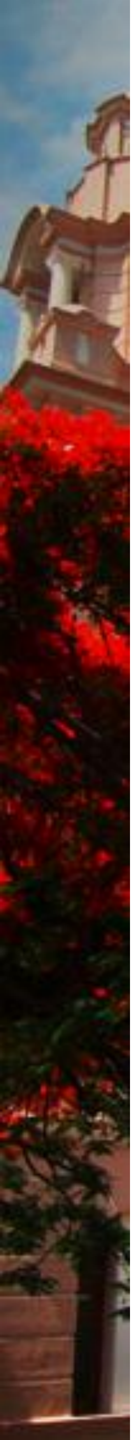
- Suppose A and B are mutually exclusive and $P(A) = 0.4$. Then P(B) = ?

- (T or F) If events A and B are dependent, it is impossible to find event C that is independent of A but not independent of B.

The Monty Hall Problem

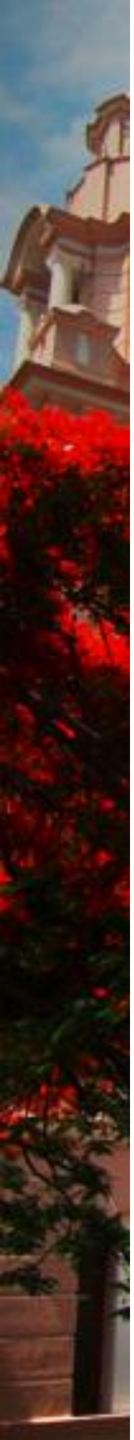
- Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?





Solve with Bayes' Theorem

You picked door 1	The car is behind door 1	The car is behind door 2	The car is behind door 3	Marginal
Host opens door 1				
Host opens door 2				
Host opens door 3				
Marginal				

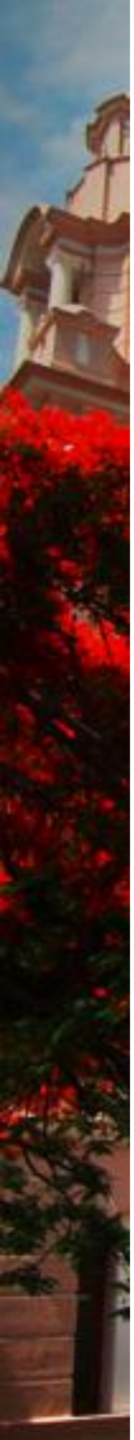


Café du Donut

- The Café buys donuts each day for $40 per carton of 20 dozen donuts. Any cartons not sold are thrown away at the end of the day. If a carton is sold, the revenue is $60.

- The salesperson is faced with two kinds of demand situations and needs to decide order size each day.

- Suppose that the order size (Q) can only be between 6 and 7 due to the storage capacity and delivery capacity.



Café du Donut

- The salesperson's initial belief is that two demand situations are equally likely.

DAILY DEMAND (CARTONS)	PROBABILITY UNDER LOW DEMAND	PROBABILITY UNDER HIGH DEMAND	MARGINAL PROBABILITY
4	0.25	0.05	0.15
5	0.20	0.10	0.15
6	0.15	0.10	0.125
7	0.15	0.15	0.15
8	0.10	0.15	0.125
9	0.10	0.20	0.15
10	0.05	0.25	0.15

- On the first day, should the order size be 6 or 7?

Café du Donut

- Monetary Payoff (Profit) Table

	D = 4	D = 5	D = 6	D = 7	D = 8	D = 9	D = 10	EMV
Q = 6	0	60	120	120	120	120	120	93
Q = 7	-40	20	80	140	140	140	140	87.5
Prob.	0.15	0.15	0.125	0.15	0.125	0.15	0.15	

- If the salesperson finds that the demand in first day is 8, should he increase the order size from 6 to 7?
  - Here, we assume that the distribution of the demand situation on the second day is the same as the distribution of the situation on the first day.

Café du Donut

- The salesperson updates his belief:

	Low Demand	High Demand
$D_1 = 8$	0.1*0.5	0.15*0.5

DAILY DEMAND (CARTONS)	PROBABILITY UNDER LOW DEMAND	PROBABILITY UNDER HIGH DEMAND
4	0.25	0.05
5	0.20	0.10
6	0.15	0.10
7	0.15	0.15
8	0.10	0.15
9	0.10	0.20
10	0.05	0.25

Café du Donut

- The salesperson then updates monetary payoff (profit) table

	D = 4	D = 5	D = 6	D = 7	D = 8	D = 9	D = 10
Q = 6	0	60	120	120	120	120	120
Q = 7	-40	20	80	140	140	140	140
Prob.							

Tossing Coins

- What is the probability of getting a head?
- Suppose there are three possible cases: 1/3, 1/2, and 2/3.

- What if we can toss the coin many times?
- Please refer to the “TossCoin.py”.