เฉลยคำตอบพร้อมวิธีทำ

Midterm Exam วิชา 2110205 Statistics for Computer Engineering

AUTHOR

Problem 1

Which statement is correct about Bayesian's view of probability?

1) The probability of an event can be found by experimenting for a large amount of times.
2) The probability of an event can be different depending on the person.
3) Both are correct.
4) Both are incorrect.

Solution

มุมมองความน่าจะเป็นของ Bayesian เป็นการมองโจทย์ปัญหาทางหลักการนับและสถิติโดยยึดความเชื่อมั่นจากความรู้และสิ่งที่เชื่อ
มุมมองของ Frequentist เป็นการมองโจทย์ปัญหาทางหลักการนับและสถิติโดยยึดจากการทดลองและสังเกตการณ์เป็นจำนวนมาก
สำหรับข้อนี้จึงควรเลือกตอบข้อที่ 2) เพราะเป็นการหาความน่าจะเป็นจากสิ่งที่ขึ้นอยู่กับแต่ละคน
ตอบ $\boxed{2)}$

Problem 2

Which student is applying statistical analysis?

1) Somnuke picks the horse that is most likely to win by using data from the past 3 years.
2) Pawit assumes each horse having equal chance of winning. Then, he writes a computer program that will determine the best betting strategy via simulation.
3) Alan always pick the horse which is his lucky number.
4) Bart uses a 10-sided die to determine the horse to bet on.

Solution

Statistical Analysis คือการวิเคราะห์ทางสถิติซึ่งต้องอาศัยผลการทดลอง/สังเกต/การบันทึกของที่ผ่านมาเพื่อวิเคราะห์ข้อมูลเพิ่มเติมให้นำมาซึ่งคำตอบ ซึ่งตรงกับข้อที่ 1) เนื่องจากข้ออื่นเป็นการใช้ข้อมูลที่ไม่จำเป็นต้องเก็บข้อมูลเพิ่มเติมเพื่อหาคำตอบและยังมีความน่าจะเป็นที่เท่ากันในทุกเหตุการณ์อย่างแน่นอน
ตอบ $\boxed{1)}$

Problem 3

In pandas, what does the function .unique() do?

1) Returns a random unique value for every row.
2) Remove all non-unique values in a dataframe.
3) Return a list of unique values in a column.
4) Sorts the dataframe using the unique values.

Solution

ตอบ $\boxed{3)}$

Problem 4

Given a dataframe

	Type	Speed
Falcon	0	200
	1	240
Parrot	0	100
	1	110

If we apply the command, `df.groupby("Type").mean()` what would be the output?

	Speed		Speed		Type	Speed			Type
Speed		Falcon	220		0	150		Falcon	0.5
162.5		Parrot	105		1	175		Parrot	0.5
1)		**2)**			**3)**			**4)**	

Solution

การทำงานของคำสั่ง `.groupby("Type")` จะเป็นการนำทุกค่าที่อยู่ใน `"Type"` เดียวกันมา Aggregation (การรวมกลุ่ม) โดยการหาค่า Mean ทำให้สามารถคำนวณออกมาได้ดังนี้:

$$\text{Type}_0 = \frac{200 + 100}{2} = 150$$
$$\text{Type}_1 = \frac{240 + 110}{2} = 175$$

ตอบ $\boxed{3)}$

Problem 5

According to the following code, what is the probability that the speed is 100?

```
import random
speed = 50
if random.randint(0,2): # uniform discrete from 0 to 2
    speed = speed*2
if random.random() > 1: # uniform continuous from 0 to 1
    speed = speed*2
print(speed)
```

1) 25%
2) 33%
3) 50%
4) 67%

Solution

เมื่อเราอ่าน Code ที่ปรากฏข้างต้นเราจะทราบว่ามีเพียง If-block แรกทำงานเท่านั้น เพราะ If-block ที่ 2 เป็นการสุ่มค่าในลักษณะของ continuous ที่เป็นไปได้ตั้งแต่ $[0.0, 1.0]$ ซึ่งการที่ If-block ที่ 2 ทำงานได้นั้นค่าต้อง > 1 **<u>ซึ่งไม่มีทางเกิดขึ้นจริง</u> จึงไม่ต้องนำมาพิจารณา** ฉะนั้นจะมีเพียงแค่เหตุการณ์ที่สุ่มออกมาแล้วทำให้ `speed = 100` นั้นมีเพียง 2 เหตุการณ์จากทั้งหมด 3 นั่นคือ 1, 2 ทำให้ทราบว่าความน่าจะเป็นที่จะเกิดเหตุการณ์ดังกล่าวคือ $\frac{2}{3} \approx 0.66\overline{6} \approx 67\%$

ตอบ $\boxed{4)}$

Problem 6

100 students were asked to fill the following survey. 40% answer "C++", 20% answer "Java", and 10% answer "Both". How many students can code in at least one language?

Survey: *Which language can you code in? (Choose one answer.)*

☐ C++ ☐ Java ☐ Both ☐ Neither

1) 30
2) 50
3) 70
4) 90

Solution

โจทย์ข้อนี้ได้มีการระบุข้อจำกัดในแบบสำรวจแล้วว่าคน ๆ หนึ่งเลือกตอบได้เพียง 1 คำตอบเท่านั้น นั่นทำให้เราเก็บข้อมูลแยกรายตัวเลือกได้ทันทีโดยไม่ต้องทำการคำนวณที่ซับซ้อน

<u>Solution-1</u>

สามารถคำนวณได้โดยตรงนำยอดที่กรอกแต่ละแบบมาบวกได้โดยตรง ซึ่งคิดได้จาก $40 + 20 + 10 = 70$ คน

(มีต่อหน้าถัดไป...)

Solution-2
หากมองให้เป็นลักษณะของ Venn Diagram จะได้ว่า...

กำหนดให้ $A :=$ คนที่เขียนภาษา C++ ได้

$B :=$ คนที่เขียนภาษา Java ได้

จากสิ่งที่กำหนดทำให้ทราบเพิ่มเติมว่า $n(A-B)=40, n(B-A)=20, n(A\cap B)=10$ โจทย์ต้องการทราบว่า $n(A\cup B)$ มีค่าเท่าใด ทำให้สามารถคิดโดยนำมาบวกได้โดยตรง คิดจาก $40+20+10=70$ คน ตอบ $\boxed{3)}$

Problem 7

Consider the following Python code.

```
s=1
for i in range(1,n):
  s=s*i
for i in range(1,k):
  s=s/i
print(s)
```

What does this code compute?

1) ${}_nP_k$
2) ${}_{n-1}P_{k-1}$
3) ${}_nP_{n-k}$
4) ${}_{n-1}P_{n-k}$

Solution

ทำการวิเคราะห์ว่าจาก Python Code ที่ระบุข้างต้นทำงานได้อย่างไร โดยเริ่มต้นจาก...

- For-block แรกทำงานตั้งแต่ $[1, n-1]$ ทำให้เรามีค่าตัว s อยู่ที่ $1\times 2\times ...\times (n-1)=(n-1)!$
- For-block ที่สองทำงานตั้งแต่ $[1, k-1]$ ทำให้เรามีค่าตัว s อยู่ที่ $1\times 2\times ...\times (k-1)=(k-1)!$
- เมื่อเขียนเป็นรูปสมการคณิตจะได้ว่า $\frac{1\times 2\times ...\times (n-1)}{1\times 2\times ...\times (k-1)}=\frac{(n-1)!}{(k-1)!}$

Permutation ของ ${}_nP_k$ สามารถถูกคำนวณได้โดย $\frac{n!}{(n-k)!}$ ให้ทำการจัดรูปของสมการที่อยู่ในแต่ละ Choice เพื่อเทียบว่าตัวเลือกใดตรงกับ $\frac{(n-1)!}{(k-1)!}$ โดยทำได้ดังนี้

(i) ${}_nP_k=\frac{n!}{(n-k)!}$

(ii) ${}_{n-1}P_{k-1}=\frac{(n-1)!}{((n-1)-(k-1))!}=\frac{(n-1)!}{(n-k)!}$

(iii) ${}_nP_{n-k}=\frac{n!}{(n-(n-k))!}=\frac{n!}{k!}$

(iv) ${}_{n-1}P_{n-k}=\frac{(n-1)!}{((n-1)-(n-k))!}=\frac{(n-1)!}{(-1+k)!}=\frac{(n-1)!}{(k-1)!}$ $\square$

(มีบอกคำแนะนำเพิ่มเติมหน้าถัดไป...)

เนื่องจากการจัดรูปเป็นวิธีที่ให้ผลถูกต้องที่สุดแต่อาจก่อให้เกิดความสับสนหรือทำช้ามาก ๆ เพื่อความรวดเร็วระหว่างการสอบ แนะนำแทนค่าสำหรับตัวแปร n, k มีค่าห่างที่ห่างกัน ณ ที่นี้ขอสมมติค่าเป็น $n = 8, k = 5$ จากรูปสมการเศษส่วนด้านบนจึงได้ว่า $\frac{7!}{4!}$ ถัดไปทำการหาว่าตัวเลือกในข้อสอบใดมีค่าตรงกับเศษส่วนข้างต้น ซึ่งแทนที่ตัวเลขได้ดังนี้...

(i) ${}_nP_k = {}_8P_5 = \frac{8!}{3!}$

(ii) ${}_{n-1}P_{k-1} = {}_7P_4 = \frac{7!}{3!}$

(iii) ${}_nP_{n-k} = {}_8P_3 = \frac{8!}{5!}$

(iv) ${}_{n-1}P_{n-k} = {}_7P_3 = \frac{7!}{4!}$

ตอบ $\boxed{4)}$

Problem 8

In September, 18 days are sunny, 14 days are rainy, and 5 days are neither sunny nor rainy. How many days in September are both sunny and rainy?

1) 3
2) 5
3) 7
4) 9

Solution

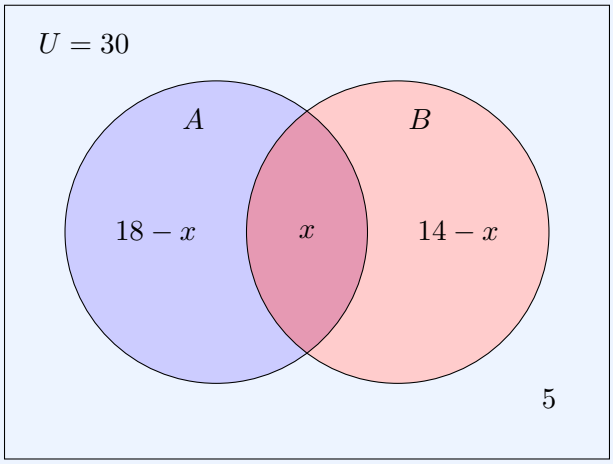
กำหนดให้ $n(A) :=$ จำนวนวันที่แดดจ้า

$n(B) :=$ จำนวนวันที่ฝนตก

โดยโจทย์ระบุข้อมูลแต่ละส่วนดังนี้ $n(A) = 18, n(B) = 14, n((A \cup B)') = 5$ โจทย์ต้องการหา $n(A \cap B)$ สามารถหาได้โดยใช้หลักการ Inclusion-Exclusion Principle นั่นคือ

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

จากข้อมูลที่มีในปัจจุบันเมื่อทำการวาด Venn Diagram จะได้ว่า



จากการที่เราทราบค่าของ $n((A \cup B)')$ จึงทำให้ทราบค่า $n(A \cup B) = n(U) - n((A \cup B)') = 30 - 5 = 25$

จึงทำให้เราหา $n(A \cap B)$ ได้โดยคำนวณดังต่อไปนี้

$$\begin{aligned} n(A \cup B) &= n(A) + n(B) - n(A \cap B) \\ 25 &= 18 + 14 - n(A \cap B) \\ n(A \cap B) &= 32 - 25 \\ n(A \cap B) &= 7 \end{aligned}$$

ตอบ $\boxed{3)}$

Problem 9

How many integers from 1 to 1,000 are divisible by 4 but not divisible by 5?

1) 150
2) 200
3) 230
4) 400

Solution

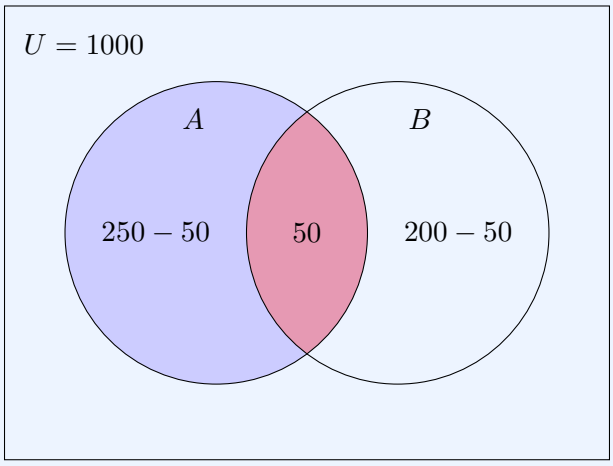
กำหนดให้ $n(A) :=$ จำนวนตัวเลขที่หาร 4 ลงตัว

$n(B) :=$ จำนวนตัวเลขที่หาร 5 ลงตัว

จากที่โจทย์ระบุ ทำให้ได้ข้อมูลแต่ละส่วนดังนี้ $n(A) = \lfloor \frac{1000}{4} \rfloor = 250, n(B) = \lfloor \frac{1000}{5} \rfloor = 200, n(A \cap B) = \lfloor \frac{1000}{lcm(4,5)} \rfloor = 50$ โจทย์ต้องการหา $n(A \cap B')$ สามารถหาได้โดย...

$$\begin{aligned} n(A \cap B') &= n(A) - n(A \cap B) \\ &= 250 - 50 \\ &= 200 \end{aligned}$$

สาเหตุที่สามารถหาโดยใช้ $n(A) - n(A \cap B)$ ได้นั่นเพราะความหมายของ $n(A \cap B')$ คือเอาของที่มีใน A แต่ไม่มีใน B ซึ่งหากลองวาดภาพ Venn Diagram ก็จะพบว่าสามารถคำนวณได้ด้วยวิธีข้างต้น



ตอบ $\boxed{2)}$

Problem 10

There are 5 men and 3 women in the CEDT student council. They want to select a president, first vice president, second vice president, and secretary. The first vice president must be a woman and the second vice president must be a man, while the president and secretary can be any gender. How many ways can they select the positions if each person can hold at most one position?

1) 360
2) 390
3) 420
4) 450

Solution

จากโจทย์ระบุว่า

- มีผู้ชาย 5 คน และผู้หญิง 3 คน
- ต้องการเลือกตำแหน่งโดยที่แต่ละคนมีได้มากสุด 1 ตำแหน่ง
- First vice president ต้องเป็นผู้หญิง และ Second vice president ต้องเป็นผู้ชาย

ทำการหยิบเลือกบุคคลที่มีข้อจำกัดมากที่สุดก่อนนั่นคือ First/Second vice president และอีก 2 ตำแหน่งที่เหลือไม่ระบุเงื่อนไขแต่สามารถมีได้เพียงแค่คนเดียว ทำให้คำนวณได้ดังนี้...

$$\underset{\text{(เลือก } 1^{st}\text{VP)}}{3} \times \underset{\text{(เลือก } 2^{nd}\text{VP)}}{5} \times \underset{\text{(เลือกคนเป็นประธาน)}}{6} \times \underset{\text{(เลือกคนเป็นเลขา)}}{5} = 450 \text{ วิธี}$$

สังเกตได้ว่าในช่วงแรก ๆ จะเลือก 2 ตำแหน่งที่มีข้อจำกัดมาก่อนจึงจัดวิธีได้ 3×5 ทำให้เราทราบว่ายังมีคนที่ยังรอเลือกอีก $8 - 1 - 1 = 6$ คน และอีก 2 ตำแหน่งถัดจากนี้ไม่มีข้อจำกัดใด ๆ ทำให้เราสามารถคำนวณต่อได้ด้วย 6×5 ซึ่งมาจาก 6 คนสุดท้ายในตำแหน่งที่ (3) และตำแหน่งที่ (4) ที่เหลือเพียง 5 คนเพราะว่าตำแหน่งที่ (3) ได้รับการเลือกเสร็จสิ้น
ตอบ $\boxed{4)}$

Problem 11

There are 5 men and 5 women in a class. They want to select 4 students to attend a conference, which must consist of at least one man and at least one woman. How many ways can they select students to attend the conference?

1) 140
2) 200
3) 210
4) 700

Solution

จากโจทย์ระบุว่า

- มีผู้ชาย 5 คน และผู้หญิง 5 คน
- มีคนต้องถูกเลือกไปประชุมทั้งสิ้น 4 คน
- ต้องการเลือกผู้ชายและผู้หญิงพร้อมกันอย่างน้อย 1 คนไปประชุม จัดได้กี่วิธี

ใช้หลักการลบแบบ Complement ซึ่งมีหลักดังต่อไปนี้

$$n(\text{กรณีที่สนใจ}) = n(\text{วิธีจัดกรณีไม่มีข้อจำกัด}) - \sum_{k=1}^{n(\text{กรณีที่ผิด})} n(\text{กรณีที่ผิด}_k)$$

ในที่นี้มีกรณีที่ผิดกฎอยู่ 2 แบบคือเลือกเพียงผู้ชายเท่านั้นกับเลือกเพียงผู้หญิงเท่านั้นซึ่งสามารถคำนวณได้ดังนี้

$$\text{กรณีเลือกโดยไม่มีข้อจำกัด} = \binom{10}{4} = 210 \text{ วิธี}$$
$$\text{กรณีเลือกแต่ผู้ชาย} = \binom{5}{4} = 5 \text{ วิธี}$$
$$\text{กรณีเลือกแต่ผู้หญิง} = \binom{5}{4} = 5 \text{ วิธี}$$

$\therefore n(\text{กรณีที่สนใจ}) = 210 - 5 - 5 = 200$ วิธี
ตอบ $\boxed{2)}$

Problem 13

The CU equestrian club has 12 members. 3 of them are first-year students, 4 are second-year students, and 5 are third-year students. If we randomly select 3 club members, what is the probability that they are all in the same year?

1) 1/10
2) 1/20
3) 3/11
4) 3/44

Solution

จากโจทย์ระบุว่า

- ชมรมมีสมาชิกทั้งสิ้น 12 คน
- เป็นนักเรียนปี 1 จำนวน 3 คน
- เป็นนักเรียนปี 2 จำนวน 4 คน
- เป็นนักเรียนปี 3 จำนวน 5 คน
- หาความน่าจะเป็นของต้องการเลือก 3 คนและมีแต่ปีเดียวกัน

ทำให้เราสามารถหาค่า $n(S) = \binom{12}{3} = 220$ และเราสามารถหาค่าของ $n(E)$ ได้โดยการรวมทุกกรณีย่อย ซึ่งสามารถคิดแต่ละกรณีย่อยได้ดังนี้...

$$\text{เลือกเพียงปี } 1 = \binom{3}{3} = 1 \text{ วิธี}$$
$$\text{เลือกเพียงปี } 2 = \binom{4}{3} = 4 \text{ วิธี}$$
$$\text{เลือกเพียงปี } 3 = \binom{5}{3} = 10 \text{ วิธี}$$

จึงรวมค่า $n(E) = 1 + 4 + 10 = 15 \therefore$ ความน่าจะเป็นจึงมีค่า $= \frac{15}{220} = \frac{3}{44}$
ตอบ $\boxed{4)}$

Problem 14

4 players are playing Among Us. Each player can vote for one other player (but cannot vote for him/herself). If every player votes randomly, what is the probability that there is a player who receives exactly 3 votes.

1) 1/81
2) 1/27
3) 4/27
4) 4/81

Solution

จากโจทย์ระบุว่า

- มีผู้เล่นในเกม Among Us ทั้งสิ้น 4 คน
- แต่ละคนโหวตได้เพียงแค่ 1 คนเท่านั้น (โหวตตัวเองไม่ได้)
- หาความน่าจะเป็นที่ผู้เล่นคนหนึ่งถูกโหวต 3 ครั้งพอดี

จากข้อมูลที่ได้รับทำให้ทราบว่า $n(S) = 3^4 = 81$ การหาค่า $n(E)$ สามารถหาได้จากการนับจำนวน**วิธีการถูกโหวต**ของผู้เล่นหมายเลขนั้น ๆ ซึ่งคำนวณได้จาก

$$n(\text{โหวตแต่ผู้เล่น No.1}) = \underset{\text{(ผู้เล่นคนที่ 1 เลือกได้)}}{3} \times \underset{\text{(ผู้เล่นที่เหลือเลือกได้)}}{1} = 3$$

$$\vdots \qquad\qquad \vdots$$

$$n(\text{โหวตแต่ผู้เล่น No.4}) = \underset{\text{(ผู้เล่นคนที่ 4 เลือกได้)}}{3} \times \underset{\text{(ผู้เล่นที่เหลือเลือกได้)}}{1} = 3$$

นำกรณีที่โหวตแต่ละผู้เล่นมารวมกันจะได้ 3×4 (จำนวนผู้เล่นทั้งหมด) $= 12$ วิธี
$\therefore$ ความน่าจะเป็นจึงมีค่า $\frac{12}{81} = \frac{4}{27}$
ตอบ $\boxed{3)}$

Problem 15

There are 3 dice. Each one have an equal probability to land on each of its face.

- The first die (singular form of dice) has 6 faces, with numbers 1, 2, 3, 4, 5, 6.
- The second die has 5 faces, with numbers 1, 2, 3, 4, 5.
- The third die has 4 faces, with numbers 1, 2, 3, 4.

Mint rolls these dice. What is the probability that the three shown numbers are all different?

1) 1/3
2) 2/3
3) 7/15
4) 8/15

Solution

จากสิ่งที่โจทย์ระบุเราทราบว่าการออกแต่ละหน้าในแต่ละลูกเต๋ามีโอกาสที่เท่ากันเสมอ (กล่าวง่าย ๆ คือ ลูกเต๋าแต่ละลูกมีโอกาสออกแต่ละหน้าที่ $\frac{1}{6}, \frac{1}{5}, \frac{1}{4}$ ตามลำดับเสมอ) จึงสามารถหา $n(S)$ ได้ว่า $4 \times 5 \times 6 = 120$ วิธี เพราะว่าแต่ละตัวที่คูณนั้นมาจากผลลัพธ์ที่ลูกเต๋าแต่ละลูกสามารถออกมาได้

ส่วนถัดไปคือการหาค่า $n(E)$ สามารถหาได้โดยใช้ 2 มุมมองนั่นคือหยิบลูกเต๋าหมายเลข (1) มาคิดก่อน กับหยิบลูกเต๋าลูกเต๋าหมายเลขที่ติดข้อจำกัดมากที่สุดก่อนในที่นี้คือหมายเลข (3)

..... *** .....

<u>มุมมองที่ 1</u>

หากเราคำนวณโดยเริ่มจากหยิบลูกเต๋าหมายเลข (1) มาก่อนนั้น จะต้องพิจารณาถึงกรณีย่อย ๆ มากพอสมควร สำหรับการเฉลย Solution นี้จะใช้สัญลักษณ์ดังต่อไปนี้ (X, Y, Z) โดยที่หมายถึงหมายเลขหน้าลูกเต๋าของหมายเลข (1), (2), (3) ตามลำดับ ซึ่งสามารถที่จะแบ่งกรณีออกได้ดังนี้...

$$(6,\ 5,\ 1\text{-}4);\quad \underset{\text{ลูกเต๋า (1) เลือกได้}}{1} \times \underset{\text{ลูกเต๋า (2) เลือกได้}}{1} \times \underset{\text{ลูกเต๋า (3) เลือกได้}}{4} = 4 \text{ วิธี}$$

$$(6,\ 1\text{-}4,\ 1\text{-}4);\quad \underset{\text{ลูกเต๋า (1) เลือกได้}}{1} \times \underset{\text{ลูกเต๋า (2) เลือกได้}}{4} \times \underset{\text{ลูกเต๋า (3) เลือกได้}}{3} = 12 \text{ วิธี}$$

$$(5,\ 1\text{-}4,\ 1\text{-}4);\quad \underset{\text{ลูกเต๋า (1) เลือกได้}}{1} \times \underset{\text{ลูกเต๋า (2) เลือกได้}}{4} \times \underset{\text{ลูกเต๋า (3) เลือกได้}}{3} = 12 \text{ วิธี}$$

$$(1\text{-}4,\ 5,\ 1\text{-}4);\quad \underset{\text{ลูกเต๋า (1) เลือกได้}}{4} \times \underset{\text{ลูกเต๋า (2) เลือกได้}}{1} \times \underset{\text{ลูกเต๋า (3) เลือกได้}}{3} = 12 \text{ วิธี}$$

$$(1\text{-}4,\ 1\text{-}4,\ 1\text{-}4);\quad \underset{\text{ลูกเต๋า (1) เลือกได้}}{4} \times \underset{\text{ลูกเต๋า (2) เลือกได้}}{3} \times \underset{\text{ลูกเต๋า (3) เลือกได้}}{2} = 24 \text{ วิธี}$$

โดยสรุปคือเราจะหยิบกรณีย่อยเพื่อแบ่งกรณีให้รัดกุมและครอบคลุมที่สุดเพื่อป้องกันปัญหา Overcounting (การนับซ้ำ) โดยเราจะเลือกให้ลูกเต๋าหมายเลข (1) เลือกที่จะหยิบเลขใด ๆ ก็ได้แต่เพื่อความง่ายและลดความสับสน จึงเลือกที่จะหยิบหมายเลขที่อนาคตจะก่อให้เกิดข้อจำกัดมากที่สุดไปยังน้อยที่สุดก่อน เพื่อให้สามารถตรวจสอบได้ง่ายมากขึ้น ฉะนั้นเมื่อนำแต่ละกรณีรวมกันจะได้ว่า $n(E) = 4 + 12 + 12 + 12 + 24 = 64$ วิธี

..... *** .....

<u>มุมมองที่ 2</u> ให้เริ่มหยิบดูผลลัพธ์ที่สามารถออกได้ของลูกเต๋าที่มีข้อจำกัดมากที่สุดก่อน ในที่นี้คือลูกเต๋าหมายเลข (3), (2), (1) ตามลำดับ ซึ่งสามารถคำนวณวิธีได้ดังนี้

$$\underset{\text{ลูกเต๋า (3) ออกผลได้}}{4} \times \underset{\text{ลูกเต๋า (2) ออกผลได้}}{4} \times \underset{\text{ลูกเต๋า (1) ออกผลได้}}{4} = 4^3 = 64 \text{ วิธี}$$

เหตุผลที่ลูกเต๋าแต่ละลูกออกได้เพียง 4 หมายเลขในทุกลูกเต๋านั้นเป็นเพราะว่า

(i) ลูกเต๋าหมายเลข (3): เลือกหมายเลขได้ $1 \sim 4$

(ii) ลูกเต๋าหมายเลข (2): เลือกหมายเลขได้ $1 \sim 5$ แต่ต้องเอาออก 1 ตัวที่เคยถูกเลือกแล้ว ทำให้เลือกได้จริง $5 - 1 = 4$

(iii) ลูกเต๋าหมายเลข (1): เลือกหมายเลขได้ $1 \sim 6$ แต่ต้องเอาออก 2 ตัวที่เคยถูกเลือกแล้ว ทำให้เลือกได้จริง $6 - 2 = 4$

เมื่อเราทราบค่า $n(E), n(S)$ แล้วจึงสามารถหาค่าความน่าจะเป็นได้ $\frac{64}{120} = \frac{8}{15}$

ตอบ $\boxed{4)}$

Problem 17

63% of CEDT students pass both the midterm and final exams of statistics class. Among those who pass the midterm exam, 90% of them also pass the final exam. What percentage of CEDT students pass the midterm exam?

1) 53%
2) 56.7%
3) 70%
4) 73%

Solution

กำหนดให้ $M :=$ นิสิตสอบผ่าน Midterm

$F :=$ นิสิตสอบผ่าน Final

จากโจทย์ระบุว่า $P(F \cap M) = 0.63, P(F|M) = 0.9$ โจทย์ต้องการทราบ $P(M)$ โดยสามารถหาได้โดยใช้ Bayes Theorem ได้ดังนี้

$$\begin{aligned} P(F|M) &= \frac{P(M|F)P(F)}{P(M)} \\ &= \frac{P(M \cap F)}{P(M)} \end{aligned}$$

ทำการแทนค่าตัวแปรต่าง ๆ และแก้สมการจึงจะได้ว่า

$$\begin{aligned} 0.9 &= \frac{0.63}{P(M)} \\ P(M) &= 0.7 \text{ (หรือ } 70\%) \end{aligned}$$

ตอบ $\boxed{3)}$

Problem 19

Among all companies in the stock market, 20% are large companies, 30% are medium companies, and 50% are small companies. Last year, the stock price of 80% of large companies, 70% of medium companies, and 30% of small companies went up. In total, the stock price of what percentage of companies in the stock market went up last year?

1) 52%
2) 56%
3) 60%
4) 64%

Solution

Solution-1: หลักการคูณอัตราส่วน (Ratio)
จากโจทย์ที่กำหนด สมมติว่ามีบริษัทอยู่ทั้งสิ้น 100 แห่ง (เพราะค่าทุกอย่างอยู่ในรูปแบบร้อยละ) ซึ่งมีบริษัทขนาดใหญ่ 20 แห่ง บริษัทขนาดกลาง 30 แห่ง และบริษัทขนาดเล็ก 50 แห่ง จากนั้นคำนวณจำนวนบริษัทที่ราคาหุ้นเพิ่มขึ้นในแต่ละขนาด ได้ว่าดังนี้

$$\begin{aligned} n(\text{บริษัทขนาดใหญ่ที่หุ้นเพิ่มขึน}) &= 20(0.8) = 16 \text{ บริษัท} \\ n(\text{บริษัทขนาดกลางที่หุ้นเพิ่มขึน}) &= 30(0.7) = 21 \text{ บริษัท} \\ n(\text{บริษัทขนาดเล็กที่หุ้นเพิ่มขึน}) &= 50(0.3) = 15 \text{ บริษัท} \\ \therefore n(\text{จำนวนบริษัทที่หุ้นเพิ่มขึ้น}) &= 16 + 21 + 15 = 52 \text{ บริษัท} \end{aligned}$$

จากบริษัททั้งหมด 100 แห่งเมื่อคิดเป็นร้อยละจะทำให้เราทราบยอดร้อยละของบริษัทที่มีหุ้นเพิ่มขึ้นดังนี้

$$\frac{\text{จำนวนบริษัทที่หุ้นเพิ่มขึ้น}}{\text{จำนวนบริษัททั้งหมด}} = \frac{52}{100} = 52\% \quad \square$$

Solution-2: หลักการ Law of Total Probability & Bayes Theorem

กำหนดให้ $L :=$ บริษัทขนาดใหญ่ $\quad M :=$ บริษัทขนาดกลาง

$S :=$ บริษัทขนาดเล็ก $\quad U :=$ หุ้นบริษัทเพิ่มขึ้น

โจทย์ต้องการทราบค่า $P(U)$ หาได้โดยใช้หลักการ Law of Total Probability ดังนี้

$$\begin{aligned} P(U) &= P(U|L)P(L) + P(U|M)P(M) + P(U|S)P(S) \\ &= 0.8(0.2) + 0.7(0.3) + 0.3(0.5) \\ &= 0.52 \text{ (หรือ } 52\%) \end{aligned}$$

ตอบ $\boxed{1)}$

Questions 29–30. *Students in a particular university are either in the CEDT or CP program. 60% of CEDT students and 30% of CP students can write JavaScript. Overall, 48% of students in this university can write JavaScript.*

Problem 29

If a student can write JavaScript, what is the probability that this student is in CEDT program?

1) 3/4
2) 4/5
3) 5/6
4) 6/7

Solution

กำหนดให้ $A :=$ นิสิต CEDT

$B :=$ นิสิต CP

$J :=$ เขียน JavaScript ได้

จากโจทย์ระบุว่า $P(J) = 0.48, P(J|A) = 0.6, P(J|B) = 0.3$ สำหรับข้อนี้โจทย์ต้องการหา $P(A|J)$ ใช้ทฤษฎีบทของ Bayes Theorem ในการหาค่าความน่าจะเป็นย้อนกลับจาก ซึ่งได้ว่า $P(A|J) = \frac{P(J|A)P(A)}{P(J)}$ แต่ตอนนี้เรายังไม่ทราบค่าของ $P(A)$ จึงยังทำให้หาค่าไม่ได้ สำหรับโจทย์ข้อนี้วิธีหา $P(A)$ ที่ง่ายที่สุดคือหาจาก $P(J)$ โดยใช้หลักการของ Law of Total Probability ทำให้สามารถคำนวณออกมาได้ดังนี้

$$\begin{aligned}
P(J) &= P(J|A)P(A) + P(J|B)P(B) \\
0.48 &= 0.6x + 0.3(1-x) \\
0.48 &= 0.6x + 0.3 - 0.3x \\
0.3x &= 0.18 \\
x &= \frac{0.18}{0.3} = 0.6
\end{aligned}$$

โดยจากการแก้สมการข้างต้นนั้นได้แทนที่ $x = P(A)$ และ $P(B) = 1 - x$ เพราะ $P(B)$ สามารถหาได้โดยการนำจำนวนยอดร้อยละทั้งหมดมาหักหับยอดความน่าจะเป็นของการเกิด $P(A)$ เมื่อรู้เช่นนั้นแล้วจึงแก้สมการออกมาและได้ผลลัพธ์ $P(A) = 0.6$

เมื่อได้ค่า $P(A)$ แล้วจึงนำกลับมาแทนค่าในสมการของ Bayes Theorem ต่อดังนี้

$$\begin{aligned}
P(A|J) &= \frac{P(J|A)P(A)}{P(J)} \\
&= \frac{0.6 \times 0.6}{0.48} \\
&= 0.75 \text{ (หรือ } \frac{3}{4}\text{)}
\end{aligned}$$

ตอบ $\boxed{1)}$

Problem 30

If a student cannot write JavaScript, what is the probability that this student is in CEDT program?

1) 3/13
2) 4/13
3) 5/13
4) 6/13

Solution

โจทย์ข้อนี้ต้องการทราบ $P(A|J')$ ซึ่งสามารถคิดได้จาก Bayes Theorem ดังนี้

$$\begin{aligned} P(A|J') &= \frac{P(J'|A)P(A)}{P(J')} \\ &= \frac{0.4 \times 0.6}{0.52} \\ &\approx 0.46153 \left(\text{หรือ } \frac{6}{13}\right) \end{aligned}$$

ตอบ $\boxed{4)}$

Notes: เราสามารถหาค่าของ $P(J'|A)$ ได้หากทราบค่า $P(J|A)$ โดยการนำ $1-P(J|A) = 1-0.6 = 0.4$

Problem 31

Which of the following is NOT a random variable?

1) The face value of a die (singular form of dice).
2) The hand of a poker.
3) The result of a quadratic formula.
4) The lottery outcome.

Solution

Random Variable (หรือตัวแปรสุ่ม) ฟังก์ชันจาก Sample Space ของการทดลองสุ่มไปยัง Set ของจำนวนจริง ซึ่ง Choice 3) ไม่สามารถ mapping ไปยัง Set ของจำนวนจริงได้โดยตรงเพราะ**ผลลัพธ์ของสมการกำลังสองนั้นมีค่าที่เป็นทั้งบวกและลบ**

ตอบ $\boxed{3)}$

Problem 32

What does the following random variable represent?

X	1	2	3	4	5	6
pmf	1/6	1/6	1/6	1/6	1/6	1/6

1) A die (singular form of dice).
2) A coin.
3) A card in a deck.
4) A telephone number.

Solution

ข้อนี้ตรงกับเหตุการณ์ของ Choice 1) มากที่สุดเพราะมีจำนวนเหตุการณ์ที่เกิดขึ้นได้มากที่สุด 6 เหตุการณ์ และทุกเหตุการณ์มีโอกาสเกิดเท่ากัน นั่นคือลูกเต๋าแต่ละหน้ามีโอกาสออกแต่ละแบบอยู่ที่ $\frac{1}{6}$

ตอบ $\boxed{1)}$

Questions 33–36. *Use the following table.*

X	1	2	3	4	5	6
pmf	1/12	2/12	1/12	5/12	a	1/12
cdf	1/12	b	4/12	9/12	11/12	12/12

Problem 33

What is the value of a?

1) 1/12
2) 2/12
3) 3/12
4) 4/12

Solution

หาค่าของ a ได้โดยอาศัยสมบัติของการเปลี่ยนค่าตัวแปรสุ่มระหว่าง $4 \Rightarrow 5$ ของ *cdf* จึงทำให้ทราบได้ว่า *pmf* ของ $P(X = 5)$ คิดได้จาก *cdf* โดย $\frac{11}{12} - \frac{9}{12} = \frac{2}{12}$
ตอบ $\boxed{2)}$

Problem 34

What is the value of b?

1) 1/12
2) 2/12
3) 3/12
4) 4/12

Solution

หาค่าของ a ได้โดยอาศัยสมบัติของการเปลี่ยนค่าตัวแปรสุ่มระหว่าง $1 \Rightarrow 2$ ของ *pmf* จึงทำให้ทราบได้ว่า *cdf* ของ $P(X = 2)$ คิดได้จาก *pmf* โดย $\frac{2}{12} - \frac{1}{12} = \frac{1}{12}$
ตอบ $\boxed{1)}$

Problem 35

What is the value of pmf $p(X = 3.5)$?

1) $-1/6$
2) 0
3) 1/6
4) None of above.

Solution

ตารางดังกล่าวเป็น Discrete Random Variable ทำให้ไม่มีค่า ณ ตัวเลขที่ไม่ได้ปรากฏในตาราง
ตอบ $\boxed{4)}$

Problem 36

What is the expected value of this random variable?

1) 3.667
2) 7.333
3) 3.250
4) 6.500

Solution

สามารถคำนวณค่า Expected Value (หรือ $E[X]$) โดยใช้สมการทางคณิตศาสตร์คือ...

$$E[X] = \sum_{\text{x} \in \text{range(X)}} x \cdot P(X = x)$$

ดังนั้นจึงสามารถคำนวณได้ดังนี้...

$$\begin{aligned} E[X] &= 1(\frac{1}{12}) + 2(\frac{2}{12}) + 3(\frac{1}{12}) + 4(\frac{5}{12}) + 5(\frac{2}{12}) + 6(\frac{1}{12}) \\ &= \frac{1}{12} + \frac{4}{12} + \frac{3}{12} + \frac{20}{12} + \frac{10}{12} + \frac{6}{12} \\ &= \frac{44}{12} \\ &\approx 3.666\overline{6} \ (\approx 3.667) \end{aligned}$$

ตอบ $\boxed{1)}$

Problem 37

Which random variable CANNOT be modeled with Bernoulli distribution?

1) HEAD or TAIL of a coin.
2) Whether a person has no phone, a feature phone, or a smart phone.
3) Throw a die (singular form of dice). Did you get HIGH face (4,5,6) or LOW face (1,2,3)?
4) Whether you answer this question correctly or not.

Solution

Bernoulli distribution คือการแจกแจงความน่าจะเป็นของเหตุการณ์ที่เกี่ยวกับผลลัพธ์การทดลอง/สังเกต ซึ่งต้องมีผลลัพธ์เพียง 2 อย่างคือสำเร็จ/ล้มเหลวเท่านั้น
จาก Choice ที่มีให้จึงพบว่า **2)** ไม่สามารถนำมาใช้กับโจทย์ประเภท Bernoulli distribution มากที่สุด เพราะ**ผลลัพธ์ที่ได้จากบันทึก/สังเกตมีมากกว่า** 2 **แบบ**
ตอบ $\boxed{2)}$

Problem 38

Which random variable CANNOT be modeled with Binomial distribution?

1) HEAD or TAIL of a coin.
2) Number of HEADs when throwing multiple coins.
3) The number of people having COVID-19 in the entire population.
4) All of above can be modelled using Binomial distribution.

Solution

Binomial distribution คือการแจกแจงความน่าจะเป็นของการเกิดเหตุการณ์ที่จะสำเร็จ X ครั้ง จากการทดลอง n ครั้ง และ**โอกาสในการลองที่จะลองต้องมีโอกาสที่เท่ากันในทุกเหตุการณ์เสมอ**
จาก Choice ที่มีให้จึงพบว่าทุกข้อสามารถใช้ Binomial distribution ทั้งหมดเพราะทุกอันข้างต้นสามารหาและใช้ถูกหาโดยอาศัยความน่าจะเป็นที่จะสำเร็จ/ล้มเหลวเท่ากันเสมอในทุกเวลา
ตอบ $\boxed{4)}$

Problem 39

Can the number of question YOU answered correctly on this test be modeled using a Binomial distribution?

1) Yes. It is a sum of multiple trials.
2) Yes. It is the number of successful challenges.
3) No. Each problem has different probability of you answering correctly.
4) No. Binomial can only be modelled when the probability of the challenge is known.

Solution

ข้อนี้ Choice **3)** เป็นเพียงข้อเดียวไม่สามารถใช้งาน Binomial distribution ได้เพราะโอกาสในการตอบแต่ละข้อมีค่าที่ไม่เท่ากัน (ตอบถูก +1 ตอบผิด -0.5)
ตอบ $\boxed{3)}$

Problem 40

Which random variable CANNOT be modeled using Uniform distribution?

1) The result of throwing a die (singular form of dice).
2) The sum of the results from throwing two dice.
3) HEAD or TAIL of a coin.
4) The result of a lottery.

Solution

Uniform distribution คือการแจกแจงความน่าจะเป็นในทุก ๆ เหตุการณ์มีค่าเท่ากันเสมอตลอดเวลา (เช่นทอยเหรียญและโยนลูกเต๋าที่ยุติธรรม)
จาก Choice ที่มีให้จึงพบว่า **2)** ไม่สามารถนำมาใช้กับโจทย์ประเภท Uniform distribution มากที่สุด เพราะผลลัพธ์ที่เกิดขึ้นจากการรวมหน้าบนลูกเต๋าแต่ละชุดมีค่าไม่เท่ากัน เช่น $n(E)$ ของ $(1,3) \neq n(E)$ ของ $(2,4)$
ตอบ $\boxed{2)}$

Questions 41–42. *Use the following table.*

X	1	2	3	4
pmf	1/6	2/6	1/6	2/6

Problem 41

What is $E(X)$?

1) 2.333
2) 2.500
3) 2.667
4) 3.000

Solution

สามารถคำนวณค่า Expected Value (หรือ $E[X]$) โดยใช้สมการทางคณิตศาสตร์คือ...

$$E[X] = \sum_{\text{x} \in \text{range(X)}} x \cdot P(X = x)$$

ดังนั้นจึงสามารถคำนวณได้ดังนี้...

$$\begin{aligned} E[X] &= 1(\frac{1}{6}) + 2(\frac{2}{6}) + 3(\frac{1}{6}) + 4(\frac{2}{6}) \\ &= \frac{1}{6} + \frac{4}{6} + \frac{3}{6} + \frac{8}{6} \\ &= \frac{16}{6} \\ &\approx 2.666\overline{6} \ (\approx 2.667) \end{aligned}$$

ตอบ $\boxed{3)}$

Problem 42

What is $E(X^2)$?

1) 6.333
2) 7.333
3) 8.333
4) 9.333

Solution

สามารถคำนวณค่า Expected Value (หรือ $E[X]$) โดยใช้สมการทางคณิตศาสตร์คือ...

$$E[X] = \sum_{\text{x} \in \text{range(X)}} \boldsymbol{x^2} \cdot P(X = x)$$

ดังนั้นจึงสามารถคำนวณได้ดังนี้...

$$\begin{aligned} E[X] &= 1^2(\frac{1}{6}) + 2^2(\frac{2}{6}) + 3^2(\frac{1}{6}) + 4^2(\frac{2}{6}) \\ &= \frac{1}{6} + \frac{8}{6} + \frac{9}{6} + \frac{32}{6} \\ &= \frac{50}{6} \\ &\approx 8.333\overline{3} \end{aligned}$$

ตอบ $\boxed{3)}$

Problem 43

Given a random variable X and $Y = X^2$. Assume that $E(X) = 5$ and $E(Y) = 25$, what is $\text{Var}(X)$?

1) 0
2) 20
3) 30
4) Not enough information.

Solution

สามารถคำนวณค่า Variance (หรือ $Var(X)$) โดยใช้สมการทางคณิตศาสตร์คือ...

$$Var(X) = E[X^2] - (E[X])^2$$

จากที่โจทย์กำหนดเทำให้ตอนนี้เราทราบว่า $E(X) = 5, E(Y) = E(X^2) = 25$ ดังนั้นเราสามารถคำนวณ $Var(X)$ ดังนี้

$$\begin{aligned} Var(X) &= 25 - (5)^2 \\ &= 0 \end{aligned}$$

ตอบ $\boxed{1)}$

Problem 44

Given random variable X with $E(X) = 5$ and $\text{Var}(X) = 2$. Find $E(X + 2X)$.

1) 10
2) 15
3) 25
4) Not enough information.

Solution

จากสิ่งที่โจทย์ต้องการหาเมื่อใช้คุณสมบัติ Linearlity ที่กล่าวว่า...

$$E[aX+b] = aE[X]+b \qquad ;\text{ (เมื่อ } a, b \text{ เป็นค่าคงที่)}$$
$$E[X+Y] = E[X]+E[Y] \qquad ;\text{ (เมื่อ } X, Y \text{ เป็นตัวแปรสุ่ม)}$$

จึงสามารถใช้สมบัติแยกออกมาได้ว่า $E[X]+2E[X]$ ทำให้สามารถคำนวณหาตอบได้ว่า

$$\begin{aligned} E[X+2X] &= E[X]+2E[X] \\ &= 5+2(5) \\ &= 15 \end{aligned}$$

ตอบ $\boxed{2)}$

Problem 47

What is the expected value of X if X follows Binomial$(10, 0.8)$?

1) 8
2) 16
3) 80
4) 0.8

Solution

จากโจทย์กำหนดให้ทำให้ทราบว่า $n = 10, p = 0.8$ สามารถคำนวณ Expected value สำหรับ Binomial distribution ได้ดังนี้

$$\begin{aligned} E[X] &= np \\ &= 10(0.8) \\ &= 8 \end{aligned}$$

ตอบ $\boxed{1)}$

Problem 48

If $X \sim$ Binomial$(5, 0.5)$, calculate $P(X = 3)$.

1) 1.25
2) 0.625
3) 0.3125
4) Not enough information.

Solution

จากโจทย์กำหนดให้ทำให้ทราบว่า $n = 5, p = 0.5$ สามารถคำนวณความน่าจะเป็นเมื่อ $X = 3$ ได้ดังนี้

$$\begin{aligned} \text{จาก } P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ P(X = 3) &= \binom{5}{3} (0.5)^3 (1-0.5)^2 \\ &= 10 \times 0.125 \times 0.25 \\ &= 0.3125 \end{aligned}$$

ตอบ $\boxed{3)}$

Problem 49

Given $X \sim \text{Bernoulli}(0.3)$. Calculate $\text{Var}(X + X)$.

1) 0.21
2) 0.42
3) 0.63
4) 0.84

Solution

จากสิ่งที่โจทย์ต้องการหาเมื่อใช้คุณสมบัติ Scaling and Shifting ที่กล่าวว่า...

$$Var(aX + b) = a^2 Var(X) \text{ ; (เมื่อ } a, b \text{ เป็นค่าคงที่ และ } X \text{ เป็นตัวแปรสุ่ม)}$$

ทำให้เมื่อใช้สมบัติทำให้สรุปได้ว่าเราต้องหาค่าของ $Var(X + X)$ ได้โดยที่แปลงตามสมบัติดังนี้

$$\begin{aligned} Var(X + X) &= Var(2X) \\ &= 2^2 Var(X) \end{aligned}$$

ทำการหาค่า $Var(X)$ โดยการคำนวณจากสิ่งที่โจทย์กำหนดให้ โดยเราทราบว่าเป็น Bernoulli distribution และ $p = 0.3$ โดยคำนวณได้จาก

$$\begin{aligned} Var(X) &= 0.3(1 - 0.3) \\ &= 0.3(0.7) \\ &= 0.21 \\ Var(2X) &= 2^2 Var(X) \\ &= 4(0.21) \\ &= 0.84 \end{aligned}$$

ตอบ $\boxed{4)}$

Problem 55

Which statement best fits the memoryless property?

1) Students who have failed a course are as likely to fail the course the second time.
2) Somchai rolled no SSRs for the past 100 rolls. If the SSR chance is 1%, he should get one SSR in the next 10 rolls.
3) A fair die has the same probability for all sides.
4) For the Thai government lottery, the previous last-two-digit winning number cannot be the same as the next last-two-digit winning number.

Solution

Memoryless Property คือคุณสมบัติของค่าตัวแปรสุ่มหนึ่งที่สามารถ model ได้ด้วย Geometric distribution โดยที่มีค่าความน่าจะเป็นของอนาคตที่ไม่มีความเกี่ยวโยงกับความน่าจะเป็นในอดีต

(i) **ถูกต้อง**: สามารถ model ได้ด้วย Geometric distribution และ ผลลัพธ์ที่ได้จากในอดีตไม่ได้ส่งผลต่ออนาคต

(ii) **ผิด**: ความน่าจะเป็นที่จะเกิดในอนาคตขึ้นอยู่กับผลลัพธ์จากอดีต

(iii) **ผิด**: เพราะโอกาสการออกหน้าลูกเต๋ายุติธรรมนั้นมีค่าเสมอต้น-ปลาย เหมาะกับการใช้ Uniform distribution ซึ่งขัดต่อคุณสมบัติของ Memoryless ที่ต้องเป็น Geometric distribution

(iv) **ผิด**: ความน่าจะเป็นที่จะเกิดในอนาคตขึ้นอยู่กับผลลัพธ์จากอดีต

ตอบ $\boxed{1)}$