Chapter 1 Overview and Descriptive Statistic

1.1 Populations, Samples, and Processes

Populations → ประชากร

Samples → ตัวอย่าง

- การรวบรวมข้อมูลต่างๆ จากประชากรทั้งหมด (Census หรือ การทำสำมะโนประชากร) มีข้อจำกัดในเรื่อง เวลา, เงิน และปัจจัยอื่นๆอีก ทำให้การทำ Census ไม่เหมาะที่จะนำมาใช้ทางด้านสถิติ
- ดังนั้นในการศึกษามักใช้วิธีสุ่มกลุ่มตัวอย่าง (Sample) ที่เราสนใจออกมาจากกลุ่มประชากรแทน (Population)
- Variable คือ คุณสมบัติหรือลักษณะอะไรซักอย่างที่เราสนใจ
  - Numerical → เก็บข้อมูลที่เป็นตัวเลข
  - Categorical → เก็บข้อมูลที่ไม่ใช่ตัวเลข

  เช่น x = แบรนด์ของโทรศัพท์ที่ นศ. นิยมใช้

  y : จำนวนผู้ติดเชื้อ Covid-19

- Univariate → data ซึ่งได้มาจากการสำรวจ ตัวแปร 1 ตัวแปร
  - ตัวอย่างเช่น ส่วนสูงของนักเรียนในห้อง D, ประเภทของรถยนต์ etc.
- Bivariate → data ซึ่งได้มาจากการสำรวจ ตัวแปร 2 ตัวแปร
  - ตัวอย่างเช่น ส่วนสูง และน้ำหนักของคน. etc
- Multivariate → ≥ 2 ตัวแปรขึ้นไป (Bivariate ก็คือเป็น Multivariate รูปแบบหนึ่ง)

Branches of Statistics

->Descriptive Statistics(สถิติเชิงพรรณนา)

consists of methods for

- Organizing, Displaying
- decribe data by using table, graphs → Histogram, Scatter plot, Box plot
- Summery measures.
- Mean, Median, Mode, SD

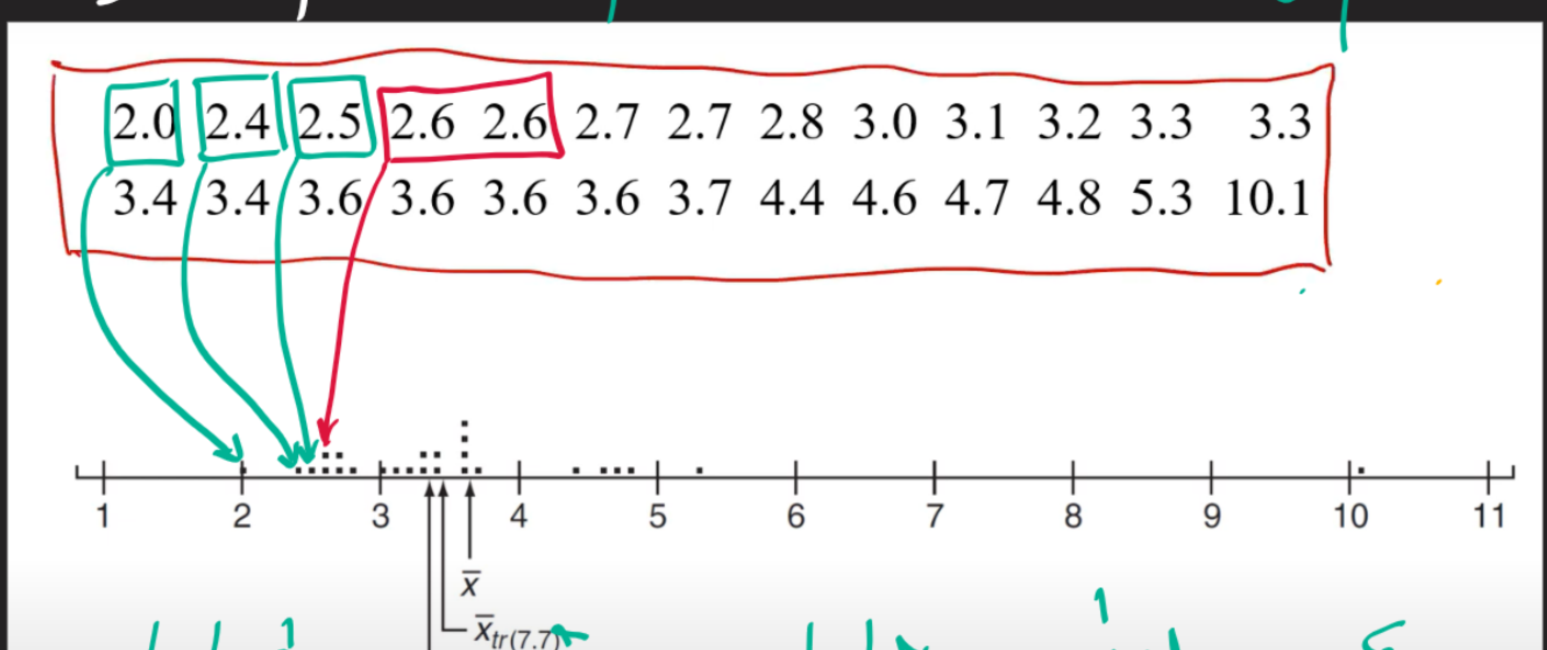
  Correlation Coefficient

->Inferential Statistics(สถิติเชิงอนุมาน)

describing population based on Sample Results.

- point estimation
- Hypothesis testing
- Estimation by confidence intervals

- Dot plot : plot บนเส้นจำนวนตรงๆ เลย



- plot ง่ายเหมาะกับการ plot ข้อมูลที่มีขนาดเล็กๆ

- Stem-and-Leaf Diagram

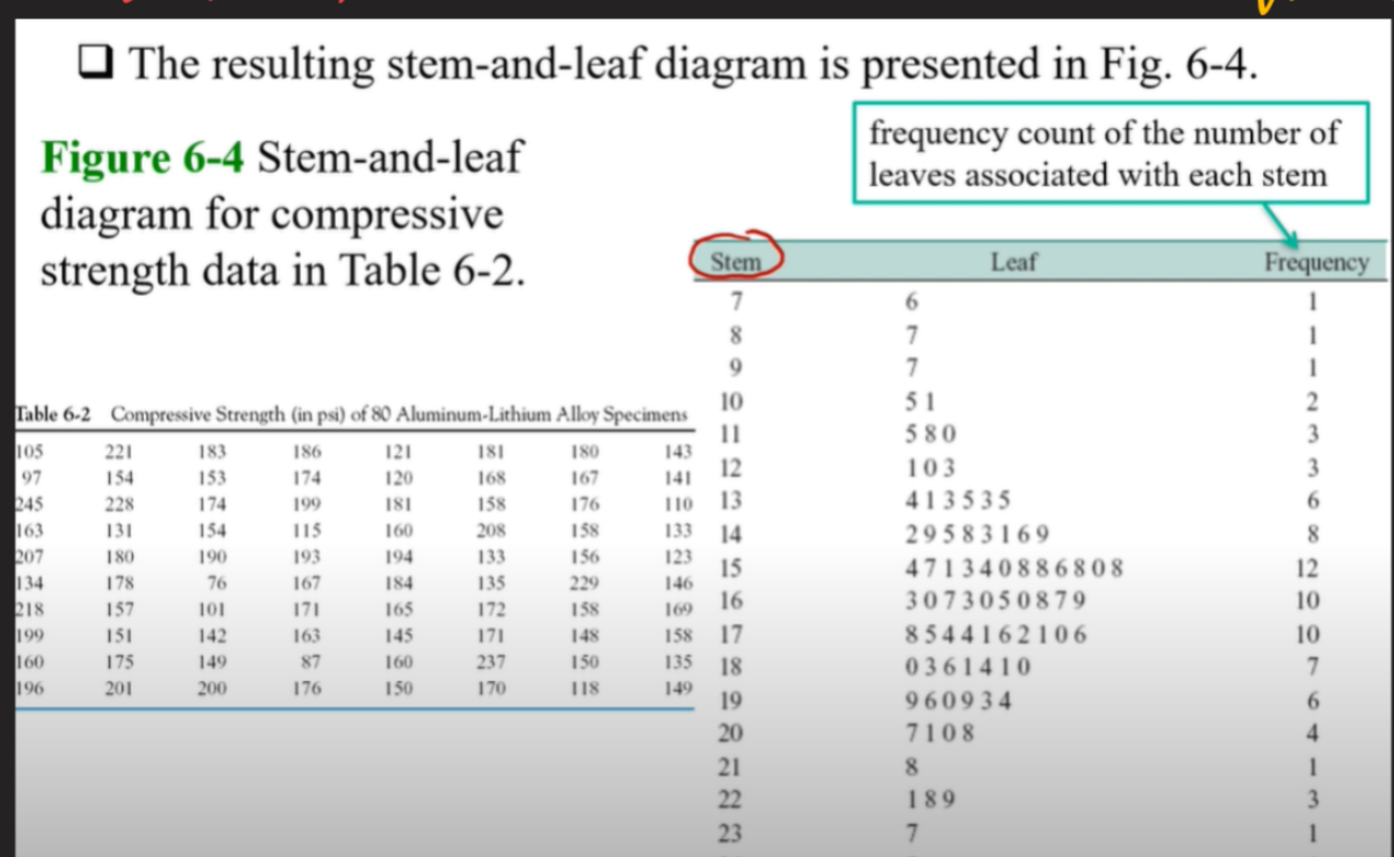
ขั้นตอนการสร้าง S&L Diagram

1) แบ่งตัวเลข x แต่ละตัวเป็น 2 ส่วน
   - Stem → เลขหลักที่นำมาด้านหน้า ≥ 1 หลัก
   - leaf → หลักที่เหลือ

   Ex 10 25 18 11 29 26 35 48 51

   • : stem • : leaf
2) เอาค่าของ stem ทั้งหมดมาไปสร้างเป็น column แนวตั้ง
3) เอาค่าของ leaf มาต่อกับ stem ของมันไปทางด้านขวา
4) เรียงค่าความถี่ไว้ด้านข้าง

น้อย → มาก



❑ The resulting stem-and-leaf diagram is presented in Fig. 6-4.

Figure 6-4 Stem-and-leaf diagram for compressive strength data in Table 6-2.

frequency count of the number of leaves associated with each stem

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

เสริม

- ถ้า plot ปกติแล้วยังไม่เห็นลักษณะการกระจาย สามารถซอย stem ให้ละเอียดขึ้นได้

ซอยเพิ่มด้วย U, L

L → 0 - 4

U → 5 - 9

Lower, Upper

ซอยเพิ่มด้วย

z → 0, 1

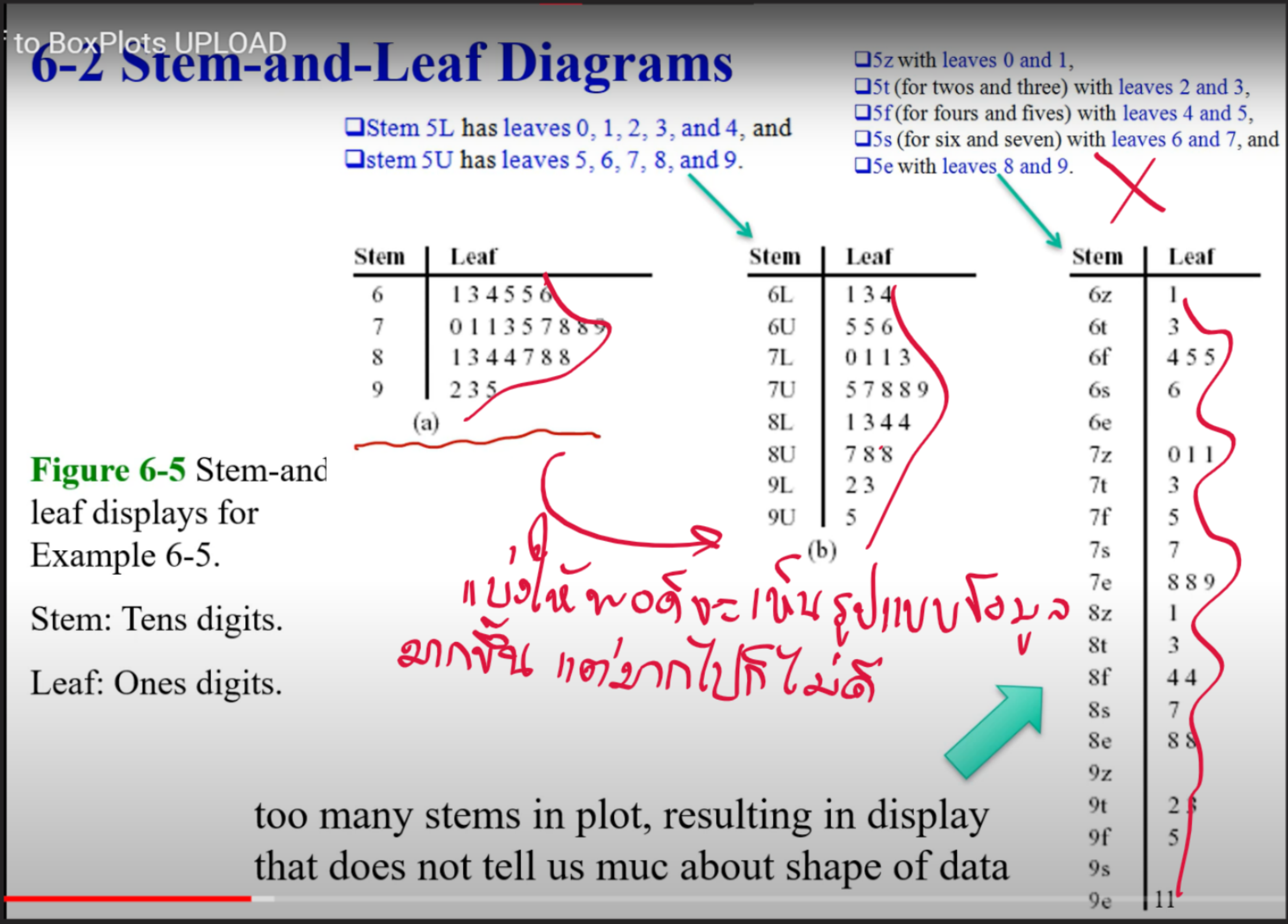
t → 2, 3

f → 4, 5

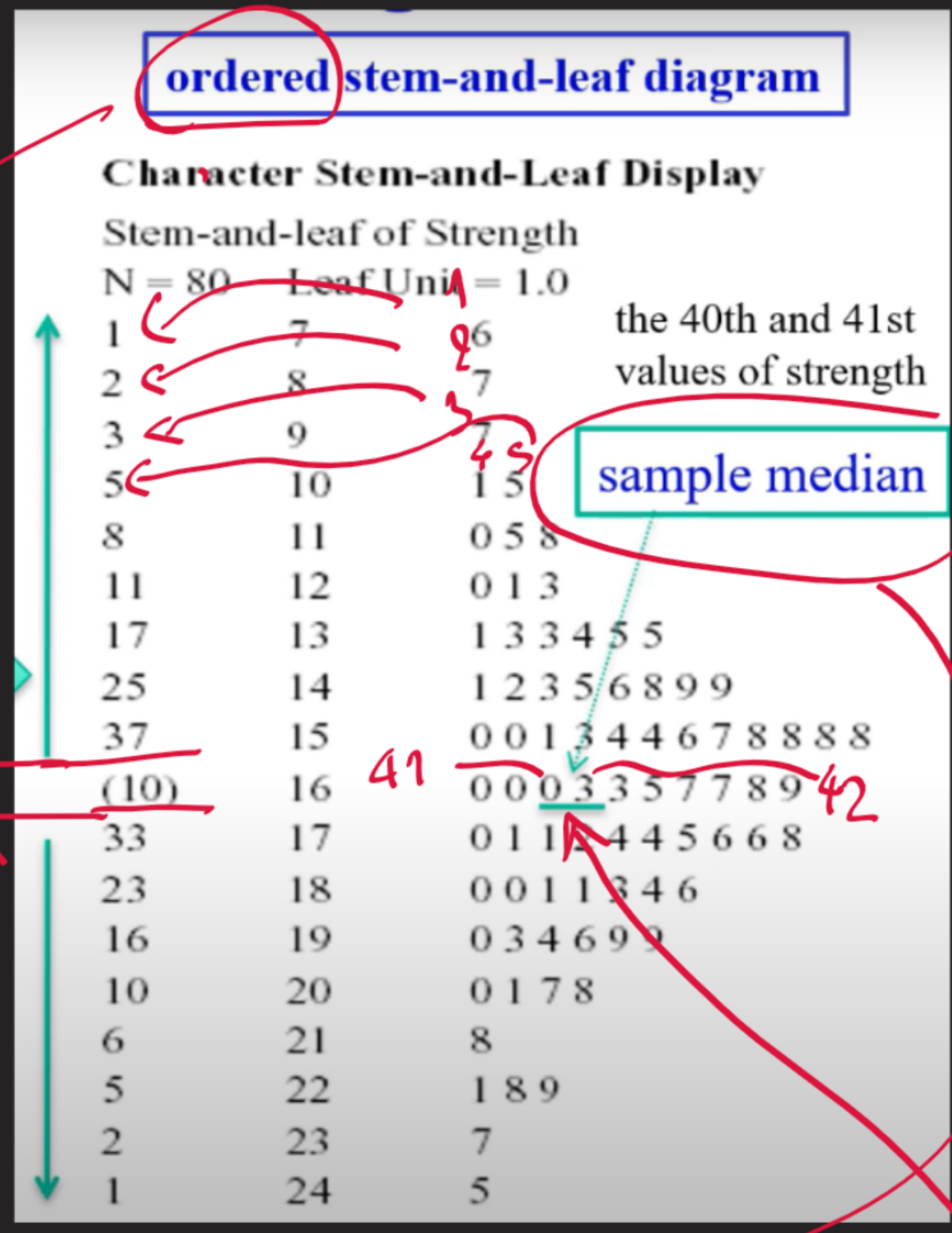
s → 6, 7

e → 8, 9

Zero Two Four Six Eight



เรียง leaves



อยู่ระหว่าง

- Ordered Stem and leaf (เอา S&L มาเรียงลำดับน้อย → มาก)

พอเรียงลำดับแล้วทำให้หาค่าต่อไปนี้ได้

- Median (วิธีหาด้านขวามือ) →
- Percentile, Quatile → แบ่งข้อมูลออกเป็น 4 ส่วน

  25% → q_1 50% → q_2 75% → q_3

  วิธีหาค่าเหมือนกับ Median เปลี่ยนแค่ค่าที่นำมาคูณตอนหาตำแหน่ง

  $q_1 = \frac{1}{4}$ $q_2 = \frac{1}{2}$ $q_3 = \frac{3}{4}$

  ตำแหน่ง = $(N+1)(q_i)$ → พอได้ตำแหน่งก็ไปหาค่าปกติ
- Range

  ↳ $R = max - min$
- IQR (Inter Quatile)

  ↳ $IQR = Q_3 - Q_1$

1) หาตำแหน่ง Median $= \frac{(N+1)}{2}$

$= \left(\frac{80+1}{2}\right) = 41.5$

2) หาค่าของ Median

$Med = X_{41} + (X_{42} - X_{41})(0.5)$

$= 160 + (3)(0.5)$

$= 161.5$

Frequency Distribution

↳ การสร้างตารางแจกแจงความถี่ (อันตรภาคชั้น)

นำมา plot → and

Histogram

↳ การสร้าง ฮิสโตแกรม

1) หาจำนวนอันตรภาคชั้น (ขึ้นอยู่กับจำนวนข้อมูล ปกติจะอยู่ในช่วง 5 → 20 ชั้น)

- กำหนดเองก็ได้ตามความเหมาะสม
- ใช้สูตร

จำนวนอันตรภาค = $\sqrt{N}$ → จำนวน data

2) หาความกว้างของอันตรภาคชั้น

สูตร ความกว้าง = $\frac{Range}{จำนวนชั้น}$ = $\frac{max - min}{จำนวนชั้น}$

ถ้าหารไม่ลงเพิ่มให้หารลงได้

70 90 145 150

3) สร้างตาราง แบ่งช่วงตามความกว้างที่คำนวณได้ โดยเริ่มต้นจากค่าที่ต่ำที่สุด

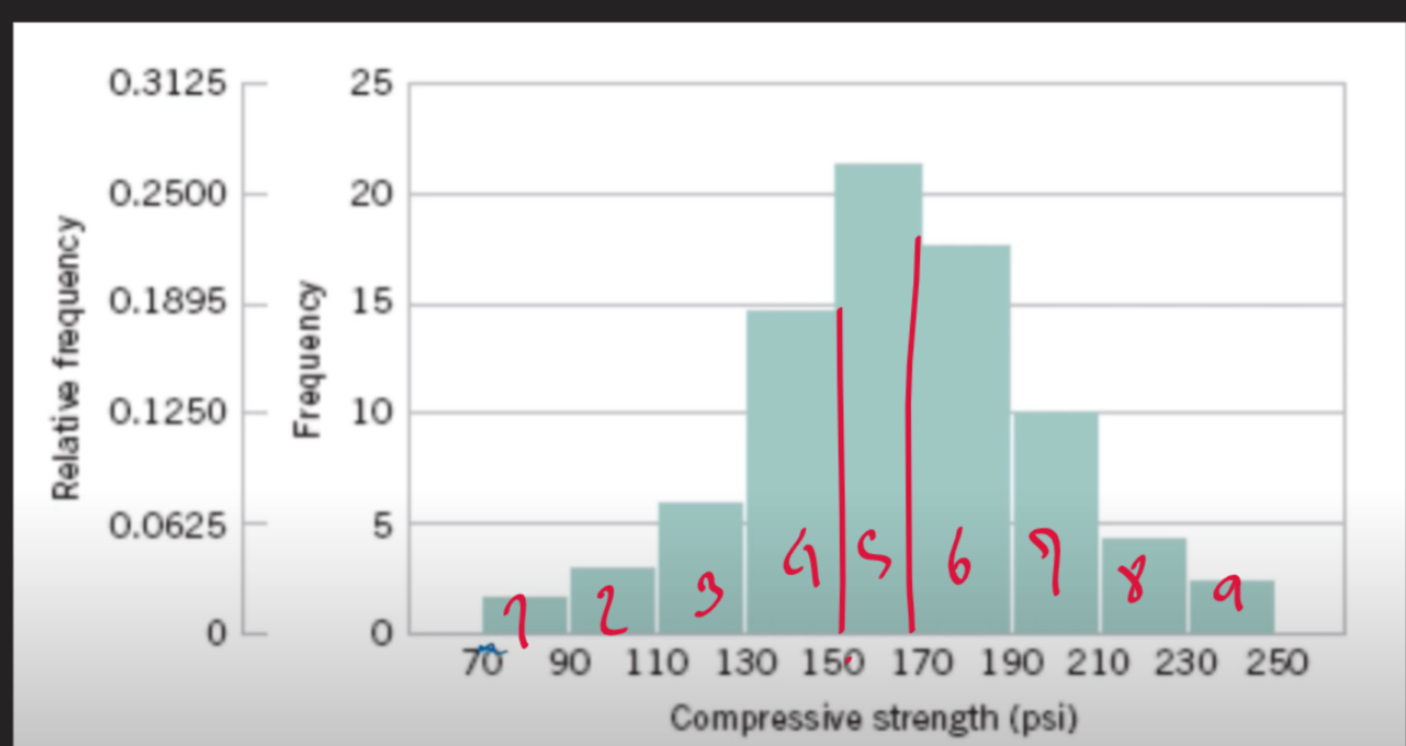
4) เขียนค่าความถี่สัมพัทธ์/ความถี่สัมพัทธ์สะสม

ตัวอย่าง ตารางแจกแจงความถี่

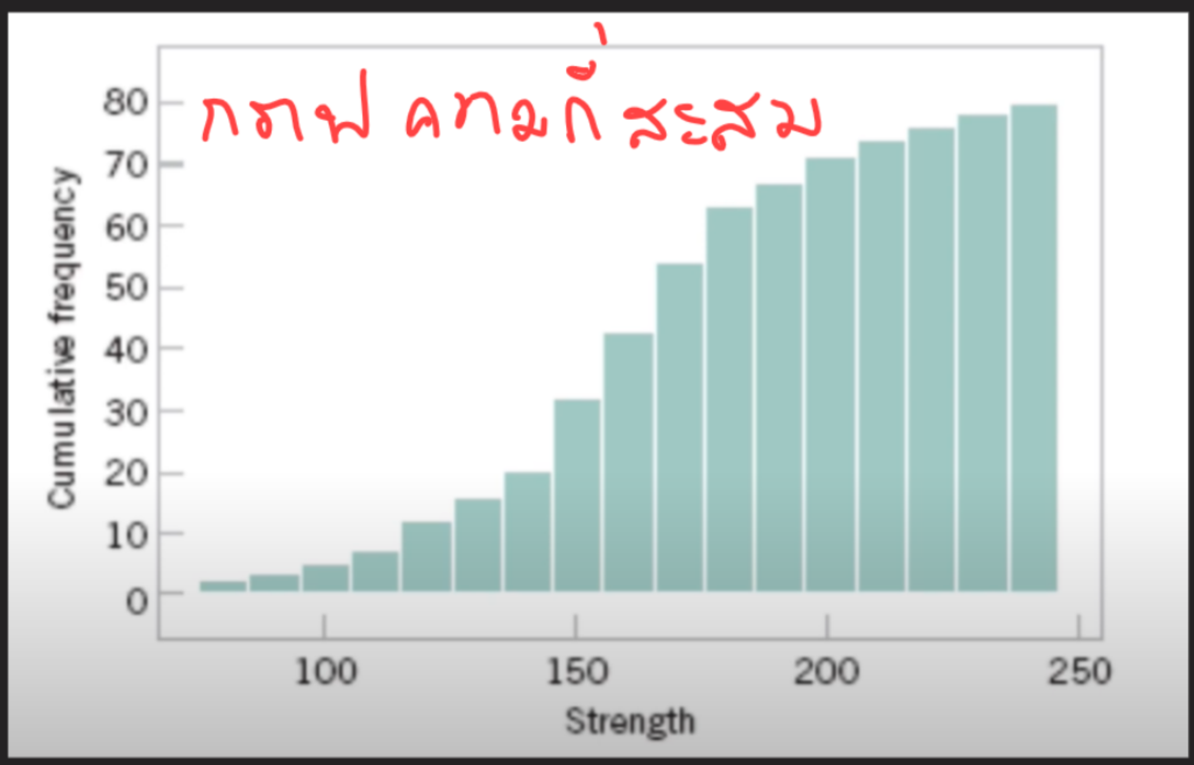
Class	$70 \le x < 90$	$90 \le x < 110$	$110 \le x < 130$	$130 \le x < 150$	$150 \le x < 170$	$170 \le x < 190$	$190 \le x < 210$	$210 \le x < 230$	$230 \le x < 250$
Frequency	2 (80)	3	6	14	22	17	10	4	2
Relative frequency	0.0250	0.0375	0.0750	0.1750	0.2750	0.2125	0.1250	0.0500	0.0250
Cumulative relative frequency	0.0250	0.0625	0.1375	0.3125	0.5875	0.8000	0.9250	0.9750	1.0000

1 2 3 4 5 6 7 8 9

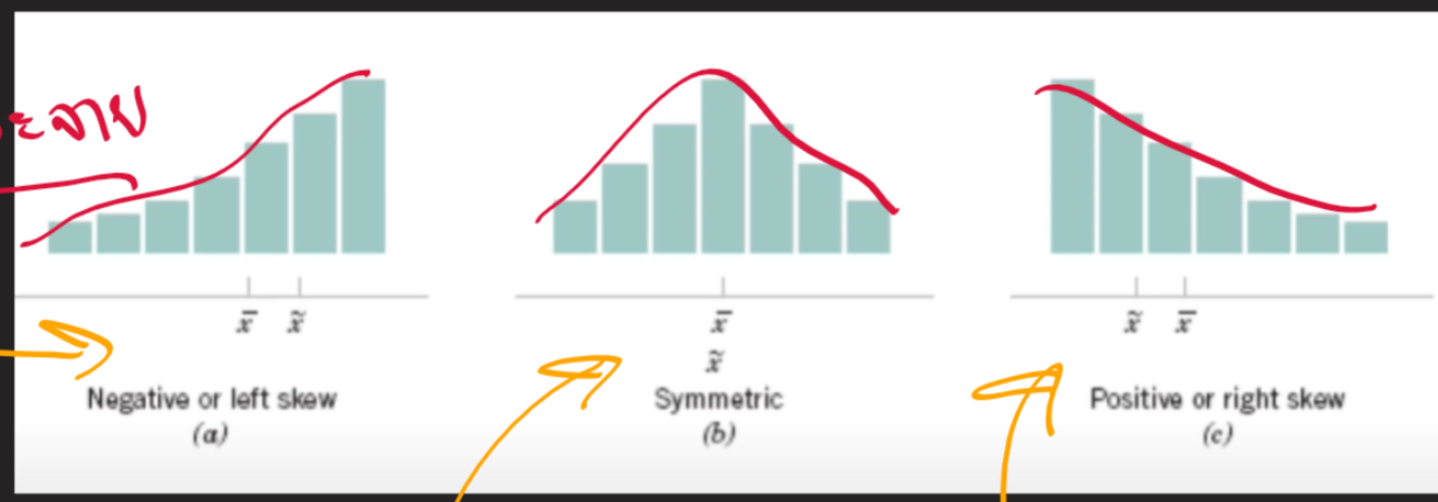
นำมา plot Histogram



กราฟ ความถี่สะสม



ลักษณะการกระจาย



a) เบ้ซ้าย
Mean < Median

b) สมมาตร
mean = median

c) เบ้ขวา
mean > median

- Box plot

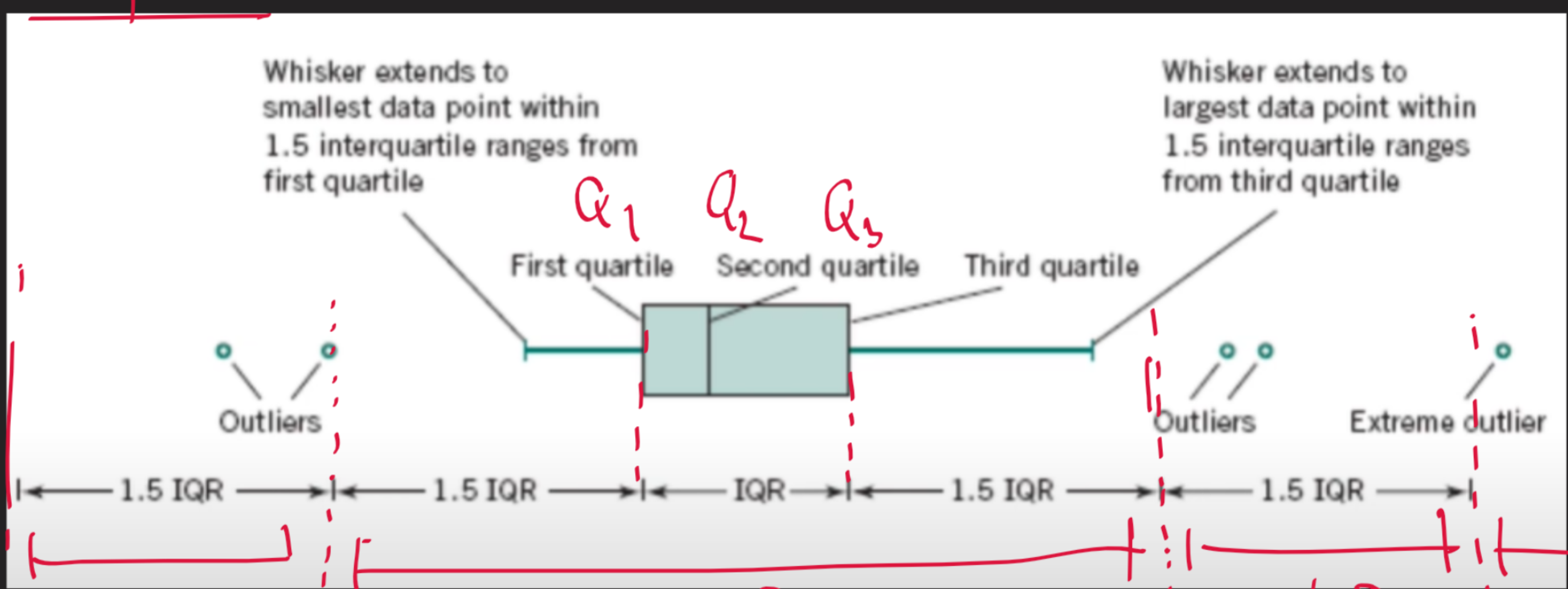
Feature
- center (ค่าสู่ศูนย์กลาง Mean ....)
- spread (ลักษณะการกระจาย)
- departure from symmetry (ห่างจากความสมมาตรขนาดไหน)
- ทำให้เห็นถึงข้อมูลที่อยู่ห่างจากข้อมูลส่วนมาก (ข้อมูล Extreme)

❑ **Box plot** is graphical display that simultaneously describes several important features of a data set, such as

- center,
- spread,
- departure from symmetry, and
- identification of observations that lie unusually far from the bulk of the data.

- **Whisker**
- **Outlier**
- **Extreme outlier**

วิธี plot



ผิดปกติมากๆ

ผิดปกติ

ปกติ

ผิดปกติ

ผิดปกติมากๆ

Chapter 2 Probability → การศึกษาสิ่งที่เป็นเกิดขึ้นอย่างสุ่ม มีความไม่แน่นอน

↳ Method ที่ทำให้เราหาโอกาสที่เหตุการณ์ใดเหตุการณ์หนึ่งจะสามารถเกิดขึ้นได้

2.1) Sample Spaces and Events → เหตุการณ์ต่างๆ ที่เราสนใจใน Sample space (เป็น subset ของ sample space)

↳ Set of all possible outcome

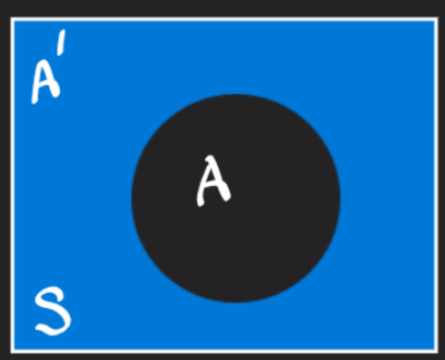
- Experiment (การทดลอง) eg. การโยนเหรียญ, การทอยลูกเต๋า
  - ↳ $S = \{H, T\}$
  - ↳ $S = \{1,2,3,4,5,6\}$

Simple Event → Event ที่มีผลลัพธ์เพียง 1 outcome

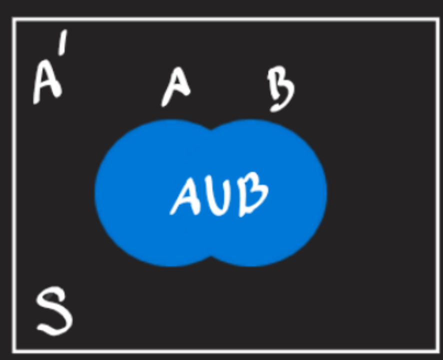
Compound Event → Event ที่มีผลลัพธ์ > 1 outcomes

->Some Relations from Set Theory

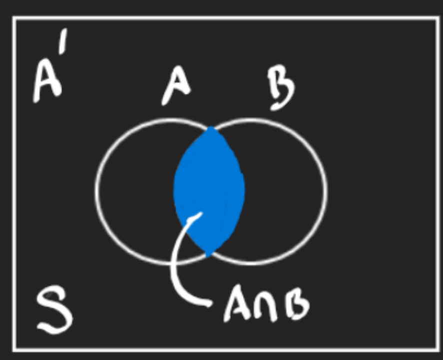
1) Complement (A')



2) Union ($A \cup B$)



3) Intersection ($A \cap B$)



->Mutually Exclusive / Disjoint Event

↳ เหตุการณ์ที่เป็น Mutually Exclusive ต่อกันคือ เหตุการณ์ 2 เหตุการณ์ A และ B ซึ่ง $A \cap B = \phi$ → (Empty set)

2.2) Axioms, Interpretations

->Axioms (สัจพจน์) (สัจพจน์ : คำกล่าวที่ถือว่าเป็นความจริงโดยไม่ต้องพิสูจน์)

กำหนดให้ A = เหตุการณ์ใดๆ

Axiom 1 $\boxed{P(A) \geq 0}$ โอกาสที่เหตุการณ์ A จะเกิดขึ้น ≥ 0 เสมอ ติดลบไม่ได้

Axiom 2 $\boxed{P(S) = 1}$ ความน่าจะเป็นของ sample space = 1 เสมอ

Axiom 3 ถ้ามี Event $A_1, A_2, A_3 \ldots A_n$ ซึ่งเหตุการณ์เหล่านี้เป็น Disjoint Event แล้ว

$$\boxed{P(A_1 \cup A_2 \cup A_3 \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i)}$$

- ถ้าเหตุการณ์ A และ B Disjoint กัน และ $A \cup B = S$ แล้ว

$$P(S) = P(A) + P(B)$$
$$1 = P(A) + P(B)$$
$$\boxed{P(B) = 1 - P(A)}$$

↳ เช่น A กับ A'

ดังนั้น $P(A') = 1 - P(A)$ เช่นกัน

->Interpreting probability (การกำหนดค่าความน่าจะเป็น)

การบอกว่า Prob ของ Event ใด Event ใด Event หนึ่ง มีค่า เท่านั้นเท่านี้ เราจะรู้ได้ยังไง ?
จริงๆหลักการหาค่า Prob ของเหตุการณ์หนึ่ง สามารถใช้ concept ของ "ความถี่สัมพัทธ์" ได้
(Relative frequencies)

วิธีการ เพื่อได้มาของค่า Prob ของ Event ใดๆ มีดังนี้

1) ทำการทดลอง (Experiment) หนึ่ง ซ้ำๆไปเรื่อยๆ
2) ในแต่ละครั้งให้บันทึกค่า และ คำนวณ ความถี่สัมพัทธ์โดยนับเฉพาะเหตุการณ์ที่เราสนใจ $\frac{n(A)}{N}$

จำนวนครั้งที่เกิด Event A

N → จำนวนครั้งการทดลอง

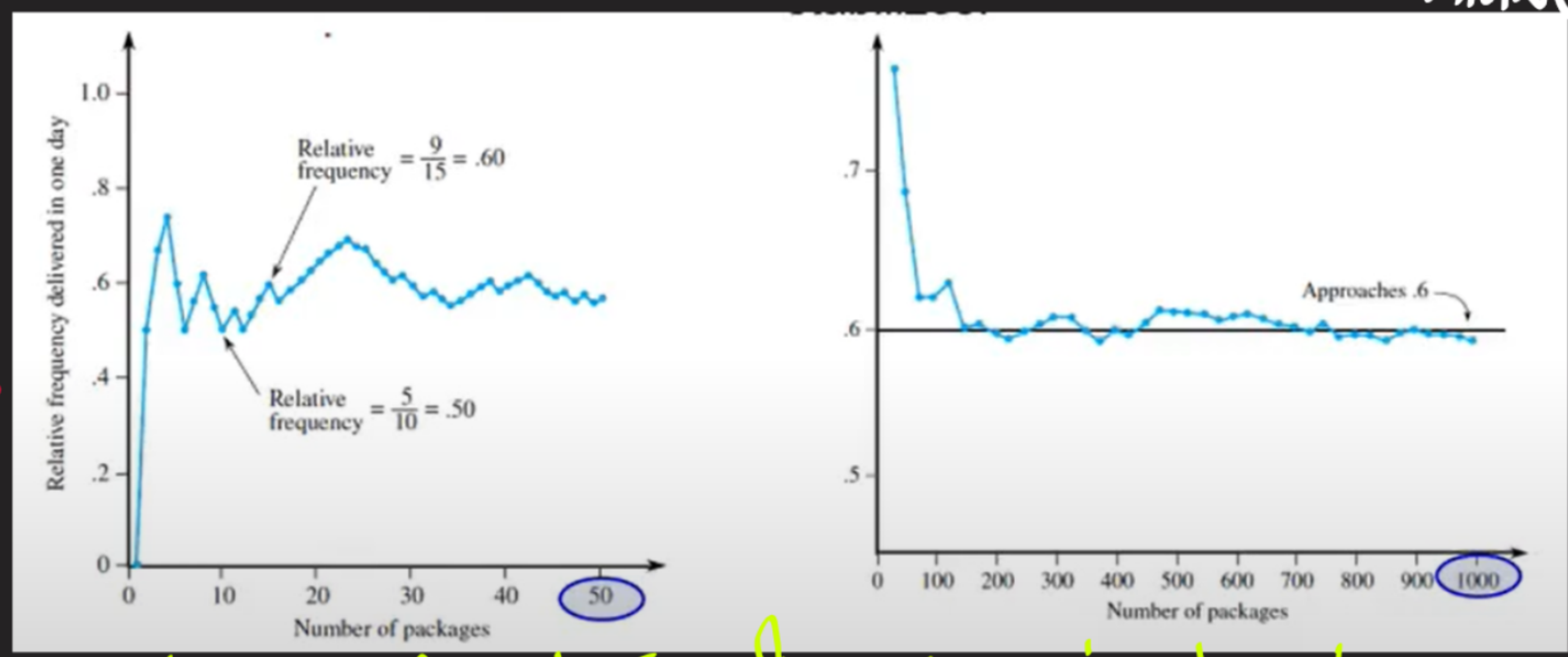
- For example, let *A* be event that package sent within the state of California for 2nd day delivery actually arrives within one day.

A = เหตุการณ์ที่พัสดุที่ส่ง ถึงผู้รับภายในกำหนด

- Results from sending 10 such packages (the first 10 replications) are as follows:

Package #	1	2	3	4	5	6	7	8	9	10
Did A occur?	N	Y	Y	Y	N	N	Y	Y	N	N
Relative frequency of A	0	.5	.667	.75	.6	.5	.571	.625	.556	.5
$\frac{n(A)}{n}$	$\frac{0}{1}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{3}{5}$	$\frac{3}{6}$	$\frac{4}{7}$	$\frac{5}{8}$	$\frac{5}{9}$	$\frac{5}{10}$

นำความถี่สัมพัทธ์ไป plot



* จะเห็นว่าเมื่อ N↑ ค่าก็เข้าใกล้ค่าคงที่ค่าหนึ่ง

* ดังนั้น Prob ของ A ใดๆ จึงมาจากค่าความถี่สัมพัทธ์ของเหตุการณ์ A เมื่อทำการทดลองซ้ำไปซ้ำมา (long run) จนค่าความถี่สัมพัทธ์เข้าใกล้ค่าคงที่ค่าหนึ่งนั่นเอง เราเรียกวิธีการหา Prob แบบนี้ว่า Limitting Relative freq.

->Properties of Probability

Proposition

- For any event *A*,

$$P(A)+P(A')=1$$
$$P(A)=1-P(A')$$

Proposition

- For any event *A*, $P(A) \le 1$.
- This is because

$$1 = P(A)+P(A') \ge P(A); \quad \text{since } P(A') \ge 0$$

P(A∪B) เมื่อ A และ B Disjoint กัน

- When events *A* and *B* are mutually exclusive, $P(A \cup B) = P(A) + P(B)$.

- For events that are not mutually exclusive, adding *P*(*A*) and *P*(*B*) results in "doublecounting" outcomes in intersection. The next result shows how to correct for this.

Proposition

- For any two events *A* and *B*,

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

P(A∪B) เมื่อ A และ B ไม่ Disjoint

P(A∪B∪C) ใช้สูตรนี้

- Probability of a union of more than two events can be computed analogously.

For any three events *A*, *B*, and *C*,

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$

->Determining Probabilitys systematically (การพิจารณา)ความน่าจะเป็นอย่างเป็นระบบ)

Sample Space มีได้ 2 แบบ คือ

1 finite (จำกัด) เช่น การโยนเหรียญ 2 ครั้ง $S = \{(H,H), (H,T), (T,H), (T,T)\}$

2 countably infinite (นับได้อย่างไม่จำกัด)

เช่น การทดลองหยิบก้านไฟฉายไปเรื่อยๆจนกว่าจะเจออันที่สมบูรณ์

F = ไม่สมบูรณ์ S = สมบูรณ์

$E_1 = \{S\}$

$E_2 = \{FS\}$ → Sample space มีอยู่อย่างไม่จำกัด แต่สามารถนับไปเรื่อยๆได้

$E_3 = \{FFS\}$

⋮

การกำหนดค่า prob ให้กับเหตุการณ์ใดๆ การคำนึงถึง

- Prob เหตุการณ์ทั้งหมดใน Sample space รวมกันต้อง = 1

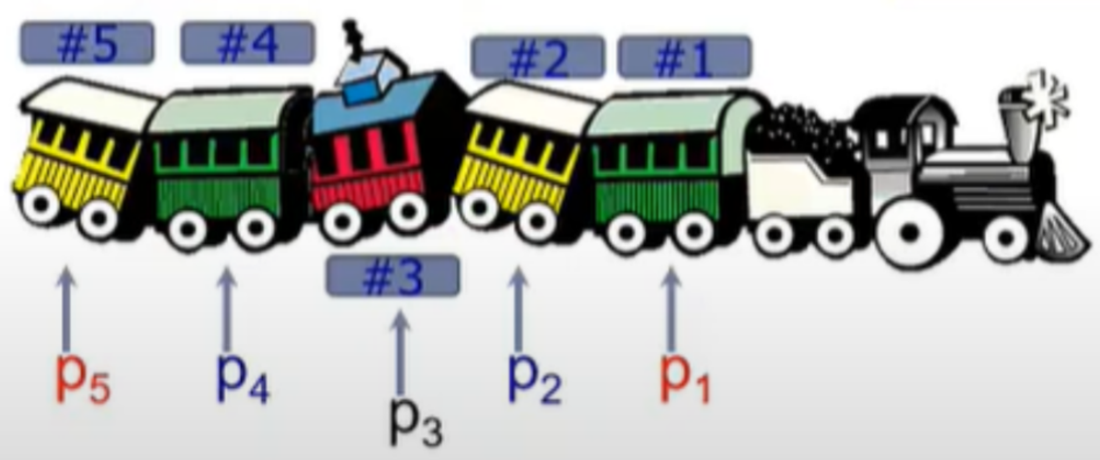
$$\sum_{i=1}^{n} P(E_i) = 1$$

Example 15

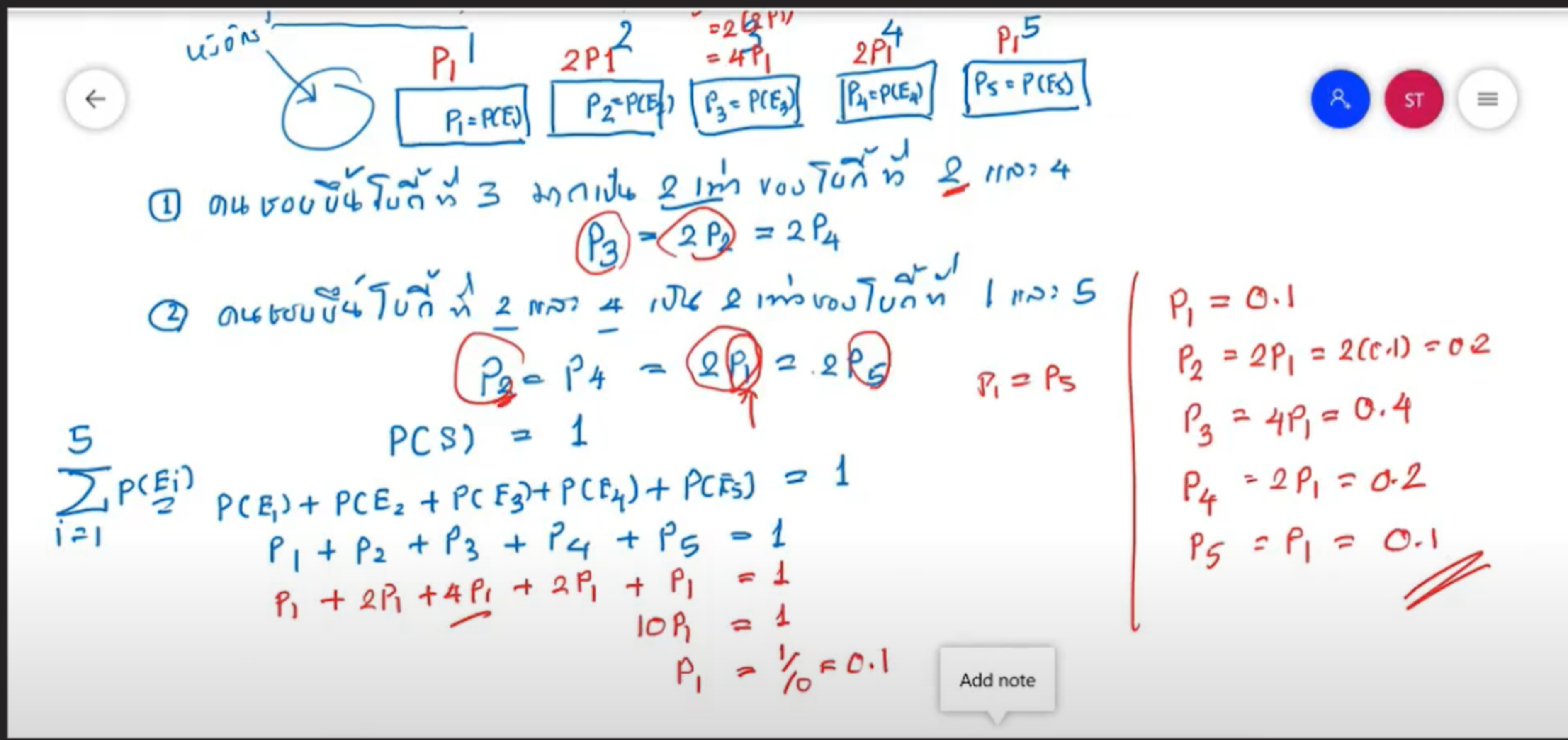
- During off-peak hours a commuter train has five cars.
- Suppose a commuter is twice as likely to select the middle car (#3) as to select either adjacent car (#2 or #4), and is twice as likely to select either adjacent car as to select either end car (#1 or #5).

$p_3 = 2p_2 = 2p_4$

$p_2 = p_4 = 2p_1 = 2p_5$



- Let $p_i = P(\text{car } i \text{ is selected}) = P(E_i)$.



Equally Likely outcomes (ผลลัพธ์ที่มีโอกาสเกิดขึ้นอย่างเท่าเทียมกัน)

- ในการทดลองที่มี Outcome อยู่ N outcomes ซึ่งแต่ละ Outcome มีโอกาสเกิดได้เท่ากัน

eg. - การโยนเหรียญ x ครั้ง
- การทอยเต๋า x ครั้ง
- การเลือกไพ่จากสำรับ มา x ใบ

ทุก outcome มีโอกาสเกิดเท่าๆกัน $P(E_i) = \frac{1}{N}$

สูตร Probability ที่ทุก outcome are equally likly

$$\boxed{P(A) = \frac{N(A)}{N}}$$

Counting Techniques (เทคนิคการนับ)

↳ ใช้ได้กับ Equally Likely Event

-> Product Rule for Ordered Pairs (กฎการคูณสำหรับคู่อันดับที่มีการเรียงลำดับ)

↳ การทำงานแบ่งเป็น state 1, state 2, state 3, ... ให้นำจำนวนวิธีในแต่ละ state คูณกัน

eg. การโยนเหรียญ 1 เหรียญ แล้วโยนลูกเต๋า 1 ลูก

state 1	state 2
{H,T}	{1,2,3,4,5,6}
2	6

state 1 (2) state 2 (6)

$2 \times 6 = 12$

Permutations (การเรียงสับเปลี่ยน)

$$\boxed{{}^{n}P_{k} = \frac{n!}{(n-k)!}}$$

เลือกของ k ชิ้นจาก n ชิ้น แบบสนใจตำแหน่ง

$(a,b) \neq (b,a)$

Combination(การจัดหมู่)

$$\boxed{{}^{n}C_{k} \,/\, \binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

เลือกของ k ชิ้นจาก n ชิ้น แบบไม่สนใจลำดับ

$(a,b) = (b,a)$

* $\binom{n}{n} = 1 \qquad \binom{n}{0} = 1 \qquad \binom{n}{1} = n$

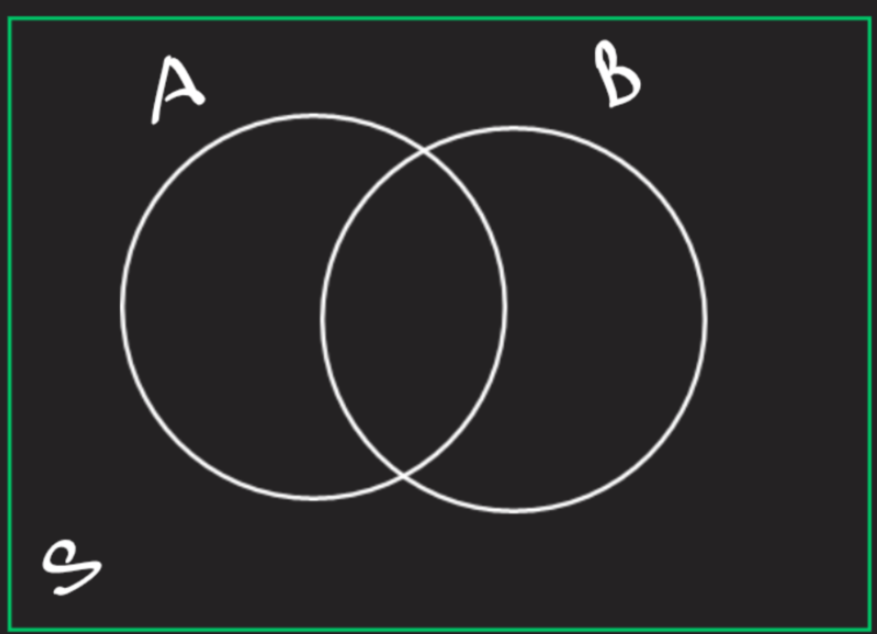
$$\frac{\cancel{n!}}{\cancel{n!}\,\cancel{(n-n)}!} = 1 \qquad \frac{\cancel{n!}}{\cancel{0!}\,\cancel{(n-0)!}} = 1 \qquad \frac{n!}{1!\,(n-1)!} = \frac{n \times \cancel{(n-1)!}}{1\,\cancel{(n-1)!}} = n$$

Conditional Probability

$P(A \mid B) \rightarrow$ conditional probability of A given that B has occurred

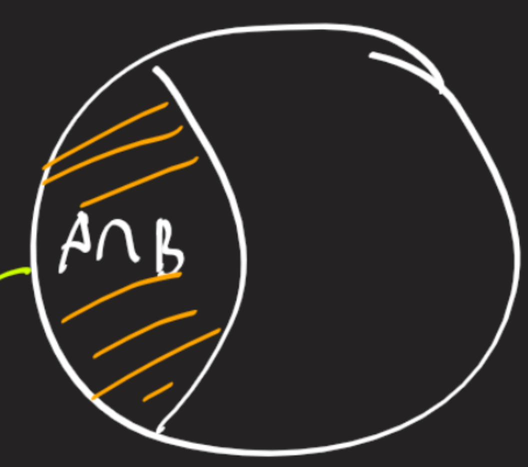
เกิดตามมา　　เกิดก่อน

ความน่าจะเป็นที่ A จะเกิดหาก B เกิดขึ้นไปแล้ว



B เกิดก่อน B เสมือนกลายเป็น S
อันใหม่

S = B



เหตุการณ์ A ที่เกิดขึ้นหลังจาก B เกิดก่อน

* ระวัง !

$P(A|B) \neq P(B|A)$

$$\boxed{\begin{aligned} P(A|B) &= \frac{n(A \cap B)}{n\,B} \\ P(A|B) &= \frac{P(A \cap B)}{P(B)} \end{aligned}}$$

* การมองประมาณหาเหตุการณ์ A ใน sample space B

อธิบายงง ลองมาโจทย์กันดู

ตัวอย่างที่ 3.15 จากการสำรวจนักศึกษาสาขาสถิติในมหาวิทยาลัยแห่งหนึ่ง พบว่าความน่าจะเป็นที่นักศึกษาจะสอบผ่านวิชาภาษาอังกฤษ (E) เป็น 0.75 ความน่าจะเป็นที่จะสอบผ่านวิชาภาษาฟิสิกส์เป็น 0.70 และความน่าจะเป็นที่จะสอบผ่านทั้งสองวิชา ($E \cap P$) เป็น 0.50 ถ้าสุ่มนักศึกษามหาวิทยาลัยแห่งนี้มา 1 คน จงหา

ก. ความน่าจะเป็นที่เขาจะสอบผ่านวิชาภาษาอังกฤษ ถ้าเขาทราบว่าวิชาฟิสิกส์สอบผ่านแล้ว

ข. ความน่าจะเป็นที่เขาจะสอบผ่านวิชาฟิสิกส์ ถ้าเขาทราบว่าวิชาภาษาอังกฤษสอบผ่านแล้ว

ภาษาฟิสิกส์ ? LOL
P

Sol^{n} $P(E)$ = Prob ผ่าน Eng = 0.75　　$P(E')$ = Prob ไม่ผ่าน Eng = 0.25

$P(P)$ = Prob ผ่าน Physic = 0.70　　$P(P')$ = Prob ไม่ผ่าน Physic = 0.30

$P(E \cap P)$ = Prob ผ่านทั้งสอง = 0.50

A) หา $P(E|P) = \dfrac{P(E \cap P)}{P(P)} = \dfrac{0.5}{0.7} = 0.714$

เท่ากัน

B) หา $P(P|E) = \dfrac{P(P \cap E)}{P(E)} = \dfrac{0.5}{0.75} = 0.667$

Multiplication Rule for $P(A \cap B)$

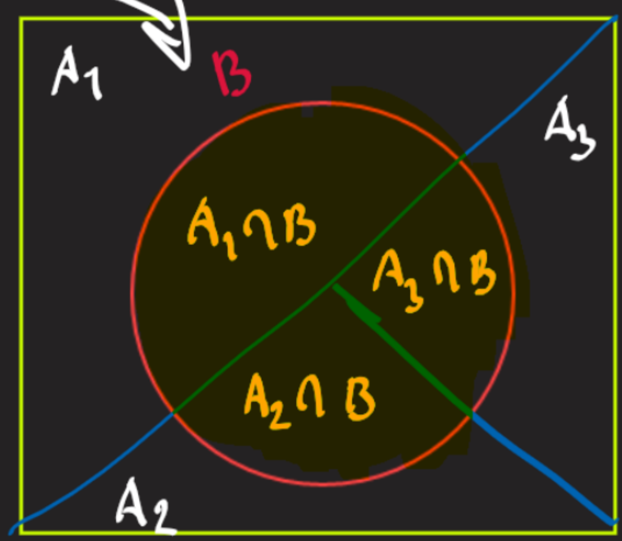
$$\boxed{P(A \cap B) = P(A|B) \times P(B)}$$

Total Probability

ซึ่งรวมกันแล้ว = Sample Space

เหตุการณ์ $A_1, A_2, A_3, \dots, A_n$ Disjoint และ มีเหตุการณ์ B เกิดหลังจาก $A_1, A_2, A_3 \dots A_n$

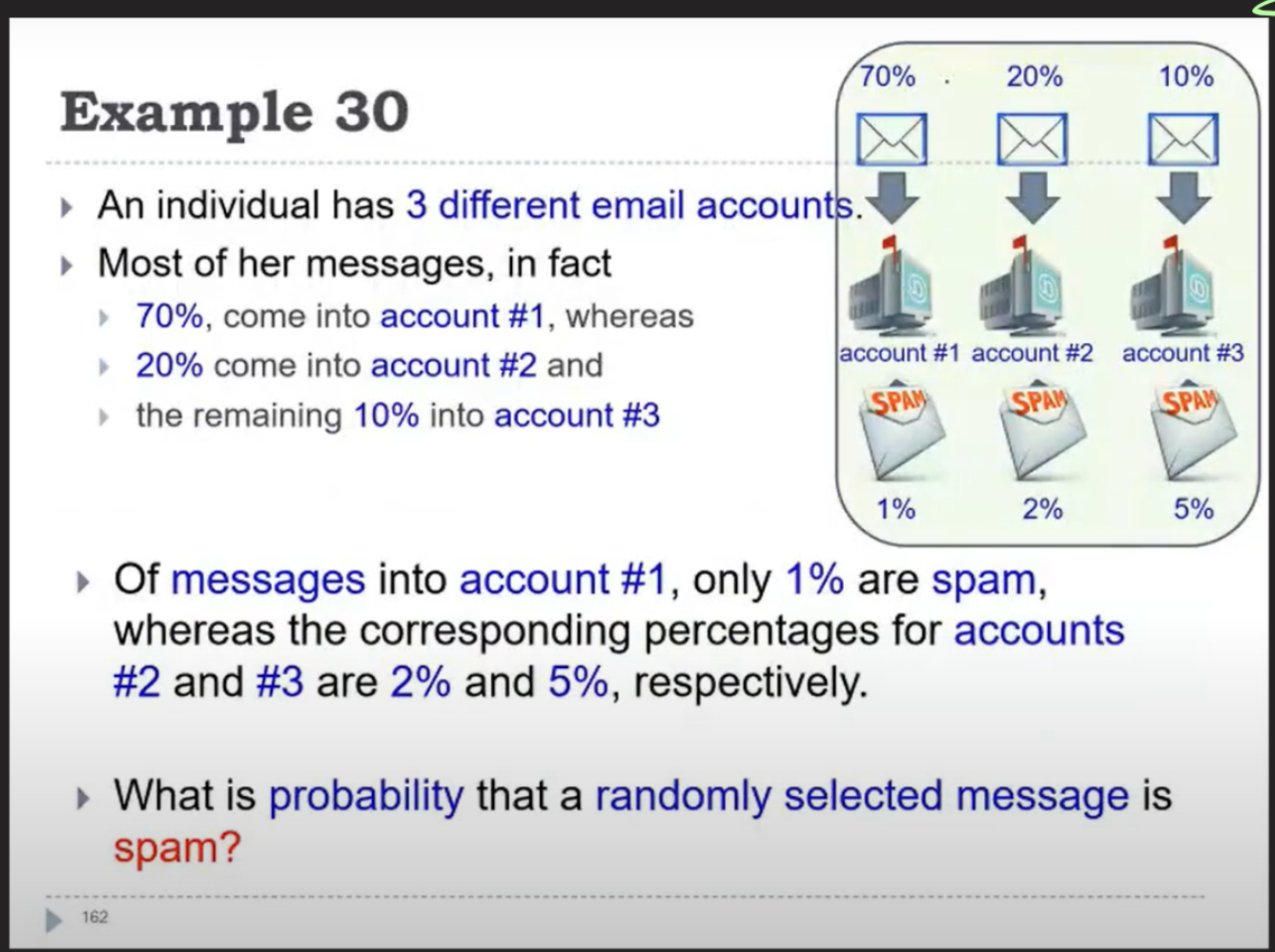
$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \dots P(A_n \cap B)$$



* ถ้าเราแทน $P(A_n \cap B) = P(B|A_n) \times P(A_n)$

$$\boxed{P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \dots P(B|A_n)P(A_n)}$$

Example



โจทย์ให้

$P(E_1)$ = Prob เมลเข้าแอ็ค 1 = 0.7

$P(E_2)$ = Prob เมลเข้าแอ็ค 2 = 0.2

$P(E_3)$ = Prob เมลเข้าแอ็ค 3 = 0.1

กำหนด S = เหตุการณ์ที่ได้รับเมล spam

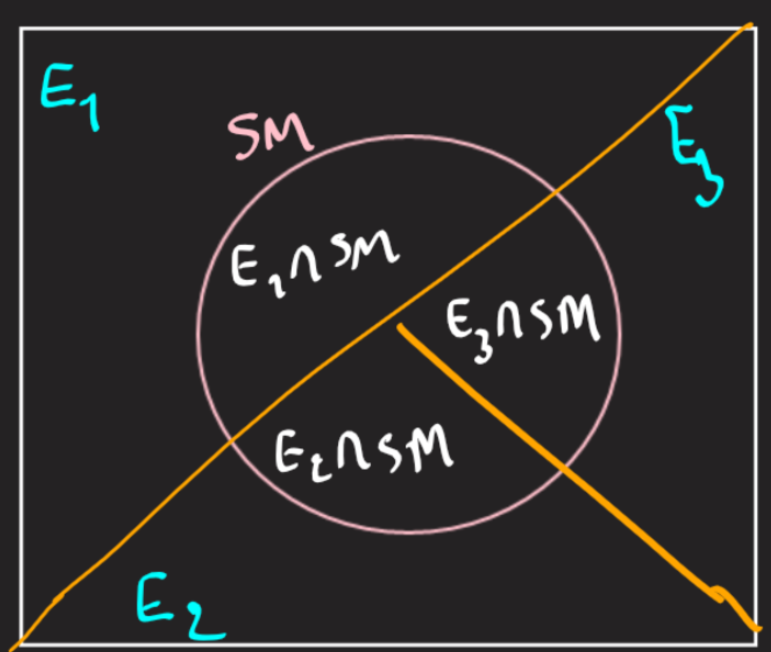
$P(SM|E_1) = 0.01$ $\quad P(SM|E_3) = 0.05$

$P(SM|E_2) = 0.02$

โจทย์ถาม $P(SM) = ?$

วิธีทำ จาก Total probability

$$P(SM) = P(SM|E_1)P(E_1) + P(SM|E_2)P(E_2) + P(SM|E_3)P(E_3)$$
$$= (0.01)(0.7) + (0.02)(0.2) + (0.05)(0.1)$$
$$= (0.007) + (0.004) + (0.005)$$
$$= 0.016 \;\#$$



Bayes's Theory

ไม่ว่าจะ $p(A|B)$ หรือ $p(B|A)$ ก็สามารถใช้หา $p(A \cap B)$ ได้เช่นกัน เพราะ $p(A \cap B) = p(B \cap A)$

* จากสูตร Condition prob ปกติ เราทราบว่า

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

นำ $p(A \cap B) = p(B|A) \times p(A)$ มาแทน

$$\boxed{p(A|B) = \frac{p(B|A)\,p(A)}{p(B)}}$$

นำ $p(B)$ จากทฤษฎี Total prob มาแทนจะได้

$$\boxed{p(A|B) = \frac{p(B|A)\,p(A)}{P(B|A_1)P(A_1) + p(B|A_2)p(A_2) \ldots}}$$

Example

- Incidence of a rare disease.
- Only 1 in 1,000 adults in afflicted with a rare disease for which a diagnostic test has been developed.
- The test is such that when
  an individual actually has the disease, a positive result will occur **99%** of the time,
  whereas
  an individual without the disease will show a positive test result only **2%** of the time.
- If a randomly selected individual is tested and the **result is positive**, what is **probability** that **individual has the disease?**

โจทย์ให้ กำหนดให้ rd = Event ที่คนติดโรค

p = Event ที่ตรวจแล้วเป็น +

$p(rd) = 0.001$ $\quad p(rd') = 0.999$

$p(p|rd) = 0.99$

$p(p|rd') = 0.02$

โจทย์ถาม $p(rd|p) = ?$

วิธีทำ

$$p(rd|p) = \frac{\boxed{p(rd \cap p)}}{\boxed{p(p)}}$$

$p(rd \cap p) \rightarrow p(p|rd) \times p(rd)$

$p(p) \rightarrow$ Total prob

1) หา $p(rd \cap p)$ ก่อน

$$p(rd \cap p) = p(p|rd)\,p(rd) = 0.99(0.001) = 0.00099$$

2) หา $p(p)$

$$p(p) = p(p|rd)p(rd) + p(p|rd')p(rd')$$
$$= (0.99)(0.001) + (0.02)(0.999)$$
$$= 0.00099 + 0.01998$$
$$= 0.02097$$

$$p(rd|p) = \frac{0.00099}{0.02097} = 0.04721 \ \#$$

Chapter3 Discrete Random Variables and Probability distributions

Outcomes → qualitative (เชิงคุณภาพ) เช่น เกรดของนศ. {A, B⁺, B, C⁺, C, D⁺, D, F}
→ quantitative (เชิงปริมาณ) เช่น น้ำหนัก, ส่วนสูง etc.

* ในทางสถิติเราต้องการให้ข้อมูลเป็นตัวเลขเพื่อใช้ในการคำนวณค่าต่างๆ เพื่อนำมาอธิบายข้อมูลเหล่านั้น แต่ข้อมูลแบบ qualitative มันไม่ใช่ตัวเลข ทำให้เราต้องใช้สิ่งที่เรียกว่า

-> Random variables → เพื่อแปลงข้อมูลต่างๆให้อยู่ในเชิงตัวเลข โดยที่ข้อมูลจริงอาจจะไม่ใช่ตัวเลขก็ได้!
(ตัวแปรสุ่ม)

Experimental Outcomes $\xrightarrow{\text{Random Variables}}$ Numerical Outcomes.

3.1) Random Variables

↳ rv หมายถึงกฎที่เชื่อมโยงตัวเลขเข้ากับ Outcome

Random variables. คือ function ที่มี Domain เป็น sample space และมี Range เป็นจำนวนจริง (Real Number)

- rv. จะแทนด้วยลักษณ์ตัวอักษรตัวอักษรพิมพ์ใหญ่เช่น X, Y และใช้ตัวอักษรพิมพ์เล็กเพื่อแสดงค่าของตัวแปรสุ่มนั้นๆ

* ค่าของ rv. จะเป็นอะไรขึ้นอยู่กับการให้ความหมายของตัวแปรนั้น e.g.
X = จำนวน bit ของเลขฐาน 2
$X(0010) = 4$ $X(1101010) = 7$
เป็นต้น...

ตัวแปรสุ่ม X

$$\boxed{X(s) = x}$$

ค่าที่สัมพันธ์กับ outcome s

outcome s (ผลลัพธ์ที่เราสนใจ)

Example 1

When a student calls a university help desk for technical support, he/she will either immediately be able to speak to someone (S, for success) or will be placed on hold (F, for failure).

With $\mathcal{S} = \{S, F\}$, define an rv X by

$X(S) = 1$ $X(F) = 0$

The rv X indicates whether (1) or not (0) student can immediately speak to someone.

การทดลองนร.โทรหา call center เพื่อขอความช่วยเหลือ

sample space = {s = โทรติด, f = โทรไม่ติด}

กำหนด RV. โดยที่ $X(s) = 1$, $X(f) = 0$

Example 3.2

Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer (such devices are used extensively by polling organizations), and define an rv Y by

$$Y = \begin{cases} 1 & \text{if the selected number is unlisted} \\ 0 & \text{if the selected number is listed in the directory} \end{cases}$$

For example

If 5282966 appears in the telephone directory, then Y(5282966) = 0, whereas Y(7727350) = 1 tells us that number 7727350 is unlisted

A word description of this sort is more economical than a complete listing, so we will use such a description whenever possible

In ex.3.1 and 3.2, the only possible values of this random variable were 0 and 1

Such a random variable arises frequently enough to be given a special name, after the individual who first studied it.

10

การทดลองสุ่มเบอร์โทร 7 ตัว

sample space = {0000000, 0000001, ... 9999999}

กำหนด RV $Y = \begin{cases} 1; & \text{ถ้าเบอร์โทรนั้นอยู่ในสมุดจดเบอร์} \\ 0; & \text{ถ้าเบอร์โทรนั้นไม่อยู่ในสมุดจดเบอร์} \end{cases}$

สมมติ 1111111 (อยู่ในสมุด) $Y(1111111) = 1$
0123456 (ไม่อยู่ในสมุด) $Y(0123456) = 0$

- Bernoulli Random Variable : ตัวแปรสุ่มที่มีค่าที่เป็นไปได้แค่ 2 ค่า eg.

eg. ใช่, ไม่ใช่   ซื้อ, ไม่ซื้อ   สำเร็จ, ไม่สำเร็จ   0, 1   ติด Covid, ไม่ติด covid

->Types of Random Variables

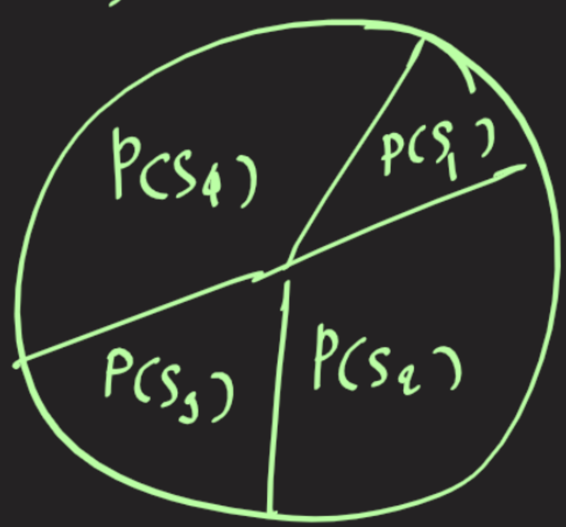
1) Discrete : (rv. ไม่ต่อเนื่อง) rv. ที่มีค่าอยู่ใน finite set หรือ Countably infinite set

2) Continuous : (rv. ต่อเนื่อง) rv. ที่มีค่าเป็นช่วงใดช่วงหนึ่งบนเส้นจำนวน หรือ $P(X=c)=0$ สำหรับทุกค่า c

3.2) Probability Distributions for Discrete Random Variables (การแจกแจงความน่าจะเป็นสำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง)

Probability Distribution of X → จะบอกถึงการกระจายค่าความน่าจะเป็นของตัวแปรสุ่มของผลลัพธ์ทั้งหมดว่า total prob = 1 เนี่ย แบ่งไปให้ $P(X=s_1)$ เท่าไร $P(X=s_2)$ เท่าไร . . . .

สมมุติ rv. X มีค่าที่เป็นไปได้ 4 แบบ คือ $\{s_1, s_2, s_3, s_4\}$
ให้ total prob เหมือนเค้กก้อนหนึ่ง ซึ่งมีน้ำหนัก = 1 kg.
เราจะแบ่งเค้ก (prob) ให้กับ $P(X=s_1), P(X=s_2), P(X=s_3), P(X=s_4)$
สมมุติแบ่งแบบนี้



อันนี้เนี่ยคือ Probability Distribution
Outcome แต่ละอันได้เค้กที่น้ำหนักต่างกันไป
แต่พอนำมารวมกันแล้ว = 1 kg. เหมือนเดิม

Probability Distribution / Probability Mass Function(pmf)

↳ การแจกแจงความน่าจะเป็น หรือ ฟังก์ชันมวลของความน่าจะเป็น (ค่าความน่าจะเป็นของตัวแปรสุ่มทุกค่าที่เป็นไปได้)

$$\boxed{p(x) = P(X=x) = P(\text{all } s \in S : X(s) = x)}$$

* โดยที่ $p(x) \geq 0$, $\sum_{\text{all posible } x} p(x) = 1$

Example 3.8

$X(S) = x$

- Six lots of components are ready to be shipped by a certain supplier.
- Number of defective components in each lot is as follows:

ชิ้นส่วนที่ใช้งานไม่ได้/ชำรุด/บกพร่อง

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

$X(1)=0$
$X(2)=2$
$X(3)=0$
$X(4)=1$
$X(5)=2$
$X(6)=0$

- One of these lots is to be randomly selected for shipment to particular customer $x=\{0,1,2\}$

ตัวแปรสุ่ม   จำนวนของชิ้นส่วนที่ใช้งานไม่ได้ ใน lot ที่ถูกเลือก

- Let X be number of defectives in selected lot
- Three possible X values are 0, 1, and 2 ⇨ $X(1,3,6)=0)$, $X(4)=1)$, $X(2,5)=2)$

$p(x) = P(X=x) = P(all\ s \in \{1,2,3,4,5,6\} : X(s) = x)$

32

Example 3.8 $p(x) = P(X=x) = P(all\ s \in \{1,2,3,4,5,6\} : X(s) = x)$

- Of the six equally likely simple events,

Three result in X=0   One in X=1

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

Two in X=2

- Then

$p(0) = P(X=0) = P(lot\ 1\ or\ 3\ or\ 6\ is\ sent) = \frac{3}{6} = 0.500$ ⇨ Probability of 0.5 is distributed to X value 0

$p(1) = P(X=1) = P(lot\ 4\ is\ sent) = \frac{1}{6} = 0.167$ ⇨ Probability of 0.167 is placed on X value 1

$p(2) = P(X=2) = P(lot\ 2\ or\ 5\ is\ sent) = \frac{2}{6} = 0.333$ ⇨ Probability of 0.333 is associated with X value 2

- Value of X along with their probabilities collectively specify the pmf
- If this experiment were repeated over and over again, in the long run
  - X = 0 occur one-half of the time,
  - X = 1 one-sixth of the time, and
  - X = 2 one-third of the time

33

Example 3.10 $p(y) = P(Y = y) = P(\text{all } s \in \{a, b, c, d, e\} : Y(s) = y)$

- Consider a group of five potential blood donors – a, b, c, d, and e – of whom only a and b have type O+ blood.

O+

- Five blood samples, one from each individual, will be typed in random order until and $\boxed{\text{O+ individual}}$ is identified
- Let rv Y = Number of typings necessary to identify an O+ individual.

จำนวนครั้งของการเจาะเลือด จนกระทั่งสามารถระบุคนที่เป็น O+ ได้

- Then the pmf of Y is

$$p(1) = P(Y = 1) = P(a \text{ or } b \text{ typed first}) = \frac{2}{5} = 0.4$$

$$p(2) = P(Y = 2) = P(c, d, \text{ or } e \text{ first, and then } a \text{ or } b)$$

$$= P(c, d, \text{ or } e \text{ first}) \cdot P(a \text{ or } b \text{ next} \mid c, d, \text{ or } e \text{ first}) = \frac{3}{5} \cdot \frac{2}{4} = 0.3$$

$$p(3) = P(Y = 3) = P(c, d, \text{ or } e \text{ first and second, and then } a \text{ or } b)$$

$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = 0.2$$

$$p(4) = P(Y = 4) = P(c, d, \text{ and } e \text{ all done first}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = 0.1$$

$$p(y) = 0 \quad \text{if } y \neq 1, 2, 3, 4$$

$\underline{\text{Sol}^n}$ มีคน 5 คน group O+ 2 คน ที่เหลือไม่ใช่

ให้ Y = จำนวนครั้งที่ตรวจจนกระทั่งพบ O+ คนแรก

Y จะมีค่าได้ตรงแค่

1 → (O+)

2 → (NO+, O+)

3 → (NO+, NO+, O+)

4 → (NO+, NO+, NO+, O+)

มากสุดแค่ 4 ครั้ง เพราะคนที่ Not O+ มีแค่ 3 คน

$\underline{\text{Sol}^n}$ หา pmf ของ Y

มีคน O+ 2 คนไง

$p(Y=1) = O^+ = \frac{2}{5}$

$p(Y=2) = NO^+, O^+ = p(NO^+) \cdot p(O^+ \mid NO^+)$

$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)$

เลือกไปคนนึงแล้วจาก 5 → 4

$p(Y=3) = NO^+, NO^+, O^+ = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right)$

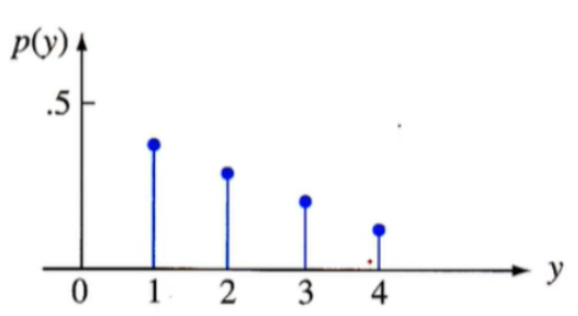
$p(Y=4) = NO^+, NO^+, NO^+, O^+ = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right)\left(\frac{2}{2}\right)$

Example 3.10

- In tabular form, the pmf is

y	1	2	3	4
p(y)	0.4 +	0.3 +	0.2 +	0.1 = 1

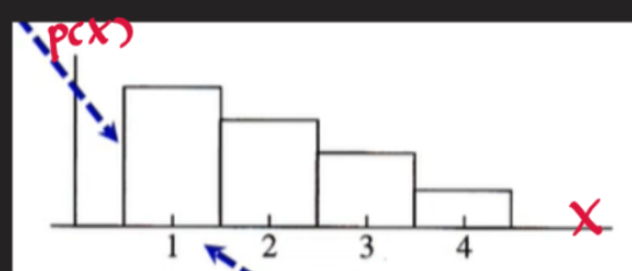
- where any y value not listed receives zero probability



Line graph for pmf in Example 3.10

หา pmf ของ Y แล้ว plot กราฟ

probability histogram



Probability histogram for Example 3.10

Probability Distribution แบบต่างๆ

-> Bernoulli Distribution

การทดลองมีผลลัพท์สองแบบ สำเร็จ กับ ไม่สำเร็จ

$p(1) = \alpha \longrightarrow$ parameter ของ Bernoulli $\quad \alpha$: ความน่าจะเป็นที่จะสำเร็จ p(success)

$p(0) = 1-\alpha$

$$\boxed{\therefore\ p(x,\alpha) = \begin{cases} 1-\alpha & ; x = 0 \\ \alpha & ; x = 1 \\ 0 & ; \text{otherwise} \end{cases}}$$

-> Geometric Distribution

การทดลองที่ผลลัพท์สองแบบ สำเร็จ กับ ไม่สำเร็จ

การทดลองที่ทำไปเรื่อยๆจนกว่าจะได้ผลลัพท์ที่สำเร็จ

X = จำนวนครั้งที่ทำการทดลองจนได้ผลลัพธ์ที่สำเร็จ ค่าที่เป็นไปได้ $1, 2, 3 \ldots$

parameter $\rightarrow$ p = ความน่าจะเป็นที่จะสำเร็จ p(success)

$$\boxed{p(x) = \begin{cases} (1-p)^{x-1}p & ; x = 1, 2, 3 \ldots \\ 0 & ; \text{otherwise} \end{cases}}$$

-> Cumulative Distribution Function

ความน่าจะเป็นที่ rv X มีค่าอย่างมาก x

$$\boxed{F(x) = P(X \leq x) = \sum_{y:\, y \leq x} p(y)}$$

$$p(X \leq 4) = p(0) + p(1) + p(2) + p(3) + p(4)$$

->สูตรCDF ใน Geometric distribution

$$F(x) = \sum_{y;\, y \leq x} p(y) = \sum_{y=1}^{x} p(1-p)^{x-1} = p\sum_{y=0}^{x-1}(1-p)^{y}$$

$\therefore$ จาก สูตรด้านขวา $a = p$, $r = (1-p)$, $n = x$

$$p\sum_{y=0}^{x-1}(1-p)^{y} = p\,\frac{(1-(1-p)^{x})}{1-(1-p)} = \cancel{p}\,\frac{(1-(1-p)^{x})}{\cancel{p}}$$

$$= 1-(1-p)^{x}$$

$$\boxed{F(x) = \begin{cases} 0 & ; x < 1 \\ 1-(1-p)^{[x]} & ; x = 1, 2, 3 \ldots \end{cases}}$$

$[x]$ = ค่าที่มากที่สุดที่ $\leq x$ เช่น $[2.7] = 2$

Sequences and Series Formulas

Arithmetic	
Sequence formula of n^{th} term	$a_n = a + (n-1)d$
Series formula for the sum of n terms	$S_n = \frac{n}{2}[2a + (n-1)d]$
Geometric	
Sequence formula of n^{th} term	$a_n = ar^{(n-1)}$
* Series formula for the sum of n terms	$S_n = \frac{a(1-r^n)}{(1-r)}$
Series formula for sum of infinite terms	$S_n = \frac{a}{(1-r)}$ when $\lvert r \rvert < 1$
Harmonic	
Sequence formula of n^{th} term	$a_n = \frac{1}{a+(n-1)d}$
Series formula for the sum of n terms	$S_n = \frac{1}{d}\ln\left[\frac{2a+(2n-1)d}{2a-d}\right]$

CDF (cumulative distribution function)

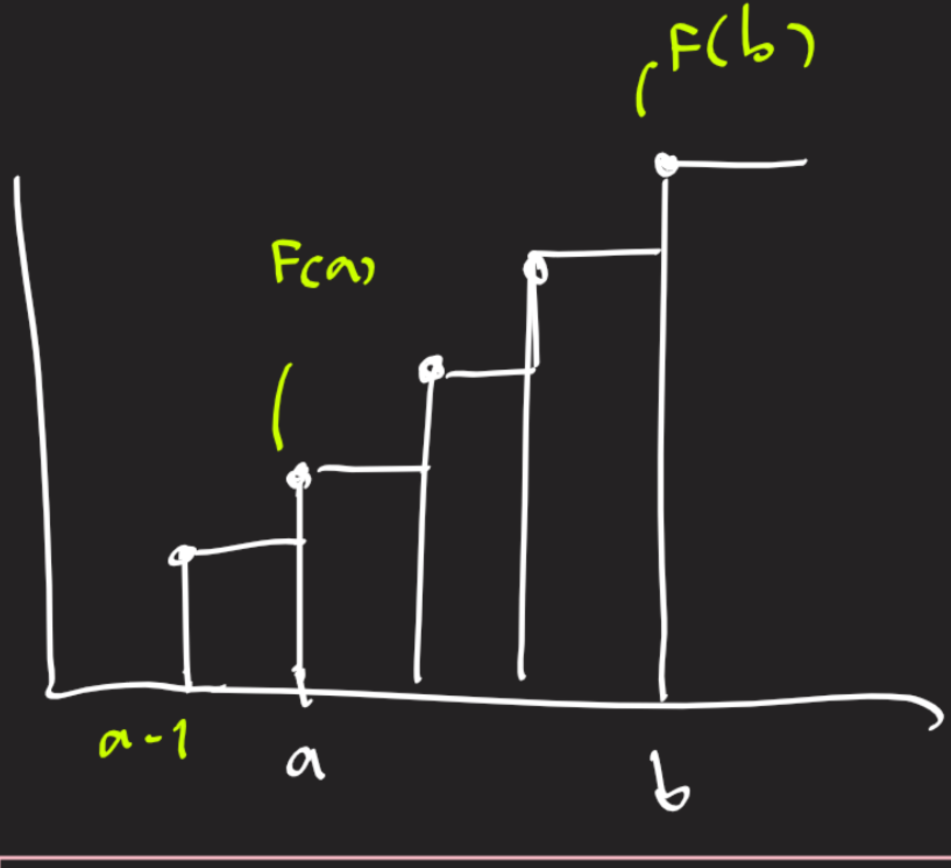
↳ มีกราฟลักษณะเป็นขั้นบันได (step function)

- การหาค่า Prob ระหว่างค่า a, b

$$P(a \leq X \leq b) = F(b) - F(a-1)$$

- ความน่าจะเป็นที่ $X = a$

$$P(X = a) = F(a) - F(a-1)$$



$$P(a < X \leq b) = P(b) - P(a)$$

3.3) Expected Values : ค่าคาดหวัง/ค่าเฉลี่ย

↳ ข้อดี : สามารถหา ค่าเฉลี่ยตัวแปรสุ่มของประชากรได้โดยรู้แค่ pmf ไม่ต้องรู้จำนวนประชากรทั้งหมด

$$E(x) = \underbrace{\mu_x}_{\text{(ค่าเฉลี่ยประชากร)}} = \sum_{x \in D} x \cdot p(x)$$

D = เซ็ตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X

- Expected value for Bernoulli RV.

$$p(x) = \begin{cases} 1-p \; ; & x = 0 \\ p \; ; & x = 1 \\ 0 \; ; & x \neq 1, x \neq 0 \end{cases}$$

prob ของ Bernoulli RV

จากสูตร $E(x) = \sum_{x \in D} x \cdot p(x)$

$$= 0 \times (1-p) + 1 \times p$$

$$= p$$

$$\therefore \quad E(x)_{\text{Bernoulli}} = p$$

$$E(x)_{GD} = \frac{1}{p}$$

Expected value for Geometric Distribution Function

$$p(x) = \begin{cases} p(1-p)^{x-1} \; ; & x = 1,2,3,4 \ldots \\ 0 \; ; & \text{otherwise} \end{cases}$$

prob ของ GDF

จาก $E(x) = \sum_{x \in D} x \cdot p(x)$

ค่าณที่

$$= \sum_{x=1}^{\infty} x \cdot p(1-p)^{x-1}$$

$$= p \sum_{x=1}^{\infty} x(1-p)^{x-1}$$

จัดรูปใหม่ จาก $\frac{d(1-p)^x}{dp}$ — chain rule

$$\frac{d(1-p)^x}{dp} = x(1-p)^{x-1} \frac{d(1-p)}{dp}$$

$$\frac{d(1-p)^x}{dp} = x(1-p)^{x-1}(-1)$$

$$x(1-p)^{x-1} = -\frac{d(1-p)^x}{dp}$$

แทนกลับ

$$= p \sum_{x=1}^{\infty} -\frac{d(1-p)^x}{dp}$$

$$= -p \sum_{x=1}^{\infty} \frac{d(1-p)^x}{dp} = -p \frac{d}{dp} \sum_{x=1}^{\infty} (1-p)^x$$

สลับลำดับได้

$$\sum_{x=0}^{\infty} (1-p)^x = (1-p)^0 + \sum_{x=1}^{\infty} (1-p)^x$$

$$\sum_{x=1}^{\infty} (1-p)^x = \sum_{x=0}^{\infty} (1-p)^x - 1$$

$$= -p \frac{d}{dp} \left[\sum_{x=0}^{\infty} (1-p)^x - 1 \right]$$

อนุกรมเรขาคณิต (อนันต์) $\frac{a}{1-r}$; $a = 1$ $r = (1-p)$

$$= -p \frac{d}{dp} \left(\frac{1}{1-(1-p)} - 1 \right)$$

$$= -p \frac{d}{dp} \left(\frac{1}{p} - 1 \right) = -p \left(\frac{\cancel{p \frac{d1}{dp}}^{0} - 1 \cancel{\frac{dp}{dp}}^{1}}{p^2} - 0 \right)$$

$$= \cancel{-p} \left(-\frac{1}{\cancel{p^2}} \right) = \frac{1}{p}$$

Variance of X : ความแปรปรวนของตัวแปรสุ่ม X

- ค่า Expected value ไม่สามารถบ่งบอกถึงการกระจายของข้อมูลได้ (บอกแต่ค่ากึ่งกลางของข้อมูลนั้นๆ)

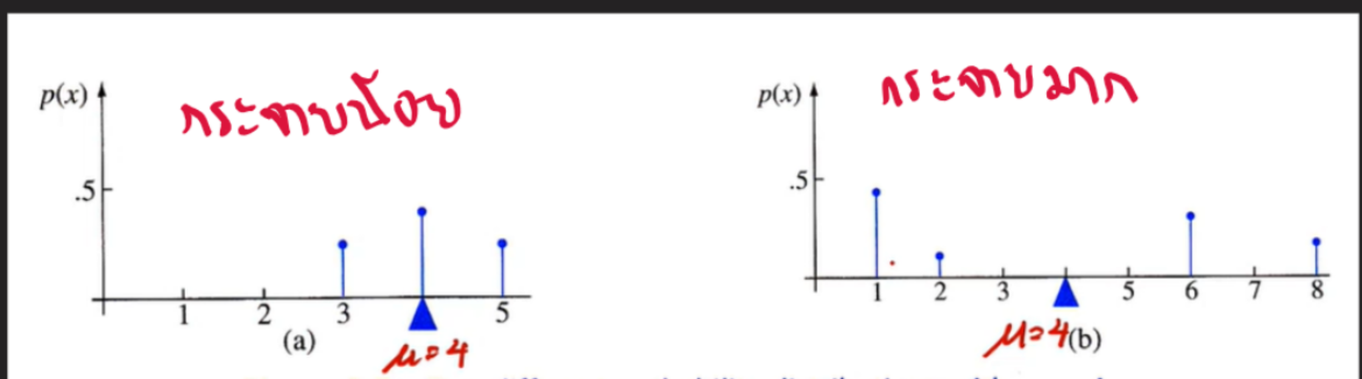


Figure 3.7 Two different probability distributions with $\mu = 4$

→ μ เท่ากันแต่การกระจายต่างกัน

* ในการคำนวณหาค่าการกระจายของข้อมูลเราจะใช้ Variance แทน

ให้ X เป็นตัวแปรสุ่มที่มี pmf $p(x)$ และ expected value μ.
ค่า Variane ของ X แทนด้วยสัญลักษณ์ $V(X)$ หรือ σ_X^2 หรือ σ^2

สูตร

$$\boxed{V(X) = \sum_{D} (x-\mu)^2 \cdot p(x) = E\left[(X-\mu)^2\right]}$$

Standard Deviation (SD) ของ X

$$\boxed{\sigma_X = \sqrt{\sigma_X^2}}$$

* สัญลักษณ์เปลี่ยนได้ ขึ้นอยู่กับว่า คำนวณ ข้อมูลของใคร

	ประชากร	ตัวอย่าง
Variance	σ^2	SD^2, s^2
Standard Deviation	σ	SD, s

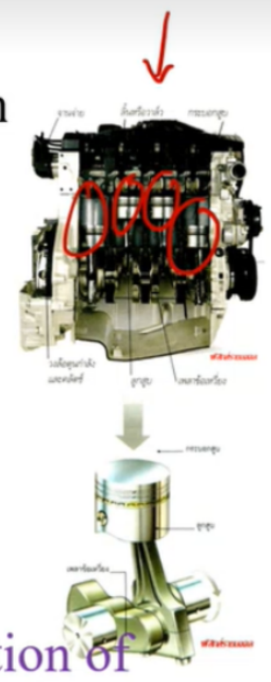
Example 3.24

x	4	6	8
$p(x)$	.5	.3	.2

Prob ที่เกิดขึ้นของแต่ละตัว ถูกประเมินแล้วครบ.

If X is the number of cylinders on the next car to be tuned at a service facility

[p(4)=0.5, p(6) = 0.3, p(8) = 0.2, from which $\mu = 5.4$], then



$$V(X) = \sum_{D} (x-\mu)^2 \cdot p(x) = E\left[(X-\mu)^2\right]$$

$$V(X) = \sigma^2 = \sum_{x=4}^{8} (x-5.4)^2 \cdot p(x)$$

$$= (4-5.4)^2(.5) + (6-5.4)^2(.3) + (8-5.4)^2(.2) = 2.44$$

The standard deviation of X is $\sigma = \sqrt{2.44} = 1.562$.

When pmf $p(x)$ specifies mathematical model for distribution of population values, both σ^2 and σ measure the spread of values in population; σ^2 is population variance, and σ is population standard deviation.

98

$\underline{\text{Sol}^n}$

x	4	6	8
$p(x)$	.5	.3	.2

$\mu = 5.4$

$$\sigma_X^2 = V(X) = \sum_{D} (x-\mu)^2 \cdot p(x)$$

$$= (4-5.4)^2(0.5) + (6-5.4)^2(.3) + (8-5.4)^2(0.2)$$

$$= 2.44$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2.44} = 1.562$$

สูตรลัดหา Variance

$$\boxed{\sigma_X^2 = V(X) = \left[\sum_{D} x^2 \cdot p(x)\right] - \mu^2 = E(X^2) - (E(X))^2}$$

3.4 Binomial Probability Distribution

- การทดลองแบบ Binomial

เงื่อนไข

1) ต้องรู้จำนวนครั้งของการทดลองทั้งหมด (ต้องรู้ n)
   eg โยนเหรียญ n ครั้ง, โยนหมุด n ครั้ง, ทอยเต๋า n ครั้ง etc.
2) ผลของแต่ละครั้งของการทดลองมีค่าที่เป็นไปได้แค่ 2 กรณี
   → success (S), failure (F)
   eg. ถามนักศึกษา 3 คน ว่าฉีดวัคซีนแล้วหรือยัง
   ผลลัพธ์มี 2 อย่าง {ฉีดแล้ว, ยังไม่ฉีด}
   ★ ตัวไหนจะเป็น S หรือ F ขึ้นอยู่กับนิยามของตัวแปรสุ่ม X
3) การทดลองในแต่ละการทดลอง ต้อง Independent กัน (เป็นอิสระต่อกัน)
4) รู้ค่าความน่าจะเป็นของผลลัพธ์ที่สำเร็จ ($P(S)$) (ค่า $p(S)$ ต้องเท่าเดิมเสมอ ตลอดการทดลอง)

* โดยปกติ การทดลองที่มีผลลัพธ์ 2 แบบ ที่ทำการสุ่มตัวอย่างออกมา โดยไม่ใส่คืน จะไม่ใช่ การทดลองแบบ Binomial เว้นแต่ว่า ! → $\frac{n}{N} \leq 0.05$ (sample ไม่เกิน 5% ของ population)
ถึงจะสามารถใช้ Binomial ประมาณค่าได้

n = จำนวน sample
N = จำนวน population

- ในการทดลองแบบ Binomial ส่วนใหญ่เรามักจะสนใจเฉพาะ จำนวนครั้งที่ได้ผลลัพธ์ที่สำเร็จ (S)

- Binomial Random Variable X : ตัวแปรสุ่ม X ที่สัมพันธ์กับการทดลองแบบ Binomial n ครั้ง
  โดยกำหนดให้

X = จำนวนครั้งที่ได้ผลลัพธ์ที่สำเร็จ (S) ในการทดลอง n ครั้ง

$x = 0, 1, 2, 3, \ldots, n$

n = จำนวนครั้งของการทดลอง

p = prob ของผลลัพธ์ที่สำเร็จ

มันมีแบบเป็นตาราง สำเร็จแล้ว ก็มีแต่อยู่ในรูป CDF ($P(X \leq n)$) แต่ตอนสอบไม่น่ามีให้

สูตร ความน่าจะเป็นที่กระจายแบบ Binomial

$$b(x, n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & ; x = 0, 1, 2, 3 \ldots n \\ 0 & ; \text{otherwise} \end{cases}$$

$E(X)_{Bin} = np$

$V(X)_{Bin} = np(1-p) = npq$

$\sigma_X = \sqrt{npq}$

3.5 Hypergeomatric and Negative Binomial Distribution

→ Hypergeometric Distribution

เงื่อนไข

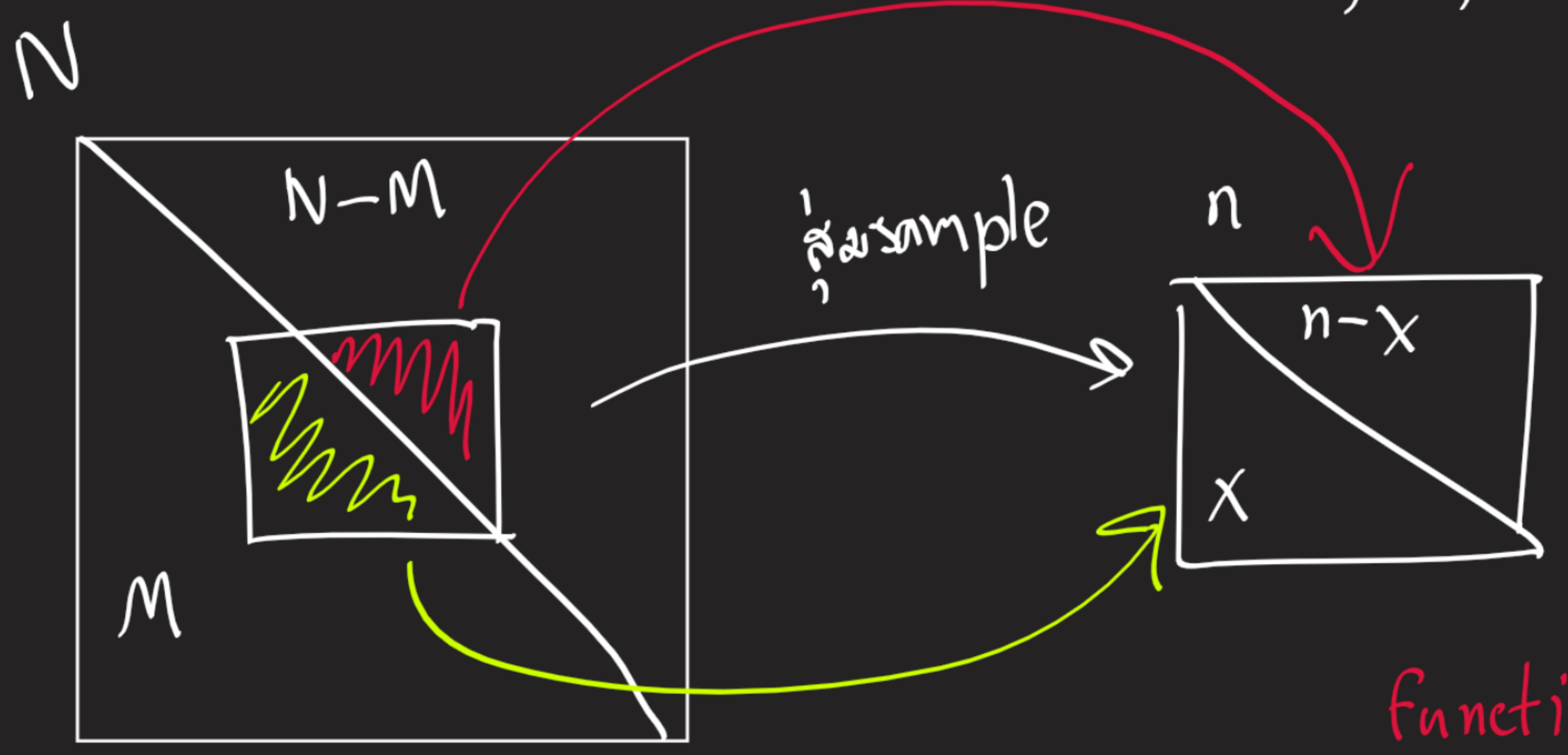
- 1) Population หรือ Sample มีขนาด N หน่วย
- 2) มีผลลัพธ์ 2 แบบ สำเร็จ กับ ไม่สำเร็จ และใน Population มีจำนวนที่สำเร็จอยู่ M หน่วย
- 3) สุ่มเลือกตัวอย่างออกมา n หน่วย (ไม่ใส่คืน)

กำหนด ตัวแปรสุ่ม X = จำนวนเหตุการณ์ S (สำเร็จ) ในกลุ่มตัวอย่างที่สุ่มมา (sample)

$P(X = x) = h(x : n, M, N)$

จาก pool สำเร็จ M เลือกมา x

จาก pool ไม่สำเร็จ N−M เลือกมาเติมให้เต็ม (n−x)



$$h(x : n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Sample space

function return ค่ามากสุดระหว่าง 0 กับ n−N+M

ค่าของ x ที่เป็นไปได้ $\max(0, n-N+M) \le x \le \min(n, M)$

fuction return ค่าต่ำสุดระหว่าง n, M

- Expected Value

$$E(X) = n \cdot \frac{M}{N} = n \cdot p \qquad p = \frac{M}{N}$$

- Variance

$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

$$= \left(\frac{N-n}{N-1}\right) \cdot np(1-p)$$

- สูตรประมาณจำนวนประชากร

$$\hat{N} = \frac{M \times n}{x}$$

Negative Binomial Distribution

เงื่อนไข

1) การทดลองจะทดลองเป็นลำดับไปเรื่อยๆ และ เป็นการทดลองแบบอิสระต่อกัน (Independent)
2) แต่ละการทดลอง จะมีผลลัพธ์ได้ แค่ 2 แบบ คือ สำเร็จ (S) และไม่สำเร็จ (F)
3) รู้ความน่าจะเป็นที่จะสำเร็จ P(S) และ P(S) ต้องเป็นค่าคงตัว
4) การทดลองจะทำซ้ำไปเรื่อยๆ จนกว่าจะได้ ผลลัพธ์ที่สำเร็จจำนวน r ครั้ง

กำหนด ตัวแปรสุ่ม

X = จำนวนครั้งที่ได้ผลลัพธ์ไม่สำเร็จ ก่อนที่จะเจอ ผลลัพธ์สำเร็จครบ r ครั้ง

สูตร

$$\boxed{nb(x;r,p) = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad x = 0,1,2,3,\dots}$$

Ex $nb(7,3,p) \rightarrow$ fail 7 รอบ ก่อน เจอ success ครบ 3 รอบ

วิธีคิดแบบเป็นภาพ ทำการทดลองทั้งหมด 10 ครั้ง มี 10 ตำแหน่ง

_ _ _ _ _ _ _ _ _ S — ตำแหน่งสุดท้ายเป็น S เสมอ

มี 9 ตำแหน่งที่เหลือ ต้องการ วาง S อีก 2 ตัว $\binom{9}{2}$ วิธี prob คือ $\binom{9}{2}p^2$

อีก 7 ตำแหน่งที่เหลือวาง F 7 ตัว $\binom{7}{7} = 1$ วิธี prob คือ $1(1-p)^7$

ตำแหน่งสุดท้ายวาง S prob คือ p

$$\therefore \binom{9}{2}p^2(1-p)^7 p = \binom{9}{2}p^3(1-p)^7$$

คิดแบบใช้สูตร

$$nb(7,3,p) = \binom{7+3-1}{3-1}p^3(1-p)^7 = \binom{9}{2}p^3(1-p)^7$$

Exected value

$$E(X) = \frac{r(1-p)}{p}$$

Variance

$$V(X) = \frac{r(1-p)}{p^2}$$

3.6)Poisson Probability Distribution

สูตร $p(x,\lambda) = \dfrac{e^{-\lambda}\cdot\lambda^{x}}{x!}$ $x = 0,1,2,3,\ldots$

λ = อัตรา เทียบต่อ เวลา หรือ ต่อ พท.

• สามารถใช้การแจกแจงแบบ Poisson มาประมาณค่า Binomial ได้ เงื่อนไข

• สมมุติมี Binomial pmf $b(x;n,p)$

1) $n \to \infty$ (จำนวนครั้งของการทดลองต้องมีค่ามากๆ)

2) $p \to 0$ (prob ของ S ต้องมีค่าน้อยๆ)

* ถ้าผ่าน 2 เงื่อนไข แล้ว $b(x;n,p) \to p(x;\lambda)$.

$$\boxed{\lambda = np}$$

* ดูง่ายๆ ว่าถ้า $n > 50$ และ $np < 5$ การประมาณค่านี้สามารถใช้ได้

Example 3.40

- If publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors,
- so that probability of any given page containing **at least one** such **error** is 0.005 and $P(S) = 0.005$
- errors are independent from page to page,
- what is probability that one of its 400-page novels will contain n
  1) **exactly** one page with errors?
  2) **at most** three pages with errors?
- With ***S*** denoting page containing at least one error and ***F*** an error-free page,
- the number ***X*** of pages containing at least one error is a binomial random variable with $n = 400$ and $p = 0.005$, so

$\lambda = np = (400)(0.005) = 2$

225

กำหนด X = จำนวนหน้าที่ error

$n > 50$: $400 > 50$ ผ่าน

$np < 5$: $400(0.005) = 2 < 5$ ผ่าน

$\lambda = np = 2$

$p(x;\lambda) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

$p(1;2) = \dfrac{e^{-2}(2)^{1}}{1!}$

$p(X\le 3;2) = \sum_{i=0}^{3} \dfrac{e^{-\lambda}(\lambda)^{x_i}}{x_i!}$

$= \sum_{i=0}^{3} \dfrac{e^{-2}(2)^{x_i}}{x_i!}$

ถ้าใช้ Binomial ปกติ คำนวณยาก

$b(x;n,p) = \binom{n}{x} p^{x}(1-p)^{n-x}$

$p(X=1) = \binom{400}{1}(0.005)^{1}(0.995)^{399}$ ← ยกกำลังสามร้อยกว่า Haiya!

* ยิ่ง n มีค่ามากๆ และ p มีค่าน้อยๆ มากเท่าไรค่าประมาณก็จะใกล้เคียง Binomial มากขึ้น

Expected value and Variance

$E(X) = V(X) = \lambda$

Standard Deviation

$\sigma = \sqrt{\lambda}$

-> Poisson Process

↳ ประยุกต์ใช้กับการเกิดขึ้นของ Event ที่เกิดแบบ overtime (ไปเรื่อยๆบนช่วงเวลา)

เช่น

คนเข้าเว็บ

เส้นเวลา

- จำนวนผู้เข้าชมเว็บไซต์เว็บไซต์หนึ่ง
- สัญญาณ pulses ที่วิ่งไปยังวงจรรับ
- email/messages ที่ถูกส่งไป address หนึ่ง
- อุบัติเหตุ ที่เกิดขึ้นในโรงงานอุตสาหกรรม

คล้าย $p(x;\lambda) = \frac{e^{-\lambda}\cdot\lambda^{x}}{x!}$

สูตร $$P_k(t) = \frac{e^{-\alpha t}\cdot(\alpha t)^k}{k!}$$

$\lambda = \alpha t$

3000ครั้งใน 5 นาที

αt = ค่าเฉลี่ยของจำนวนเหตุการณ์ที่เกิดขึ้นในช่วงเวลา t

α = ค่าเฉลี่ยต่อ 1 หน่วยเวลา eg. 5 รอบ/วิ   100 word/minute
(เหมือนอัตราอะไรซักอย่าง)

Ex รถวิ่งด้วยความเร็ว 100 km/hr วิ่งไป 5 hr แล้วจะไปได้ 500 km.

α αt

Example 42

α

- Suppose pulses arrive at counter at average rate of **six per minute**.
- Find probability that in 0.5-min interval **at least** one pulse is received.

t $X \le 1$

Solution

$\alpha = 6/\text{min}$

$t = 0.5$ min

$\alpha t = 3$

Sol/n
$$P_{k\ge 1}(0.5) = 1 - P_0(0.5)$$
$$= 1 - \frac{e^{-(3)}(3)^0}{0!} = 0.950$$

Example

$\alpha = 8.6/hr.$

- Customers arrive to a bank according to Poisson Process having a constant average rate of 8.6 customers per hour.
- Suppose we begin observing the bank at some point in time.

a) What is the expected value of the number of customers that arrive in the first 30 min.? — $t = 0.5$ hr.

b) What is the probability that 3 customers arrive in the first 30 min.? $t = 0.5$ hr.

a) ถามหา $\alpha t = ?$

$\alpha t = (8.6)(0.5) = 4.3$

b) $P_3(0.5) = \frac{e^{-4.3}(4.3)^3}{3!} = 0.18$ or 18 %