บทที่ 3
ตัวแปรเชิงสุ่ม

3.1 ตัวแปรเชิงสุ่มแบบไม่ต่อเนื่อง และตัวแปรเชิงสุ่มแบบต่อเนื่อง

Discrete and Continuous Random Variables

ตัวแปรเชิงสุ่ม คือตัวแปรซึ่ง การเกิดขึ้นของค่าต่างๆ ที่เป็นตัวเลขทั้งหลายนั้น ถูกกำหนดโดยโศลก หรือ โชค (chance) ใช้อักษรตัวใหญ่ แทนตัวแปร และอักษรตัวเล็ก แทนตัวเลขที่สังเกตได้

ตัวอย่างที่ 1 ให้ Y คือจำนวนเรือประมงที่เกิดอุบัติเหตุในทะเลอ่าวไทยใน 1 เดือน Y เป็นตัวแปรเชิงสุ่มเพราะมีค่าที่เปลี่ยนแปลงทุกเดือน สมมติเดือนนี้ มีจำนวนอุบัติเหตุ 5 ลำ ดังนั้น $y = 5$

ตัวอย่างที่ 2 X คือจำนวนยาเป็น ซีซีที่ใช้ควบคุมอาการโรคลมชัก X จะเป็นตัวแปรเชิงสุ่ม เนื่องจากปริมาณที่ใช้กับคนไข้แต่ละคนจะไม่เท่ากัน เพราะคนไข้มีการดูดซึม และอาการของโรคไม่เหมือนกัน แม้ในคนไข้คนเดียวกันก็ยังใช้ขนาดของยาต่างกันเมื่อเวลาเปลี่ยนไป

นิยาม 3.1 ตัวแปรเชิงสุ่ม X จะเป็น**แบบไม่ต่อเนื่อง** ถ้ามีค่าต่างๆ เป็นแบบ**นับถ้วน** (finite)

นิยาม 3.2 ตัวแปรเชิงสุ่ม X จะเป็น**แบบต่อเนื่อง** ถ้ามีค่าต่างๆ อยู่ในช่วงเลขจำนวนจริง และโอกาสที่จะมีค่าเฉพาะค่าใดค่าหนึ่ง = 0

จากตัวอย่างที่ 1 Y เป็นตัวแปรแบบไม่ต่อเนื่อง และ X เป็นตัวแปรแบบต่อเนื่อง ค่าของ Y จะเป็นจำนวนนับถ้วน ส่วนค่าของ X สมมติอยู่ระหว่าง 0 - 0.3 ซีซี [0, .3] **ข้อกำหนดว่า P(ค่าใดๆ) = 0** สำหรับตัวแปรแบบต่อเนื่อง หมายความว่า ก่อนการกำหนดขนาดของยา ถ้าจะหาโอกาสที่จะฉีดยาขนาด .25981725 ซีซี เป็นเท่าใด ต้องตอบว่า 0 เพราะไม่สามารถตวงขนาดของยาขนาด .25981725 ซีซี โดยไม่มีความผิดพลาด โดยปกติ ตัวแปรแบบไม่ต่อเนื่อง จะสัมพันธ์กับข้อมูลแบบ**แจงนับ** (count data) ส่วนตัวแปรแบบต่อเนื่องจะสัมพันธ์กับข้อมูลแบบวัดค่า (measurement data)

$$E[H(X)] = \sum_{\text{all } x} H(x).f(x)$$

เช่น ถ้า $H(X) = X$, $E[X] = \sum_{\text{all } x} x\, f(x)$

ถ้า $H(X) = X^2$, $E[X^2] = \sum_{\text{all } x} x^2 f(x)$

ถ้า $H(X) = (X-C)^2$, $E[(X-C)]^2 = \sum_{\text{all } x} (x-C)^2.f(x)$, C เป็นค่าคงที่

<u>**ตัวอย่างที่ 3**</u> ถ้าโยนลูกเต๋าสมดุลย์ 1 ลูก X คือหน้าของลูกเต๋า X จะมีการแจกแจง**แบบยูนิฟอร์ม** และมี density ดังนี้

x	1	2	3	4	5	6
f(x)	1/6	1/6	1/6	1/6	1/6	1/6

$$E[X] = \sum x_i f(x_i)$$

$$= 1(1/6) + 2(1/6) + \ldots + 6(1/6) = 3.5 = \mu_x$$

หมายความว่า ถ้าโยนลูกเต๋าที่สมดุลย์ซ้ำๆ กันหลายๆ ครั้ง (ในระยะยาว) ซึ่งในแต่ละครั้งจะได้ค่า 1, 2, ..., 6 ค่าใดค่าหนึ่ง ดังนั้น ค่าเฉลี่ยของแต่ละครั้งคือ 3.5

ทฤษฎีที่ 3.1 กฎการหาค่าคาดหมาย

ถ้า X และ Y เป็นตัวแปรเชิงสุ่ม C เป็นเลขจำนวนจริงใดๆ

1. $E[C] = C$
2. $E[CX] = C\,E[X]$
3. $E(X + Y) = E(X) + E(Y)$

<u>ตัวอย่างที่ 4</u> เปรียบเทียบยา 2 ชนิด ซึ่งใช้ควบคุมการเต้นของหัวใจของผู้เป็นโรคหัวใจ ให้ผลการรักษา โดยนับอัตราการเต้น ใน 1 นาที ดังนี้

x	40	60	68	70	72	80	100
f(x)	.01	.04	.05	.80	.05	.04	.01

.90

y	40	60	68	70	72	80	100
f(y)	.40	.05	.04	.02	.04	.05	.40

.10

3.2 Density Functions

นิยาม 3.3 **Discrete Density**

ถ้า X เป็นตัวแปรแบบไม่ต่อเนื่อง จะมี density function ดังนี้

$P(X = x) = p(x) = f(x)$ หมายถึงความน่าจะเป็นที่ตัวแปรเชิงสุ่ม X จะมีค่า x เช่น f(2) คือ มีค่าเป็น 2

คุณสมบัติของ f(x) คือ

1. $f(x) > 0$
2. $\sum_{\text{all } x} f(x) = 1$

density function ของตัวแปรแบบไม่ต่อเนื่อง นิยมเขียนในรูปตาราง

นิยาม 3.4 **Continuous Density**

ถ้า X เป็นตัวแปรแบบต่อเนื่อง density function จะมีคุณสมบัติ ดังนี้

1. $f(x) > 0$
2. พื้นที่ระหว่างแกน X และ f(x) รวมกัน = 1
3. $P[a \leq X \leq b]$ คือพื้นที่ภายใต้กราฟ f ระหว่างค่า $x = a$ และ $x = b$
4. $f(x) = 0$ สำหรับค่าเฉพาะใดๆ ของ x

3.3 ค่าคาดหมายและพารามิเตอร์

นิยาม 3.5 **ค่าคาดหมาย** (Expected Value)

ถ้า X เป็นตัวแปรเชิงสุ่ม E(X) หมายถึงค่าของ X **ในระยะยาว** ซึ่งคือค่า เฉลี่ยของ X นั่นเอง

นิยาม 3.6 ถ้า X เป็นตัวแปรแบบไม่ต่อเนื่องด้วย density f(x) และ H(x) เป็น function ของ X

จากตัวอย่างที่ 4 ทั้ง f(x) และ f(y) สมมาตร ดังนั้น $\mu_x = 70$, $\mu_y = 70$ แม้ว่า ยา 2 ชนิดจะมีค่าเฉลี่ยเท่ากัน แต่ X จะให้ผลดีกว่าเพราะ ค่าตรงกลางคือ 68 - 72 มีโอกาสเกิดสูงถึง .90 ส่วนค่าผิดปกติ คือ สูงไปหรือต่ำไป มีโอกาสเกิดเพียง 10% ในขณะที่ Y จะให้ผลด้อยกว่าเพราะ อัตราการเต้น ระหว่าง 68 - 72 มีโอกาสเกิดเพียง 10% แต่ค่าผิดปกติมีโอกาสสูงถึง 90% ดังนั้น การมีความรู้เฉพาะค่าเฉลี่ยเพียงอย่างเดียว จึงไม่เพียงพอ จำเป็นต้องทราบความ ผันแปรของตัวแปรแต่ละตัวด้วย

นิยาม 3.7 **ความแปรปรวน**

ถ้า X เป็นตัวแปรเชิงสุ่ม ซึ่งมีค่าเฉลี่ย μ และความแปรปรวน σ^2

$V(X) = \sigma^2 = E[(x - \mu)^2]$ สูตรนิยาม

$= E(X^2) - \mu^2$ สูตรเครื่องคำนวณ

โดยที่ $\mu = E(X) = \sum_{\text{all } x} x.f(x)$

<u>ตัวอย่างที่ 5</u> จากตัวอย่างที่ 4

$V(X) = \sum (x - 70)^2 f(x)$

$= (-30)^2(.01) + (-10)^2(.04) + (-2)^2(.05) + \ldots + (30)^2(.01) = 26.4$

$V(Y) = \sum (y - 70)^2 f(y)$

$= 730.32$

$V(X) < V(Y)$ แสดงว่า X ให้ผลค่อนข้างคงที่ แต่ Y มีความผันแปรสูง

ข้อสังเกต

1. ค่า 26.4 เมื่อยังไม่ได้เปรียบเทียบกับค่าอื่น จะยังตัดสินไม่ได้ว่าสูงหรือต่ำ
2. หน่วยของ σ^2 ไม่มีความหมาย เช่น "กำลังสองของหัวใจ" ดังนั้นในการรายงานค่า σ^2 จึงไม่ต้องกำกับด้วยหน่วย
3. ถ้าจะรายงานผลโดยมีหน่วยกำกับ จะต้องใช้ ค่าเบี่ยงเบนมาตรฐาน (σ)

<u>ตัวอย่างที่ 6</u> จากตัวอย่างที่ 4

$\sigma(X) = \sqrt{V(X)} = \sqrt{26.4} = 5.4$ ครั้ง/นาที

$\sigma(Y) = \sqrt{V(Y)} = \sqrt{730.32} = 27.02$ ครั้ง/นาที

ทฤษฎีที่ 3.2 กฎการหาความแปรปรวน

1. $V(C) = 0$
2. $V(CX) = C^2 V(X)$
3. ถ้า X และ Y เป็นอิสระกัน ดังนั้น $V(X \pm Y) = V(X) \pm V(Y)$

<u>ตัวอย่างที่ 7</u> แสดงการใช้สูตรคำนวณหาความแปรปรวน

x	2	3	4
f(x)	.2	.5	.3

$E(X) = \mu(X) = 2(.2) + 3(.5) + 4(.3) = 3.1$

$E(X^2) = \sum x^2 f(x) = 4(.2) + 9(.5) + 16(.3) = 10.1$

$V(X) = E(X^2) - \mu^2(X)$

$= 10.1 - (3.1)^2 = .49$

$\sigma(X) = \sqrt{.49} = .70$

3.4 ทฤษฎีของ Chebyshev's inequality เป็นทฤษฎีที่ใช้ประโยชน์ของค่าเบี่ยงเบนมาตรฐานในการหาความน่าจะเป็น ดังนี้

$P(\mu - k\sigma < X < \mu + k\sigma) = 1 - 1/k^2$, k เป็นค่าคงที่

เช่นถ้า $k = 2$, $\mu = 3$, $\sigma = 1$

$P(3\text{-}2 < X < 3+2) = 1 - 1/4 = .75$

$P(1 < X < 5) = .75$

นั่นคือโอกาสที่ X จะมีค่า 2σ ไปจากค่าเฉลี่ย = .75

แบบฝึกหัดที่ 3

1. จงระบุว่าเป็นตัวแปรเชิงสุ่มแบบใด
   1.1 V = ปริมาณปัสสาวะที่ขับถ่าย/วัน
   1.2 B = ปริมาณเลือดที่คนไข้สูญเสียเนื่องจากการผ่าตัด
   1.3 H = ความต้องการแสงสว่างของพืชชนิดหนึ่ง เป็นชม./วัน
   1.4 C = จำนวนผึ้งแรงงานต่อ 1 รัง
   1.5 R = ปริมาณน้ำฝน/วัน ในเมืองหนึ่ง
   1.6 W = น้ำหนักที่เพิ่มขึ้นต่อสัปดาห์ของมารดาขณะตั้งครรภ์
2. ผู้ผลิตยาแก้ปวดหัวไมเกรน โฆษณาว่าให้ผลการรักษา 90% ให้ X คือจำนวนผู้หายปวดหัวจากผู้ทดลองใช้ 3 คน จงหา
   2.1 $f(x)$
   2.2 $P(X \leq 1)$
   2.3 ถ้าทั้ง 3 คนไม่หายปวดหัว จะยังเชื่อว่ายานี้มีสรรพคุณตามอ้างหรือไม่?
3. จากข้อ (2) จงหา $E(X)$, $\mu(X)$, $E(X^2)$, $V(X)$, $\sigma(X)$
4. ถ้า X และ Y เป็นอิสระกัน กำหนดให้ $\mu(X) = 2$, $\mu(Y) = 6$, $V(X) = 9$, $V(Y) = 16$ จงหา
   4.1 $\sigma(X)$, $\sigma(Y)$
   4.2 $E(X^2)$, $E(Y^2)$
   4.3 $E(X+2Y)$, $V(X+2Y)$
   4.4 $E(3X-2Y-2)$, $V(3X-2Y-2)$
   4.5 $E[(X-2)/3]$, $V[(X-2)/3]$
   4.6 $E[(Y-6)/4]$, $V[(Y-6)/4]$
   4.7 เหตุใด (4.5) และ (4.6) จึงให้ผลเท่ากัน
5. ถ้า X คือปริมาณน้ำฝน/สัปดาห์ของเมืองหนึ่ง ซึ่งมี $\mu = 1.00$ นิ้ว, $\sigma = .25$ นิ้ว ถ้าในสัปดาห์หนึ่งมีฝนตกมากว่า 2 นิ้ว จะถือเป็นเรื่องผิดปกติหรือไม่ จงอธิบายโดยใช้กฎของ Chebyshev's inequality

เฉลยแบบฝึกหัดที่ 3

1. ทุกข้อเป็น Continuous ยกเว้น 1.4
2. A = หายปวดหัว $P(A) = .9$, $P(A') = .1$

$f(0) = P(A'A'A') = (.1)^3 = .001$

$f(1) = P(AA'A') + P(A'AA') + P(A'A'A) = 3(.9)(.1)^2 = .027$

$f(2) = P(AAA') + P(AA'A) + P(A'AA) = 3(.9)^2(.1) = .243$

$f(3) = P(AAA) = (.9)^3 = .729$

2.1

x	0	1	2	3
f(x)	.001	.027	.243	.729

2.2 $P(X \leq 1) = f(0) + f(1) = .028$

2.3 $f(0) = .001$ ควรสงสัยสรรพคุณที่อ้าง เพราะโอกาสของการเกิดเหตุการณ์นี้มีเพียง 1 ใน 1,000 ถ้า ยาสามารถรักษาโรคได้ 90% จริง ดังนั้น ถ้าเกิด $x = 0$ ให้สงสัยว่า $P(A) \neq .90$

3. 3.1 $E(X) = \sum xf(x) = 0(.001) + 1(.027) + 2(.243) + 3(.729)$

$= 2.7$ คน

ถ้ายาดังกล่าวมีสรรพคุณรักษาได้ 90% จริง ดังนั้น ในระยะยาวถ้ามีคนปวดไมเกรน 3 คน กินยาดังกล่าว จะมีผู้หายโดยเฉลี่ย 2.7 คน

3.2 $\mu(X) = E(X) = 2.7$ คน

3.3 $E(X^2) = 0(.001) + 1(.027) + 4(.243) + 9(.729) = 7.56$

3.4 $V(X) = E(X^2) - \mu^2(X) = 7.56 - (2.7)^2 = 0.27$

$\sigma(X) = \sqrt{.27} = .52$ คน

4. 4.1 $\sigma(X) = \sqrt{9} = 3$, $\sigma(Y) = \sqrt{16} = 4$

4.2 $E(X^2) = V(X) + \mu^2(X) = 9 + 4 = 13$

$E(Y^2) = V(Y) + \mu^2(Y) = 16 + 36 = 52$

4.3 $E(X+2Y) = E(X) + 2E(Y) = 2 + 2(6) = 14$

$V(X+2Y) = V(X) + 4V(Y) = 9 + 4(16) = 73$

4.4 $E(3X-2Y-2) = 3E(X) - 2E(Y) - 2$

$= 3(2) - 2(6) - 2 = -8$

$V(3X-2Y-2) = (V(X) + 4V(Y) - 0$

$= 9(9) + 4(16) = 145$

4.5 $E[(X-2)/3] = (1/3)E(X-2) = (1/3)[E(X) - 2]$

$= (1/3)(2-2) = 0$

$V[(X-2)/3] = (1/9)V(X-2) = (1/9)[V(X) - 0]$

$= (1/9)(9) = 1$

4.6 $E[(Y-6)/4] = (1/4)[E(Y-6)] = (1/4)[E(Y) - 6]$

$= (1/4)(6-6) = 0$

$V[(Y-6)/4] = (1/16)V[(Y-6)] = (1/16)[V(Y) - 0]$

$= (1/16)(16) = 1$

4.7 ข้อ 4.5 = 4.6 เพราะใช้กฎ

E[(ตัวแปร - ค่าเฉลี่ย)/ค่าเบี่ยงเบนมาตรฐาน] = 0

V[(ตัวแปร - ค่าเฉลี่ย)/ค่าเบี่ยงเบนมาตรฐาน] = 1

นั่นคือ $E[(X-\mu)/\sigma] = (1/\sigma)[E(X) - \mu] = (1/\sigma)(\mu - \mu) = 0$

$V[(X-\mu)/\sigma] = (1/\sigma^2)[V(X) - V(\mu)] = (1/\sigma^2)(\sigma^2 - 0) = 1$

5. $\mu = 1, \sigma = .25$ ลองแทนค่า k จนถึงค่า k = 4 จะได้

$P[1 - 4(.25) \leq X \leq 1 + 4(.25)] = 1 - 1/16$

$= P(1 - 1 \leq X \leq 1 + 1) = 15/16 = .9375$

$= P(0 \leq X \leq 2) = .9375$

$P(X > 2) = 1 - P(X \leq 2) = 1 - .9375 = .0625$ ซึ่งเป็นค่าที่ไม่โตนัก ดังนั้น การที่ฝนตกมากกว่า 2 นิ้ว/สัปดาห์ จึงถือเป็นเรื่องไม่ปกติ
