บทที่ 7

ตัวอย่างจากการทดลองหรือจากปรากฏการณ์ธรรมชาติที่สอดคล้องกับการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

ในบทนี้ได้ยกตัวอย่างจากการทดลองหรือจากปรากฏการณ์ธรรมชาติที่สอดคล้องกับการแจกแจงความน่าจะเป็นของตัวแปรสุ่มทั้งชนิดไม่ต่อเนื่องและชนิดต่อเนื่องที่สำคัญบางอัน ดังต่อไปนี้

7.1 ตัวอย่างที่สอดคล้องกับการแจกแจงแบบทวินาม

Defective items in mass production

การที่เครื่องจักรผลิตวัตถุแต่ละชิ้นออกมาเรื่อย ๆ ชิ้นต่อชิ้น แน่นอนวัตถุที่ผลิตออกมาจะต้องมีวัตถุที่เสียปนออกมาด้วย ในการที่เครื่องจักรผลิตวัตถุออกมาแต่ละครั้งถ้าเราสนใจวัตถุที่ดี และวัตถุที่เสีย ก็จะสอดคล้องกับการทดลองแบบ Bernoulli ถ้าให้ S แทนวัตถุที่ดี ดังนั้น F แทนวัตถุที่เสีย ถ้าทำการผลิต n ครั้ง ต้องการได้วัตถุที่ดี x ครั้ง ก็จะสอดคล้องกับการทดลองแบบทวินาม

สมมุติ $n = 5$ คือผลิต 5 ครั้ง ต้องการได้วัตถุดี 3 ครั้ง ดังนั้นผลที่ผลิตได้อาจจะเป็น SSFSF หรือ SSSFF หรือ SFSFS ... ถ้าความน่าจะเป็นที่จะได้ $S = p$ และความน่าจะเป็นที่จะได้ $F = q$ และการกระทำแต่ละครั้งเป็นอิสระดังนั้นความน่าจะเป็นที่จะได้ $S = 3$ คือ

$$P(SSFSF) = p.p.q.p.q = p^3q^2$$

$$P(SSSFF) = p.p.p.q.q = p^3q^2$$

$$\vdots$$

และทำการผลิต 5 ครั้งต้องการได้ s 3 ครั้ง จะทำได้ทั้งหมด = $\binom{5}{3}$ วิธี

และแต่ละวิธีมีความน่าจะเป็นเท่ากับ p^3q^2

ดังนั้นความน่าจะเป็นที่จะได้ s 3 ครั้ง เป็น $\binom{5}{3}\ p^3q^3$

ด้วยวิธีการเดียวกันนี้ ถ้าเครื่องจักรผลิตวัตถุทั้งหมด n ชิ้น ต้องการได้วัตถุดี(s) อยู่ x ชิ้น ฉะนั้นจะเป็นวัตถุเสียอยู่ $n - x$ ชิ้น ความน่าจะเป็นที่จะได้วัตถุดี x ชิ้น จะเท่ากับ $\binom{n}{x}p^xq^{n-x}$ เมื่อ $q = 1 - p$

ซึ่งเป็นฟังก์ชันของการแจกแจงทวินาม มี n, p เป็นพารามิเตอร์

7.2 <u>ตัวอย่างที่สอดคล้องกับการแจกแจงปัวส์ซอง</u>

<u>กระบวนการปัวส์ซอง</u> (Poisson Process)

พิจารณาสารกัมมันตภาพรังสีอย่างหนึ่งซึ่งกำลังปล่อยอนุภาคแอลฟ่า (α - particle) อยู่

ถ้าให้ X_t เป็นจำนวนอนุภาคที่แพร่กระจายออกมาในช่วง $[0,t)$ ดังนั้น X_t จะเป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง ที่มีค่าเป็น 0, 1, 2,... และ X_t เป็นฟังก์ชันของ t ในการหาการแจกแจงความน่าจะเป็นของ X_t เราต้องสร้างข้อสมมติของ X_t ให้สอดคล้องกับสถานการณ์จริง และอยู่ในแบบที่จะคำนวณหาค่าได้

ถ้าให้ $P_n(t) = P(X_t = n) \quad n = 0,1,2,\ldots$

ข้อสมมติเป็นดังนี้

1. จำนวนอนุภาคที่แพร่กระจายในระหว่างช่วงเวลาต่าง ๆ ที่ไม่ซ้อนกันจะเป็นตัวแปรสุ่มอิสระ

2. การแจกแจงของจำนวนอนุภาคที่แพร่กระจายในช่วงเวลาหนึ่งขึ้นอยู่กับความกว้างของช่วงเวลาเท่านั้น ไม่ขึ้นอยู่กับว่าเป็นเวลาอะไร กล่าวคือ ถ้า X_t เป็นจำนวนอนุภาคที่แพร่กระจายในช่วงเวลา $[0,t)$

ให้ Y_t เป็นจำนวนอนุภาคที่แพร่กระจายในช่วงเวลา $[t_1, t_1+t)$

ดังนั้นเมื่อ $t_1 \longrightarrow 0$ ตัวแปรสุ่ม X_t กับ Y_t จะมีการแจกแจงเป็นอย่างเดียวกัน

3. ถ้าช่วงเวลาเล็กอย่างเพียงพอ (Sufficsently small) แล้วความน่าจะเป็นที่อนุภาคจะแพร่ออกมา 1 ตัว ในช่วงเวลานั้น เป็นสัดส่วนกับความกว้างของช่วงเวลาดังกล่าว คือถ้า $\lambda > 0$ เป็นค่าคงที่ และ Δt เป็นความกว้างของช่วงเวลาหนึ่ง ซึ่งมีค่าน้อยแล้ว $P_1(\Delta t)$ จะมีค่าใกล้ $\lambda \Delta t$ หรือเขียนได้ว่า $p_1(\Delta t) \sim \lambda \Delta t$ หมายความว่า

$$p_1(\Delta t)/\lambda \Delta t \longrightarrow 1 \quad \text{เมื่อ } \Delta t \longrightarrow 0$$

4. ความน่าจะเป็นที่อนุภาคจะแพร่กระจายเกินกว่าหรือเท่ากับ 2 ตัว ในช่วงเวลาน้อย ๆ จะมีค่าใกล้ศูนย์คือ $\sum_{k=2}^{\infty} p_k(\Delta t) \sim 0$ ซึ่งจะได้ว่า

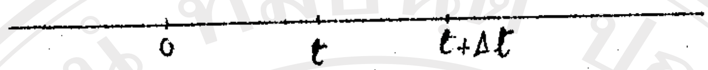
$p_k(\Delta t) \sim 0$ เมื่อ $k \geq 2$

5. ถ้าช่วงเวลาสั้นมากหรือเป็นศูนย์ จำนวนอนุภาคที่แพร่กระจายจะเป็นศูนย์ และความน่าจะเป็นที่ไม่มีอนุภาคแพร่กระจายในช่วงเวลา 0 จะเท่ากับ 1 กล่าวคือ

$$x_0 = 0 \quad \text{และ} \quad p_0(0) = 1$$

จากข้อสมมุติ 5 ข้อ สามารถดำเนินการหาการแจกแจงของ X_t ได้ดังนี้

1. จากข้อสมมติข้อ 1 และ 2 จะได้ว่าตัวแปรสุ่ม $X_{\Delta t}$ และ $X_{t+\Delta t} - X_t$ เป็นตัวแปรสุ่มอิสระ ซึ่งมีความน่าจะเป็นอย่างเดียวกัน



2. จากข้อสมมติ 3 และ 4 ได้ว่า

$$P_0(\Delta t) = 1 - P_1(\Delta t) - \sum_{k=2}^{\infty} P_k(\Delta t) \sim 1 - \lambda \Delta t \qquad \text{.........} \text{ สมการ (1)}$$

3. ได้ $P_0(t+\Delta t) = P(X_{t+\Delta t} = 0)$

$$= P\left[X_t = 0 \text{ และ } X_{t+\Delta t} - X_t = 0\right]$$

$$= P_0(t) \cdot P_0(\Delta t)$$

$$\sim P_0(t)\left[1 - \lambda \Delta t\right]$$

4. ดังนั้นได้ $\dfrac{P_0(t+\Delta t) - P_0(t)}{\Delta t} \sim -\lambda P_0(t)$

เมื่อให้ $\Delta t \to 0$ จะได้ค่าทางซ้ายมือเป็นค่าอนุพันธ์ของฟังก์ชัน P_0 คือ

$$P_0'(t) = -\lambda P_0(t)$$

$$\frac{P_0'(t)}{P_0(t)} = -\lambda$$

อินทิเกรตทั้ง 2 ข้างเทียบกับ t จะได้ $\ln P_0(t) = -\lambda t + c$, c เป็นค่าคงที่ จากข้อสมมติ 5 เมื่อ $t = 0$ จะได้ $c = 0$ ดังนั้น

$P_0(t) = e^{-\lambda t}$ นั่นคือ $P(X_t = 0) = e^{-\lambda t}$

5. พิจารณา $P_n(t+\Delta t) = P(X_{t+\Delta t} = n)$

เหตุการณ์ที่ $X_{t+\Delta t} = n$ ก็คือเหตุการณ์ที่ $X_t = x$ และ $X_{t+\Delta t} - X_t = n - x$ เมื่อ $x = 0,1,2,\ldots, n$ ดังนั้นจะได้

$$P_n(t+\Delta t) = P(X_{t+\Delta t} = n)$$

$$= \sum_{x=0}^{n} P\left[X_t = x \text{ และ } (X_{t+\Delta t} - X_t) = n-x\right]$$

$$= \sum_{x=0}^{n} P(X_t = x) \cdot P(X_{\Delta t} = n-x)$$

$$= \sum_{x=0}^{n} P_x(t) \cdot P_{n-x}(\Delta t)$$

$$= \sum_{x=0}^{n-2} P_x(t) P_{n-x}(\Delta t) + P_{n-1}(t) P_1(\Delta t)$$

$$+ P_n(t)\; P_0(\Delta t)$$

จากข้อสมมติ 3,4 และสมการ (1) จะได้

$$P_n(t + \ t) \sim P_{n-1}(t)\lambda \Delta t + P_n(t)\left[1-\lambda \Delta t\right]$$

ดังนั้น $\dfrac{P_n(t+\Delta t) - P_n(t)}{\Delta t} \sim \lambda P_{n-1}(t) - \lambda P_n(t),\ n = 1,2,\ldots$

ให้ $\Delta t \dashrightarrow 0$ ทางซ้ายมือเป็นอนุพันธของ P_n จะได้

$$P_n'(t) = -\lambda\, P_n'(t) + \lambda\, P_{n-1}(t), \quad n = 1,2,\ldots$$

ซึ่งเป็นสมการดิฟเฟอเรนเชียลแบบเชิงเส้น ถ้ากำหนดฟังก์ชัน

$q_n(t) = e^{\lambda t}\, P_n(t)$ แล้วหาค่าอนุพันธ์จะได้

$q_n'(t) = \lambda q_{n-1}(t), \quad n = 1,2,\ldots$

เพราะว่า $P_0(t) = e^{-\lambda t}$ ดังนั้นได้ว่า $q_0(t) = 1$

และจาก $q_n(0) = 0$ เมื่อ $n > 0$ จะได้สูตรออกมาเป็น

$q_1'(t) = \lambda$ ดังนั้น $q_1(t) = \lambda t$

$q_2'(t) = \lambda q_1(t) = \lambda^2 t$ ดังนั้น $q_2(t) = \frac{(\lambda t)^2}{2}$

ในกรณีทั่วไปได้ $q_n'(t) = \lambda q_{n-1}(t)$

ดังนั้น $q_n(t) = \frac{(\lambda t)^n}{n!}$

จาก q_n ที่กำหนด จะได้

$$p_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \quad , n = 0,1,2,\ldots$$

ซึ่งเป็นฟังก์ชันของการแจกแจงแบบปัวส์ซอง

<u>ดังนั้นจึงสรุปได้ว่า</u> จำนวนอนุภาคที่แพร่กระจายในช่วงเวลา $[0,t)$ จากสารกัมมันตภาพรังสี มีการแจกแจงเป็นแบบปัวส์ซอง มี λt เป็นพารามิเตอร์

7.3 <u>ตัวอย่างที่สอดคล้องกับการแจกแจงแบบไฮเปอร์จีออเมตริก</u>

<u>การประมาณจำนวนเหตุการณ์สถิติชีพโดยวิธีของจันทราเสการ์และเดมมิง</u>

สถิติชีพ (Vital Statistics) หมายถึงสถิติที่เกี่ยวกับการเกิด การตาย การเจ็บป่วย การสมรส การรับโอนบุตรบุญธรรม ตลอดจนความสัมพันธ์ของตัวแปรเหล่านั้น

การประมาณเหตุการณ์สถิติชีพ ทำขึ้นเพื่อจะแก้ปรับจำนวนหรืออัตราต่าง ๆ เกี่ยวกับสถิติชีพ วิธีการประมาณของจันทราเสการ์และเดมมิง (Chandraseka and Deming) มีหลักการดังนี้

1. จากระบบที่มีอยู่แล้ว (Existing System) คือระบบจดทะเบียนสถิติชีพ จะต้องสร้างระบบอีกระบบหนึ่งที่ไม่ขึ้นต่อระบบที่มีอยู่ เรียกว่าระบบบันทึกทวิ (Dual record system) หมายความว่าทั้ง 2 ระบบมีการบันทึกข้อมูลคนละอย่างในเหตุการณ์เดียวกัน

ระบบที่สร้างขึ้นมี 2 วิธีการคือ โดยการสำรวจด้วยตัวอย่าง และโดยการสร้างระบบการรายงานเหตุการณ์โดยวิธีอื่นเช่น วิธีการใช้บุคคลากรอาสาสมัครสาธารณสุขประจำหมู่บ้านเป็นผู้ทำรายงานการเกิด การตาย การย้ายที่อยู่ เป็นต้น ในระบบทวินี้เหตุการณ์เดียวกัน จะต้องทำการรวบรวมด้วยการสังเกตต่อเนื่อง ดังนั้นระบบทวิจึงช่วยให้สามารถทำการเปรียบเทียบจำนวนเหตุการณ์

2. เมื่อรวบรวมข้อมูลสถิติชีพจากทั้ง 2 ระบบภายในช่วงระยะเวลาที่กำหนดได้แล้วก็นำข้อมูลนี้มาเปรียบเทียบกัน โดยวิธีการจับคู่เปรียบเทียบ นับจำนวนดูว่ามีจำนวนเหตุการณ์ที่พบในบันทึกทั้ง 2 ระบบอยู่เท่าใด การเปรียบเทียบดังกล่าวจะช่วยให้สามารถประเมินคุณภาพของระบบ และอาจทำการแก้ไขข้อบกพร่องได้ วิธีการเปรียบเทียบจะต้องกำหนดหลักเกณฑ์ในการเปรียบเทียบที่สมเหตุสมผล เพื่อขจัดข้อบกพร่อง และความเฉเอียงอันอาจจะเกิดขึ้นได้ในการเปรียบเทียบ และจะได้ค่าประมาณที่ไม่ต่ำหรือสูงเกินความเป็นจริง

3. ทำการประมาณค่ารวมของเหตุการณ์สถิติชีพ โดยอาศัยหลักการณ์ที่ว่าวิธีการทั้ง 2 จะต้องมีอิสระต่อกัน จากนั้นจึงสามารถประมาณความสมบูรณ์ของการรายงานของทั้ง 2 ระบบได้

สูตรของจันทราเสการ์-เดมมิง ในการประมาณจำนวนเหตุการณ์สถิติชีพ มีดังนี้

$$\hat{N} = X.Y/C \quad \text{หรือ} \quad \hat{N} = (N_1-C)(N_2-C)/C$$

$$= C + N_1 + N_2 + N_1N_2/C$$

เมื่อ X คือจำนวนเหตุการณ์ที่พบในบันทึกในระบบที่ 1

Y คือจำนวนเหตุการณ์ที่พบในบันทึกในระบบที่ 2

C คือจำนวนเหตุการณ์ที่พบในบันทึกระบบที่ 1 และระบบที่ 2

N_1 คือจำนวนเหตุการณ์ที่พบในบันทึกระบบที่ 1 แต่ไม่พบในระบบที่ 2

N_2 คือจำนวนเหตุการณ์ที่พบในบันทึกระบบที่ 2 แต่ไม่พบในระบบที่ 1

N คือค่าประมาณของจำนวนเหตุการณ์ทั้งหมด

$\frac{N_1N_2}{C}$ คือค่าประมาณของจำนวนเหตุการณ์ที่ไม่พบในบันทึกทั้ง 2 ระบบ

ถ้าเหตุการณ์ในระบบที่ 1 และระบบที่ 2 ไม่ขึ้นต่อกันจริงแล้ว ค่าคาดหวังของ C ซึ่งเป็นค่าที่ได้จากการเก็บข้อมูลควรจะมีค่าเท่ากัน

ในระบบที่ 1 ที่มีเหตุการณ์ X มี C เป็นตัวแปรสุ่ม จะได้ฟังก์ชันการแจกแจงของ C เป็นแบบ ไฮเปอร์จีออเมตริก

$$f(c) = \frac{\binom{X}{c}\binom{N-X}{Y-c}}{\binom{N}{Y}} \quad \text{เมื่อ} \quad c = A, A+1, \ldots B$$

$$A = \max(0, X + Y - N), \quad B = \min(X,Y)$$

ได้ $E(c) = XY/N$

$$V(c) = N(X/N)(1-X/N)(Y/N)(1-Y/N)$$

พิจารณา $f(c) = \dfrac{\binom{X}{c}\binom{N-X}{Y-c}}{\binom{N}{Y}}$ เปรียบเทียบกับ $f(x) = \dfrac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$

ซึ่งเป็นฟังก์ชันของการแจกแจงแบบไฮเปอร์จีออเมตริก

ได้ว่า $X = m, \quad Y = k$ และ $N = m + n$

$$E(x) = \frac{mk}{m+n} \quad \text{ดังนั้น} \quad E(c) = \frac{XY}{N}$$

$$V(x) = \frac{mnk(m+n-k)}{(m+n)^2(m+n-1)} \quad \text{ดังนั้น}$$

$$V(c) = \frac{XY(N-X)(N-Y)}{N^2(N-1)}$$

$$= \left(\frac{N}{N}\right)\left(\frac{XY}{N}\right)\left(\frac{N-X}{N}\right)\left(\frac{N-Y}{N}\right)\left(\frac{N}{N}\right)$$

$$= N\left(\frac{X}{N}\right)\left(\frac{Y}{N}\right)\left(1-\frac{X}{N}\right)\left(\frac{N-Y}{N}\right)\left(\frac{N}{N-1}\right)$$

$$= N(X/N)(1-X/N)(Y/N)(1-Y/N)(N/N-1)$$

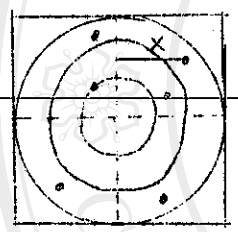
เนื่องจาก N มีขนาดใหญ่ ดังนั้น $\frac{N}{N-1}$ ประมาณ 1

ฉะนั้น $V(C) = N(X/N)(1-X/N)(Y/N)(1-Y/N)$

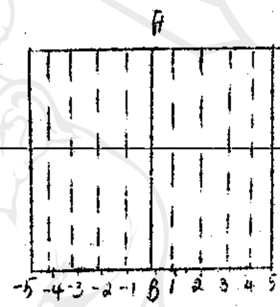
7.4 ตัวอย่างที่สอดคล้องกับการแจกแจงปกติ

พิจารณาการซ้อมยิงปืน ต้องการให้ถูกศูนย์กลางเป้ามากที่สุด หลังจากซ้อมยิงหลาย ๆ ครั้งจนชำนาญขึ้น ซึ่งคิดว่าจะไม่พลาดอีก แต่ถ้ายังพลาดก็อาจเป็นเพราะสาเหตุอื่นเช่น น้ำหนักลูกปืนไม่เท่ากัน หรือความชื้น หรืออุณหภูมิของสถานที่นั้น

หลังจากซ้อมยิงเป้าไปแล้วหลายครั้ง สมมติผลการยิงถูกเป้าปรากฏดังรูป ก.



รูป ก.



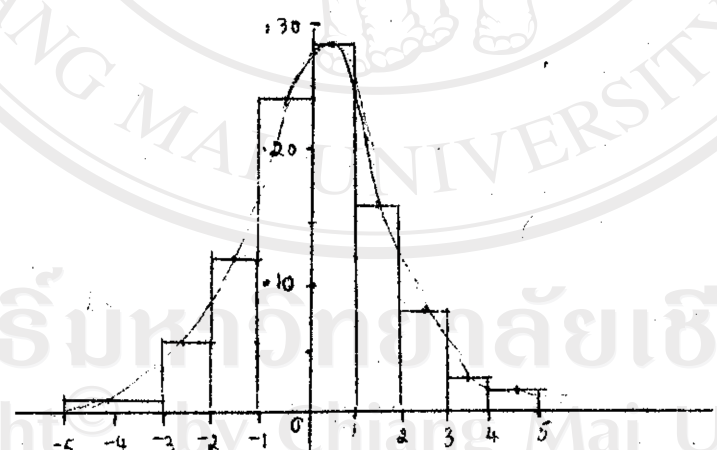
รูป ข.

ให้ x เป็นตัวแปรสุ่มมีค่าเท่ากับระยะทางที่ลูกปืนถูกเป้าห่างจากเส้นแบ่งครึ่งเป้าตามแนวดิ่ง จะเห็นว่าเป็นการยากที่จะวัด x ให้ถูกต้องจริง ๆ เพื่อหลีกเลี่ยงข้อผิดพลาดเนื่องจากการวัดที่เป็นทศนิยม จึงประมาณค่า X เป็นช่วง ๆ ถ้ายิงเป้าหลายรอบ ๆ ละ 100 ครั้ง โดยการใช้เป้ารูปสี่เหลี่ยมจัตุรัสขนาด 10" + 10" แบ่งเป้าเป็นรูปสี่เหลี่ยมผืนผ้ากว้างรูปละ 1 นิ้ว ดังรูป ข.

ให้ x มีค่าเป็นลบเมื่อลูกปืนถูกเป้าด้านซ้ายของเส้นแบ่งครึ่ง และเป็นบวกเมื่อลูกปืนถูกเป้าทางขวาของ AB ถ้าผลการยิงเป้า 100 ครั้ง ปรากฏตามตาราง

ช่วงของ x	จำนวนครั้ง	ความถี่สัมพัทธ์
$-5 \leq x \leq -4$	1	0.01
$-4 < x \leq -3$	1	0.01
$-3 < x \leq -2$	6	0.06
$-2 < x \leq -1$	13	0.13
$-1 < x \leq 0$	24	0.24
$0 < x \leq 1$	27	0.27
$1 < x \leq 2$	16	0.16
$2 < x \leq 3$	7	0.07
$3 < x \leq 4$	3	0.03
$4 < x \leq 5$	2	0.02

นำค่าจากตารางมา plot histogram จะได้ดังรูป



พื้นที่รูปผืนผ้าแต่ละรูปบนช่วง x มีค่าเท่ากับความถี่สัมพัทธ์ ถ้ายิงเป้ารอบที่สองอีก 100 ครั้ง ผลการยิงรอบที่สองค่า x อาจจะแตกต่างจากรอบที่ 1 บ้าง แต่กราฟก็ยังมีลักษณะเป็นรูประฆังคว่ำ ถ้ายิงหลาย ๆ รอบผลของ x ก็จะมีการกระจายแบบเดียวกันอีก จึงสรุปได้ว่าการกระจายของ x สอดคล้องกับการแจกแจงปกติ