# บทที่ 4
# การแจกแจงทวินาม ปัวซอง และ Normal

**4.1 การแจกแจงแบบทวินาม** ลองพิจารณาตัวอย่างต่อไปนี้

**ตัวอย่างที่ 1** ทั้งสามีและภริยา มียีนด้อย ตาสีฟ้า และยีนข่ม ตาสีน้ำตาล อยากทราบว่า ถ้าเขามีบุตร 2 คน จะมีบุตรที่ตาสีฟ้า 0, 1, 2 คน ด้วยโอกาสเท่าใด

**ตัวอย่างที่ 2** 10% ของผู้ที่เป็นพาหะของเชื้อวัณโรค สามารถแพร่เชื้อให้ผู้ใกล้ชิดได้ ถ้าวันหนึ่งผู้เป็นพาหะเชื้อวัณโรคไปติดต่อกับบุคคล 10 คน คาดว่าจะมีกี่คนที่ได้รับเชื้อจากเขา?

**ตัวอย่างที่ 3** กระทรวงเกษตรพัฒนาข้าวโพดพันธุ์ใหม่ ซึ่งให้เปอร์เซนต์งอกสูงถึง 90% ถ้าชาวไร่คนหนึ่งนำเมล็ดที่ได้รับแจก 20 เมล็ดไปเพาะในแปลงที่ดินมีสภาพเหมือน สถานีทดลองเกษตร เขาควรคาดหมายว่าจะงอกกี่ต้น? และถ้างอก 15 ต้น จะสนับสนุนตัวเลข 90% หรือไม่?

**ตัวอย่างที่ 4** 80% ของผู้ที่ถูกราชการเวนคืนที่เพื่อสร้างทางด่วนจะไม่พอใจ ค่าชดเชยที่ดิน ถ้าสุ่มผู้ถูกเวนคืนสำหรับสร้างทางสายหนึ่งมา 15 คน จงหาความน่าจะเป็นที่คนส่วนใหญ่จะคัดค้าน และจงหาโอกาสที่จะมีผู้คัดค้าน 10-14 คน

เมื่อพิจารณาตัวอย่าง ทั้ง 4 นี้ จะพบว่ามีข้อที่เหมือนกัน ดังนี้

1. **การกระทำซ้ำๆ กัน n ครั้ง** ตัวอย่างที่ 1-4 มี n = 2, n = 10, n = 20 และ n = 15 ตามลำดับ
2. **มีผลตอบแทนเพียง 2 อย่าง** ซึ่งจะเรียกว่า Success และ Failure โดยที่ Success หมายถึงลักษณะที่สนใจหรือต้องการนับ เช่น จำนวนบุตรตาสีฟ้า, จำนวนคนติดเชื้อวัณโรค, จำนวนงอก, จำนวนผู้ต่อต้าน
3. **การกระทำ n ครั้งนี้ เป็นอิสระกัน** นั่นคือ ผลตอบแทนจากการกระทำแต่ละครั้งไม่มีผลกระทบต่อการกระทำครั้งต่อไป จึงทำให้ $P(S) = p$ เป็นค่าคงที่ตลอด n ครั้ง เช่น $p = 1/4$ สำหรับลูกทุกคน, $p = .10$ สำหรับผู้ใกล้ชิดผู้เป็นวัณโรคทุกคน, เมล็ดข้าวโพดทุกเมล็ดมีโอกาสงอก = .10 ข้อสุดท้ายแม้จะเป็นการสุ่มแบบไม่แทนที่ แต่เนื่องจากประชากรที่ถูกเวนคืนมีจำนวนมาก การสุ่มมา 15 คนแบบไม่แทนที่ จะให้ผลไม่ต่างกับการสุ่มแบบมีการแทนที่ นั่นคือ $p = .80$ สำหรับทุกคน

4. **ตัวแปร X คือจำนวนความสำเร็จจากการกระทำ n ครั้ง** และเรียก X ว่าตัวแปรเชิงสุ่มแบบ **ทวินาม ด้วย** พารามิเตอร์ n และ p

จากตัวอย่างที่ 1 ให้ b = ยีน (ด้อย) ของตาสีฟ้า B = ยีน (ข่ม) ของตา สีน้ำตาล P(b) = 1/4, P(B) = 3/4, ครอบครัวที่มีลูก 2 คน จะมี S = bb, bB, Bb, BB ถ้า X คือจำนวนลูก ตาสีฟ้า จะมี ค่า $x_i$ = 2, 1, 1, 0 ; $p(x_i)$ = 1/16, 3/16, 3/16, 9/16 จะได้ density f(x) ดังนี้

| x | 0 | 1 | 2 |
|---|---|---|---|
| f(x) | 9/16 | 6/16 | 1/16 |

$$P(Bb) = P(B).P(b)$$
$$= (3/4)(1/4) = 3/16$$

**ทฤษฎีที่ 4.1** ถ้า X เป็นตัวแปรเชิงสุ่มแบบทวินาม ด้วย พารามิเตอร์ n, p

X จะมี density function ดังนี้

$$f(x) = {}^{n}C_{x}p^{x}(1-p)^{n-x} \quad , \ x = 0, 1, 2, ..., n$$

**ทฤษฎีที่ 4.2** ถ้า X เป็นตัวแปรเชิงสุ่มแบบทวินาม ด้วย พารามิเตอร์ n และ p

ดังนั้น E(X) = np, V(X) = np(1-p)

จากตัวอย่างที่ 3 n = 10, p = .10, E(X) = 10(.10) = 1

นั่นคือคาดว่าจะมีผู้ติดเชื้อวัณโรค 1 คน

จากตัวอย่างที่ 4 n = 20, p = .90, E(X) = 20(.90) = 18

ถ้าข้าวโพดมีอัตรางอก 90% จริง ถ้าปลูก 20 เมล็ด ย่อมคาดหมายได้ว่ามีจำนวนงอกประมาณ 18 เมล็ด ถ้างอกอย่างมาก 15 เมล็ด

$$P(X \leq 15) = \sum_{x=0}^{15} f(x) = \sum_{x=0}^{15} ({}^{20}C_{x})(.9)^{x}(.1)^{20-x} = .0432 = \text{p-value}$$

p-value < .05 จึงน่าสงสัยว่าตัวเลข 90% น่าจะสูงเกินไป เพราะถ้าอัตรางอก 90% จริง เหตุการณ์ที่จะงอกสูงสุด 15 เมล็ด มีโอกาสเกิดขึ้นเพียง 4.32% เท่านั้น

จากตัวอย่างที่ 4 $n = 14, p = .8$

P(คนส่วนใหญ่คัดค้าน) = P(คัดค้านเกิน 50% หรือ 8 คนขึ้นไป)

$P(X \geq 8) = 1 - P(X \leq 7) = 1 - .0042 = .9958$

นั่นคือเหตุการณ์ที่จะมีผู้คัดค้านเกินครึ่ง มีโอกาสเกิดขึ้นสูงมาก

P(ผู้คัดค้าน 10-14 คน) = $P(10 \leq X \leq 14)$

$= P(X \leq 14) - P(X \leq 9)$

$= .9648 - .0611 = .9037$ ซึ่งสูงมากเช่นกัน

**หมายเหตุ** ถ้า $n = 1$ เรียกว่า การแจกแจงแบบ Bernoulli ซึ่งจะมี $E(X) = p$ และ $V(X) = p(1-p)$

**4.2 การแจกแจงแบบปัวซอง** การแจกแจงนี้ตั้งตามชื่อนักคณิตศาสตร์ฝรั่งเศส ชื่อ Simeon Denis Poisson (1781 - 1840) ตัวแปรแบบปัวซอง จะเกิดจาก**กระบวนการแบบปัวซอง** หรือ Poisson Process ซึ่งหมายถึงการสังเกตเหตุการณ์แบบไม่ต่อเนื่อง (ซึ่งนับได้) ที่สัมพันธ์กับช่วงความต่อเนื่อง หรือ **interval** เช่น ช่วงเวลา ช่วงระยะทาง ขอบเขต (region, interval, space) เช่น การสังเกตเซลล์เม็ดเลือดขาวในหยดเลือด เหตุการณ์แบบไม่ต่อเนื่อง คือ การนับเซลล์เม็ดเลือดขาว ช่วงความต่อเนื่องคือ หยดเลือด

**นิยาม 4.1 กระบวนการแบบปัวซอง** (พารามิเตอร์ $\lambda > 0$) หมายถึง การกระทำที่สังเกตหรือนับได้ใน continuous interval of time, length, area หรือ space โดยมีเงื่อนไขดังนี้

1. ใน interval ขนาด h ซึ่งเล็กมาก จะมีเหตุการณ์อุบัติเพียง 1 ครั้ง ด้วยโอกาส h ซึ่งหมายถึง ความน่าจะเป็นของเหตุการณ์จะเป็นสัดส่วนผันแปรตามขนาดของ interval
2. เพื่อความสะดวก ให้ P(เกิดเหตุการณ์ 2 ครั้งขึ้นไป ใน interval ขนาด h ใดๆ ก็ตาม) = 0 คือไม่มีโอกาสเกิด
3. การเกิดเหตุการณ์ ใน interval ขนาด h ใดๆ ไม่มีผลต่อกัน (เป็นอิสระกัน)

ตัวแปร X คือ เหตุการณ์ใน interval ขนาด S หน่วย เรียกว่า ตัวแปรเชิงสุ่มแบบปัวซอง ด้วยพารามิเตอร์ $\lambda S$

**ทฤษฎี 4.3** ถ้า X เป็นตัวแปรเชิงสุ่มแบบปัวซองด้วย พารามิเตอร์ $\lambda S$ X จะมี density function ดังนี้

$$f(x) = (e^{-\lambda S}(\lambda S)^x)/x! \quad , x = 0, 1, 2, \ldots$$

$$e \approx 2.71828$$

**ทฤษฎี 4.4** ถ้า X เป็นตัวแปรเชิงสุ่มแบบปัวซองด้วย พารามิเตอร์ $\lambda S$ ดังนั้น

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

**ข้อสังเกต** จำนวนครั้งโดยเฉลี่ย ใน interval ขนาด S หน่วย $= \lambda S$

ดังนั้น จำนวนครั้งโดยเฉลี่ย ใน interval ขนาด I หน่วย $= \lambda S/S = \lambda$

ดังนั้น $\lambda$ จึงหมายถึง จำนวนเฉลี่ย ต่อ 1 หน่วย

**ตัวอย่างที่ 5** บุคคลปกติจะมีเซลล์เม็ดเลือดขาวโดยเฉลี่ย ไม่ต่ำกว่า 6000 หน่วยต่อ 1 มม$^3$ ดังนั้น ถ้าจะนับจากหยดเลือดขนาด .001 มม$^3$ คาดว่าจะต้องมีระดับปกติเท่าใด

$\lambda S = 6000(.001) = 6$ นั่นคือ คนปกติ มี $E(X) = 6$ ต่อ .001 มม$^3$

$(\lambda = 6000, S = .001)$

ถ้าพบเซลล์เม็ดเลือดแดง 2 หน่วย/.001 มม$^3$ จะเป็นเหตุการณ์ผิด ปกติหรือไม่

$$P(X \leq 2) = \sum_{x=0}^{2} e^{-6}6^x/x!$$

เปิดตารางปัวซอง ที่ $\lambda S = 6$ $P(X \leq 2) = .062 = $ p-value

.062 ไม่เล็กมาก แต่ก็ไม่สูงมาก จึงไม่สามารถสรุปอย่างชัดเจนว่าผิดปกติ หรือไม่

### 4.3 ความสัมพันธ์ระหว่างการแจกแจงแบบปัวซองและการแจกแจงแบบทวินาม

สำหรับการแจกแจงแบบทวินาม ที่ $n \geq 20$ และ $p \leq .05$ จะสามารถใช้การแจกแจงแบบปัวซองประมาณค่าได้อย่างมีประสิทธิภาพ โดยให้ $\lambda S = np$ ประสิทธิภาพจะสูงมาก ถ้า $n \geq 100$, $np \leq 10$ เนื่องจาก ทวินาม ที่ p เล็กมาก หมายถึงเหตุการณ์ที่ไม่เกิดได้บ่อยนัก (rare event) ซึ่งหมายถึงการแจกแจงแบบปัวซอง

**ตัวอย่างที่ 6** เซลล์แบคทีเรียชนิดหนึ่งที่ตรวจพบในระบบการย่อย มีการผ่าเหล่า (mutation) คือ พบว่ามีอัตรา 1 ต่อ $10^9$ เซลล์ ที่ต้านยา Streptomycin จึงทำให้บุคคลนั้น ต้านยาปฏิชีวนะ Streptomycin ถ้ามีการตรวจเซลล์จำนวน 2 หมื่นล้านเซลล์ ($2 \times 10^9$) จงหาโอกาสของ (1) ไม่พบเซลล์ผ่าเหล่า (2) พบอย่างน้อย 1 เซลล์

กรณีนี้ ทราบค่า n และ p $n = 2 \times 10^9$ , $p = 1/10^9$ จึงมีการแจกแจงแบบทวินาม แต่ p เล็กมาก เมื่อเทียบกับ n จึงใช้ปัวซองประมาณค่าด้วย $\lambda S = np = (2 \times 10^9)(1/10) = 2$ จากตารางปัวซอง $P(X = 0) = .135$, P(อย่างน้อย 1 เซลล์ที่ผ่าเหล่า) = $P(X > 1) = 1 - P(X \leq 0) = 1 - .135 = .865$

**4.4 การแจกแจงแบบปกติ** ตัวแปรแบบนี้เป็นตัวแปรแบบต่อเนื่อง De Moivre (1773) เป็นผู้ริเริ่มสร้างการแจกแจงนี้โดยพัฒนาจากการแจกแจงแบบทวินาม ที่ n มีค่าไม่จำกัดจำนวน แต่ไม่เป็นที่รู้จักแพร่หลาย จน 50 ปีต่อมา Laplace และ Gauss ได้ร่วมกันสร้าง density function เพื่อใช้แสดงความผิดพลาดในกระบวนการวัดค่าต่างๆ ทางดาราศาสตร์ (Astronomical measurement) ต่อมา มีการใช้การแจกแจงแบบปกติบ่อยมากโดยเฉพาะข้อมูลทางวิทยาศาสตร์

**นิยาม 4.2 ตัวแปรเชิงสุ่มแบบปกติ** ตัวแปรเชิงสุ่ม X จะมีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ $\mu$ และ $\sigma^2$ ถ้ามี density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(1/2)(x-\mu)^2/\sigma^2} \quad , -\infty \leq x \leq \infty$$

$$-\infty \leq \mu \leq \infty , \sigma > 0$$

$$e \approx 2.1828,$$

$$\pi = 3.1416$$

**และมีคุณสมบัติเบื้องต้น ดังนี้**

1. **กราฟของ f(x) เป็นรูประฆังคว่ำที่สมมาตร โดยมี $\mu$ เป็นค่ากึ่งกลาง หรือที่ตั้งของโค้ง**
2. **จะมีจุดหักเหของโค้งที่ $\mu \pm \sigma$ จุดหักเหจะแสดงลักษณะของโค้ง ถ้า $\sigma$ มีค่าสูง จุดหักเหจะ**

อยู่ไกล $\mu$ ทำให้ ได้โค้งค่อนข้างเตี้ยราบ (ไม่สูงโด่ง)

3. เนื่องจาก X เป็นตัวแปรแบบต่อเนื่อง ดังนั้น $f(x) \geq 0$ และพื้นที่ทั้งหมดภายใต้โค้ง $f(x) = 1$ (พื้นที่ คือความน่าจะเป็น)
4. จะมีพื้นที่ประมาณ 68% ที่อยู่ระหว่าง $\mu \pm \sigma$
   จะมีพื้นที่ประมาณ 95% ที่อยู่ระหว่าง $\mu \pm 2\sigma$
   จะมีพื้นที่ประมาณ 99.8% ที่อยู่ระหว่าง $\mu \pm 3\sigma$
5. การหาความน่าจะเป็นหรือพื้นที่ภายใต้โค้งปกติใดๆ จะต้องแปลงตัวแปรเชิงสุ่ม X ให้เป็น Z ซึ่งคือตัวแปรเชิงสุ่มแบบปกติ**มาตรฐาน** โดยที่ Z มี $\mu = 0, \sigma = 1$ แล้วหาพื้นที่จากตารางโค้งปกติมาตรฐาน หรือตาราง Z

**ทฤษฎี 4.5 Standardization Theorem** ถ้า X เป็นตัวแปรเชิงสุ่มแบบปกติ ด้วยค่าเฉลี่ย $\mu$ และความแปรปรวน $\sigma^2$ ดังนั้น ตัวแปรเชิงสุ่ม $(X - \mu)/\sigma$ คือตัวแปรเชิงสุ่มปกติมาตรฐาน (standard normal)

**<u>ตัวอย่างที่ 7</u>** X คือปริมาณไฮโดรคาร์บอน เป็นกรัม จากท่อไอเสียที่วิ่งระยะทาง 1 ไมล์ สมมติ X มีการแจกแจงแบบปกติด้วย $\mu = 1$ กรัม $\sigma = .25$ กรัม

ก) จงหาโอกาสที่รถยนต์คันหนึ่ง จะมีปริมาณไฮโดรคาร์บอน 1.0-1.5 กรัม

ข) จงหาโอกาสที่จะพบในปริมาณ $\mu \pm 2\sigma$

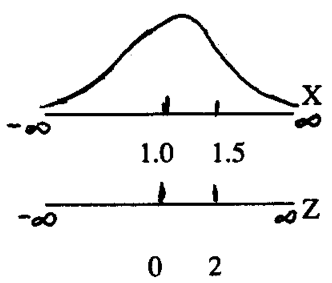
ค) อยากทราบว่าปริมาณไฮโดรคาร์บอน ระดับที่สูงสุด 5% คือระดับใด

**วิธีทำ** ต้องแปลงค่า X เป็น Z ดังนี้

(ก) $x_1 = 1, Z_1 = (1 - 1)/.25 = 0$

$x_2 = 1.5, Z_2 = (1.5 - 1)/.25 = 2$

$$P(1 \leq X \leq 1.5) = P(0 \leq Z \leq 2)$$
$$= P(Z \leq 2) - P(Z \leq 0)$$
$$= .9972 - .5000 = .4972$$




(ข) $\mu \pm 2\sigma = 2 \pm 2(.25) = .50, 1.50$

เมื่อ $x = .50, Z = (.50 - 1)/.25 = -2$

$x = 1.50,\ Z = (1.50 - 1)/.25 = 2$

$$P(.50 \leq X \leq 1.50) = P(-2 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq -2)$$

$$= .9972 - .0028 = .9544$$

(ค) พื้นที่ 5% ด้านขวามือ จะพบเมื่อ $Z = 1.645$ จะให้ $P(Z \geq 1.645) = .05$

จาก $Z = (X - \mu)/\sigma$ ดังนั้น $X = \sigma Z + \mu$

นั่นคือ $Z = 1.645,\ X = \sigma Z + \mu = (.25)(1.645) + 1 = 1.41$

นั่นคือ ระดับไฮโดรคาร์บอน สูงสุด 5% คือระดับ 1.41 กรัมขึ้นไป

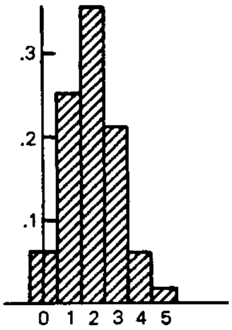
**4.5 การใช้การแจกแจงปกติประมาณค่าการแจกแจงแบบอื่นๆ**

เราทราบว่า การแจกแจงแบบปัวซองสามารถใช้ประมาณค่า การแจกแจงทวินามที่มี n สูง และ p เล็กมาก ในทำนองเดียวกัน เราสามารถใช้โค้งปกติ ประมาณการแจกแจงทวินามได้อย่างดี เมื่อ n สูง และ p ไม่เล็กเกินไป หรือไม่ใหญ่ เกินไป จากรูป 4.1 จะแสดงการแจกแจงทวินาม ที่มี p = .40 และ n = 5, 10, 15, 20 ซึ่งพบว่า เมื่อ n = 20 เส้นโค้ง และกราฟแท่งสอดคล้องกันดี

**รูปที่ 4.1**
**แสดงความสัมพันธ์ระหว่างการแจกแจงแบบทวินาม และโค้งปกติ**

$n = 5$, $p = .4$

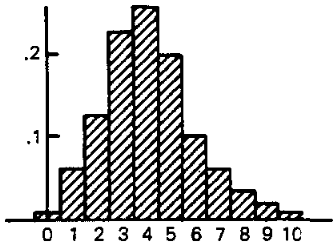
| $x$ | $f(x)$ |
|---|---|
| 0 | .0778 |
| 1 | .2592 |
| 2 | .3456 |
| 3 | .2304 |
| 4 | .0768 |
| 5 | .0102 |




*(a)*

$n = 10$, $p = .4$

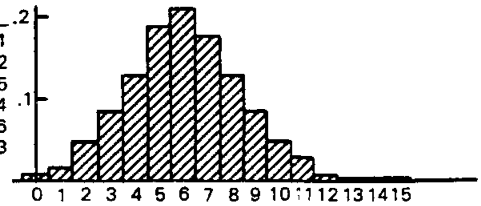
| $x$ | $f(x)$ |
|---|---|
| 0 | .0060 |
| 1 | .0404 |
| 2 | .1209 |
| 3 | .2150 |
| 4 | .2508 |
| 5 | .2007 |
| 6 | .1114 |
| 7 | .0425 |
| 8 | .0106 |
| 9 | .0016 |
| 10 | .0001 |




*(b)*

$n = 15$, $p = .4$

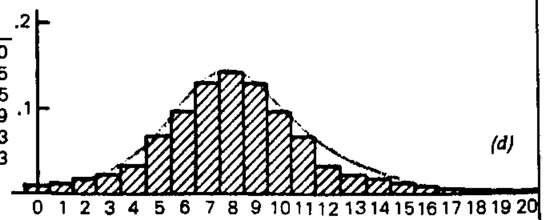
| $x$ | $f(x)$ | $x$ | $f(x)$ |
|---|---|---|---|
| 0 | .0005 | 8 | .1181 |
| 1 | .0047 | 9 | .0612 |
| 2 | .0219 | 10 | .0245 |
| 3 | .0634 | 11 | .0074 |
| 4 | .1268 | 12 | .0016 |
| 5 | .1859 | 13 | .0003 |
| 6 | .2066 | 14 | ~0 |
| 7 | .1771 | 15 | ~0 |




*(c)*

$n = 20$, $p = .4$

| $x$ | $f(x)$ | $x$ | $f(x)$ |
|---|---|---|---|
| 0 | ~0 | 11 | .0710 |
| 1 | .0005 | 12 | .0355 |
| 2 | .0031 | 13 | .0145 |
| 3 | .0124 | 14 | .0049 |
| 4 | .0350 | 15 | .0013 |
| 5 | .0746 | 16 | .0003 |
| 6 | .1244 | 17 | ~0 |
| 7 | .1659 | 18 | ~0 |
| 8 | .1797 | 19 | ~0 |
| 9 | .1597 | 20 | ~0 |
| 10 | .1172 | | |




*(d)*

Density for $X$ binomial: (*a*) $n = 5$, $p = .4$; (*b*) $n = 10$, $p = .4$; (*c*) $n = 15$, $p = .4$; (*d*) $n = 20$, $p = .4$.

**ทฤษฎีที่ 4.6** ถ้า X มีการแจกแจงแบบทวินาม ด้วยพารามิเตอร์ n และ p เมื่อ n มีค่าสูงมาก X จะมีการแจกแจงโดยประมาณแบบโค้งปกติด้วย $\mu$ = np และ $\sigma^2$ = np(1-p) โดยมี np > 5 และ n(1-p) > 5

**<u>ตัวอย่างที่ 7</u>** การศึกษาความสัมพันธ์ระหว่างมารดาที่สูบบุหรี่ขณะตั้งครรภ์ และจำนวนทารกที่คลอดผิดปกติ พบว่า 40% ของมารดาสูบบุหรี่ระหว่างตั้งครรภ์ ถ้าพบทารกที่คลอดผิดปกติ 20 คน จงหาโอกาสที่จะมาจากมารดาที่สูบบุหรี่ขณะตั้งครรภ์ 12 คนขึ้นไป

**วิธีที่ 1** ใช้การแจกแจงทวินาม

$n = 20,\ p = .4,\ P(X \geq 12) = 1 - P(X \leq 11) = 1 - .9435$

$= .0565$

**วิธีที่ 2** ใช้การแจกแจงแบบปกติ

$np = 20(.4) = 8,\ n(1-p) = 20(.6) = 12$

ทั้ง np และ n(1-p) มากกว่า 5 จึงใช้โค้งปกติประมาณได้

โดยมี $\mu = np = 8,\ \sigma^2 = np(1-p) = 20(.4)(.6) = 4.8$

$P(X \geq 12)\quad P(Y \geq 11.5)$ (0.5 คือการปรับความต่อเนื่อง)

เมื่อ $y = 11.5,\ Z = (11.5 - 8)/ 4.8 = 1.59$

$P(Y \geq 11.5) = P(Z \geq 1.59) = 1 - P(Z \leq 1.59)$

$= 1 - .9441 = .0559$

(ซึ่งใกล้เคียง .0565)

**ทฤษฎีที่ 4.7** ถ้า X มีการแจกแจงแบบปัวซองด้วยพารามิเตอร์ $\lambda s$ สำหรับ $\lambda s$ ที่มีค่าสูง X จะมีการแจกแจงโดยประมาณแบบโค้งปกติ ด้วยค่าเฉลี่ย $\lambda s$ และความแปรปรวน $\lambda s$

**<u>ตัวอย่างที่ 8</u>** ในชายที่มีสุขภาพสมบูรณ์จะมีเม็ดเลือดแดง เฉลี่ย 5,400,000 เซลล์ ต่อเลือด 1 มม$^3$ จงหาโอกาสที่จะพบเซลล์เม็ดเลือดแดง 500-580 เซลล์ จากหยดเลือดขนาด 1/10,000 มม$^3$

จะเห็นว่า X มีการแจกแจงแบบปัวซอง ด้วย $\lambda s = 5{,}400{,}000(1/10{,}000) = 540$

$P(500 \leq X \leq 580) = P(499.5 \leq Y \leq 580.5)$ (ปรับค่าความต่อเนื่อง)

เมื่อ $y = 499.5,\ Z = (499.5 - 540)/ 540 = -1.74$

$y = 580.5$, $Z = (580.5 - 540)/\sqrt{540} = 1.70$

$P(-1.74 \leq Z \leq 1.74) = .9591 - .0409 = .9182$

## แบบฝึกหัดที่ 4

1. ตัวแปรต่อไปนี้ถ้ามีการแจกแจงแบบทวินาม ให้ระบุค่า พารามิเตอร์

1.1 ต้องการรับบริจาคเลือด AB negative ทราบว่าประชากรในเมืองนั้น มีผู้ที่มีเลือด AB negative เพียง 0.6% X คือจำนวนผู้มีเลือดกลุ่มนี้ จากผู้บริจาค 5 ราย

1.2 ผู้วิจัยอ้างว่ากระบวนการวิเคราะห์ปฏิกิริยาเคมีแบบใหม่ ให้ผลดีกว่าแบบเดิม 90% ของจำนวนครั้งทั้งหมด ให้ X คือจำนวนครั้งที่วิธีใหม่ให้ผลดีกว่าวิธีเดิม จากการทดลอง 10 ครั้ง

1.3 นักวิจัยเตรียมต้นไม้ไว้ 8 ต้น โดยไม่ทราบว่ามี 3 ต้นที่เป็นโรค จึงไม่เหมาะสมกับการนำไปทดลอง ถ้าเขาหยิบแบบสุ่มมา 4 ต้น ให้ X คือจำนวนต้นที่เป็นโรค

1.4 ในการศึกษาการย้ายถิ่นฐานของนกชนิดหนึ่ง โดยสามารถจับมาติดแถบโลหะ แล้วปล่อยไปได้ประมาณ 5% ของประชากรทั้งหมด ถ้าในการจับครั้งต่อไป จำนวน 8 ตัว X คือจำนวน นกที่ติดแถบโลหะ

1.5 สามีภริยาคู่หนึ่งต้องการมีลูกสาว 1 คน จึงตกลงกันว่าจะมีลูกไปเรื่อยๆ จนได้ลูกสาวคนแรกจึงจะหยุด ให้ X คือจำนวนลูกชายที่เกิดก่อนได้ลูกสาวคนแรก

2. จงหา density function ค่าเฉลี่ย และความแปรปรวน ของ X สำหรับข้อ 1.1

3. ให้ R และ Y เป็นโครโมโซมเพศ โดยทุกคนจะมีโครโมโซม R เฉพาะเพศชาย จึงจะมีโครโมโซม Y ดังนั้น RR คือเพศหญิง RY คือเพศชาย โรคตาบอดสี เกิดจาก recessive allelle บนโครโมโซม R แต่จะไม่มี recessive allelle ของโรคตาบอดสี บนโครโมโซม Y ดังนั้น เพศหญิงจะมี genotype 3 อย่าง คือ RR (ปกติ), Rr(carrier) และ rr (ตาบอดสี) ส่วนเพศชายจะมี genotype เพียง 2 อย่างคือ RY (ปกติ) และ rY (ตาบอดสี) บุตรจะได้รับโครโมโซมแบบเชิงสุ่มจากบิดา และมารดาคนละ 1 อัน

ถ้ามารดาเป็น carrier ของตาบอดสี และบิดาไม่เป็นโรคตาบอดสี

3.1 ถ้ามีบุตร 1 คน จงหาโอกาสที่จะได้บุตรชายที่ตาปกติ

3.2 ถ้ามีบุตร 1 คน จงหาโอกาสที่จะได้บุตรชายที่ตาบอดสี

3.3 ถ้ามีบุตร 3 คน จงหาโอกาสที่จะได้บุตรชายที่ตาบอดสี 2 คน

3.4 ถ้ามีบุตร 5 คน จงหาโอกาสที่จะได้บุตรชายที่ตาบอดสี ไม่เกิน 2 คน และหาจำนวนคาดหมายของบุตรชายที่ตาบอดสี

3.5 ถ้ามีบุตร 5 คน จงหาโอกาสที่จะมีบุตรชายตาบอดสี 3 คนขึ้นไป

4. โรงงานนิวเคลียร์แห่งหนึ่งมีรังสีรั่วไหลโดยเฉลี่ย 2 ครั้ง/เดือน

   4.1 จงหาโอกาสที่จะไม่มีรังสีรั่วไหล ใน 3 เดือน

   4.2 จำนวนครั้งที่รังสีรั่วไหลโดยเฉลี่ย ใน 3 เดือน

   4.3 ถ้ามีรังสีรั่วไหล 12 ครั้ง ภายใน 3 เดือน จะสนับสนุนว่ามีจำนวนเฉลี่ย 2 ครั้ง/เดือน หรือไม่

5. ในเมืองหนึ่งมีจำนวนประชากรที่ตายด้วยโรคมะเร็งในปอด 12 คน/ปี ถ้าอัตราการตายมีการแจกแจงแบบปัวซอง จงหา

   5.1 โอกาสที่จะมีผู้ตาย 10 คน/ปี

   5.2 โอกาสที่จะมีผู้ตาย 15 คนขึ้นไป/ปี

   5.3 โอกาสที่จะมีผู้ตาย ไม่เกิน 10 คน/ปี

6. ในการเพาะเลี้ยงจุลินทรีย์ชนิดหนึ่ง พบว่ามีเซลล์ซึ่งเป็นสาเหตุของโรคไข้รากสาดใหญ่ โดยเฉลี่ย 5 หน่วย/20 ตารางไมโครเมตร (1/10,000 ซ.ม.) ถ้าใช้พื้นที่ขนาด 16 ตารางไมโครเมตร ท่านคาดว่าจะพบเซลล์ชนิดนี้จำนวนเท่าใด โอกาสที่จะไม่พบเลยเป็นเท่าใด และโอกาสที่จะพบอย่างน้อย 9 เซลล์ เป็นเท่าใด

7. ในเมืองหนึ่งพบว่ามีสารปรอทตกค้างในไก่ฟ้า สันนิษฐานว่า สารปรอทปนเปื้อนอยู่กับเมล็ดพืชที่ชาวไร่ฉีด methyl mercury ให้ X คือระดับสารปรอทที่พบในไก่/ล้านส่วน และค่า X มีการแจกแจงแบบปกติ ด้วย $\mu = .25$ และ $\sigma = .08$ จงหา $P(X \leq .3)$, $P(X \geq .17)$, $P(.2 \leq X \leq .4)$ และ $P(.01 \leq X \leq .49)$

8. ให้ X คือระดับน้ำตาลในเลือดของผู้เป็นโรคเบาหวาน และมีการแจกแจงแบบปกติ ด้วย $\mu = 106$ mg/100 ml และมี $\sigma = 8$ mg/100 ml จงหา

   8.1 $P(X \leq 120$ mg/100 ml)

   8.2 เปอร์เซนต์ผู้มีน้ำตาลระดับ 90-120 mg/100 ml

   8.3 $P(106 \leq X \leq 110)$

   8.4 $P(X \geq 121$ mg/100 ml)

   8.5 จุด $X_0$ ที่แสดงว่ามีอยู่ 25% ที่มีระดับน้ำตาลในเลือดต่ำกว่า $x_0$

9. ถ้า X มีการแจกแจงแบบทวินามด้วย $n = 20, p = .3$ จงใช้โค้งปกติประมาณความน่าจะเป็นต่อไปนี้

   1) $P(X \leq 3)$  2) $P(3 \leq X \leq 6)$  3) $P(X \geq 4)$  4) $P(X = 4)$

10. ถ้า 10% ของประชากรเป็นโรคภูมิแพ้ ถ้าสุ่มมา 100 คน จงหาโอกาสที่จะมีผู้เป็นโรคภูมิแพ้ อย่างน้อย 12 ราย? อย่างมาก 8 ราย?

11. อัตราผู้ตายเนื่องจากกินยาคุมกำเนิดชนิดหนึ่ง เป็น 3 ต่อ 100,000 ถ้ามีผู้ใช้ยานั้น 1 ล้านคน จงหาจำนวนคาดหมายผู้ตายเนื่องจากยาดังกล่าว, โอกาสที่จะมีผู้ตายอย่างมาก 25 ราย? และระหว่าง 25-35 ราย

12. ถ้า X มีการแจกแจงแบบปัวซองด้วย พารามิเตอร์ $\lambda S = 100$ จงหา

1) $P(X \geq 95)$ 2) $P(X \leq 80)$ 3) $P(90 \leq X \leq 110)$

4) $P(X = 99)$

# เฉลยแบบฝึกหัดที่ 4

1. 1.1 ทวินาม $n = 5, p = .006$

   1.2 ทวินาม $n = 10, p = .90$

   1.3 ไม่ใช่ทวินาม เพราะเป็นการหยิบแบบไม่ใส่คืน

   1.4 ทวินาม $n = 8, p = .05$

   1.5 ไม่ใช่ทวินาม เพราะไม่ทราบค่า n

2. $f(x) = ({}^5C_x)(.006)^x(.994)^{5-x}$ , $x = 0, 1, ..., 5$

   $E(X) = np = 5(.006) = .03$

   $V(X) = np(1-p) = 5(.006)(.994) = .02982$

3.

| แม่ | พ่อ | | prob. |
|---|---|---|---|
| r | R | หญิง (พาหะตาบอดสี) | 1/4 |
| r | Y | ชาย (ตาบอดสี) | 1/4 |
| R | R | หญิง (ปกติ) | 1/4 |
| R | Y | ชาย (ปกติ) | 1/4 |

   $P(r) = P(R) = 1/2, P(Y) = 1/2$

   3.1 P(บุตรชายตาปกติ) = P(RY) = 1/4

   3.2 P(บุตรชายตาบอดสี) = P(rY) = 1/4

   3.3 $n = 3$, P(บุตรชายตาบอดสี 2 คน) = ${}^3C_2(1/4)^2(3/4) = .14$

   3.4 $n = 5$, P(บุตรชายตาบอดสี $\leq 2$) = $P(X \leq 2) = .8965$, $E(X) = 5(1/4) = 1.25$ คน

   3.5 $n = 5$, $P(X \geq 3) = 1 - P(X \leq 2) = 1 - .8965 = .1035$

4. $\lambda = 2, S = 3, \lambda S = 6$

   4.1 $P(X = 0) = .002$

   4.2 $E(X) = \lambda S = 6$ ครั้ง/3 เดือน

   4.3 $P(X \geq 12) = 1 - P(X \leq 11) = 1 - .98 = .02$ = p-value

   ถ้า $\lambda = 2$ จริง โอกาสการรั่วไหล 12 ครั้ง/3 เดือน จะเกิดเพียง .02 ซึ่งนับว่าน้อยมาก จึงน่าสงสัยว่า $\lambda \neq 2$ น่าจะสูงกว่า 2

5. $\lambda = 12$/ปี

   5.1 $P(X = 10) = F(10) - F(9) = .347 - .242 = .105$

   5.2 $P(X \geq 15) = 1 - P(X \leq 14) = 1 - .772 = .228$

5.3 $P(X \leq 10) = F(10) = .347$

6. $\lambda = 5$ หน่วยต่อ 20 ตารางไมโครเมตร $= 5/20 = .25/1$ ไมโครเมตร

ถ้า 16 ตารางไมโครเมตร $S = 16$, $\lambda S = .25(16) = 4$, $E(X) = \lambda S = 4$, $P(X = 0) = .018$,

$P(X \geq 9) = 1 - P(X \leq 8) = 1 - .979 = .021$

7. $X \sim N(\mu = .25, \sigma = .08)$

เมื่อ $X = .3$, $Z = (.3 - .25)/.08 = .625$, $P(Z \leq .625) = .7340$

เมื่อ $X = .17$, $Z = (.17 - .25)/.08 = -1.0$,

$P(Z \geq -1.0) = 1 - P(Z \leq -1.0) = 1 - .1587 = .8413$

$P(.2 \leq X \leq .4) = P(-.625 \leq Z \leq 1.875) = .9696 - .2660 = .7036$

$P(.01 \leq X \leq .49) = P(-3 \leq Z \leq 3) = .9987 - .0013 = .9974$

8. $X \sim N(\mu = 106$ mg/100 ml, $\sigma = 8$ mg/100 ml$)$

8.1 $P(X \leq 120) = P(Z \leq 1.75) = .9599$

8.2 $P(90 \leq X \leq 120) = P(-2 \leq Z \leq 1.75) = .9599 - .0228 = .9371$

8.3 $P(106 \leq X \leq 110) = P(0 \leq< Z \leq .5) = .6915 - .5000 = .1915$

8.4 $P(X \geq 121) = P(Z \geq 1.875) = 1 - P(Z \leq 1.875) = 1 - .9696 = .0304$

8.5 $P(Z \leq x_0) = .25$, $x_0 = -.675$

$Z = (x_0 - \mu)/\sigma$ , $x_0 = \sigma Z + \mu = 8(-.675) + 106 = 100.6$

9. $n = 20$, $p = .3$ X เป็นทวินาม, $\mu = 20(.3) = 6$, $\sigma = \sqrt{20(.3)(.7)} = 2.049$

1) $P(X \leq 3) = P(Y \leq 2.5) = P(Z \leq -1.22) \quad .1112$

(ถ้าใช้ทวินาม $P(X \leq 3) = .1071$)

2) $P(3 \leq X \leq 6) = P(2.5 \leq Y \leq 6.5)$

$= P(-1.708 \leq Z \leq 0.24) = .5948 - .0436 = .5512$

(ถ้าใช้ทวินาม $= .5725$)

3) $P(X \geq 4) = 1 - P(y \leq 3.5) = 1 - P(Z \leq -1.22) = 1 - .1112 = .8888$

(ทวินาม $= .8929$)

4) $P(X = 4) = P(3.5 \leq Y \leq 4.5) = P(-1.22 \leq Z \leq -.73)$

$= .2327 - .1112 = .1215$

(ทวินาม = .1304)

10. $n = 100$, $p = .10$, $\mu = 100(.1) = 10$, $\sigma = \sqrt{100(.1)(.9)} = 3$

$P(X \geq 12) = P(Y \geq 11.5) = P(Z \geq 0.5) = 1 - P(Z \leq 0.5)$

$= 1 - .6915 = .3085$

$P(X \leq 8) = P(Y \leq 8.5) = P(Z \leq - 0.5) = .3085$

11. $p = 3/100{,}000$, $n = 1{,}000{,}000$, $\mu = np = 30$, $\sigma = 5.477$

1) $E(X) = 30$

2) $P(X \leq 25) = P(Y \leq 25.5) = P(Z \leq -.82) = .2061$

3) $P(25 \leq X \leq 35) = P(24.5 \leq Y \leq 35.5) = P(-1.0 \leq Z \leq 1.0)$

$= .8413 - .1587 = .6826$

12. $\mu = \lambda s = 100$, $\sigma^2 = \lambda s = 100$, $\sigma = 10$

1) $P(X \geq 95) = P(Y \geq 94.5) = P(Z \geq -.55)$

$= 1 - P(Z \leq -.55) = 1 - .2912 = .7088$

2) $P(X \leq 80) = P(Y \leq 80.5)$

$= P(Z \leq -1.95)$

$= .0256$

3) $P(90 \leq X \leq 110) = P(89.5 \leq Y \leq 110.5)$

$= P(-1.05 \leq Z \leq 1.05)$

$= .8531 - .1469 = .7062$

4) $P(X = 99) = P(98.5 \leq Y \leq 99.5)$

$= P(-.15 \leq Z \leq -.05)$

$= .4801 - .4404 = .0397$