บทที่ 4
การแจกแจงของตัวแปรสุ่ม การเลือกตัวอย่าง และการแจกแจงของตัวสถิติ
(Special Parametric Families of Univeriate Distribution, Random Sampling and Sampling Distribution)

4.1 บทนำ

ในบทที่ 2 และบทที่ 3 ได้กล่าวถึงการแจกแจงของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ รวมถึงการแจกแจงของตัวสถิติที่เนื่องมาจากกลุ่มประชากรดังกล่าวและเทคนิคสำหรับหา pdf ของตัวสถิติเหล่านั้นไว้แล้ว แต่อย่างไรก็ตาม ความรู้เท่าที่กล่าวถึงเป็นสิ่งที่เกี่ยวข้องอยู่กับการแจกแจงปกติเท่านั้น ยังไม่เพียงพอที่จะนำไปใช้ในงานทดสอบข้อสมมุติฐานซึ่งเป็นวัตถุประสงค์ของหนังสือเล่มนี้ ทั้งนี้เพราะในสถานการณ์จริงเราไม่อาจแน่ใจได้ว่าสถานการณ์ที่กำลังศึกษาอยู่นั้นจะเป็นลักษณะของตัวแปรสุ่มแบบปกติหรืออนุโลมให้เป็นตัวแปรสุ่มแบบปกติได้เสมอไป สถานการณ์ ๆ หนึ่ง ๆ ย่อมมีลักษณะของการแจกแจงจำเพาะของตนเสมอ มีบ้างที่บางครั้งสามารถนำความรู้เรื่องการแจกแจงแบบปกติมาใช้หรืออนุโลมใช้ได้ แต่บางครั้งก็ไม่อาจนำมาใช้ได้เพราะถ้านำมาใช้จะได้ผลลัพธ์ค่อนข้างไม่น่าเชื่อถือ หรือยืนอยู่บนความเสี่ยงของข้อตกลง (Basic Assumption) มากเกินไป กล่าวคือผลลัพธ์เช่นผลงานวิจัยหรืองานทดลองเป็นจริงตามข้อตกลงเท่านั้น หากสถานการณ์ที่เป็นอยู่จริงมิได้เป็นไปตามข้อตกลง ผลงานนั้นก็มิอาจเชื่อถือได้ โดยเฉพาะอย่างยิ่งในงานเกี่ยวกับการทดสอบข้อสมมุติฐานซึ่งเป็นเรื่องที่ละเอียดอ่อนเพราะ

---

[1] คำว่า Parametric Families ของ Density Function หมายถึงกลุ่มของ Density Function ที่ชี้บ่งหรือระบุ (index) ไว้ด้วยปริมาณอันหนึ่ง ที่เรียกว่า พารามิเตอร์

เช่น $f_X(x) = \lambda e^{-\lambda x}$; $x \geq 0, \lambda > 0$

หรือ $f_X(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2, \ldots$

จะเห็นได้ว่าในที่นี้พารามิเตอร์ λ เป็นตัวบ่งบอกการแจกแจงของประชากร

ต้องเสี่ยงต่อความผิดพลาดสำคัญถึง 2 ประการคือ α-error และ β-error ถ้าเพิ่มความเสี่ยงเรื่องข้อตกลงหรือข้อสมมุติฐานเกี่ยวกับการแจกแจงของประชากรเข้าไปอีก จะเป็นเหตุให้เพิ่มความเสี่ยงในงานโดยไม่จำเป็น

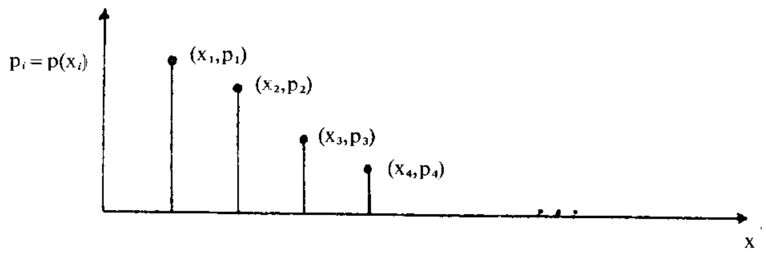
ด้วยเหตุผลดังกล่าว จึงเป็นความจำเป็นต้องกล่าวถึงตัวแปรสุ่มที่สำคัญ ๆ รวมถึงสถานการณ์ที่ชี้หรือบ่งให้เห็นลักษณะจำเพาะตัวของตัวแปรสุ่มไว้ในที่นี้ด้วย แม้ว่าเป็นเรื่องที่ "ฟุ่มเฟือย" เกินวัตถุประสงค์ของหนังสือเล่มนี้ก็ตาม

การเสนอเรื่องการแจกแจงของตัวแปรสุ่มจะแบ่งการศึกษาออกเป็น 2 ส่วนคือส่วนที่ว่าด้วยตัวแปรสุ่มแบบตัดตอนและตัวแปรสุ่มแบบต่อเนื่อง โดยจะกล่าวถึงค่าคาดหมาย ความแปรปรวน รวมถึงการแจกแจงของตัวสถิติหรือตัวแปรสุ่มเหล่านั้นควบกันไป การกล่าวถึงการแจกแจงของตัวแปรสุ่มควบไปพร้อมกันนี้ เป็นเรื่องที่ผู้เขียนจำเป็นต้องทำเพราะปรารถนาจะรวบรัด ไม่ปรารถนาจะกล่าวแยกไว้ต่างหาก เพราะจะทำให้หนังสือหนาเกินพอดี อีกประการหนึ่งเห็นว่านักศึกษามีความรู้พื้นฐานในเรื่องนี้มาแล้วในบทที่ 1-3 คงไม่ยากเกินไปที่จะจำแนกหรือทำความเข้าใจว่า อะไรคือการแจกแจงของตัวแปรสุ่ม อะไรคือการแจกแจงของตัวสถิติ อนึ่งเรื่องนี้จำเป็นต้องนำไปใช้ในการทดสอบข้อสมมุติฐานจำเป็นต้องกล่าวถึงไว้อีกครั้งเพื่อเป็นการประกันความรู้ความเข้าใจของนักศึกษาทั้งที่เรื่องนี้น่าจะเป็นเรื่องของ "ความน่าจะเป็น" มิใช่ "ทฤษฎีสถิติ" ตามเจตนารมณ์ดังได้กล่าวแล้วในตอนต้น

4.2 การแจกแจงของตัวแปรสุ่มแบบตัดตอน (Dicrete Univariate Distributions)

ตัวแปรสุ่มแบบตัดตอน คือตัวแปรสุ่มที่มีค่าที่เป็นไปได้ทั้งหมดเป็นจำนวนที่นับได้ (Countable) อาจเป็น finite หรือ infinite (ที่นับได้) ก็ได้ เช่นตัวแปรสุ่มคือ X ค่าของ X อาจเป็น $x_1, x_2, \ldots, x_n$ หรือ $x_1, x_2, \ldots$ ก็ได้

การแจกแจง (Probability Distribution) ของตัวแปรสุ่มประเภทนี้สามารถหาได้โดยอาศัยกลุ่ม (Collection) ของคู่ลำดับ $(x_i, P(x_i))$; $i = 1, 2, \ldots, N$ หรือ $i = 1, 2, \ldots$ ดังภาพ



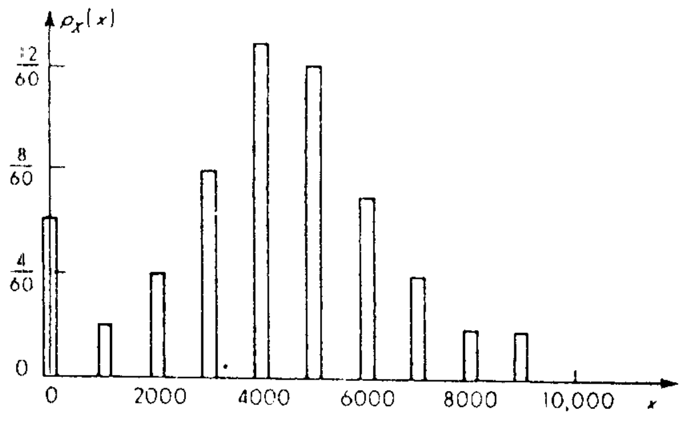
$f_X(x) = P(X = x)$, ในทุกค่าของ x เรียกว่าฟังก์ชั่นความน่าจะเป็นของตัวแปรสุ่ม X

โดยทั่วไปจะนิยมเรียกฟังก์ชั่นความน่าจะเป็นของตัวแปรสุ่มแบบตัดตอนว่า Probability Mass Function (pmf) และนิยมเรียกฟังก์ชั่นความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่องว่า Probability Density Function (pdf) แต่บ่อยครั้งที่เราเรียกว่าเป็น pdf อย่างเดียวกันเพื่อปัดความยุ่งยาก

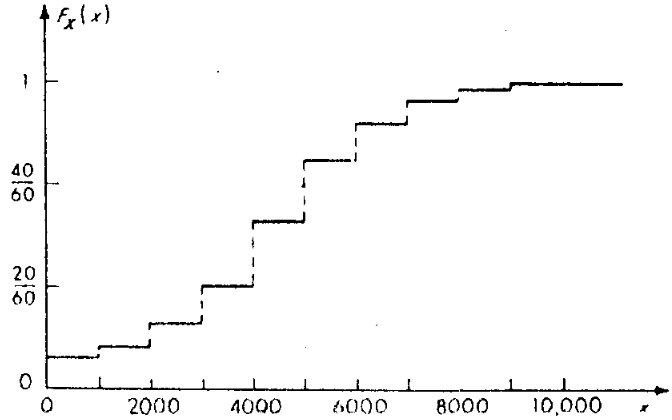
pmf ต้องมีคุณสมบัติดังต่อไปนี้

1. $0 \leqslant f_X(x) \leqslant 1$
2. $\sum\limits_{\text{all}x} f_X(x) = 1$
3. $\Pr(a \leqslant X \leqslant b) = \sum\limits_{x \geqslant a}^{x < b} f_X(x)$

และโดยทั่วไป ภาพของ pmf และ cdf จะปรากฏดังภาพ



(a)



(b)

ภาพแสดง (a) pmf และ (b) cdf ของตัวแปรสุ่มแบบตัดตอน

4.2.1 การแจกแจงแบบยูนิฟอร์ม (Discrete Uniform Distribution)

นิยาม 4.1

ตัวแปรสุ่ม X[1] มีฟังก์ชั่นความน่าจะเป็น ดังนี้คือ

$$f_X(x) = f_X(x; N) = \frac{1}{N} \quad \text{เมื่อ } x = 1, 2, \ldots, N$$

$$= 0 \quad \text{เมื่อ } x \text{ มีค่าเป็นอย่างอื่น}$$

โดยที่พารามิเตอร์ N มีค่าเป็นจำนวนเต็มบวก แล้วตัวแปรสุ่ม X จะมีการแจกแจงแบบยูนิฟอร์ม

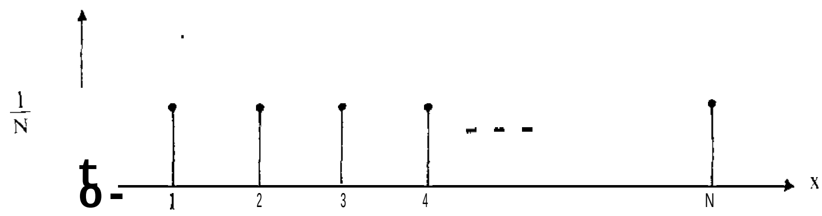
การแจกแจงประเภทนี้นิยมใช้มากในทฤษฎีของการสุ่มตัวอย่าง โดยเฉพาะกรณี Sampling with Replacement (หรือกรณี Sampling without Replacement ก็ได้ถ้ากรอบตัวอย่าง (Sampling Frame) มีขนาดใหญ่) โดยถือว่าหน่วยสำรวจ (Sampling Unit) ทุกหน่วยมีโอกาสได้รับเลือกเป็นตัวอย่างได้ทัดเทียมกัน เช่นในชั้นเรียนหนึ่งมีนักเรียน 45 คน ถ้าต้องการเลือกนักเรียนมาเป็นกลุ่มตัวอย่างเพื่อตอบแบบสอบถามเรื่องใดเรื่องหนึ่งโดยสุ่ม (สมมุติเรื่อง "ความคาดหวังในชีวิต") นักเรียนทุกคนจะมีโอกาสได้รับเลือกเท่ากันคือเท่ากับ $\frac{1}{45}$ ดังนี้เป็นต้น กรณี

---

[1] คำว่าตัวแปรสุ่ม มาจากคำว่า Random Variable หรือ Variate บางครั้ง (ในตำราหลายเล่ม) เรียกว่า Chance Variable หรือ Stocastic Variable ภาษาเยอรมันเรียกว่า Zofallige ภาษาฝรั่งเศส เรียกว่า Variable ale'atire

ที่เห็นได้ง่าย อีกตัวอย่างหนึ่งคือเรื่องของการทอดลูกเต๋า ลูกเต๋าแต่ละหน้าจะมีโอกาสหงายขึ้นเท่ากับ $\frac{1}{6}$ เมื่อหน้าลูกเต๋าคือ $x = 1, 2, 3, \ldots, 6$

ภาพของการแจกแจงแบบ Discrete Uniform ปรากฏดังนี้



ทฤษฎี 4.1 ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบยูนิฟอร์ม (Discrete Uniform) แล้ว จะพบว่า

$$M_X(t) = \sum_{i=1}^{N} e^{it} \cdot \frac{1}{N}, \quad E(X) = \frac{N+1}{2} \quad \text{และ} \quad V(X) = \frac{(N^2 - 1)}{12}$$

การพิสูจน์ทฤษฎีสามารถจะทำได้โดยง่าย จะขอเว้นไว้เป็นแบบฝึกหัด แต่ขอให้คำแนะนำสำหรับการหา V(X) ไว้เล็กน้อยคือ เราต้องใช้อนุกรม $\sum_{u=1}^{N} u^2 = \frac{N(N+1)(2N+1)}{6}$

4.2.2 การแจกแจงแบบเบอร์นุลลิ

(Bernoulli Distribution **หรือ** Point Binomial)

การแจกแจงแบบเบอร์นุลลิ เป็นวิธีการที่ใช้ได้กับสถานการณ์การทดลองหรือการตรวจสอบประเมินค่าหรือคุณภาพที่ง่ายที่สุดที่ต้องการผลลัพธ์ออกมาเพียง 2 ประการคือ "ใช้ได้" กับ "ใช้ไม่ได้" อย่างใดอย่างหนึ่งอาจเป็น "Success" กับ "Failure" หรือ "ใช่" กับ "ไม่ใช่" หรืออื่น ๆ ในทำนองนี้ เช่น ในการสำรวจทัศนะของผู้มีสิทธิเลือกตั้งเกี่ยวกับวิธีการเลือกตั้ง ผลลัพธ์ที่ต้องการอาจเป็นชอบวิธี "เลือกแบ่งเขต" กับ "ชอบวิธีเลือกตั้งแบบรวมเขต" อย่างใดอย่างหนึ่ง การตรวจสอบคุณภาพของสินค้าส่งมอบผลลัพธ์ที่ต้องการอาจเป็น "ตรงตามมาตรฐานที่กำหนด" กับ "ไม่ตรงตามมาตรฐานที่กำหนด" อย่างใดอย่างหนึ่ง การตรวจความหนาแน่นของการจราจรบริเวณสี่แยกผลลัพธ์ที่ต้องการอาจเป็น "รถเลี้ยวซ้าย" กับ "ไม่เลี้ยวซ้าย" อย่างใดอย่างหนึ่ง ดังนี้เป็นต้น

การแจกแจงแบบเบอร์นุลลิ เป็นหลักการขั้นพื้นฐานของการแจกแจงแบบทวินาม การแจกแจงแบบอนุกรมเรขาคณิต และการแจกแจงแบบนิเสธทวินาม ซึ่งจะมีประโยชน์อย่างยิ่งสำหรับการประมาณค่าพารามิเตอร์และการทดสอบข้อสมมุติฐานเกี่ยวกับสัดส่วน (p) และจำนวนครั้งของการทดลองและอื่น ๆ ซึ่งจะได้กล่าวถึงโดยละเอียดต่อไป

โดยปกติเราจำเป็นต้องกำหนดผลลัพธ์ "ใช้ได้" กับ "ใช้ไม่ได้" ออกมาเป็นตัวเลขเพื่อประโยชน์ในการคำนวณ อาจให้เป็นเลข 1 สำหรับผลลัพธ์ประเภท "ใช้ได้" และ 0 สำหรับผลลัพธ์ประเภท "ยังใช้ไม่ได้" หรือกลับกัน หรือแม้แต่จะให้เป็นตัวเลขอย่างอื่น ๆ ก็ได้ แต่นิยมใช้ 0 กับ 1 มากกว่าเพราะทำให้ผลของความน่าจะเป็น p กับ q ดังที่เคยพบเห็นกันทั่วไป สำหรับกรณีที่จะใช้เป็นตัวเลขอื่น ผลของความน่าจะเป็น อาจออกมาในรูปอื่น ซึ่งเมื่อตีความหมายแล้วก็ย่อมให้ผลสรุปตรงกัน แต่อาจเพิ่มภาระให้แก่ผู้อื่นที่ต้องศึกษาเพราะไม่ใช่รูปหรือลักษณะที่เป็นลักษณะสากล

นิยาม 4.2 ตัวแปรสุ่ม X จะมีการแจกแจงแบบเบอร์นูลลิ ถ้าตัวแปรสุ่ม X ให้ผลลัพธ์ออกมาเพียง 2 ประการ และมีฟังก์ชั่นความน่าจะเป็นดังนี้คือ

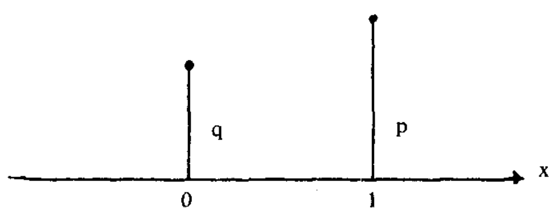
$$f_X(x) = f_X(x;p) = p^x q^{1-x} \quad \text{เมื่อ } x = 0 \text{ หรือ } 1$$
$$= 0 \quad \text{เมื่อ } x \text{ มีค่าเป็นอย่างอื่น}$$

โดยที่ p เป็นพารามิเตอร์ที่สอดคล้องกับเงื่อนไข $0 \leqslant p \leqslant 1$ และ $q = 1 - p$

ข้อสังเกต

1. p คือ Probabity of Success q คือ Probability of Failure

2. จาก $f_X(x)$ ตามนิยาม ถ้าให้ $x = 0$ จะได้ $f_X(x) = q$ ถ้าให้ $x = 1$ จะได้ $f_X(x) = p$ กราฟการแจกแจงแบบ Bernoulli ปรากฏดังภาพต่อไปนี้



ทฤษฎี 4.2 ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบเบอร์นูลลิแล้ว จะพบว่า

$$M_X(t) = q + pe^t, \quad E(X) = p, \quad V(X) = pq$$

พิสูจน์ จาก $f_X(x) = p^x q^{1-x}$; $x = 0, 1$

$$\Rightarrow \quad M_X(t) = E(e^{tx}) = \sum_{x=0}^{1} e^{tx} p^x q^{1-x}$$

$$= q + pe^t$$

และ $$E(X) = M_X'(t)\Big|_{t=0} = pe^t\Big|_{t=0} = p$$

หรือ $$E(X) = \sum_{x=0}^{1} xp^x q^{1-x} = 0 + p = p$$

และ $$V(X) = M_X''(t)\Big|_{t=0} - p^2 = p - p^2 = p(1-p) = pq$$

หรือ $$V(X) = \sum_{x=0}^{1} (x - E(X))^2 p^x q^{1-x} = \sum_{x=0}^{1} (x-p)^2 p^x q^{1-x}$$

$$= p^2q + q^2p = pq(p+q) = pq(1) = pq$$

ตัวอย่าง 4.1 เด็กนักเรียน 20 คนได้รับคำสั่งให้ช่วยกันเดินหาปากกาที่ครูทำหล่นภายในบริเวณสนามหญ้าหน้าโรงเรียน โดยที่นักเรียนแต่ละคนจะมีโอกาสหาพบได้เพียง 10% เท่านั้น เพราะสนามรกมาก จงคาดหมายว่าจะมีนักเรียนเป็นผู้พบเห็นปากกาดังกล่าว (ถือว่านักเรียนมีความสามารถในการค้นหาไม่ต่างกัน)

วิธีทำ ให้ X_i; $i = 1, 2, \ldots, 20$ คือผลลัพธ์จากการค้นหาของนักเรียนคนที่ i

โดยที่ $X_i = 0$ เมื่อหาไม่พบ

$= 1$ เมื่อหาพบ

$p_i = .10$; $i = 1, 2, \ldots, 20$

จะเห็นได้ว่า $X_i \sim$ Bernoulli

สิ่งที่ต้องการคือ $E(\sum_{i}^{20} X_i) = ?$

พิจารณาตัวแปรสุ่ม $Y = \sum_{i}^{20} X_i$ จะพบว่า

$$M_Y(t) = \prod_{i}^{20} M_{X_i}(t) = (q + pe^t)^{20} = (.90 + .10e^t)^{20}$$

$$\Rightarrow \quad E(Y) = M_Y'(t)\Big|_{t=0} = 20(.90 + .10e^t)(.10e^t)\Big|_{t=0}$$

$$= 20(.10)$$

$$= 2$$

แสดงว่าในบรรดานักเรียนทั้ง 20 คนนั้น คาดว่าจะมีอยู่ 2 คนที่สามารถหาปากกาพบ (อาจจะเป็นในลักษณะที่พบหรือเห็นพร้อมกัน)

ขอให้สังเกต mgf ของตัวแปรสุ่ม Y เปรียบเทียบกับ mgf ของตัวแปรสุ่มทวินามที่จะกล่าวถึงในลำดับต่อไป

4.2.2 การแจกแจงแบบทวินาม

(Binomial Distribution)

การแจกแจงแบบทวินามเป็นการแจกแจงที่ถูกนำไปใช้อย่างกว้างขวาง โดยเฉพาะอย่างยิ่งในกรณีที่ต้องการทดสอบข้อสมมุติฐานหรือการประมาณค่าสัดส่วน (Proportion)

การแจกแจงแบบทวินามพัฒนาขึ้นมาจากการแจกแจงแบบเบอร์นูลลิ กล่าวคือเมื่อทำการทดลองเบอร์นูลลิ (Bernoulli Trial) ติดต่อกันไปเรื่อย ๆ ถึง n ครั้ง โดยที่การทดลองเบอร์นูลลิแต่ละครั้งเป็นอิสระต่อกัน[1] (หมายความว่าผลการทดลองแต่ละครั้งไม่กระทบกระเทือนถึงกัน เช่นการทอดเหรียญครั้งก่อนออกหัว ครั้งต่อไปก็ไม่จำเป็นจะต้องออกหัวตาม) และความน่าจะเป็น p จะต้องคงที่ตลอดไปสำหรับการทดลองทุกครั้ง ผลรวมของการทดลองทั้ง n ครั้ง จะมีการแจกแจงแบบทวินาม[2] ผลรวมที่กล่าวถึงนี้อาจหมายถึง Success หรือ Failure ก็ได้ ทั้งนี้สุดแต่ความมุ่งหมายของผู้ทดลองหรือผู้สังเกต[3]

เหตุที่เรียกการแจกแจงแบบนี้ว่า**ทวินาม** (Binomial) เพราะ "ผลการทดลอง" แต่ละครั้งมีได้เพียง 2 ประการ (Bi = ทวิ, nomial = เทอม, พวก) ทำให้การทดลองทั้ง n ครั้งมีลักษณะต้องกันกับการกระจายทวินาม (Binomial Expansion) ในวิชาคณิตศาสตร์เบื้องต้นที่ทุกคนทราบกันดีอยู่แล้ว ดังนี้

---

[1] สำหรับกรณีการสุ่มตัวอย่าง การแจกแจงแบบทวินามใช้กับกรณีสุ่มตัวอย่างแบบเลือกแล้วใส่กลับคืน (Draw with Replacement) การเลือกแล้วใส่คืนทำให้หน่วยตัวอย่างเป็นอิสระต่อกัน โอกาสที่ได้รับเลือกจะคงที่มิได้แปรปรวนไปเพราะหน่วยอื่นถูกเลือกไปก่อนแล้ว

[2] เมื่อตัวแปรสุ่ม X มีการแจกแจงแบบเบอร์นูลลิ ตัวแปรสุ่ม $Y = X_1 + X_2 + \dots + X_n$ จะมีการแจกแจงแบบทวินาม (ดูตัวอย่าง 4.1 และ 4.2)

[3] การแจกแจงแบบทวินามมิได้หมายถึงเฉพาะสถานการณ์ทดลองเท่านั้น แต่รวมไปถึงการสังเกตปรากฏการณ์ที่เกิดขึ้นโดยสุ่ม ตัวอย่างเช่น การเลี้ยวของรถตรงทางแยก การเกิดน้ำท่วมในแต่ละปีและอื่น ๆ

$$(p+q)^n = \binom{n}{n}p^n + \binom{n}{n-1}p^{n-1}q + \binom{n}{n-2}p^{n-2}q^2 + \dots$$

$$+ \binom{n}{1}pq^{n-1} + \binom{n}{0}q^n$$

$$= \sum_{x=0}^{n}\binom{n}{x}p^{n-x}q^x$$

ตัวอย่างเช่น

$$(p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

$$= \binom{4}{4}p^4 + \binom{4}{3}p^3q + \binom{4}{2}p^2q^2 + \binom{4}{1}p^1q^3 + \binom{4}{0}q^4$$

เมื่อนำการกระจายทวินามมาใช้ในลักษณะของความน่าจะเป็น โดยที่ p = Probability of Success[1] และ $q = 1 - p$ = Probability of Failure จึงพบว่า

$$(p+q)^n = 1 = \sum_{x=0}^{n}\binom{n}{x}p^xq^{n-x}$$

ซึ่งหมายความว่าแต่ละเทอมของ Sum ก็คือความน่าจะเป็นที่จะได้ Success x ครั้ง ในการทดลองอิสระ n ครั้งนั่นเอง

ตัวอย่าง เช่น ตัวต้านทานชนิด 20 Ω ในกล่องใบหนึ่งมีตัวที่เสื่อมคุณภาพปนอยู่ 10% สุ่มตัวอย่างตัวต้านทานดังกล่าวมา 5 ตัวโดยเลือกแล้วใส่คืน[2] (Draw with Replacement) ผลของความน่าจะเป็นจะปรากฏดังนี้ ($p = .10$, $q = .90$)

$$(p+q)^5 = (.10 + .90)^5 = (.1)^5 + 5(.1)^4(.9) + 10(.1)^3(.9)^2 + 10(.1)^2(.9)^3$$

$$+ 5(.1)(.9)^4 + (.9)^5$$

---

[1] p อาจหมายถึง Probability of Success หรือ Probability of Failure ก็ได้ การจะให้ p หมายถึงอะไรย่อมขึ้นอยู่กับสถานการณ์เป็นเรื่อง ๆ ไป นอกจากนี้ยังขึ้นอยู่กับความต้องการของผู้ทดลองแต่ละคนอีกด้วย

[2] สุ่มมาทีละตัว ตรวจสอบเสร็จแล้วใส่คืนลงในกล่องตามเดิม เคล้าให้ทั่วแล้วสุ่มตัวต่อไปมาตรวจสอบ

$(.1)^5$ หมายถึงความน่าจะเป็นที่ตัวต้านทานที่สุ่มมาทั้ง 5 ตัวนั้นเสื่อมคุณภาพหมดทุกตัว

$5(.1)^4(.9)$ หมายถึงความน่าจะเป็นที่ตัวต้านทานที่สุ่มมาทั้ง 5 ตัวนั้นมี 4 ตัวเสื่อมคุณภาพอีก 1 ตัวยังมีสภาพดี

$10(.1)^3(.9)^2$ หมายถึงความน่าจะเป็นที่ตัวต้านทานที่สุ่มมาทั้ง 5 ตัวนั้นมี 3 ตัวเสื่อมคุณภาพอีก 2 ตัวยังมีสภาพดี

.
.
.

จึงเห็นได้ว่าแต่ละเทอมของ Sum หมายถึงความน่าจะเป็นที่ต้องการสำหรับแต่ละเรื่อง เช่น $10(.1)^2(.9)^3$ หมายถึงความน่าจะเป็นที่กลุ่มตัวอย่างของตัวต้านทาน 5 ตัวนั้นจะมีตัวต้านทานที่เสื่อมคุณภาพปนอยู่ 2 ตัว ดังนี้เป็นต้น

สำหรับสัมประสิทธิ์ $\binom{n}{n}$ เช่น 1, 5, 10, 10, 5, 1 หรืออื่น ๆ นั้นหมายถึงจำนวนหนทาง (วิธี) ที่พึงเป็นไปได้ที่จะเกิดเหตุการณ์นั้น ๆ เช่น $10(.1)^3(.9)^2$ หมายความว่าหนทางที่กลุ่มตัวอย่างตัวต้านทานที่สุ่มมา 5 ตัวและจะมีตัวที่เสื่อมคุณภาพปนมาถึง 3 ตัวนั้นจะมีหนทางที่พึงเป็นไปได้ถึง $\binom{5}{3} = 10$ หนทางคือ {DDDNN, DDNDN, DDNND, DNDDN, DNNDD, DNDND, NNDDD, NDDDN, NDNDD, NDDND} อย่างใดอย่างหนึ่ง[1] คือเมื่อสุ่มตัวอย่างตัวต้านทานมา 5 ตัว โดยหวังว่าจะพบตัวที่เสื่อมสภาพปนมา 3 ตัว อาจพบเหตุการณ์ (ตัวอย่าง) DDDND หรือ DDNDN หรือ ... หรือ NDDND แต่ละเหตุการณ์มีค่าความน่าจะเป็นเท่ากันเท่ากับ $p^3q^2 = (.1)^3(.9)^2$ ดังนั้น สำหรับเหตุการณ์ที่หวังว่าจะพบตัวต้านทานที่เสื่อมคุณภาพปนมา 3 ตัว จึงมีค่าความน่าจะเป็นเท่ากับ

$$\underbrace{(.1)^3(.9)^2 + \dots + (.1)^3(.9)^2}_{10 \text{ ครั้ง}} = 10(.1)^3(.9)^2$$

นิยาม 4.3 ตัวแปรสุ่ม X จะมีการแจกแจงแบบทวินามที่มีพารามิเตอร์ n และ p ถ้าหากว่า X มีฟังก์ชั่นความน่าจะเป็นดังนี้คือ

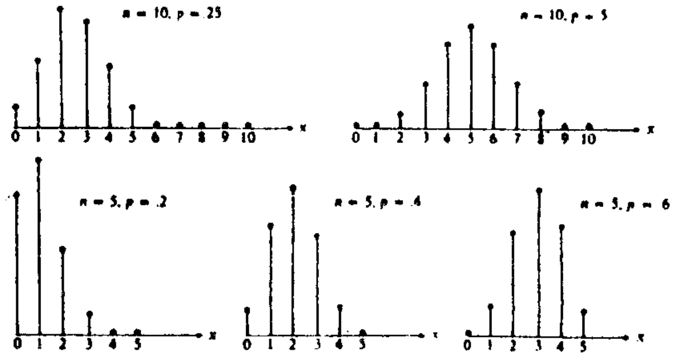
$$f_X(x) = b(x; n, p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & ; \quad x = 0, 1, 2, \dots, n \\ 0 & \text{เมื่อ } x \text{ มีค่าเป็นอย่างอื่น} \end{cases}$$

---

[1] D = Defective = เสื่อมคุณภาพ N = Nondefective = ยังมีสภาพดี P(D) = p = .1, P(N) = q = .9 DDNDN อ่านว่าตัวต้านทานที่สุ่มมา 2 ตัวแรกเสื่อมสภาพตัวที่สามสภาพดี ตัวที่สี่เสื่อมสภาพ ตัวสุดท้ายสภาพดี ชุดอื่น ๆ ก็อ่านได้ในทำนองเดียวกัน

โดยที่ตัวแปรสุ่ม X หมายถึงจำนวนครั้งของ Success (หรือ Failure) และ $0 \leq p \leq 1$, $q = 1 - p$

ภาพการกระจายของตัวแปรสุ่ม X สำหรับค่า p บางค่าปรากฏดังนี้



Binomial densities.

จากภาพขอให้สังเกตว่า เมื่อ p มีค่าเท่ากับ 0.5 กราฟการแจกแจงของตัวแปรสุ่ม X จะมีลักษณะสมมาตรกัน ถ้าลองลากเส้นผ่านปลายแท่ง (Bar Chart) แต่ละแท่งจะได้โค้งปกติ ส่วนกรณีเมื่อ $p > 0.5$ กราฟจะมีลักษณะเบ้ซ้าย (Left Hand Skewness) หมายความว่าความถี่ส่วนใหญ่จะตกอยู่ด้านซ้ายของค่าคาดหมาย $(E(X) = np)$ ถ้า $p < 0.5$ กราฟจะมีลักษณะเบ้ขวา (Right Hand Skewness) หมายความว่าความถี่ส่วนใหญ่จะตกอยู่ด้านขวาของค่าคาดหมาย

ข้อสังเกตข้างต้นเป็นสิ่งที่นักศึกษาจำเป็นต้องทราบ เพราะมีอยู่หลายกรณีที่เราไม่อาจหาค่าความน่าจะเป็นของตัวแปรสุ่มทวินามได้โดยตรงโดยเฉพาะอย่างยิ่งเมื่อ $n \rightarrow \infty$ กรณีเช่นนี้ เราจะหาค่าประมาณได้โดยอาศัยทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง (CLT) กล่าวคือประมาณค่าความน่าจะเป็นของตัวแปรสุ่มทวินาม (Binomial Variate) ด้วยตัวแปรสุ่มปกติ (Normal Variate) ซึ่งการประมาณค่าดังกล่าวจะได้ผลถูกต้องค่อนข้างสมบูรณ์ก็เฉพาะเมื่อตัวแปรสุ่มทวินามมีลักษณะการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติเท่านั้น นั่นคือการประมาณค่าจะได้ผลดีเมื่อ p มีค่าเท่ากับ 0.5 หรือใกล้เคียง 0.5 ส่วนในกรณี $p < 0.5$ หรือ $p > 0.5$ ผลการประมาณค่าจะคลาดเคลื่อนมากให้หันไปใช้การแจกแจงแบบพัวซอง (Poisson Distribution) มาประมาณค่าแทน

ทฤษฎี 4.3 ถ้าตัวแปรเชิงสุ่ม X มีการแจกแจงแบบทวินามแล้ว X จะมี mgf ค่าคาดหมายและความแปรปรวนดังนี้

$$M_X(t) = (q + pe^t)^n \,;\; E(X) = np,\; V(X) = npq$$

พิสูจน์

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} \\
&= (q + pe^t)^n \quad : \text{อาศัยการกระจายทวินาม (Binomial Expansion)}
\end{aligned}$$

$$\Rightarrow \quad E(X) = M_X'(t)\Big|_{t=0} = \frac{d}{dt}(q + pe^t)^n\Big|_{t=0}$$

$$= n(q + pe^t)^{n-1} pe^t\Big|_{t=0} = np$$

และ

$$\begin{aligned}
V(X) &= M_X''(t)\Big|_{t=0} - (np)^2 \\
&= \{np^2e^t(n-1)(q + pe^t)^{n-2}e^t + np(q + pe^t)^{n-1}e^t\}\Big|_{t=0} - n^2p^2 \\
&= np^2(n-1) + np - n^2p^2 = n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 = np(1-p) \\
&= npq
\end{aligned}$$

ข้อสังเกต

1. จากทฤษฎี ถ้าให้ $n = 1$ จะได้

$$\begin{aligned}
f_X(x) &= \binom{1}{x} p^x q^{1-x} = p^x q^{1-x} \,;\; x = 0, 1 \\
M_X(t) &= (q + pe^t)^1 = q + pe^t \\
E(X) &= 1.p
\end{aligned}$$

และ $V(X) = 1.p.q = pq$

หมายความว่าการแจกแจงแบบทวินามจะแปลงรูปเป็นการแจกแจงแบบเบอร์นุลลิเมื่อ $n = 1$ และโดยนัยกลับกัน การแจกแจงแบบเบอร์นุลลิจะกลายเป็นการแจกแจงแบบทวินามเมื่อ $n = 1$

2. $E(X) = np$ ค่าคาดหมายของจำนวนครั้งของ "ความสมหวัง" (Success) ที่คาดว่าจะเกิดขึ้นเมื่อทำการทดลอง n ครั้ง แต่ละครั้งมีโอกาสสมหวังคงที่เท่ากับ p

ตัวอย่าง 4.2 เมื่อตัวแปรสุ่ม $X_1, X_2, \ldots, X_{10}$ เป็นอิสระต่อกันและต่างก็มีการแจกแจงแบบเบอร์นุลลิ มีพารามิเตอร์ p เดียวกัน จงหาการแจกแจงของตัวแปรสุ่ม $Y = X_1 + X_2 + \ldots + X_{10}$

วิธีทำ เมื่อ $X_i \sim$ Bernoulli (p) ; $i = 1, 2, \ldots, 10$ และเป็นอิสระต่อกัน

$$=> \quad f_{X_i}(x) = p^x q^{1-x} \; ; \quad x = 0, 1 : i = 1, 2, \ldots, 10$$

$$M_{X_i}(t) = q + pe^t$$

ดังนั้น

$$M_Y(t) = \sum_{i=1}^{10} M_{X_i}(t) = (q + pe^t)^{10}$$

แสดงว่าตัวแปรสุ่ม Y มีการแจกแจงเบบทวินาม มีพารามิเตอร์เท่ากับ 10 และ p ตามลำดับ หรือนัยหนึ่ง

$$=> \quad Y \sim b(y; 10, p)$$

ตัวอย่างนี้ชี้ให้เห็นอย่างเด่นชัดว่าการแจกแจงแบบทวินามพัฒนามาจากการทดลองเบอร์นุลลิ ขณะเดียวกัน เมื่อมีการทดลองเบอร์นุลลิเพียงครั้งเดียว $(n = 1)$ การแจกแจงแบบทวินามจะลดรูปเป็นการแจกแจงแบบเบอร์นุลลิ ดังนั้นบางครั้งเราจึงเรียกการแจกแจงแบบเบอร์นุลลิว่า Point Binaomial

ทฤษฎี 4.4 ถ้าตัวแปรสุ่ม $X_1, X_2, \ldots, X_r$ เป็นอิสระต่อกัน โดยที่ $X_1 \sim b(n_1, p), X_2 \sim b(n_2, p), \ldots, X_r \sim b(n_r, p)$ แล้ว ตัวแปรสุ่ม $Y = X_1 + X_2 + \ldots + X_r$ จะมีการแจกแจงแบบทวินาม มีพารามิเตอร์เป็น $(n_1 + n_2 + \ldots + n_r)$ และ p และมี mgf ค่าคาดหมายและความแปรปรวนดังนี้

$$M_Y(t) = (q + pe^t)^{n_1+n_2+\ldots+n_r} \; ; \quad E(Y) = (n_1 + n_2 + \ldots + n_r)p$$

$$V(Y) = (n_1 + n_2 + \ldots + n_r)pq$$

ทฤษฎีนี้แสดงให้เห็นถึงคุณสมบัติที่สำคัญประการหนึ่งของการแจกแจงแบบทวินาม คือความสามารถรวมตัวกันได้ของตัวแปรสุ่มทวินาม เรียกว่า Convolution Property หรือ Reproductive Property เรานำไปใช้มากในด้านการทดสอบข้อสมมุติฐาน การประมาณค่าพารามิเตอร์ และการหาค่าความน่าจะเป็น ทฤษฎีนี้สามารถพิสูจน์ได้โดยง่ายจึงขอเว้นไว้เป็นแบบฝึกหัด

ตัวอย่าง 4.3 ถ้าตัวแปรเชิงสุ่ม X มีการแจกแจงแบบทวินาม มีพารามิเตอร์เป็น n และ p

ก. จงหาการแจกแจงของตัวแปรสุ่ม $Y = n - X$

ข. จงหามัธยฐาน (median) ของ X ถ้ากำหนดให้ $n = 5$ และ $p = .5$

ค. ถ้า mgf ของ X เป็น $(\frac{1}{3} + \frac{2}{3}e^t)$ จงหา P(X = 2 หรือ 3)

วิธีทำ

ก. เนื่องจาก $X \sim b(x; n, p)$

$$=> \quad M_X(t) = (q + pe^t)^n$$

เมื่อกำหนดให้ $Y = n - X$

$$=> \quad M_Y(t) = M_{n-X}(t) = e^{nt}(q + pe^{-t})^n = (p + qe^t)^n$$

นั่นคือ $Y = n - X$ จะมีการแจกแจงแบบทวินาม มีพารามิเตอร์เป็น n และ q หรือ

$$=> \quad Y \sim b(y; n, q)$$

ข. กำหนดให้ $n = 5$ และ $p = .5$

$$=> \quad f_X(x) = \binom{5}{x}(.5)^x(.5)^{5-x} \ ; \ x = 1, 2, \ldots, 5$$

จากนิยามของมัธยฐาน "med (X)" จะสอดคล้องกับสมการ

$$\int_{-\infty}^{\text{med}(X)} f_X(x)dx = \frac{1}{2} = \int_{\text{med}(X)}^{\infty} f_X(x)dx$$: เมื่อ X เป็นตัวแปรสุ่มแบบต่อเนื่อง

และ

$$\sum^{\text{med}(X)} f_X(x) = \frac{1}{2} = \sum_{\text{med}(X)} f_X(x)$$: เมื่อ X เป็นตัวแปรสุ่มแบบดัดตอน

ดังนั้น med(X) กรณีที่ X เป็นตัวแปรสุ่มทวินามตามตัวอย่างจึงปรากฏดังนี้

$$\sum_{x=0}^{\text{med}(X)} \binom{5}{x}(.5)^x(.5)^{5-x} = \frac{1}{2}$$

$$=> \sum_{x=0}^{med(X)} \binom{5}{x}(.5)^x(.5)^{5-x} = \sum_{x=0}^{2} \binom{5}{x}(.5)^x(.5)^{5-x}$$

$$=> \quad med(X) = 2$$

ค. กำหนดให้ $M_X(t) = (\frac{1}{3} + \frac{2}{3}e^t)^5$

$$=> n = 5, p = \frac{2}{3}$$

ดังนั้น $P(X = 2$ หรือ $3)$

$$= P(X = 2) + P(X = 3)$$

$$= \binom{5}{2}(\frac{2}{3})^2(\frac{1}{3})^3 + \binom{5}{3}(\frac{2}{3})^3(\frac{1}{3})^2$$

$$= .494$$

ตัวอย่าง 4.4 ให้ X เป็นจำนวนของผู้ตอบว่า "เห็นด้วย" ในการสัมภาษณ์ชาย 50 คน ในการสำรวจความคิดเห็นเรื่องการทำหมันชาย ซึ่งการตอบว่า "เห็นด้วย" และ "ไม่เห็นด้วย" ว่าจะมีโอกาสเกิดขึ้นพอ ๆ กัน จงคำนวณหา

ก. Pr (มีผู้ตอบ "เห็นด้วย" 25 คน)
ข. Pr (มีผู้ตอบ "เห็นด้วย" ไม่เกิน 30 คน)
ค. Pr (มีผู้ตอบ "เห็นด้วย" ระหว่าง 21 ถึง 30 คน)
ง. Pr (มีผู้ตอบ "เห็นด้วย" มากกว่า 20 คนแต่ไม่ถึง 30 คน)
จ. Pr (มีผู้ตอบ "เห็นด้วย" มากกว่า 26 คน)
ฉ. Pr (มีผู้ตอบ "เห็นด้วย" อย่างมาก 30 คน)
ช. Pr (มีผู้ตอบ "เห็นด้วย" มากกว่าไม่เห็นด้วย)

วิธีทำ

เนื่องจากโอกาสที่จะตอบ "เห็นด้วย" กับ "ไม่เห็นด้วย" มีค่าพอ ๆ กัน ดังนั้น $p = q = .50$ และ $n = 50$ ดังนั้น $E(X) = np = (50)(.50) = 25$, $V(X) = npq = 12.5$, $\sigma_X = 3.536$ ดังนั้น

ก. Pr (มีผู้ตอบ "เห็นด้วย" 25 คน) $= Pr(X = 25) = \binom{50}{25}(.5)^{25}(.5)^{25}$

$$\cong Pr(Z \leqslant \frac{25 - 25}{3.536}) = Pr(Z \leqslant 0) = 0.5$$

ข. Pr (มีผู้ตอบ "เห็นด้วย" ไม่เกิน 30 คน) = Pr(X $\leqslant$ 30) = $\sum_{x=0}^{30} \binom{50}{x}(.5)^x(.5)^{50-x}$

$\simeq$ $Pr(Z \leqslant \frac{30-25}{3.536})$ = Pr(Z $\leqslant$ 1.414) "0.92073

ค. Pr (มีผู้ตอบ "เห็นด้วย" ระหว่าง 21 ถึง 30 คน) = Pr(21 6 X $\leqslant$ 30)

$= \sum_{x=21}^{30} \binom{50}{x}(.5)^x(.5)^{50-x} \simeq Pr(\frac{21-25}{3\ 536} \leqslant z \leqslant \frac{30-25}{3\ 536})$

$\simeq$ Pr(-1.131 $\leqslant$ Z $\leqslant$ 1.414) $\simeq$ 0.92073 0.1292 $\simeq$ 0.79153

ง. Pr (มีผู้ตอบ "เห็นด้วย" เกิน 20 คนแต่ไม่ถึง 30 คน) = Pr(20 < X < 30)

$= \sum_{x=21}^{29} \binom{50}{x}(.5)^x(.5)^{50-x} \simeq Pr(\frac{20-25}{3\ 536} < Z < \frac{30-25}{3\ 536})$

$\simeq$ Pr(− 1.414 < Z < 1.414) $\simeq$ 0.92073 − 0.07927

$\simeq$ 0.84146

9. Pr (มีผู้ตอบ "เห็นด้วย" มากกว่า 26 คน) = Pr(X > 26)

$= \sum_{x=27}^{50} \binom{50}{x}(.5)^x(.5)^{50-x}$

$\simeq Pr(Z \geqslant \frac{26-25}{3.536}) \simeq Pr(Z \geqslant 0.283)$

$\simeq$ 1 − 0.6103 $\simeq$ 0.3897

a. Pr (มีผู้ตอบ "เห็นด้วย" อย่างมาก 30 คน) = Pr(X $\leqslant$ 30)

$= \sum_{x=0}^{30} \binom{50}{x}(.5)^x(.5)^{50-x}$

$\simeq Pr(Z \leqslant \frac{30-25}{3.536}) \simeq Pr(Z \leqslant 1.414)$

$\simeq$ **0.92073**

ช. Pr (มีผู้ตอบ "เห็นด้วย" มากกว่าตอบว่า "ไม่เห็นด้วย") = Pr(X > 25)

$= \sum_{x=26}^{50} \binom{50}{x}(.5)^x(.5)^{50-x}$

$\simeq Pr(Z \geqslant \frac{2s-2s}{3.536}) \simeq Pr(Z \geqslant 0) \simeq 0.50$

ตัวอย่าง 4.5 ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบทวินาม b(10.0.6) จงคำนวณหา

ก. $E(X - 6)$ ข. $E(\frac{X-6}{10})$

ค. $E(\frac{X}{10})$ ง. $E((X - 6)^2)$

จ. $E(X^2)$

วิธีทำ เนื่องจากตัวแปรสุ่ม X มีพารามิเตอร์ $n = 10, p = 0.6 \Rightarrow E(X) = 10(0.6) = 6$

ก. $E(X - 6) = E(X) - 6 = np - 6 = 6 - 6 = 0$

ข. $E(\frac{X-6}{10}) = E(\frac{X}{10}) - \frac{6}{10} = \frac{1}{10}E(X) - \frac{6}{10} = \frac{6}{10} - \frac{6}{10} = 0$

ค. $E(\frac{X}{10}) = \frac{1}{10}E(X) = \frac{6}{10}$

ง. $E((X - 6)^2) = E(X - E(X))^2 = V(X) = (10)(.6)(.4) = 2.4$

จ. $E(X^2) = npq + n^2p^2 = (10)(.6)(.4) + (10)^2(.6)^2 = 2.4 + 36 = 38.4$

ตัวอย่าง 4.6 ผู้จำหน่ายเมล็ดพันธุ์ถั่วทำการตรวจสอบเมล็ดพันธุ์พบว่า 5% ของเมล็ดพันธุ์เหล่านั้นฝ่อไม่อาจเพาะให้งอกได้ เขาจำหน่ายเมล็ดพันธุ์เป็นถุง ๆ ละ 200 เมล็ด และประกันไว้หน้าซองว่า "รับประกันงอก 90%" สมมุติว่าท่านซื้อเมล็ดพันธุ์มาถุงหนึ่ง จงหาความน่าจะเป็นที่ถั่วถุงนั้นไม่เป็นไปตามคำรับรอง

วิธีทำ $n = 200, p = .95, q = .05$

ให้ X แทนจำนวนถั่วที่เพาะแล้วงอก

ดังนั้นความน่าจะเป็นที่ต้องการคือ $Pr(X < 180)$

$$
\begin{aligned}
Pr(X < 180) &= \sum_{x=0}^{179} \binom{200}{x}(.95)^x(.05)^{200-x} \\
&\simeq Pr(Z \leqslant \frac{180 - (200)(.95)}{\{(200)(.95)(.05)\}^{1/2}}) \\
&\simeq Pr(Z \leqslant \frac{-10}{3.082}) \\
&\simeq Pr(Z \leqslant -3.245) \\
&\simeq 0.0020
\end{aligned}
$$

4.2.3 การแจกแจงแบบอนุกรมเรขาคณิต

(Geometric Distribution)

การแจกแจงแบบอนุกรมเรขาคณิต เป็นเรื่องของการแจกแจงที่นำมาใช้คำนวณหาจำนวนครั้งของการทดลองทั้งหมด (จำนวนการทดลองเบอร์นุลลิ) ที่จะต้องดำเนินการซ้ำ ๆ ไปจนกว่าจะประสบเหตุการณ์ที่มุ่งหวัง (Success) เป็นครั้งแรกโดยที่ความน่าจะเป็นของการประสบ-เหตุการณ์เช่นนั้นจะต้องคงที่เสมอ เช่นจำนวนครั้งทั้งหมดของการหางานจนกระทั่งได้งานทำ จำนวนครั้งทั้งหมดของการยิงเป้าบินจนกว่าจะถูก จำนวนครั้งทั้งหมดของการเปลี่ยนเหยื่อจนกระทั่งตกปลาได้เป็นตัวแรกสำหรับนักตกปลาสมัครเล่น จำนวนครั้งทั้งหมดของการทอดลูกเต๋าจนกว่าจะหงายหน้าเอี่ยว (หรือหน้าใด ๆ ที่สนใจ) และอื่น ๆ

จะเห็นได้ว่าตัวแปรสุ่มที่มีการแจกแจงแบบอนุกรมเรขาคณิตเป็นตัวแปรสุ่มที่แสดงถึงการรอคอย (Discrete Waiting Time) ในรูปของจำนวนครั้งของความล้มเหลว (Failure) ก่อนที่จะประสบความสำเร็จหรือได้สิ่งที่ต้องการ (Success)

เมื่อกล่าวถึงการรอคอย เราจะพบว่าเรามีการรอคอยได้ 2 ลักษณะ คือ รอคอยจนกว่าจะได้สิ่งที่ต้องการเป็นครั้งแรก กับรอคอยจนกว่าจะได้สิ่งที่ต้องการซ้ำมากกว่า 1 ครั้ง เช่นจำนวนสลากกินแบ่งที่จะต้องซื้อ (ซื้องวดละ 1 ฉบับ) จนกว่าจะถูกรางวัลที่ 5 และจำนวนสลาก-กินแบ่งที่จะต้องซื้อจนกว่าจะถูกรางวัลที่ 5 อีกเป็นครั้งที่ 3 เป็นต้น ตัวแปรสุ่มที่แสดงการรอคอยจนกระทั่งประสบสิ่งที่ต้องการเป็นครั้งแรก เรียกว่าตัวแปรสุ่มอนุกรมเรขาคณิต (Geometric Variate) ส่วนตัวแปรสุ่มที่แสดงการรอคอยจนกว่าจะประสบสิ่งที่ต้องการเกินกว่า 1 ครั้ง เรียกว่าตัวแปรสุ่มนิเสธทวินาม (Negative Binomial Variate) ตัวแปรสุ่มทั้งสองสัมพันธ์กันในลักษณะรูปทั่วไปและรูปเฉพาะของกันและกันกล่าวคือตัวแปรสุ่มอนุกรมเรขาคณิตเป็นรูปเฉพาะของตัวแปรสุ่มนิเสธทวินาม และตัวแปรสุ่มนิเสธทวินามเป็นรูปทั่วไปของตัวแปรสุ่มอนุกรมเรขาคณิต ขอให้สังเกตว่า การรอคอยแต่ละครั้ง (การทดลอง การสังเกตหรือการเฝ้าติดตาม) มีการแจกแจงแบบเบอร์นุลลิ เพราะมีผลเป็น 2 ประการเท่านั้นคือ "สมหวัง" กับ "ไม่สมหวัง" อย่างใดอย่างหนึ่ง แต่ทั้งนี้ต้องมีข้อจำกัดที่ต้องขอย้ำไว้ในที่นี้คือ ค่าความน่าจะเป็น p ของการรอคอยหรือทดลองแต่ละครั้งจะต้องคงที่เสมอ

อนึ่งในเรื่องของการรอคอยเรามักมองไปถึงเรื่องของเวลา เช่นเวลาที่ต้องรอจนกว่ารถเมล์ที่ต้องการจะมาถึง เวลาที่จะต้อง (ทน) รอจนกว่าเครื่องไฟฟ้า (เช่นโทรทัศน์) จะเสีย หรืออื่น ๆ เรื่องเหล่านี้เป็นการรอคอยเช่นกัน แต่ขอให้สังเกตว่าตัวแปรเวลา (หรืออื่น ๆ) มีลักษณะเป็นตัวแปรแบบต่อเนื่องตัวแปรสุ่มที่เกี่ยวกับการรอคอยสำหรับสถานการณ์เช่นนี้ (Continuous

Waiting Time) คือตัวแปรสุ่มเอกโพเนนเชียล (Exponential Variate) และตัวแปรสุ่มแกมม่า (Gamma Variate) ซึ่งจะได้กล่าวถึงในภายหลังโดยที่ตัวแปรสุ่มเอกโพเนนเชียลแสดงถึงการรอคอยจนกระทั่งประสบสิ่งที่ต้องการ (หรือไม่ต้องการ) เป็นครั้งแรกส่วนตัวแปรสุ่มแกมม่าแสดงถึงการรอคอยจนกระทั่งประสบสิ่งที่ต้องการ (หรือไม่ต้องการ) เกินกว่า 1 ครั้ง เช่นเวลาที่ต้องรอคอยโทรศัพท์จากแฟนที่จะหมุนเข้ามาเป็นครั้งที่ 5 ดังนี้เป็นต้น การที่ต้องกล่าวถึงการแจกแจงแบบนิเสธทวินาม การแจกแจงแบบเอกโพเนนเชียล และการแจกแจงแบบแกมม่าไว้ในที่นี้ก็เพราะเป็นเรื่องที่เนื่องถึงกันในลักษณะของสถานการณ์ที่คล้ายคลึงกัน และความเป็นกรณีทั่วไปของกันและกันของแต่ละคู่ทั้งนี้เพื่อจะได้มองเห็นภาพต่าง ๆ ได้เป็นกลุ่มเป็นก้อนพร้อม ๆ กันไป

ตัวแปรสุ่มอนุกรมเรขาคณิตพัฒนามาจากแนวคิดง่าย ๆ ดังนี้ จะขอเริ่มด้วยตัวอย่างเรื่องการทอดลูกเต๋า กล่าวคือสมมุติว่าปรารถนาจะทอดลูกเต๋าให้หงายหน้า 4 อยากทราบว่าจะต้องทอดลูกเต๋ากี่ครั้งจึงเกิดหน้าที่ต้องการดังกล่าว ในที่นี้จะเห็นว่าเราไม่อาจทราบว่าจะต้องหยุดทอดลูกเต๋าเมื่อไร คงทอดไปเรื่อย ๆ เมื่อใดหงายหน้า 4 ก็หยุด (ดังนี้จึงเห็นได้ว่าจำนวนครั้งของการทอดลูกเต๋าไม่เป็นจำนวนคงที่ ต่างจากกรณีของตัวแปรสุ่มทวินามที่จำนวนดังกล่าวคงที่เสมอ) ความน่าจะเป็นที่ลูกเต๋าจะหงายหน้า 4 คงที่เท่ากับ $p = \frac{1}{6}$ สมมุติทอดลูกเต๋าทั้งสิ้น x ครั้ง $x = 1, 2, \ldots$ ปรากฏผลดังนี้

$$\underbrace{F \quad F \quad F \ldots F \quad F}_{(x-1)\text{ ครั้ง}} \quad \underset{\text{ครั้งที่ } x}{S}$$

S = Success หมายถึงหงายหน้า 4, F = Failure หมายถึงหงายหน้าอื่น

ความน่าจะเป็นที่ต้องการคือ $(\frac{1}{6})(\frac{5}{6})^{x-1}$ หรือเขียนเป็นรูปทั่วไปได้เป็น pq^{x-1} จะเห็นว่า ถ้า $x = 1$ ความน่าจะเป็นจะเท่ากับ $pq^{\circ} = p$ หมายความว่าทอดครั้งแรกหงายหน้า 4 เลย ถ้า $x = 2$ จะได้ FS ความน่าจะเป็น pq หมายความว่าทอด 2 ครั้งแล้วหงายหน้า 4 ในครั้งที่ 2 ดังนี้เป็นต้น

นิยาม 4.4 ถ้าตัวแปรสุ่ม X แทนจำนวนครั้งทั้งหมดของการทดลองแบบเบอร์นุลลิ (รวมครั้งที่เกิดเหตุการณ์ที่ต้องการด้วย) ที่ดำเนินการไปเรื่อย ๆ จนกระทั่งเกิดเหตุการณ์ที่มุ่งหวังเป็นครั้งแรก และถ้า X มีฟังก์ชันการแจกแจงเป็น

$$f_X(x) = \begin{cases} f_X(x; p) = pq^{x-1}; \ x = 1, 2, \ldots \\ 0 \quad \text{เมื่อ } x \text{ มีค่าเป็นอย่างอื่น} \end{cases}$$

โดย p = ความน่าจะเป็นที่จะประสบสิ่งที่มุ่งหวัง, $q = 1 - p$

เราเรียกตัวแปรสุ่ม X ว่าตัวแปรสุ่มอนุกรมเรขาคณิต และการแจกแจงของตัวแปรสุ่มดังกล่าวเรียกว่า การแจกแจงแบบอนุกรมเรขาคณิต

จากนิยามนี้ นักศึกษาจะเห็นได้ว่า $f_X(x) = pq^{x-1}$; $x = 1, 2, \ldots$ เมื่อลองแทนค่า x แล้วนำค่าความน่าจะเป็นมาจัดในรูปชุดลำดับหรืออนุกรม จะปรากฏผลดังนี้

$$p, pq, pq^2, pq^3, \ldots \qquad \text{..........(1)}$$

และ

$$p + pq + pq^2 + pq^3 + \ldots \qquad \text{..........(2)}$$

จาก (2) จะพบว่า $p + pq + pq^2 + pq^3 + \ldots = p(1 + q + q^2 + q^3 + \ldots)$

$$= p(1 - q)^{-1} = 1 \qquad \text{..........(3)}$$

จาก (1) และ (2) จะเห็นได้ว่า (1) แสดงถึงชุดลำดับเรขาคณิต (Geometric Sequence) ส่วน (2) แสดงอนุกรมเรขาคณิต (Geometric Series) และ (3) ยืนยันให้เห็นว่า $f_X(x) = pq^{x-1}$; $x = 1, 2,$ เป็นตัวแบบของความน่าจะเป็น (Probabilistic Model) ณ. ที่นี้คงทำให้นักศึกษาเข้าใจได้แล้วว่าเพราะเหตุใดตัวแปรสุ่มตามนิยาม 4.4 นี้ถึงเรียกว่าตัวแปรสุ่มอนุกรมเรขาคณิต

ภาพของการแจกแจงแบบอนุกรมเรขาคณิตปรากฏดังนี้ ขอให้สังเกตว่าฐานนิยมของการแจกแจงจะปรากฏ ณ ค่า $x = 1$

---

[1] โดยอาศัยอนุกรมเทย์เลอร์ (Taylor Series) จะพบว่า

$$(1-y)^t = \sum_{j=0}^{\infty} \binom{t}{j}(-y)^j$$

ถ้าให้ $t = -n$ จะได้

$$(1 - y)^{-n} = \sum_{j=0}^{\infty} \binom{-n}{j}(-y)^j = \sum_{j=0}^{\infty} \binom{n+j-1}{j} y^j$$

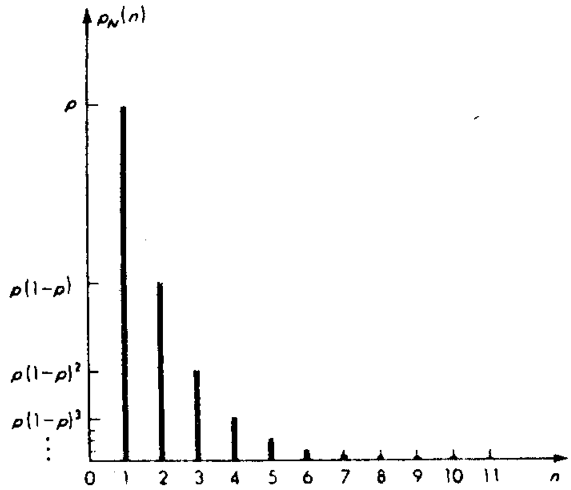
โดยที่

$$\binom{-n}{j} = \frac{(-n)(-n-1)(-n-2)\ldots(-n-j+1)}{j!} = (-1)^j \frac{n(n+1)(n+2)\ldots(n+j-1)}{j!}$$

$$= (-1)^j \binom{n+j-1}{j}$$

ถ้าให้ $n = 1$ จะได้

$$(1 - y)^{-1} = \sum_{j=0}^{\infty} \binom{j}{j} y^j = \sum_{j=0}^{\infty} y^j = 1 + y + y^2 + y^3 + \ldots$$



Geometric distribution G (p).

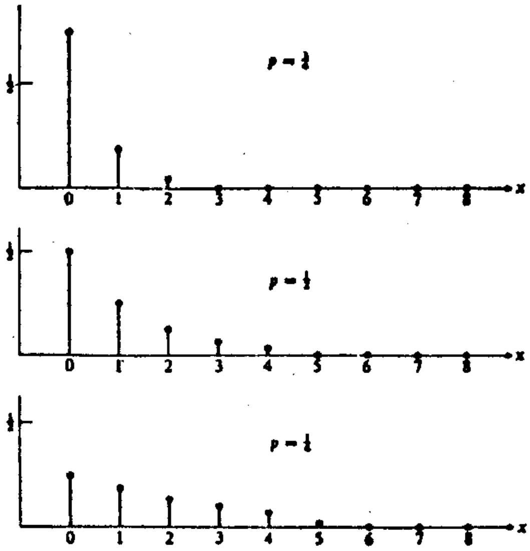
อนึ่ง เราสามารถนิยามการแจกแจงแบบอนุกรมเรขาคณิตได้ในอีกลักษณะหนึ่งคือ

$$f_x(x) \quad = \quad pq^x;\ x = 0, 1, 2, \ldots$$

รูปนี้ก็นิยมใช้กว้างขวางเช่นกัน ขอให้สังเกตค่าของ X ในที่นี้ค่าของ X จะเริ่มที่ $x = 0$ เป็นต้นไป แสดงว่า $f_x(x)$ ในที่นี้นิยามความหมายของตัวแปรสุ่ม X ต่างไปจากนิยาม 4.4 กล่าวคือ ตามนิยาม 4.4 เรานิยามว่า X หมายถึงจำนวนการทดลองทั้งหมดที่ต้องใช้ (รวมครั้งที่ Success ด้วย) จนกว่าจะได้ผลที่มุ่งหวัง ซึ่งแน่นอนที่ว่าจะต้องทำการทดลองอย่างน้อย 1 ครั้ง จึงจะมีผล "อะไร" เกิดขึ้นมาค่าของ x จึงต้องเริ่มที่ $x = 1$ แต่สำหรับ $f_x(x) = pq^x$ ในที่นี้ X หมายถึงจำนวน Failure ทั้งหมดที่เกิดขึ้นก่อนที่จะประสบความสมหวัง เมื่อ $x = 0$ ก็แสดงว่าได้ทำการทดลองไป 1 ครั้ง แล้วประสบผลตามความต้องการทันที หมายความว่าเมื่อทำการทดลองก็ไม่พบ Failure ขอให้สังเกตว่าถ้า $x = 0$ จะได้ $f_x(x) = pq^0 = p$

เรื่องนี้แม้จะใช้รูปของฟังก์ชั่นความน่าจะเป็นต่างกันแต่ก็เป็นเรื่องเดียวกันเพียงแต่มอง "เรื่องเดียวกัน" นี้ในคนละมุม การอภิปรายผลในขั้นสุดท้ายจะตรงกันเสมอ

สำหรับกราฟของ $f_x(x) = pq^x;\ x = 0, 1, 2, \ldots$ จะปรากฏดังนี้ ขอให้สังเกตว่าฐานนิยม (Mode) จะปรากฏค่า $x = 0$



Geometric densities.

ทฤษฎี 4.5 ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบอนุกรมเรขาคณิตแล้ว mgf ค่าคาดหมายและความแปรปรวนของตัวแปรสุ่มดังกล่าวจะปรากฏดังนี้

$$M_x(t) = \frac{pe^t}{1-qe^t} \; ; \; E(x) = 1/p \text{ และ } V(X) = q/p^2$$

พิสูจน์ เนื่องจาก

$$f_x(x) = pq^{x-1} \; ; \; x = 1, 2, \ldots$$

$$=> \quad M_x(t) = \sum_{x=1}^{\infty} e^{tx} pq^{x-1}$$

$$= pe^t \sum_{x=1}^{\infty} (qe^t)^{x-1}$$

$$= pe^t (1 + (qe^t) + (qe^t)^2 + \ldots)$$

$$= pe^t (1 - qe^t)^{-1}$$

$$= \frac{pe^t}{1 - qe^t}$$

การหาค่าคาดหมายและความแปรปรวนสามารถดำเนินการได้ดังนี้

วิธีที่ 1

$$E(X) = M_X'(t)\big|_{t=0}$$

$$M_X'(t) = \frac{pqe^{2t}}{(1 - qe^t)^2} + \frac{pe^t}{1 - qe^t}$$

$$=> E(X) = M_X'(t)\big|_{t=0} = (q/p) + 1 = \frac{p+q}{p} = 1/p$$

$$V(X) = M_X''(t)\big|_{t=0} - \frac{1}{p^2} = q/p^2$$

วิธีที่ 2

$$E(X) = \sum_{x=1}^{\infty} x.pq^{x-1}$$

$$= p\sum_{x=0}^{\infty} \frac{d\,q^x}{dq} \qquad ; \frac{d\,q^x}{dq} = x.q^{x-1}$$

$$= p\frac{d}{dq}\sum_{x=1}^{\infty} q^x \qquad ; \frac{d}{dq}$$ และ Σ สลับกันได้เพราะ

อนุกรมนี้เป็นชนิดลู่เข้า เมื่อ $|q| < 1$

$$= p\frac{d}{dq}(q+q^2+q^3 + \ldots)$$

$$= p\frac{d}{dq}q(1 + q + q^2 + \ldots) = p \cdot \frac{d}{dq}q(1 - q)^{-1}$$

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

$V(X) = E(X(X-1)) + E(X) - (E(X))^2$; ต้องใช้ Factorial Moment เพราะไม่อาจหา V(X) ได้โดยวิธี Raw Moment

$$E(X(X-1)) = \sum_{x=1}^{\infty} x(x - 1)\, pq^{x-1}$$

$$= q\sum_{x=2}^{\infty} x(x-1)\, pq^{x-2}$$

$$= qp\sum_{x=2}^{\infty} \frac{d^2}{dq^2} q^x = qp\frac{d^2}{dq^2}\sum_{x=2}^{\infty} q^x$$

$$= qp\frac{d^2}{dq^2}(q^2 + q^3 + q^4 + \ldots)$$

$$= pq \frac{d^2}{dq^2} (q^2(1 + q + q^2 + \ldots))$$

$$= p \quad q \quad \frac{d^2}{dq^2} (q^2(1 - q)^{-1})$$

$$= pq \left(\frac{2q^2}{(1 - q)^3} + \frac{2q}{(1 - q)^2} + \frac{2q}{(1 - q)^2} + \frac{2}{(1 - q)}\right)$$

$$= \left(\frac{2q^2 + 4pq + 2p^2}{p^2}\right) \cdot q$$

$$= \left(\frac{2(p + q)^2}{p^2}\right) = \frac{2}{p^2} \cdot q$$

$$=> V(X) = E(X(X - 1)) + E(X) - (E(X))^2$$

$$= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2}$$

$$= \frac{2q + p - 1}{p^2}$$

$$= \frac{q}{p^2}$$

หมายเหตุ ถ้าใช้ฟังก์ชั่นความน่าจะเป็นในรูป $f_x(x) = pq^x$; $x = 0, 1, 2, \ldots$

จะพบว่า $M_x(t) = \frac{p}{1 - qe^t}$; $E(X) = \frac{q}{p}$ และ $V(X) = \frac{q}{p^2}$

เห็นได้ว่า mgf และ E(X) จะอยู่ในรูปที่ต่างกัน ทั้งนี้เพราะกำหนดให้ X แทนสถานะการณ์ต่างกันแต่เราสามารถนำมาอภิปรายเชื่อมโยงเป็นเรื่องเดียวกันได้ ขอให้พิจารณาเปรียบเทียบจากตัวอย่างต่อไปนี้

ตัวอย่าง 4.6 โดยปกติแล้วในภาคใต้ของประเทศไทย จะมีฝนตกตลอดปี แต่สำหรับช่วงฤดูแล้งระหว่างเดือนมีนาคมถึงเดือนเมษายนจะมีฝนตกไม่บ่อยครั้งนัก คาดว่าโอกาสที่จะมีฝนตกสำหรับแต่ละวันในช่วงเดือนดังกล่าวมีค่าประมาณ 10% (ถือว่าการเกิดฝนตกในแต่ละวันมิได้มีอิทธิพลเนื่องถึงกัน)

ก. จงหาความน่าจะเป็นที่จะมีฝนตกในวันที่ 3 เมษายน ถ้าก่อนหน้านั้นยังไม่เคยมีฝนตกเลย

ข. จงคาดหมายว่าในช่วงเดือนดังกล่าวกี่วันจึงจะมีฝนตกครั้งหนึ่ง

ก. ให้ X = จำนวนวันทั้งหมดที่ต้องรอคอยจนกระทั่งฝนตกเป็นครั้งแรก

วิธีทำ $p = 10\% = .10$ = โอกาส (ความน่าจะเป็น) ที่จะมีฝนตกในวันหนึ่ง ๆ

$q = 90\% = .90$ = โอกาส (ความน่าจะเป็น) ที่จะไม่มีฝนตกในวันหนึ่ง ๆ

$x = 34$ = จำนวนวันทั้งสิ้นตั้งแต่ 1 มีนาคม ถึง 3 เมษายน (วันที่ 3 เมษายนเป็นวันที่ 34 เมื่อเริ่มต้นนับที่วันที่ 1 มีนาคม)

ดังนั้น

$$P(X = 34) = (.10)(.90)^{34-1} = 0.003$$

นั่นคือ โอกาสที่จะมีฝนตกในวันที่ 3 เมษายนเท่ากับ 0.3%

ข. เนื่องจาก $E(X) = 1/p$ และค่าคาดหมายของจำนวนครั้งของการรอคอยทั้งหมดที่ต้องใช้ไปเพื่อจะพบสิ่งที่มุ่งหวัง (ในที่นี้สิ่งที่มุ่งหวังคือฝน) เป็นครั้งแรก

$$=> E(X) = \frac{1}{.10} = 10$$

นั่นคือคาดว่าฤดูแล้ง (มีนาคม-เมษายน) นี้จะมีฝนตกครั้งหนึ่งในทุก 10 วัน หมายความว่า 9 วัน แรกไม่มีฝนตก วันที่ 10 จะมีฝนตก เป็นดังนี้เรื่อยไปตลอดฤดูกาล

ข้อสังเกต สำหรับผลในข้อ ข. ถ้าจะคาดหมายโดยใช้ $f_X(x) = pq^x$; $x = 0, 1, 2, \ldots$ จะพบว่า $E(X) = q/p = .90/.10 = 9$ = จำนวนวันที่ไม่มีฝนตกก่อนที่จะมีฝนตกเป็นครั้งแรกในวันต่อไป

ตัวอย่าง 4.7 สมมุติว่าค่าใช้จ่ายเพื่อขุดเจาะน้ำบาดาลในพื้นที่จังหวัดภาคตะวันออกเฉียงเหนือครั้งหนึ่ง ๆ เท่ากับ 100,000 บาท ถ้าเจาะไม่พบแหล่งน้ำใต้ดินที่เพียงพอสำหรับการประปาในท้องถิ่นจะต้องโยกย้ายแหล่งใหม่และต้องเสียค่าใช้จ่ายเพื่อการโยกย้ายขนส่ง และค่าจ้างนักสำรวจรวมครั้งละ 20,000 บาท ซึ่งจะต้องดำเนินการและต้องเสียค่าใช้จ่ายเช่นนี้เรื่อยไปตราบใดที่ยังเจาะหาแหล่งน้ำใต้ดินที่เหมาะสมไม่ได้ สมมุติว่าความน่าจะเป็นที่จะเจาะพบแหล่งน้ำที่เหมาะสมเท่ากับ 0.2 จงคาดหมายว่ากว่าจะเจาะพบแหล่งน้ำที่เหมาะสมและพอเพียงแก่การประปาสาธารณะได้ต้องสิ้นค่าใช้จ่ายเท่าไร

วิธีทำ

ให้ X = จำนวนครั้งของการขุดเจาะบ่อบาดาลทั้งหมดจนกระทั่งพบบ่อที่เหมาะสม

ให้ C = ค่าใช้จ่ายในการดำเนินการทั้งหมด

ดังนั้น

$$C = 100{,}000X + 20{,}000(X-1) \quad \text{บาท} \qquad \ldots(1)$$

สมการ (1) แสดงให้เห็นว่า ถ้าต้องขุดเจาะถึง X ครั้ง จึงพบบ่อน้ำที่ต้องการ (แต่ละครั้งต้องเสียค่าใช้จ่าย 100,000 บาท) ก่อนหน้านั้น X − 1 ครั้ง คือ จำนวนครั้งของความล้มเหลว ซึ่งจะต้องเสียค่าโยกย้ายและค่าช่างสำรวจครั้งละ 20,000 บาท ค่าใช้จ่ายรวมจึงเท่ากับ 100,000X + 20,000(X − 1)

$$=> \quad C = 120{,}000X - 20{,}000 \quad \text{บาท}$$

ดังนั้น

$$E(C) = 120{,}000E(X) - 20{,}000 \quad \text{บาท}$$

$$= 120{,}000\left(\frac{1}{0.2}\right) - 20{,}000 \quad \text{บาท} \because E(X) = 1/p$$

$$= 580{,}000 \quad \text{บาท}$$

นั่นคือ ถ้าจะต้องทำการขุดเจาะบ่อบาดาลไปเรื่อย ๆ จนกระทั่งพบบ่อที่มีลักษณะตามต้องการคาดว่าจะต้องสิ้นค่าใช้จ่ายถึง 580,000 บาท ($E(X) = \frac{1}{p} = \frac{1}{0.2} = 5$ แสดงว่าโดยถัวเฉลี่ยแล้วคาดว่า จะต้องขุดเจาะประมาณ 5 บ่อจึงจะพบบ่อที่ตรงตามต้องการ)

ข้อสังเกต ถ้าจะคำนวณโดยใช้ $E(X) = q/p$ สำหรับกรณี $f_x(x) = pq^x; x = 0, 1, 2, \ldots$ ผลลัพธ์จะตรงกัน ขอเว้นไว้ให้เป็นแบบฝึกหัด

ตัวอย่าง 4.8 ในการทดลองทอดลูกเต๋าพร้อมกัน 2 ลูก และเราสนใจเฉพาะผลรวมบนหน้าลูกเต๋าที่หงายขึ้นในการทอดแต่ละครั้ง ถ้าต้องการให้ผลรวมบนหน้าลูกเต๋ามีค่าเท่ากับ 9

ก. จงคาดหมายดูว่าเราจะต้องทำการทดลองทอดลูกเต๋ากี่ครั้งจึงจะได้ผลรวมบนหน้าลูกเต๋ามีค่าเท่ากับ 9

ข. จงหาความน่าจะเป็นที่จะพบครั้งที่ผลรวมบนหน้าลูกเต๋ามีค่าเท่ากับ 9 ในการทดลองเพียงไม่ถึง 5 ครั้ง

วิธีทำ ความน่าจะเป็นที่ลูกเต๋าจะหงายหน้าออกมาให้ผลรวมเท่ากับ 9 หาได้ดังนี้

ให้ S = {ลูกเต๋ามีผลรวมหน้าหงายเท่ากับ 9}

= $\{(x, y): x + y = 9\}$ เมื่อ x = หน้าลูกเต๋าลูกที่ 1 y = หน้าลูกเต๋าลูกที่ 2

$=> S = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$

ดังนั้น $P(S) = \frac{4}{36} = \frac{1}{9} = p$ และ $P(F) = \frac{8}{9} = q$ ความน่าจะเป็นที่ผลรวมหน้าหงายจะมีค่าเป็นอย่างอื่น

ให้ X เป็นจำนวนครั้งของการทอดลูกเต๋าทั้งหมดจนกว่าจะปรากฏครั้งที่ให้ผลรวมบนหน้าหงายเท่ากับ 9

ก. $E(X) = ?$

$$E(X) = \frac{1}{p} = \frac{1}{1/9} = 9$$

นั่นคือ คาดว่าจะต้องทอดลูกเต๋าอยู่ถึง 9 ครั้งจึงจะปรากฏผลว่าผลรวมบนหน้าหงายมีค่าเท่ากับ 9 โดย 8 ครั้งแรก ปรากฏผลรวมหน้าหงายมีค่าเป็นอย่างอื่น ครั้งที่ 9 เป็นครั้งที่ได้ผลตามต้องการ

$$P(X < 5) = ?$$

$$\begin{aligned} P(X < 5) &= \sum_{k=1}^{4} P(X = k) \\ &= p + pq + pq^2 + pq^3 \\ &= p(1 + q + q^2 + q^3) \\ &= \left(\frac{1}{9}\right)\left(1 + \frac{8}{9} + \left(\frac{8}{9}\right)^2 + \left(\frac{8}{9}\right)^3\right) \\ &= .372 \end{aligned}$$

นั่นคือ โอกาสที่จะต้องทดลองลูกเต๋าไม่ถึง 5 ครั้งแล้วปรากฏหน้าที่ต้องการ มีค่าเท่ากับ .372

ตัวอย่าง 4.9 เมื่อ $X_1, X_2, \ldots, X_r$ เป็นกลุ่มตัวอย่างที่สุ่มมาจากกลุ่มประชากรที่มีการแจกแจงแบบอนุกรมเรขาคณิตที่มีค่าความน่าจะเป็น p จงหาค่าคาดหมายและความแปรปรวนของยอดรวม $\sum^{r} X_i$

วิธีทำ เพราะว่า $X_1, X_2, \ldots X_r$ เป็น Sampled Random Variable ที่สุ่มมาจากกลุ่มประชากรที่มีการแจกแจงแบบอนุกรมเรขาคณิต

ดังนั้น

$$f_{X_i}(x_i) = pq^{x_i - 1};\ x_i = 1, 2, \ldots, i = 1, 2, \ldots . I$$

$$Y = \sum^{r} X_i$$

$$\Rightarrow \quad M_Y(t) = \prod^{r} M_{X_i}(t) = \left(\frac{pe^t}{1 - qe^t}\right)^r$$

$$\begin{aligned} E(Y) &= M'_Y(t)\Big|_{t=0} \\ &= (rqe^t(pe^t)^r(1 - qe^t)^{-r-1} + r(pe^t)^r(1 - qe^t)^{-r})\Big|_{t=0} \end{aligned}$$

$$= \frac{rqp^r}{p^{r+1}} + \frac{rp^r}{p^r}$$

$$= \frac{rq}{p} + r = \frac{rq + rp}{p} = \frac{r(q + p)}{p} = \frac{r}{p}$$

$$V(Y) = E(Y^2) - (E(Y))^2$$

$$E(Y^2) = M''_Y(t) \Big|_{t=0}$$

$$M''_Y(t) = \{(rqp^r(e^t)^{r+1}(r+1)qe^t(1-qe^t)^{-r-2} + rqp^r(1-qe^t)^{-r-1}$$

$$.(r+1)(e^t)^r) + (rp^2(e^t)^r\, rqe^t(1-qe^t)^{-r-1} +$$

$$r^2p^r(e^t)^r(1-qe^t)^{-r})\}$$

$$=> \quad E(X^2) = M''_X(t) \Big|_{t=0}$$

$$= r(r+1)\frac{q^2}{p^2} + r(r+1)\frac{q}{p} + r^2\frac{q}{p} + r^2$$

$$= r^2\frac{q^2}{p^2} + r\frac{q^2}{p^2} + 2r^2\frac{q}{p} + r\frac{q}{p} + r^2$$

$$=> \quad V(X) = E(X^2) - (E(X))^2$$

$$= r^2\frac{q^2}{p^2} + r\frac{q^2}{p^2} + 2r^2\frac{q}{p} + r\frac{q}{p} + r^2$$

$$(r^2\frac{q^2}{p^2} + 2r^2\frac{q}{p} + r^2)$$

$$= r\frac{q^2}{p^2} + r\frac{q}{p} = \frac{rq}{p}\left(\frac{p+q}{p}\right)$$

$$= \frac{rq}{p^2}$$

นั่นคือ เมื่อ X_i ; $i = 1, 2, 3, \ldots, r$ มีการแจกแจงแบบอนุกรมเรขาคณิตแล้วตัวสถิติ $Y = \sum_i^r X_i$ จะมีค่าคาดหมายและความแปรปรวนเป็น $(\frac{rq}{p} + r) = \frac{r}{p}$ และ $\frac{rq}{p^2}$ ตามลำดับ

4.2.4 **การแจกแจงแบบนิเสธทวินาม** (Negative Binomial Distribution)[1]

การแจกแจงแบบนิเสธทวินามเป็นกรณีทั่วไปของการแจกแจงแบบอนุกรมเรขาคณิต กล่าวคือ ในเรื่องของการแจกแจงเบบนิเสธทวินามนั้น เรามุ่งที่จะศึกษาถึงการแจกแจงของตัวแปรสุ่มที่แสดงจำนวนครั้งของการทดลองเบอร์นุลลิทั้งหมดจนกระทั่งประสบเหตุการณ์ที่มุ่งสนใจซ้ำ ๆ กันครบจำนวนครั้งที่ตั้งเป้าหมายไว้ r ครั้ง โดยที่ $r \geq 2$ เช่น จำนวนสลากกินแบ่งที่จะต้องซื้อ (งวดละ 1 ฉบับ) จนกระทั่งถูกรางวัลที่ 1 ครบ 3 งวด จำนวนครั้งที่จะต้องแทงไฮโลทั้งหมดจนกระทั่งถูก "สามต่ำ" ครบ 7 ครั้ง หรือจำนวนครั้งของการทดลองส่งยานอวกาศไปลงบนดาวอังคารทั้งหมดจนกระทั่งสามารถนำลงสู่ผิวดาวอังคารได้ครบ 4 ครั้ง ดังนี้เป็นต้น การแจกแจงแบบนิเสธทวินามส่วนใหญ่จะถูกนำไปใช้ในเรื่องที่เกี่ยวกับการเกิดอุบัติเหตุ การเกิดโรคภัยไข้เจ็บ การคมนาคมและเรื่องที่เกี่ยวกับการนับและการรอคอยต่าง ๆ ขอให้เป็นที่สังเกตุว่า การทดลอง (หรือการสังเกตุ) แต่ละครั้งเป็นการทดลองแบบเบอร์นุลลิ และความน่าจะเป็นที่จะประสบผลสำเร็จ p จะต้องคงที่เสมอไปในทุกการทดลอง

เหตุที่เรียกการแจกแจงชนิดนี้ว่า "นิเสธทวินาม" เพราะการแจกแจงชนิดนี้มีลักษณะสวนทางกับการแจกแจงทวินามทั้งที่สถานการณ์แวดล้อมคล้ายคลึงกันกล่าวคือ ในการแจกแจงแบบทวินามนั้นเป็นเรื่องของการหาจำนวนครั้งและความน่าจะเป็นของ "ความสมหวัง" (X) เมื่อกำหนดจำนวนครั้งของการทดลองทั้งหมด n ให้ ส่วนการแจกแจงแบบนิเสธทวินามเป็นเรื่องของการหาจำนวนครั้งของการทดลอง และความน่าจะเป็นที่จะได้จำนวนครั้งของการทดลองทั้งหมด (X) เมื่อกำหนดจำนวนครั้งของ "ความสมหวัง" r ให้คงที่

การแจกแจงแบบนิเสธทวินามพัฒนาขึ้นมาจากแนวคิดง่าย ๆ เช่นกัน จะขอเริ่มด้วยตัวอย่างเกี่ยวกับการทอดลูกเต๋าดังนี้

สมมุติว่าเราปรารถนาจะทอดลูกเต๋าให้หงายหน้า "เอี่ยว" (หน้า 1) ครบ 10 ครั้ง อยากทราบว่าจะต้องทดลองทอดลูกเต๋าทั้งสิ้นกี่ครั้งจึงจะได้ผลรวมตามที่ต้องการ

กรณีนี้จะเห็นได้ว่า การทดลองจะกระทำไปเรื่อยโดยไม่ทราบว่าจะหยุดเมื่อใด ทราบแต่เพียงว่าถ้าทอดไปจนลูกเต๋าหงายหน้าเอี่ยวครบ 10 ครั้งเมื่อใดก็หยุดเมื่อนั้น ไดอะแกรมสำหรับการแจกแจงจึงอาจปรากฏดังนี้ (S = Success หงายหน้าเอี่ยว; F = Failure หงายหน้าอื่น)

$$\underbrace{S \quad F \quad S \quad F \quad E}_{\text{มี S ปนอยู่ ๙ ครั้ง}} \ldots \underbrace{S}_{\text{S ครั้งที่ 10}}$$

[1] บางครั้งเรียกว่า Pascal Distribution

จะเห็นว่าในบรรดาการทดลองทั้งหมดก่อนที่พบ S ครั้งที่ 10 จะมี S อยู่ 9 ครั้ง และมี F อยู่จำนวนหนึ่งที่ไม่ทราบจำนวนแน่นอน และ S กับ F เหล่านั้นสามารถเกิดสลับที่กันได้มากมายหลายวิธี

สมมุติว่าทำการทดลองอยู่ x ครั้ง x = 1, 2, ... แล้วได้ S ครบ 10 ครั้งพอดี อาจจะปรากฏไดอะแกรม ดังนี้

$$\underbrace{\text{F F} \ldots \text{S S F S F F F} \ldots \text{S S}}_{x-1 \text{ ครั้ง}} \; \underbrace{\text{S}}$$

(มี S ปนอยู่ 9 ครั้ง) S ครั้งที่ 10 (การทดลองครั้งที่ X)

ในบรรดาการทดลองทั้ง (x – 1) ครั้งมี S ปนอยู่ 9 ครั้ง ดังนั้น S กับ F สามารถสลับที่กันได้ถึง $\binom{x-1}{9}$ วิธี และเนื่องจากแต่ละเหตุการณ์เช่นนี้ จะมีค่าความน่าจะเป็นเท่ากับ $p^{10}q^{x-10}$ ดังนั้น ความน่าจะเป็นที่ต้องการมีค่าเท่ากับผลรวมในทุก ๆ วิถีทางที่เป็นไปได้ของความน่าจะเป็นของแต่ละเหตุการณ์

นั่นคือ Pr (ลูกเต๋าหงายหน้าเอี่ยวครบ 10 ครั้งในการทดลอง x ครั้ง)

$$= \underbrace{p^{10}q^{x-10} + \ldots + p^{10}q^{x-10}}_{\binom{x-1}{9} \text{ ครั้ง}}$$

$$=> \Pr(X = x) = \binom{x-1}{9} p^{10}q^{x-10}; x = 1, 2, \ldots$$

เช่นปรารถนาจะให้ได้หน้าเอี่ยว 2 ครั้ง สมมุติว่าทอดไป 5 ครั้งแล้วได้หน้าเอี่ยวครบ 2 ครั้งตามต้องการ ผลการทดลองที่สอดคล้องกับเรื่องนี้จึงเป็นไปได้ถึง 4 เหตุการณ์อย่างใดอย่างหนึ่งคือ SFFFS หรือ FSFFS หรือ FFSFS หรือ FFFSS

ความน่าจะเป็นของแต่ละเหตุการณ์มีค่าเท่ากันตลอดเท่ากับ $(pq^3)\,p = (\frac{1}{6})(\frac{5}{6})^3(\frac{1}{6})$ ดังนั้นความน่าจะเป็นที่ลูกเต๋าจะหงายหน้าเอี่ยวครบ 2 ครั้งในการทอดลูกเต๋า 5 ครั้งจึงมีค่าดังนี้

$$\begin{aligned}\Pr(X = 5) &= (pq^3)\,p + (pq^3)\,p + (pq^3) + (pq^3)\,p \\ &= 4(pq^3)\,p \\ &= \binom{4}{1}(pq^3)\,p\end{aligned}$$

$$= \binom{5-1}{2-1} p^{2-1} q^{5-2} . p$$

$$= \binom{5-1}{2-1} p^2 q^{5-2}$$

$$= \binom{5-1}{2-1} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2}$$

Combination $\binom{4}{1}$ ยืนยันถึงจำนวนหนทางที่พึงเป็นไปทั้งหมดที่จะเกิด S ครบ 2 ครั้งจากการทดลอง 5 ครั้ง และชี้ให้เห็นว่าในบรรดาการทดลอง 4 ครั้งแรก[1] ซึ่งจะต้องมี S หนึ่งครั้งนั้น จะมีหนทางที่เป็นไปได้ทั้งหมดของการเกิด S ได้ทั้งสิ้น $\binom{4}{1} = 4$ วิธี อาจเกิด S ได้ในครั้งแรก 3 ครั้ง ต่อมาเกิด F หรือครั้งแรกเกิด F ครั้งต่อมาจึงเกิด S อีก 2 ครั้งเกิด F...

อนึ่งเมื่อเขียน $Pr(X = 5)$ ในลักษณะ $Pr(X = 5) = \binom{4}{1} p^2q^3$ แม้ว่าจะถูกต้องแต่ก็มิได้ชี้ให้เห็นว่าการทดลองของเราดำเนินไปกี่ครั้ง หวังจะให้เกิด S กี่ครั้ง เราจึงแปลงรูป $Pr(X)$ เสียใหม่ให้อยู่ในลักษณะที่สามารถ ผนวกสิ่งต่าง ๆ เหล่านั้นไว้ครบ $Pr(X)$ จึงกลายเป็น $Pr(X = 5) = \binom{5-1}{2-1} p^2q^3$ ซึ่งชี้ชัดลงไปว่า $Pr(X)$ นี้คือความน่าจะเป็นที่จะเกิด S ครบ 2 ครั้งในการทดลอง 5 ครั้ง

นิยาม 4.5 ถ้าตัวแปรสุ่ม X แทนจำนวนครั้งทั้งหมดของการทดลองเบอร์นูลลิซึ่งดำเนินการไปเรื่อย ๆ จนกระทั่งเกิดเหตุการณ์ที่มุ่งหวังครบ r ครั้งพอดี และถ้า X มีฟังก์ชั่นการกระจายเป็น

$$f_X(x) = f_X(x; r, p) = \begin{cases} \binom{x-1}{r-1} p^r q^{x-r}\ ; x = r, r+1, r+2, \ldots \\ 0 \text{ เมื่อ } x \text{ มีค่าเป็นอย่างอื่น} \end{cases}$$

โดยที่ p = ความน่าจะเป็น (คงที่) ที่จะเกิดเหตุการณ์ที่ต้องการ r = จำนวนครั้งของเหตุการณ์ที่ตั้งเป้าหมายไว้ให้เกิดขึ้น

[1] ครั้งที่ 5 เป็นครั้งที่ได้ S เป็นวาระที่ 2 ขอให้สังเกตุว่าความน่าจะเป็นของแต่ละเหตุการณ์เท่ากับ (pq^3) p p ตัวหลังที่เขียนแยกไว้ต่างหากนี้แสดงความน่าจะเป็นของ S ในวาระที่ 2 ซึ่งเรา (ในที่นี้) ทราบแน่นอนแล้วว่า S สำหรับวาระที่ 2 จะต้องวางอยู่ ณ ตำแหน่ง "หางแถว" นี้เสมอ ปัญหาของเรื่องจึงเหลืออยู่ว่า S วาระแรกจะสามารถวางไว้ได้ ณ ที่ตำแหน่งใดบ้าง หนทางที่จะเกิด S 1 ครั้ง ในบรรดาการทดลองทั้งสิ้น 4 ครั้ง ที่ต้องเป็น F รวมอยู่ด้วย 3 ครั้งเป็นไปได้ทั้งสิ้นเท่าไร

เราเรียกตัวแปรเชิงสุ่มที่มีฟังก์ชั่นการแจกแจงเช่นนี้ว่า ตัวแปรสุ่มนิเสธทวินาม (Negative Binomial Variate) เรียกการแจกแจงว่า การแจกแจงนิเสธทวินาม

ข้อสังเกต การที่ค่าของ X เริ่มที่ $x = r$ หมายความว่า การทดลองใด ๆ ที่หวังหรือตั้งเป้าหมายไว้ว่าจะให้เกิดเหตุการณ์ที่ต้องการครบ r ครั้ง การทดลองนั้นก็จำเป็นต้องดำเนินการทดลองอย่างน้อย r ครั้ง เช่นต้องการที่จะยิงปืนให้ถูกเป้า 5 ครั้ง นักยิงปืนก็จะต้องยิงอย่างน้อย 5 ครั้ง

อนึ่ง ถ้านิยามให้ X แทนจำนวน "ความล้มเหลว" F ทั้งหมดก่อนที่จะพบ "ความสมหวัง" S ครบ r ครั้งฟังก์ชั่นการแจกแจงจะมีรูปดังนี้

$$f_X(x) = f_X(x; r, p) = \binom{x+r-1}{x} p^{r-1} q^x \; ; x = 0, 1, 2, \ldots..$$

ขอให้สังเกตว่าค่าของ X จะเริ่มที่ $x = 0$ หมายความว่าการทดลองนั้นไม่เคยพบ "ความล้มเหลว" เลย รูปของ $f_X(x)$ ทั้งสองนี้แม้จะต่างกัน แต่เมื่ออภิปรายผลแล้วจะปรากฏผลลัพธ์ตรงกัน เช่นเดียวกับการแจกแจงแบบอนุกรมเรขาคณิต

ทฤษฎี 4.6 ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบนิเสธทวินาม (ตามนิยาม 4.5) แล้ว mgf ค่าคาดหมายและความแปรปรวนของ X จะปรากฏดังนี้

$$M_X(t) = \left(\frac{pe^t}{1 - qe^t}\right)^r$$

$$E(X) = \frac{rq}{p} + r = \frac{r}{p} \text{ และ } V(X) = \frac{rq}{p^2}$$

พิสูจน์

$$\because f_X(x) = \binom{x-1}{r-1} p^r q^{x-r} \qquad ; x = r, r+1, \ldots$$

$$M_X(t) = \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r q^{x-r}$$

$$= \frac{p^r}{q^r} \sum_{x=r}^{\infty} \binom{x-1}{r-1} (qe^t)$$

$$= \frac{p^r}{q^r} (qe^t)^r \{1 + r(qe^t) + \frac{r(r+1)(qe^t)^2}{2!} + \frac{r(r+1)(r+2)(qe^t)^3}{3!}$$

$$+ \quad \}$$

$= p^r(e^t)^r . (1 - qe^t)^{-r}$: (ดูเชิงอรรถท้ายนิยาม 4.4)

$$= \frac{(pe^t)^r}{(1 - qe^t)^r}$$

$$= \left(\frac{pe^t}{1 - qe^t}\right)^r$$

$E(X) = M'_X(t)\Big|_{t=0} = \frac{rq}{p} + r = \frac{r}{p}$ และ $V(X) = M''_X(t)\Big|_{t=0} - \frac{r^2}{p^2} = \frac{rq}{p^2}$ (ดูตัวอย่าง 4.9) นั่นคือเมื่อตัวแปรสุ่ม X มีการแจกแจงแบบนิเสธทวินาม (ตามนิยาม 4.5) แล้ว X จะมี mgf ค่าคาดหมายและความแปรปรวนเป็น $M_X(t) = \left(\frac{pe^t}{1 - qe^t}\right)^r$, $E(X) = \frac{rq}{p} + r = \frac{r}{p}$ และ $V(X) = \frac{rq}{p^2}$

หมายเหตุ ถ้านิยามให้ตัวแปรเชิงสุ่ม X แสดงจำนวนครั้งของ "ความล้มเหลว" F ทั้งหมดก่อนที่พบ "ความสมหวัง" S ครบ r ครั้ง ตามที่ตั้งเป้าหมายไว้จะพบว่าตัวสุ่ม X มีฟังก์ชั่นความน่าจะเป็น mgf ค่าคาดหมาย และความแปรปรวนดังนี้

$$f_X(x) = f_X(x; r, p) = \binom{x+r-1}{x} p^{r-1} q^x ; \; x = 0, 1, 2, \ldots$$

$$M_X(t) = \left(\frac{p}{1 - qe^t}\right)^r, E(X) = \frac{rq}{p} \text{ และ } V(X) = \frac{rq}{p^2}$$

ในกรณีนี้จะเห็นได้ว่า $E(X) = \frac{rq}{p}$ หมายถึงจำนวน F ทั้งหมดที่คาดว่าจะมีขึ้นก่อนที่จะได้ S เป็นครั้งที่ r ส่วน $E(X) = \frac{rq}{p} + r$ ตามทฤษฎี 4.6 หมายถึงจำนวนการทดลองทั้งหมดที่คาดว่าที่จะต้องกระทำเพื่อให้ได้ S ครบ r ครั้งตามที่กำหนดเป้าหมายไว้ ณ จุดนี้จึงเป็นสิ่งที่ควรสังเกตไว้ว่า กรณี $E(X) = \frac{rq}{p} + r$ นั้นแสดงว่าจำนวนครั้งของการทดลองทั้งหมดที่คาดว่าจะต้องทำจะเท่ากับจำนวนครั้งที่คาดว่าจะเกิด F ทั้งหมดก่อนจะได้ S เป็นครั้งที่ r + จำนวนครั้งของการทดลองขั้นต่ำที่สุดที่ต้องกระทำ เช่นถ้าต้องการทราบว่าจะต้องปาเป้าทั้งสิ้นกี่ครั้งจึงจะถูก "แจ๊คพอท" ครบ 4 ครั้ง กรณีเช่นนี้ก่อนอื่นผู้ปาเป้าจะต้องปาเป้าอย่างน้อย 4 ครั้ง ส่วนที่จะมากกว่า 4 ครั้งไปอีกเพียงใดขึ้นอยู่กับจำนวนครั้งที่ปาพลาด ซึ่งก็หมายถึงความแม่นยำของผู้ปาเป็นสำคัญ สมมุตินาย ก. ปาแม่นมาก มีโอกาสที่จะปาถูกถึง 80% ถ้านาย ก. ต้องการถูก

"แจ๊คพอท" 4 ครั้ง เราก็คาดว่า นาย ก. จะเสียเวลาปาเพียง $\frac{rq}{p} + r = \frac{4(.2)}{.8} + 4 = 5$ ครั้ง (พลาดเพียง $\frac{rq}{p} = 1$ ครั้งเท่านั้น)

เรื่องของการแจกแจงนิเสธทวินามเป็นเรื่องของการรอคอย เมื่อพิจารณาเชื่อมโยงไปถึงเรื่องของการนับ (Count) เช่นเดียวกับลักษณะของการแจกแจงแบบพัวซอง นักศึกษาจะพบว่าการแจกแจงแบบนิเสธทวินามสามารถสร้างขึ้นมาได้จากการผสม (Contigous หรือ Mixture) ระหว่างตัวแปรสุ่มที่มีการแจกแจงแบบพัวซองและแบบแกมม่า สำหรับเรื่องการแจกแจงแบบลูกผสม (contigous Distribution) นี้นักศึกษาจะได้พบในตอนท้ายของบทที่ 4 นี้

การแจกแจงแบบนิเสธทวินามนอกจากจะถูกนำไปใช้ประโยชน์ในด้านที่เกี่ยวกับอุบัติเหตุโรคภัยไข้เจ็บหรือสถิติที่เกี่ยวกับสุขภาพอนามัย เช่นจำนวนคนที่เป็นโรคชนิดใดชนิดหนึ่ง (การแพร่ระบาดของโรค) รังสีคอสมิค กระบวนการเรียนรู้ การจราจรและคมนาคมแล้ว ยังใช้เป็นหลักเกณฑ์ในการเลือกตัวอย่าง (Sampling Technique) อีกด้วย กล่าวคือ ถ้ามีประสบการณ์หรือสภาพการณ์ชี้ชัดว่าสัดส่วนของกลุ่มประชากร p ในเรื่องหนึ่งเรื่องใดเป็นค่าที่ทราบได้ เช่น ทราบว่า ประมาณ 8% ของคนในกรุงเทพฯ เป็นโรคจิต หรือประมาณ 40% ของบัณฑิตสาขาวิทยาศาสตร์จะไม่มีงานทำ เมื่อกำหนดว่าจะสุ่มตัวอย่างหน่วยสำรวจสำรวจในเรื่องนั้นให้ได้ครบ r หน่วยโดยจะดำเนินการสำรวจไปเรื่อย ๆ จนพบหน่วยสำรวจที่มีคุณสมบัติตรงกับเงื่อนไขตามความต้องการแล้วจำนวนหน่วยสำรวจทั้งหมด (n) ที่ได้ไปสอบถามมาจะมีการแจกแจงแบบนิเสธทวินาม

อนึ่ง เมื่อพิจารณาฟังก์ชั่นความน่าจะเป็น mgf ค่าคาดหมายและความแปรปรวนของตัวแปรสุ่มนิเสธทวินาม คือ

$$f_X(x) = \binom{x-1}{r-1} p^r q^{x-r} ; x = r, r+1, r+2, \ldots$$

$$M_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r ; \ E(X) = \frac{r}{p} \quad \text{และ } V(X) = \frac{rq}{p^2}$$

จะพบว่า

ถ้าให้ r = 1 รูปของ $f_X(x)$, mgf, E(X) และ V(X) จะกลายเป็น

$$f_X(x) = \binom{x-1}{r-1} p^1 q^{x-1} \quad pq^{x-1} ; x = 1, 2, \ldots$$

$$M_X(t) = \frac{pe^t}{1-qe^t}, E(X) = \frac{1}{p}, V(X) = \frac{q}{p^2}$$

ซึ่งก็คือ $f_X(x)$, mgf, E(X) และ V(X) ของตัวแปรสุ่มอนุกรมเรขาคณิตนั่นเอง แสดงว่า การแจกแจงแบบอนุกรมเรขาคณิต เป็นกรณีเฉพาะของการแจกแจงแบบนิเสธทวินาม (เมื่อ $r = 1$) ขณะเดียวกัน ถ้า $r \geq 1$ การแจกแจงแบบอนุกรมเรขาคณิตจะขยายรูป (Generalize) เป็นการแจกแจงแบบนิเสธทวินาม (ดูตัวอย่าง 4.9)

ตัวอย่าง 4.10 ในการส่งยานอวกาศไปดาวอังคาร โอกาสที่จะนำยานลงสู่ดาวอังคารได้เท่ากับ 0.8 (สมมุติ) ทางโครงการอวกาศกำหนดเป้าหมายไว้ว่า ถ้าสามารถทดลองนำยานลงสู่ดาวอังคารได้สำเร็จครบ 3 ครั้งแล้ว ครั้งต่อไปจะส่งมนุษย์ไปได้พร้อมกับยาน

ก. จงคาดหมายว่าการทดลองจะต้องกระทำอยู่กี่ครั้งจึงจะสามารถส่งมนุษย์ไปกับยานได้

ข. จงหาค่าความน่าจะเป็นที่จะต้องดำเนินการทดลองทั้งสิ้นไม่ถึง 6 ครั้ง ก็สามารถนำยานลงไปครบ 3 ครั้ง

ค. ถ้าการทดลองส่งยานอวกาศไปยังดาวอังคารครั้งหนึ่งจะสิ้นค่าใช้จ่ายถึง 500 ล้านเหรียญสหรัฐและถ้าครั้งใดพบความล้มเหลวคือไม่สามารถบังคับให้ยานลงสู่ดาวอังคารได้ก็จะต้องเสียค่าใช้จ่ายเพิ่มขึ้นอีกครั้งละ 10 ล้านเหรียญ จงหาค่าคาดหมายว่าโครงการนี้จะสำเร็จตามเป้าหมายจะต้องสิ้นค่าใช้จ่ายเท่าไร

วิธีทำ ให้ X = จำนวนครั้งของการทดลองทั้งหมดจนกระทั่งสามารถนำยานลงสู่ดาวอังคารได้ครบ 3 ครั้ง $p = 0.8$ ความน่าจะเป็นที่จะสามารถนำยานลงได้ และ $r = 3$

ก. $E(X) = \dfrac{r}{p} = \dfrac{3}{0.8} = 3.75 \cong 4$

นั่นคือเราสามารถคาดหมายได้ว่าโครงการอวกาศจะต้องดำเนินการทดลองทั้งสิ้น 4 ครั้ง ครั้งที่ 5 จะเป็นครั้งที่สามารถส่งมนุษย์ไปพร้อมกับยานได้

ข. $\Pr(X < 6) =$?

$$\begin{aligned}
\Pr(X < 6) &= \Pr(X = 3) + \Pr(X = 4) + \Pr(X = 5) \\
&= \binom{3-1}{3-1}(.8)^3 + \binom{4-1}{3-1}(.8)^3(.2) + \binom{5-1}{3-1}(.8)^3(.2)^2 \\
&= .512 + 3(.1024) + 6(.0205) \\
&= .942
\end{aligned}$$

นั่นคือโอกาสที่จะต้องทดลองส่งยานไปดาวอังคารไม่ถึง 6 ครั้งก็สามารถนำยานลงสู่ผิวดาวอังคารได้ครบ 3 ครั้ง มิได้ถึง 94.2%

ค. ให้ C = ค่าใช้จ่ายทั้งสิ้นของโครงการอวกาศก่อนที่จะทดลองส่งมนุษย์ไปพร้อมกับยาน

$$=> \quad C = 500{,}000{,}000X + 10{,}000{,}000\,(X-r) \quad \text{เหรียญสหรัฐ}$$

$$= 510{,}000{,}000X - 10{,}000{,}000r \quad \text{เหรียญสหรัฐ}$$

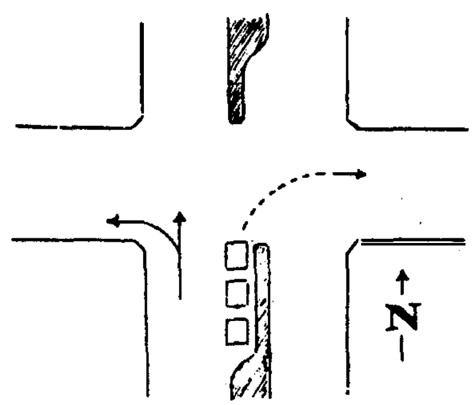
$$=> E(C) = 510{,}000{,}000E(X) - 10{,}000{,}000r \quad \text{เหรียญสหรัฐ}$$

$$= 510{,}000{,}000\left(\frac{3}{0.8}\right) - 10{,}000{,}000\,(3) \quad \text{เหรียญสหรัฐ}$$

$$= 1{,}882{,}500{,}000 \quad \text{เหรียญสหรัฐ}$$

นั่นคือ กว่าที่โครงการอวกาศนี้จะสามารถส่งมนุษย์ไปดาวอังคารได้ จะต้องเสียค่าใช้จ่ายในการทดลองทั้งสิ้น 1,882.5 ล้านเหรียญสหรัฐ

ตัวอย่าง 4.11 ณ. สี่แยกไฟแดงแห่งหนึ่ง ทุก ๆ ครั้งที่เปิดไฟเขียวรถที่มุ่งไปทางเหนือจะสามารถเลี้ยวขวาได้เพียงคราวละ 3 คันเท่านั้นทั้งนี้เพราะมีรถเลนตรงและเลนเลี้ยวซ้ายผ่านตลอดเป็นจำนวนมาก (ดูภาพประกอบ) ถ้าทราบว่าโดยปกติแล้วประมาณ 30% ของรถที่วิ่งไปทางเหนือจะเป็นรถเลี้ยวขวา



ก. จงหาความน่าจะเป็นที่จะมีรถสามารถเลี้ยวขวาได้ถึง 4 คัน ถ้ามีรถเข้าคิวในเลนเลี้ยวขวาทั้งสิ้น 6 คัน

ข. ถ้าสัญญาณไฟแดงเป็นสัญญาณอัตโนมัติ กล่าวคือ ถ้ามีรถเข้าคิวรถเลี้ยวขวาเป็นจำนวนหนึ่งสัญญาณอัตโนมัติ จะห้ามรถทุกเลนและอนุญาตให้รถเลี้ยวขวาเหล่านี้ผ่านไปทั้งหมด จงคาดหมายจำนวนรถดังกล่าว

วิธีทำ ก. จากโจทย์พบว่า $p = .30$ $r = 3$

ดังนั้น ความน่าจะเป็นที่จะมีรถสามารถ (แย่ง) เลี้ยวขวาได้ถึง 4 คันเมื่อมีรถเข้าคิวเลี้ยวขวา 6 คัน

$$= Pr(XQ6)$$

$$= \sum_{x=4}^{6} Pr(X = x) = \sum_{x=4}^{6} \binom{x-1}{4-1} (.3) (.7)^{x-4}$$

$$= \binom{4-1}{4-1} (.3)^4 + \binom{5-1}{4-1} (.3)^4 (.7) + \binom{6-1}{4-1} (.4)^5 (.7)^2$$

$$= .0081 + 4(.0057) + 10.(.00397)$$

$$= .0706$$

นั่นคือ ถ้ามีรถเข้าคิวเลี้ยวขวาอยู่ 6 คัน และเลนนี้สามารถปล่อยให้รถเลี้ยวได้คราวละ 3 คันเท่านั้น แล้วโอกาสที่จะมีรถสามารถเลี้ยวขวาได้ 4 คัน มิได้เพียงประมาณ 7%

ข. เนื่องจาก $E(X) = \frac{r}{p}$

$=> \quad E(X) = \frac{3}{.3} = 10$

นั่นคือคาดว่า ถ้ามีรถเข้าคิวเลี้ยวขวาได้ถึง 10 คันเมื่อใด สัญญาณจราจรจะห้ามรถเลนอื่นทั้งหมดและปล่อยรถเลนเลี้ยวขวานี้ให้ผ่านไปได้ทั้งหมด

ตัวอย่าง 4.12 ในการวางแผนสร้างเขื่อนหรือสร้างตึกสูง ๆ วิศวกรจำเป็นต้องคำนึงถึงปรากฏ-การณ์ธรรมชาติที่อาจเกิดขึ้นได้แม้จะไม่บ่อยนัก เช่นปริมาณน้ำเหนือเขื่อนที่ไหลบ่าจนท่วมสันเขื่อนหรือลมพายุใต้ฝุ่นขนาดหนักที่สามารถพัดให้ตึกโค่นล้มได้ ซึ่งเกี่ยวกับเรื่องนี้วิศวกรจำเป็นต้องศึกษาสถิติของการเกิดปรากฏการณ์ดังกล่าวและร้อยละหรือด้วยความน่าจะเป็นของเหตุการณ์เช่นนั้น แล้วสร้างเงื่อนไขในการสร้างตึกหรือเขื่อนให้สามารถทนทานแรงกดดันได้สูงกว่าแรงกดดันสูงสุดที่เคยเกิดขึ้นมาแล้วในอดีต แต่การก่อสร้างในลักษณะดังกล่าวแม้ว่าจะดีและควรสนับสนุน แต่ก็สิ้นเปลืองวัสดุอุปกรณ์โครงสร้างมาก โดยปกติวิศวกรมักจะพิจารณาความน่าจะเป็น p ที่จะเกิดปรากฏการณ์เหล่านั้นเสียก่อน ถ้าค่า p สูงมาก แผนก่อสร้างก็ต้องรัดกุมมากและตึกหรือเขื่อนก็ต้องรัดกุมและตึกหรือเขื่อนก็ต้องแข็งแรงทนทานต่อต้านแรงกดดันดังกล่าวได้มากครั้ง ถ้า p มีค่าต่ำก็ไม่จำเป็นต้องสร้างให้แข็งแรงทนทานนานปีมากนัก เพราะรูปทรงของสิ่งก่อสร้างตลอดจนเทคโนโลยีที่ใช้จะล้าสมัยเกินไป

ในการสร้างตึกสูงขนาด 100 ชั้นซึ่งจะต้องปะทะกับลมเบื้องบนที่แรงมากกว่าธรรมดานั้นวิศวกรคำนวณแล้วพบว่า ตึกจะสามารถทนทานแรงลมพายุที่มีความเร็วได้เพียง 1,000 ไมล์ต่อชั่วโมง และปรากฏการณ์เช่นนี้จะมีโอกาสเกิดขึ้นเพียง 1% เท่านั้น (ถือว่า $p = 0.01$ คงที่ตลอดไปในทุกปี)

ก. จงหาความน่าจะเป็นที่ตึกนี้จะมีความทนทานมากกว่า 10 ปี

ข. จงคาดหมายว่าเมื่อไรจึงจะเกิดพายุดังกล่าวอีกครั้งหนึ่ง ถ้าพายุผ่านไปเมื่อเร็ว ๆ นี้

ค. ถ้าวิศวกรออกแบบให้ตึกสามารถทนทานแรงกดดันของลมขนาดความเร็ว 1,000 ไมล์ต่อชั่วโมงได้เพียง 2 ครั้งเท่านั้น จงคาดหมายว่าตึกดังกล่าวจะพังทะลายเพราะพายุเช่นนี้ในอีกกี่ปีข้างหน้า

วิธีทำ ให้ X = จำนวนปีทั้งหมดก่อนที่จะเกิดพายุขนาดความเร็วลม 1,000 ไมล์ต่อชั่วโมง (นับรวมปีที่เกิดพายุด้วย)

$p = .01$ = ความน่าจะเป็นที่จะเกิดพายุขนาดความเร็ว 1,000 ไมล์ต่อชั่วโมง

ก. Pr (ตึกมีความทนทานมากกว่า 10 ปี) $= \Pr(X > 10) = 1 - \Pr(X \leqslant 10)$

$$= 1 - \sum_{x=1}^{10} pq^{x-1}$$

$$= 0.92$$

โอกาสนั่นคือตึกแห่งนี้จะสามารถทนทานเกินกว่า 10 ปี (หรือโอกาสที่ยังไม่เกิดพายุในช่วง 1-10 ปี) มีค่าสูงถึง 92%

ข. พายุจะเกิดอีกครั้ง (Average Return Period) เมื่อไร

$$E(X) = \frac{1}{p} = \frac{1}{0.01} = 100 \text{ ปี}$$

นั่นคือคาดว่าพายุขนาดความเร็วลม 1,000 ไมล์ต่อชั่วโมงจะเกิดขึ้นอีกครั้งใน 100 ปีข้างหน้า

ค. ตึกสามารถทนทานพายุขนาดความเร็วลม 1,000 ไมล์ต่อชั่วโมงได้เพียง 2 ครั้งเท่านั้น แสดงว่า ถ้าโดนพายุขนาดความเร็วลมนี้อีกในครั้งที่ 3 จะพังทะลายทันที

นั่นคือ $r = 3$

ดังนั้น $E(X) = \frac{r}{p} = \frac{3}{0.01} = 300$ ปี

นั่นคือตึกจะพังทะลายลงเพราะพายุเช่นนี้ในอีก 300 ปีข้างหน้า

4.2.5 การแจกแจงแบบพัวซอง (Poisson Distribution)

การแจกแจงแบบพัวซองเป็นเรื่องของการศึกษาถึงสถานการณ์ของการเกิดปรากฏการณ์ที่อุบัติขึ้นโดยสุ่ม[1] (Random occurence) ส่วนใหญ่จะเป็นอุบัติการณ์ในช่วงของเวลา (Time Interval) มากกว่าอย่างอื่น เช่นปรากฏการณ์ที่เกิดขึ้นภายในช่วงเวลาใดเวลาหนึ่ง (Fixed Time Interval) แต่ช่วงหรือขอบเขตของการเฝ้าติดตามนับ (Count) จำนวนการปรากฏการณ์ที่อุบัติขึ้นนี้มิได้หมายถึงเฉพาะช่วงหรือขอบเขตของเวลาเท่านั้น หากรวมไปถึงการเฝ้าติดตามนับจำนวนปรากฏการณ์ที่อุบัติขึ้นในขอบเขตของพื้นที่ ขอบเขตของปริมาตร ระยะทาง จุดนัดพบหรืออื่น ๆ อีกด้วย เช่น จำนวนครั้งของโทรศัพท์ที่หมุนเข้ามายังโอเปเรเตอร์ในช่วงเวลาหนึ่ง (โดยเฉพาะช่วงเวลาที่มีผู้ใช้โทรศัพท์มาก)

จำนวนอีเลคตรอนที่พุ่งออกมาจากขั้วลบในหลอดสูญญากาศ จำนวนดาวในบริเวณหนึ่ง (ปริมาตร) ของทางช้างเผือกจำนวนเซลเม็ดโลหิตที่สามารถเห็นได้จากกล้องจุลทัศน์ (จำนวนที่มองเห็นได้หมายถึงจำนวนที่ปรากฏในพื้นที่หน่วยหนึ่งที่กว้างยาวเท่ากับความกว้างยาวของเลนส์) จำนวนอนุภาค α ที่พุ่งออกมาจากแหล่งสารกัมมันตภาพรังสีในช่วงเวลาหนึ่ง จำนวนตำหนิ "ตามด" บนแผ่นไม้อัดขนาดพื้นที่หนึ่ง จำนวนปุ่มขรุขระบนเส้นลวดขนาดความยาวหนึ่ง จำนวนอุบัติเหตุบนท้องถนนสายหนึ่งในสัปดาห์หนึ่ง จำนวนอุกาบาต และสะเก็ดดาวที่หล่นลงมาเกาะบนผิวของดาวเทียมที่โคจรครบ 1 รอบ จำนวนจุลินทรีย์ในน้ำปริมาตรหนึ่ง จำนวนรอยร้าวบนผิวของแผ่นกระเบื้องขนาดพื้นที่หนึ่ง จำนวนยวดยานที่วิ่งผ่านจุดสังเกตุการณ์หนึ่ง จำนวนยวดยานที่จอดคอยสัญญาณไฟเขียว ณ สี่แยกแห่งหนึ่ง จำนวนพายุดีเปรสชั่นที่พัดผ่านตำบลหนึ่งในแต่ละปี จำนวนบุคคลที่เข้ามาติดต่อสมัครงาน ณ สำนักจัดหางาน จำนวนเรือที่เข้าเทียบท่าในแต่ละวัน ฯลฯ

จากตัวอย่างสถานะการณ์ข้างต้น นักศึกษาคงพอเข้าใจได้ว่า สถานะการณ์ที่พึงสงเคราะห์เข้ากับการกระจายแบบพัวซองนั้น

---

[1] ตัวแปรสุ่มที่ศึกษาถึงสถานการณ์ของการเกิดปรากฏการณ์ที่อุบัติขึ้นโดยสุ่มนอกเหนือจากพัวซองคือ ตัวแปรสุ่มเอกโพเนนเชียลและตัวแปรสุ่มแกมม่า โดยตัวแปรสุ่มเอกโพเนนเชียลจะศึกษาถึงระยะเวลาทั้งสิ้นจนกระทั่งเกิดอุบัติการณ์ เช่นเวลาที่รอคอยจนฟิวส์ตัวใหม่จะขาด ส่วนตัวแปรสุ่มแกมม่าศึกษาถึงระยะเวลาทั้งสิ้นที่จะต้องใช้จนกระทั่งเกิดอุบัติการณ์ครบ k ครั้ง ($k \geq 2$) ครั้ง

ตัวแปรสุ่มทั้งสองนี้เป็นตัวแปรสุ่มแสดงการรอคอย (Contineous Waiting Time) จนกระทั่งเกิดอุบัติการณ์ มีลักษณะของสถานการณ์เช่นเดียวกับตัวแปรสุ่มอนุกรมเรขาคณิตและตัวแปรสุ่มนิเสธทวินามในแง่ของการ "รอคอย" และมีลักษณะของสถานการณ์เช่นเดียวกันกับตัวแปรสุ่มพัวซองในแง่ของการเกิดเหตุการณ์ที่อุบัติโดยสุ่ม (Random Occurrence)

(1) ต้องเป็นสถานะการณ์ที่สามารถนับจำนวนปรากฏการณ์ได้ (Counting Event)

(2) อัตราเฉลี่ยของการเกิดปรากฏการณ์ (Average Rate of Occurence) จะต้องคงที่และ

(3) จะต้องกำหนดช่วงหรือเขตของการศึกษาให้แน่นอน อาจเป็นช่วงเวลา ระยะทาง จุด พื้นที่ ปริมาตรที่แน่นอน หรืออื่นใดก็ได้

การศึกษาเพื่อพัฒนาฟังก์ชั่นของความน่าจะเป็นของตัวแปรสุ่มพัวซองจำเป็นจะต้องมีข้อตกลงดังต่อไปนี้ และเพื่อความสะดวกจะขอเริ่มศึกษาด้วยขอบเขตของ "ช่วงเวลา" ส่วนช่วงหรือเขตอื่นก็สามารถสรุปผลได้โดยนัยเดียวกัน

1. จำนวนปรากฏการณ์ที่อุบัติขึ้นในช่วงเวลาคนละช่วง (Nonoverlapping Time Interval) เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน

2. ความน่าจะเป็นที่จะเกิดปรากฏการณ์เพียง 1 ครั้งในช่วงเวลา h สั้น ๆ[1] มีค่าเท่ากับ mh โดยที่ m[2] เป็นจำนวนเต็มบวกใด ๆ นัยหนึ่ง ความน่าจะเป็นที่จะเกิดปรากฏการณ์เพียง 1 ครั้งในช่วงเวลา h สั้น ๆ มีค่าเป็นสัดส่วนกับความยาวของช่วงเวลา h นั่นคือ $P_1(h) = mh$[3]

3. ความน่าจะเป็นที่จะเกิดปรากฏการณ์เกินกว่า 1 ครั้งในช่วงเวลา h สั้น ๆ มีค่าน้อยมาก (เมื่อเทียบกับ mh) จนเราสามารถตัดออกจากการพิจารณาได้ นั่นคือ $\sum_{k=2}^{n} P_k(h) \cong 0$

4. ความน่าจะเป็นที่จะไม่มีปรากฏการณ์ใดเกิดขึ้นเมื่อเริ่มต้นศึกษามีค่าเท่ากับ 1 นั่นคือ $P_0(0) = 1$

เราจะนำข้อตกลง เหล่านี้ไปใช้พัฒนาทฤษฎีของการแจกแจงแบบพัวซองดังต่อไปนี้

ทฤษฎี 4.7 ถ้า "ข้อตกลง" ทั้ง 4 ประการข้างต้นเป็นจริงจำนวนปรากฏการณ์ที่อุบัติขึ้นในช่วงเวลา t จะมีการกระจายแบบพัวซอง มีพารามิเตอร์เป็น $\lambda = mt$[4]

---

[1] แทนที่จะจะใช้ h เราอาจใช้ Δt ก็ได้ ในที่นี้ต้องการจะใช้ h เพื่อป้องกันความ "งง" ของนักศึกษาโดยเฉพาะในเรื่องของการตีความหมาย

[2] m คืออัตราเฉลี่ยของการเกิดปรากฏการณ์ (Mean Rate of Occurence) หมายถึงอัตราเฉลี่ยที่แสดงถึงจำนวนปรากฏการณ์ที่อุบัติขึ้นใน 1 หน่วยของเวลา (ช่วงเวลา) เช่นเรือเข้าเทียบท่าเฉลี่ยวันละ m = 5 ลำ อุบัติเหตุรถยนต์ชนกันในชั่วโมงคับขันชั่วโมงละ m = 7 คัน

[3] อ่านว่า ความน่าจะเป็นที่จะเกิดปรากฏการณ์ 1 ครั้งในช่วงแห่งระยะเวลา h มีค่าเท่ากับ mh

[4] $\lambda = mt$ แสดงว่า λ คืออัตราถัวเฉลี่ยของการเกิดปรากฏการณ์ในช่วงเวลายาว t หน่วยทั้งนี้เพราะ m คืออัตราถัวเฉลี่ยของการเกิดปรากฏการณ์ในช่วงเวลาสั้น ๆ (h) เช่น จำนวนอนุภาค α ที่หนีจากที่เก็บวินาทีละ 3 อนุภาค แสดงว่า m = 3 ถ้าเราทำการบันทึกข้อมูลคราวละ 10 วินาที แสดงว่า t = 10 ดังนั้น $\lambda = mt = 30$ = อัตราถัวเฉลี่ยของจำนวนอนุภาค α ที่หนีออกจากที่เก็บในทุก ๆ 10 วินาที

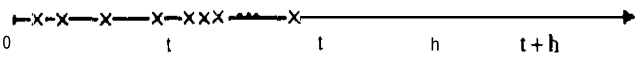
นั่นคือ ถ้าให้ตัวแปรสุ่ม X แทนจำนวนปรากฏการณ์ที่อุบัติขึ้นในช่วงเวลา t แล้ว

$$Pr(X = x) = f_X(x) \quad = \quad \frac{e^{-mt}(mt)^x}{x!} \; ; x = 0, 1, 2, \ldots$$

พิสูจน์ วิธีที่ 1

ให้ t เป็นกำหนดเวลาหลังเวลาเริ่มต้นที่ทำให้ช่วงเวลา (0, t] มีความยาว t หน่วย และช่วงเวลา (t, t+h) มีความยาว h หน่วย ดังภาพ

ให้ $P_n(t)$ เป็นฟังก์ชั่นแทนความน่าจะเป็นที่จะมีปรากฏการณ์ที่สนใจอุบัติขึ้น n ครั้ง ในช่วงเวลาที่มีความยาว t หน่วย



เครื่องหมาย × แสดงจำนวนปรากฏการณ์ที่อุบัติในช่วงแห่งเวลาที่มีความยาว t หน่วย

การพิสูจน์จะดำเนินการหาความน่าจะเป็นที่จะเกิดปรากฏการณ์ในช่วงเวลา (0, t+h) ตั้งแต่ 0, 1, 2, ... ครั้งเรื่อยไปตามลำดับดังนี้

$P_0(t+h)$ = Pr (ไม่มีปรากฏการณ์ใดอุบัติขึ้นในช่วงเวลา (0, t+h])

= Pr (ไม่มีปรากฏการณ์ใดอุบัติอุบัติขึ้นในช่วงเวลา (0, t] หรือไม่มีปรากฏการณ์อุบัติขึ้นในช่วงเวลา (t, t+h])

= Pr (ไม่มีปรากฏการณ์ใดอุบัติขึ้นในช่วงเวลา (t, t+h]) + Pr (ไม่มีปรากฏการณ์ใดอุบัติขึ้นในช่วงเวลา (t, t+h])

นั่นคือ $P_0(t+h) = P_o(t)P_o(h)$: ข้อตกลงที่ 1

แต่ $P_o(h)$ = 1 − Pr (มีปรากฏการณ์อุบัติขึ้นในช่วงเวลา (t, t+h] อย่างน้อย 1 ครั้ง)

= 1 − {Pr (มีปรากฏการณ์อุบัติขึ้น 1 ครั้งในช่วงเวลา (t, t+h]) + Pr (มีปรากฏการณ์อุบัติขึ้นตั้งแต่ 2 ครั้งขึ้นไปในช่วงเวลา (t, t+h])}

$\cong 1 - P_1(h) - \sum_{k=2}^{\infty} p_k(h)$

นั่นคือ $P_o(h) \cong 1 - mh$: ข้อตกลงที่ 2 และที่ 3

ดังนั้น $P_o(t + h) \cong P_o(t)(1 - mh)$

$\cong P_o(t) - mhP_o(t)$

$$=> \frac{P_o(t + h) - P_o(t)}{h} \cong -mP_o(t)$$

$$\Rightarrow \lim_{h\to 0} \frac{P_o(t+h) - P_o(t)}{h} = P_o'(t) \quad \because \lim_{x\to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = f'(x)$$

$$\Rightarrow \quad P_o'(t) = -mP_o(t)$$

$$\frac{P_o'(t)}{P_o(t)} = -m$$

$$\Rightarrow \quad \frac{d}{dt}\ln P_o(t) = -m \; : \; \frac{d}{dx}\ln f(x) = \frac{1}{f(x)}\frac{d}{dx}f(x) = \frac{f'(x)}{f(x)}$$

อินทรีเกรตทั้งสองด้านเทียบต่อ t

$$\Rightarrow \quad \int \frac{d}{dt}\ln P_o(t) = \int -m\,dt$$

$$\Rightarrow \quad \ln P_o(t) = -mt + c$$

ถ้าให้ $t = 0 \rightarrow C = 0$: ข้อตกลงที่ 4 ; $P_o(0) = 1, \ln 1 = 0, -m.0 = 0$

นั่นคือ $\ln P_o(t) = -mt$

นั่นคือ $\ln P_o(t) = -mt$

ดังนั้น $P_o(t) = e^{-mt}$ หรือนัยหนึ่งความน่าจะเป็นที่จะไม่เกิดปรากฏการณ์ใดในช่วงเวลายาว t หน่วย เมื่ออัตราเฉลี่ยของการเกิดปรากฏการณ์ใด ๆ ในช่วงเวลาความยาวนี้ (t หน่วย) มีค่าเท่ากับ mt

พิจารณา

$P_1(t+h)$ = Pr (เกิดปรากฏการณ์ 1 ครั้งในช่วงเวลา (0, t+h])

= Pr {เกิดปรากฏการณ์ 1 ครั้งในช่วงเวลา (0, t] แต่ไม่เกิดขึ้นในช่วงเวลา (t, t+h] หรือไม่เกิดปรากฏการณ์ใดขึ้นในช่วงเวลา (0, t] แต่ไปเกิดขึ้น 1 ครั้งในช่วงเวลา (t, t+h]}

$= P_1(t)\,P_o(h) + P_o(t)\,P_1(h)$

$= P_1(t)\,(1 - mh) + P_o(t)\,mh$: ข้อตกลงที่ (2) $P_1(h) = mh$ และสมการ (1)

$= P_1(t) + mh\,(-P_1(t) + P_o(t))$

$\Rightarrow P_1(t+h) = P_1(t) + mh(-P_1(t) + P_o(t))$.

$$\Rightarrow \frac{P_1(t+h) - P_1(t)}{h} = m(-P_1(t) + P_o(t))$$