บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

งานวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อหาค่าจำนวนเต็มบวกซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงที่ทำให้การแจกแจงหนึ่งลู่เข้าสู่การแจกแจงรูปแบบอื่น โดยผู้วิจัยจะทำการศึกษาการประมาณการแจกแจงที่สำคัญ 3 กรณีด้วยกัน ดังนี้ กรณีที่หนึ่ง ได้แก่ การประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงทวินาม, กรณีที่สอง ได้แก่ การประมาณการแจกแจงทวินามด้วยการแจกแจงปัวส์ซอง และกรณีสุดท้าย ได้แก่ การประมาณการแจกแจงทวินาม, การแจกแจงปัวส์ซอง, และการแจกแจงไฮเปอร์จีออเมตริก ด้วยการแจกแจงปกติ สำหรับการประมาณการแจกแจงที่สนใจศึกษาในการวิจัยครั้งนี้ มีหลักเกณฑ์ในการพิจารณาแบ่งเป็น 2 ลักษณะ ตามประเภทของการแจกแจงที่แท้จริงกับการแจกแจงที่ใช้ประมาณ ได้แก่ เกณฑ์พิจารณาในกรณีการประมาณการแจกแจงแบบไม่ต่อเนื่องด้วยการแจกแจงแบบไม่ต่อเนื่อง เกณฑ์พิจารณานี้ใช้สำหรับการประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงทวินาม และการประมาณการแจกแจงทวินามด้วยการแจกแจงปัวส์ซอง และสำหรับเกณฑ์ที่สองเป็นเกณฑ์การพิจารณาในกรณีการประมาณการแจกแจงแบบไม่ต่อเนื่องด้วยการแจกแจงแบบต่อเนื่อง เกณฑ์นี้ใช้สำหรับการประมาณการแจกแจงทวินาม, การแจกแจงปัวส์ซอง, และการแจกแจงไฮเปอร์จีออเมตริก ด้วยการแจกแจงปกติ ซึ่งรายละเอียดและขั้นตอนการประมาณจะถูกกล่าวไว้ในบทที่ 3 สำหรับในบทนี้จะกล่าวถึงทฤษฎีต่าง ๆ ที่เกี่ยวข้องกับงานวิจัย ได้แก่ ค่าคาดหวังและความแปรปรวน ซึ่งเป็นค่าที่แสดงว่าข้อมูลส่วนใหญ่อยู่ที่ไหน และบอกว่าค่าตัวแปรสุ่มจะมีค่าผิดพลาดกำลังสองโดยเฉลี่ยจากค่าของข้อมูลส่วนใหญ่เท่ากับเท่าไร, โมเมนต์และการหาฟังก์ชันที่ให้โมเมนต์ ซึ่งทำให้การหาค่าเฉลี่ยและค่าความแปรปรวนง่ายขึ้น, รายละเอียดเกี่ยวกับการแจกแจงต่าง ๆ ที่ถูกกล่าวถึงในงานวิจัยนี้ ได้แก่ การแจกแจงแบร์นูลลี, การแจกแจงทวินาม, การแจกแจงปัวส์ซอง, การแจกแจงไฮเปอร์จีออเมตริก, และการแจกแจงปกติ, รายละเอียดเกี่ยวกับการประมาณการแจกแจง, และในส่วนสุดท้ายของบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับการทดสอบเทียบความกลมกลืนกันโดยใช้ไคสแควร์ ซึ่งมีรายละเอียดต่าง ๆ ดังนี้

2.1 ค่าคาดหวัง (Expectation)

ให้ X เป็นตัวแปรสุ่มที่ไม่ต่อเนื่องซึ่งมีฟังก์ชันความน่าจะเป็น f และให้ $C = \{x_1, x_2, \ldots\}$ เป็นเซตที่นับถ้วนหรือนับได้ซึ่งทำให้ $P(X \in C) = 1$ ดังนั้นเราอาจกำหนดค่าคาดหวังเป็นค่าดังนี้

$$E(X) = \sum_{x \in C} x.f(x)$$

โดยที่ค่าทางขวามือของสมการมีค่าลู่เข้าสัมบูรณ์ (Converges absolutely)

ในทำนองเดียวกัน ให้ X เป็นตัวแปรสุ่มที่ต่อเนื่องซึ่งมีฟังก์ชันความหนาแน่น f จะได้ค่าคาดหวังของ X เป็นดังนี้

$$E(X) = \int_{-\infty}^{\infty} x.f(x)dx$$

โดยที่ค่าทางขวามือของสมการมีค่าลู่เข้าสัมบรูณ์ (Converges absolutely)

2.2 ความแปรปรวน (Variance)

ความแปรปรวนเป็นการอธิบายลักษณะของข้อมูลที่เกี่ยวกับประชากรเราอาจจะต้องทราบว่าข้อมูลมีการกระจัดกระจายมากหรือน้อยเพียงไรจากค่าเฉลี่ย

ให้ X เป็นตัวแปรสุ่ม เราเรียก σ_x^2 ว่าเป็นความแปรปรวนของ X ถ้า

$$\sigma_x^2 = E\left[(X-\mu)^2\right] \qquad \text{โดยที่} \quad \mu = E(X)$$

- ถ้า X เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$\sigma_x^2 = \sum_x (x-\mu)^2.f(x)$$

- ถ้า X เป็นตัวแปรสุ่มต่อเนื่อง

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu)^2.f(x)\,dx$$

2.3 ไมเมนต์ (Moment)

2.3.1 ไมเมนต์รอบจุดศูนย์กลาง (Moment about the origin)

ให้ X เป็นตัวแปรสุ่ม เราเรียก ว่าเป็นไมเมนต์ที่ r รอบจุดศูนย์กลาง (The r^{th} moment about the origin) ถ้า

$$\mu_r' = E(X^r) = \sum_x x^r . f(x)$$

สำหรับ $r = 0,1,2,3,...$ เมื่อ X ไม่ต่อเนื่อง และ

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r . f(x) dx$$

เมื่อ X ต่อเนื่อง

สังเกตได้ว่า μ_1' เรียกว่า ค่าเฉลี่ยของการแจกแจงของ X (mean of the distribution) หรือ ค่าเฉลี่ยของ X ใช้สัญลักษณ์ว่า μ

2.3.2 ไมเมนต์รอบค่าเฉลี่ย (Moment about the mean)

ให้ X เป็นตัวแปรสุ่ม เราเรียก ไมเมนต์ที่ r รอบค่าเฉลี่ย (The r^{th} moment about the mean) ถ้า

$$\mu_r = E[(X-\mu)^r] = \sum_x (x-\mu)^r . f(x)$$

สำหรับ $r = 0,1,2,3,....$ เมื่อ X ไม่ต่อเนื่อง และ

$$\mu_r = E[(X-\mu)^r] = \int_{-\infty}^{\infty} (x-\mu)^r . f(x)\, dx$$

เมื่อ X ต่อเนื่อง

สังเกตได้ว่า μ_2 เรียกว่า ความแปรปรวนของการแจกแจงของ X (Variance of distribution) หรือ ความแปรปรวนของ X ใช้สัญลักษณ์ว่า σ^2, หรือ $V(X)$

2.4 <u>ฟังก์ชันที่ให้โมเมนต์</u> (Moment-Generating Function)

ฟังก์ชันที่ให้โมเมนต์จะช่วยในการหาโมเมนต์ต่าง ๆ รอบจุดศูนย์กลางได้ง่ายขึ้น

ให้ X เป็นตัวแปรสุ่ม เราเรียก $M_X(t)$ ว่าเป็น ฟังก์ชันที่ให้โมเมนต์ (moment generating function)

$$M_X(t) = E(e^{tX}) = \sum_x e^{tx}.f(x)$$

เมื่อ X ไม่ต่อเนื่อง และ

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}.f(x)\,dx$$

เมื่อ X ต่อเนื่อง

ให้ X เป็นตัวแปรสุ่มที่มีฟังก์ชันที่ให้โมเมนต์ซึ่งมีค่าจำกัดบนช่วงเปิดใด ๆ ที่คลุมศูนย์ กล่าวคือ $M_X(t)<\infty$, $t \in (-h,h)$ เมื่อ $h>0$ จะได้ว่า

$$\mu_r' = M_X^{(r)}(0) \qquad ; \quad r = 1,2,3,\ldots$$

โดยที่ $M_X^{(r)}(0)$ เป็นค่าอนุพันธ์อันดับที่ r ของฟังก์ชันที่ให้โมเมนต์ ณ จุดศูนย์ :

$$\begin{aligned} M_X^{(r)}(t) &= \frac{d^r}{dt^r} M_X(t) \\ &= \frac{d^r}{dt^r} E(e^{tX}) \\ &= E\left(\frac{d^r}{dt^r} e^{tX}\right) \\ &= E(X^r e^{tX}) \\ M_X^{(r)}(0) &= E(X^r) = \mu_r' \qquad ; \quad r = 1,2,3,\ldots \end{aligned}$$

2.5 การแจกแจงที่เกี่ยวข้อง

2.5.1 การแจกแจงแบร์นูลลี่ (Bernoulli Distribution)

การทดลองแบบแบร์นูลลี่เป็นการทดลองที่ในแต่ละครั้งของการทดลองจะประกอบด้วยผลที่เป็นไปได้เพียง 2 ลักษณะเท่านั้น ซึ่งมักจะเรียกว่า ความสำเร็จ (Sucess) กับ ความไม่สำเร็จ (Failure) โดยที่ความน่าจะเป็นของการเกิดความสำเร็จในการทดลองแต่ละครั้งเป็น p ความน่าจะเป็นของการเกิดความไม่สำเร็จเป็น $(1-p)$ ในแต่ละครั้งของการทดลองแบบแบร์นูลลี่ ตัวอย่างเช่น การโยนเหรียญ 1 อัน 1 ครั้ง ถ้ากำหนดให้เหรียญขึ้นหัวเป็นความสำเร็จ จะได้ว่าเหรียญขึ้นก้อยเป็นความไม่สำเร็จ เป็นต้น

ตัวแปรสุ่ม X เรียกว่า ตัวแปรสุ่มแบบแบร์นูลลี่ ถ้า

$$X = \begin{cases} 1, & \text{ความสำเร็จ (Succes)} \\ 0, & \text{ความล้มเหลว (Failure)} \end{cases}$$

โดยที่ $P(X=1) = p \ , \quad 0<p<1$

และ $P(X=0) = 1-p$

ซึ่งฟังก์ชันความน่าจะเป็นอยู่ในรูปแบบของ

$$P(X=x) = p^x.(1-p)^{1-x} \qquad ; \quad x=0,1$$

ฟังก์ชันที่ให้โมเมนต์ของการแจกแจงแบร์นูลลี่

(Moment-denerating function of the Bernulli Distribution)

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{1} e^{xt} p^x (1-p)^{1-x} \\ &= (1-p)+e^t p \\ &= 1-p+e^t p \\ &= 1+p(e^t-1) \qquad , \quad -\infty<t<\infty \end{aligned}$$

สามารถแสดงได้ว่า

$$E[X^r] = \mu_r'(X) = p \qquad ; \quad r=1,2,3,\ldots$$

ค่าคาดหวังของการแจกแจงแบร์นูลลี่

$$\mu = E[X] = p$$

ค่าความแปรปรวนของการแจกแจงแบร์นูลลี่

$$\sigma^2 = Var[X] = E\left[X^2\right]-[E[X]]^2 = pq$$

2.5.2 การแจกแจงทวินาม (Binomial Distribution : $B(n,p)$)

ในทางปฏิบัติเราอาจสนใจเกี่ยวกับจำนวนผลสำเร็จในการทดลอง n ครั้ง ที่เป็นอิสระซึ่งกันและกัน ถ้ากำหนดให้ตัวแปรสุ่ม X เป็นจำนวนครั้งของการเกิดความสำเร็จในการทดลองแบบแบร์นูลลี่ n ครั้ง เราเรียก X ว่าเป็นตัวแปรสุ่มทวินาม

การทดลองแบบทวินาม (Binomial Distribution) มีลักษณะดังนี้

1). เป็นการทดลองที่กระทำซ้ำ ๆ กัน n ครั้ง ($n>1$)

2). การทดลองทั้ง n ครั้ง เป็นอิสระซึ่งกันและกัน

3). ผลของการทดลองแต่ละครั้งจะเกิดได้เพียง 2 ลักษณะ คือ สำเร็จ (success) หรือ ล้มเหลว (failure)

4). ความน่าจะเป็นของการเกิดความสำเร็จในการทดลองแต่ละครั้งมีค่าคงที่ นั่นคือ

P (เกิดความสำเร็จในแต่ละครั้ง) $= p$ และ

P (ไม่เกิดความสำเร็จในแต่ละครั้ง) $= 1-p = q$

5). ตัวแปรสุ่ม X คือ จำนวนครั้งของความสำเร็จจากการทดลองทั้ง n ครั้ง

ฉะนั้น X จะมีฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad ; \quad x=0,1,2,\ldots,n$$

และ $0<p<1$

อธิบายฟังก์ชันได้ดังนี้ เมื่อจบการทดลองครั้งที่ n เราจะมีจำนวนผลสำเร็จ = x ครั้ง และ ความล้มเหลว = $n-x$ ครั้ง ซึ่งมีความน่าจะเป็น = $p^x(1-p)^{n-x}$ โดยใช้ความเป็นอิสระของการทดลองแบบแบร์นูลลี่ นอกจากนั้นเราสามารถเลือกตำแหน่งที่เกิดผลสำเร็จ x ครั้ง ด้วยจำนวนวิธี = $\binom{n}{x}$

ตัวอย่างเช่น การโยนเหรียญ 1 อัน 5 ครั้ง แต่ละครั้งผลลัพธ์อาจเป็นหัว หรือก้อย ถ้ากำหนดให้เหรียญขึ้นหัวเป็นความสำเร็จ จะได้ว่าเหรียญขึ้นก้อยเป็นความไม่สำเร็จ และเนื่องจากใช้เหรียญอันเดียวกันในการโยน ดังนั้น ความน่าจะเป็นของการขึ้นหัวย่อมคงที่ และการขึ้นหัวหรือก้อยในแต่ละครั้งไม่มีผลต่อการขึ้นหัวหรือก้อยในการโยนครั้งอื่น ดังนั้น การโยนเหรียญจึงเป็นการทดลองสุ่มตามลักษณะดังกล่าวข้างต้น เรียกว่า การทดลองสุ่มแบบทวินาม (binomial experiment) สิ่งที่ต้องการหาในการทดลองนี้ คือ ความน่าจะเป็นที่จะได้หัว x ครั้ง โดยที่ x อาจจะเป็น 0, 1, ... หรือ 5 ครั้งก็ได้

ฟังก์ชันที่ให้โมเมนต์ของการแจกแจงทวินาม

(Moment-generating function ของ Binomial Distribution)

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}(e^t p)^x(1-p)^{n-x} \\ &= \left[e^t p+(1-p)\right]^n \\ &= \left[1+e^t p-p\right]^n \\ &= \left[1+p\left[e^t-1\right]\right]^n, \qquad -\infty<t<\infty \end{aligned}$$

สามารถแสดงได้ว่า

$$E[X] = \mu_1'(X) = np$$

$$E\left[X^2\right] = \mu_2'(X) = np+n(n-1)p^2$$

$$E\left[X^3\right] = \mu_3'(X) = np+3n(n-1)p^2+n(n-1)(n-2)p^3$$

$$E\left[X^4\right] = \mu_4'(X) = np+7n(n-1)p^2+6n(n-1)(n-2)p^3+n(n-1)(n-2)(n-3)p^4$$

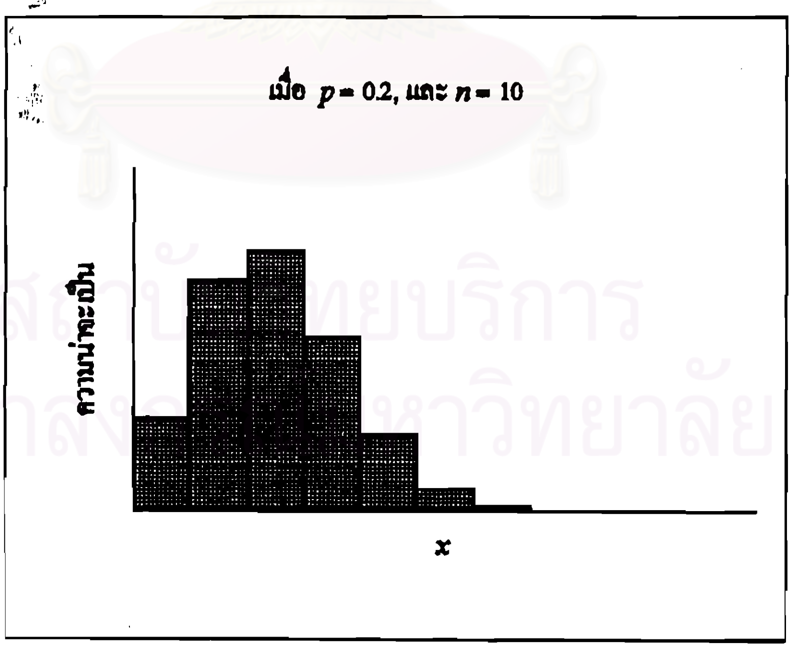
ค่าคาดหวังของการแจกแจงทวินาม

$$\mu = E[X] = np$$

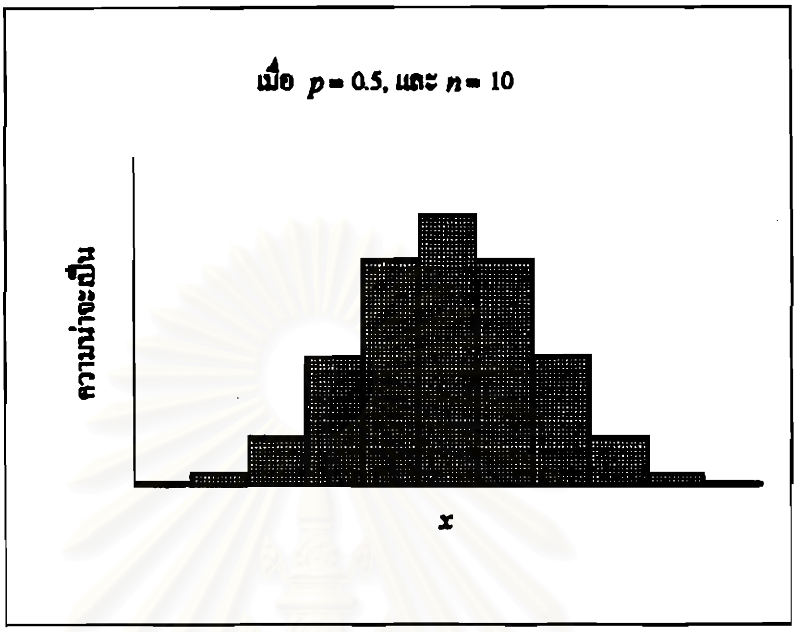
ค่าความแปรปรวนของการแจกแจงทวินาม

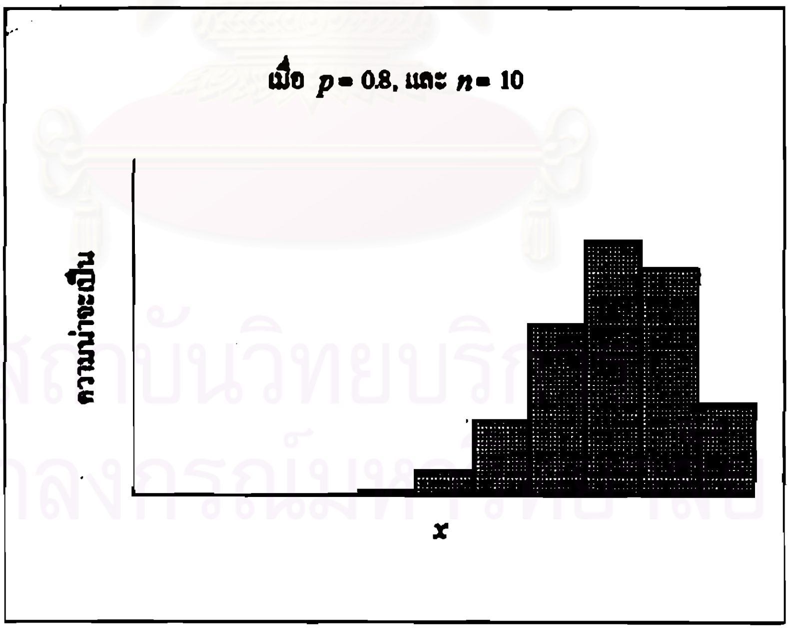
$$\sigma^2 = Var(X) = E[X^2] - [E[X]]^2 = npq$$

ฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามจะขึ้นกับค่าความน่าจะเป็น p และ ขนาดตัวอย่าง n ซึ่งเป็นพารามิเตอร์ของการแจกแจง กล่าวคือ ณ ค่าคงที่ n เมื่อค่าความน่าจะเป็น p มีค่าน้อยกว่า 0.5 กราฟแสดงฟังก์ชันความน่าจะเป็นจะมีลักษณะเบ้ขวา เมื่อค่าความน่าจะเป็น p เข้าใกล้ 0.5 กราฟจะมีลักษณะเข้าใกล้สมมาตร และ เมื่อค่าความน่าจะเป็น p มีค่ามากกว่า 0.5 กราฟแสดงฟังก์ชันความน่าจะเป็นจะมีลักษณะเบ้ซ้าย ดังแสดงในรูปที่ 2.1 อย่างไรก็ตามถึงแม้ว่า ค่าความน่าจะเป็นจะไม่ใช่ 0.5 แต่เมื่อเพิ่มขนาดตัวอย่าง n มากขึ้น กราฟแสดงลักษณะการแจกแจงก็จะเข้าใกล้สมมาตรมากขึ้น ดังจะกล่าวไว้ในหัวข้อการประมาณการแจกแจงทวินามด้วยการแจกแจงปกติ



<u>รูปที่ 2.1</u> แสดงฟังก์ชันความน่าจะเป็นของการแจกแจงทวินาม





<u>รูปที่ 2.1</u> (ต่อ)

2.5.3 <u>การแจกแจงปัวส์ซอง</u> (Poisson Distribution : $Poi(\lambda)$)

การแจกแจงความน่าจะเป็นไม่ต่อเนื่องที่สำคัญอีกแบบหนึ่งเกี่ยวข้องกับการทดลองสุ่มแบบปัวส์ซอง (Poisson experiment) ซึ่งเป็นการทดลองที่มีสิ่งที่สนใจเกิดขึ้นในช่วงเวลา หรือในขอบเขตที่กำหนดให้ ช่วงเวลาอาจเป็นวินาที นาที ชั่วโมง วัน หรือปีก็ได้ ขอบเขตอาจเป็นบนเส้นตรง พื้นที่ หรือปริมาตรก็ได้ จากการทดลองนี้เราสนใจจำนวนครั้งของความสำเร็จ ตัวอย่างเช่น จำนวนสัญญาณโทรศัพท์ในช่วง 1 นาที จำนวนคำผิดในหนังสือ 1 หน้า จำนวนผลิตภัณฑ์ชำรุดที่ผลิตต่อวัน เป็นต้น ถ้าให้ X แทนจำนวนครั้งของความสำเร็จ จะเรียก X เป็น ตัวแปรสุ่มแบบปัวส์ซอง

ถ้า X เป็นจำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลา หรือขอบเขตที่กำหนดให้ แล้วจำนวนครั้งของความสำเร็จจะประมาณได้ด้วยขบวนการปัวส์ซองที่มีพารามิเตอร์ $\lambda > 0$, ถ้ามีคุณสมบัติครบทั้ง 3 ข้อนี้

1. เราสามารถแบ่งช่วงเวลา หรือขอบเขตที่กำหนด ให้สั้นมาก ๆ ได้โดยมีความน่าจะเป็นของการเกิดความสำเร็จหนึ่ง ๆ ในช่วงเวลาหรือขอบเขต สั้น ๆ นั้น เป็นสัดส่วนโดยตรงกับช่วงเวลาหรือขอบเขตนั้น นั่นคือ ความน่าจะเป็นที่จะเกิดความสำเร็จหนึ่งครั้งในช่วงเวลาหรือขอบเขตสั้น ๆ h จะประมาณได้ด้วย λh

2. ความน่าจะเป็นที่จะเกิดความสำเร็จมากกว่าหนึ่งครั้งในช่วงเวลาหรือขอบเขตสั้น ๆ จะประมาณได้ด้วยศูนย์ (มีค่าน้อยมาก)

3. จำนวนครั้งของการเกิดความสำเร็จในช่วงเวลาหรือขอบเขต สั้น ๆ h จะไม่มีผลกระทบต่อการเกิดหรือไม่เกิดของความสำเร็จในช่วงเวลาหรือขอบเขตอื่น ๆ (เป็นอิสระต่อกัน)

ถ้า X เป็นจำนวนครั้งของความสำเร็จในการทดลองแบบปัวส์ซอง ซึ่งมีค่าเฉลี่ยในช่วงเวลาหรือขอบเขตที่กำหนดเป็น λ เราเรียก X ว่า เป็นตัวแปรสุ่มที่มีการแจกแจงปัวส์ซองด้วยพารามิเตอร์ λ และมีฟังก์ชันความน่าจะเป็น ดังนี้

$$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad ; \; x=0,1,2,\ldots$$

ฟังก์ชันที่ให้โมเมนต์ของการแจกแจงปัวส์ซอง

(Moment- generating function ของ Poisson distribution)

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} e^{xt}\lambda^x e^{-\lambda}/x! \\ &= e^{-\lambda}\sum_{x=0}^{\infty}\left(\lambda e^t\right)^x/x! \\ &= e^{-\lambda}.e^{\lambda e^t} \\ &= e^{\lambda(e^t-1)}, \qquad -\infty<t<\infty \end{aligned}$$

จาก Maclaurin's series

$$\sum_{x=0}^{\infty}\left(\lambda e^t\right)^x/x! = e^{\lambda e^t}$$

สามารถแสดงได้ว่า

$$\begin{aligned} E(X) &= \mu_1'(X) = \lambda \\ E(X^2) &= \mu_2'(X) = \lambda+\lambda^2 \\ E(X^3) &= \mu_3'(X) = \lambda+3\lambda^2+\lambda^3 \\ E(X^4) &= \mu_4'(X) = \lambda+7\lambda^2+6\lambda^3+\lambda^4 \end{aligned}$$

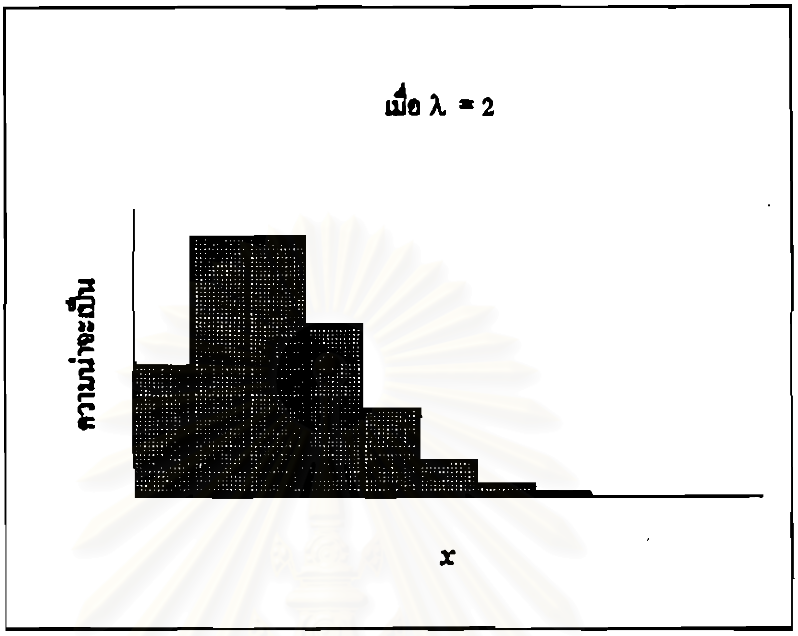
ค่าคาดหวังของการแจกแจงปัวส์ซอง

$$\mu = E[X] = \lambda$$

ค่าความแปรปรวนของการแจกแจงปัวส์ซอง

$$\sigma^2 = Var(X) = E[X^2]-[E[X]]^2 = \lambda$$

ฟังก์ชันความน่าจะเป็นของการแจกแจงปัวส์ซอง จะขึ้นกับค่าพารามิเตอร์ λ กราฟแสดงฟังก์ชันความน่าจะเป็นจะมีลักษณะเบ้ขวา ดังแสดงในรูปที่ 2.2 ทั้งนี้เมื่อค่า λ มีค่ามากขึ้น กราฟแสดงลักษณะการแจกแจงก็จะเข้าใกล้สมมาตรมากขึ้น ดังจะกล่าวไว้ในหัวข้อการประมาณการแจกแจงปัวส์ซองด้วยการแจกแจงปกติ



รูปที่ 2.2 แสดงฟังก์ชันความน่าจะเป็นของการแจกแจงปัวส์ซอง

2.5.4 การแจกแจงไฮเปอร์จีออเมตริก

(Hypergeometric Distribution : $H(N,M,n)$)

การแจกแจงทวินาม ขึ้นอยู่กับข้อสมมติเบื้องต้นที่ว่าการทดลองแต่ละครั้งจะต้องเป็นอิสระต่อกัน ซึ่งหมายความว่าความน่าจะเป็นที่จะเกิดความสำเร็จในแต่ละครั้งมีค่าคงที่ไม่เปลี่ยนแปลง นั่นคือ การทดลองในครั้งหลังจะต้องไม่มีผลอันเนื่องมาจากการทดลองครั้งที่ผ่านมา

สำหรับการแจกแจงแบบไฮเปอร์จีออเมตริกนั้น การทดลองแต่ละครั้งไม่เป็นอิสระกัน การทดลองในครั้งหลังขึ้นอยู่กับจำนวนครั้งของความสำเร็จในการทดลองที่ผ่านๆ มา กล่าวได้อีกอย่างหนึ่งในเทอมของการชักสิ่งตัวอย่าง ได้ว่า การทดลองแบบทวินามเป็นการสุ่มแบบคืนที่ แต่การทดลองแบบไฮเปอร์จีออเมตริกเป็นการสุ่มไม่คืนที่ ยกตัวอย่างการทดลองสุ่มแบบไฮเปอร์จีออเมตริก เช่น การสุ่มหยิบบอลจากถุงใบหนึ่งที่มีบอลสีแดง 3 ลูก และสีขาว 2 ลูก โดยที่กำหนดให้การสุ่มหยิบได้บอลสีแดงเป็นความสำเร็จ ถ้าทำการหยิบลูกบอล 2 ลูก

จากถุงแบบไม่คืนที่ กล่าวคือ เมื่อได้ลูกที่หนึ่งแล้วไม่ใส่กลับในถุง หยิบลูกที่สองที่เหลือ ความน่าจะเป็นที่จะได้ลูกที่สองเป็นสีแดงขึ้นอยู่กับลูกที่หนึ่งว่าเป็นสีแดงหรือสีขาว จะเห็นว่าการหยิบทั้งสองครั้งไม่เป็นอิสระกัน

ถ้า X แทนจำนวนสิ่งของพวกแรกในการสุ่มตัวอย่างขนาด n สิ่ง จากของทั้งหมด N สิ่ง ซึ่งประกอบด้วยของ 2 พวก พวกแรกมี M สิ่ง พวกที่สองมี $N-M$ สิ่ง แล้วฟังก์ชันความน่าจะเป็นของ X คือ

$$P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad ; \quad x = 0,1,2,\ldots,\text{Min}(n,M),$$

$$x \le M \quad \text{และ} \quad n-x \le N-M$$

สำหรับการแจกแจงแบบไฮเปอร์จีออเมตริกนั้นการหาฟังก์ชันที่ให้โมเมนต์จะมีรูปแบบที่ยุ่งยากมาก ดังนั้นการหาค่าคาดหวัง และค่าความแปรปรวนจะหาจากบทนิยามโดยตรง แสดงได้ดังต่อไปนี้

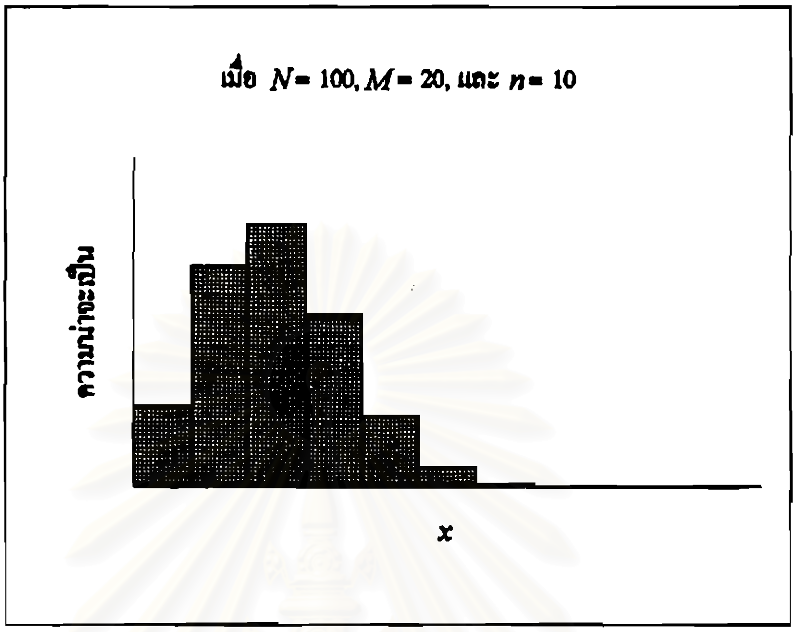
ค่าคาดหวังของการแจกแจงไฮเปอร์จีออเมตริก

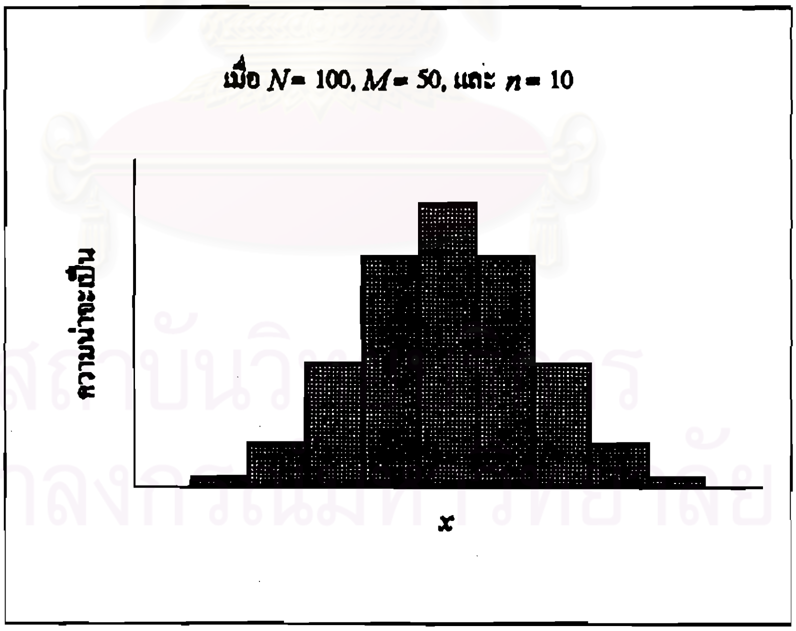
$$\begin{aligned}
\mu &= \sum_{x=0}^{n} x \frac{\binom{M}{x}\binom{M-N}{n-x}}{\binom{N}{n}} \\
&= \frac{M}{\binom{N}{n}} \sum_{x=1}^{n} \binom{M-1}{x-1}\binom{N-M}{n-x} \\
&= \frac{M}{\binom{N}{n}} \cdot \binom{N-1}{n-1} \\
&= \frac{nM}{N}
\end{aligned}$$

ค่าความแปรปรวนของการแจกแจงไฮเปอร์จีออเมตริก

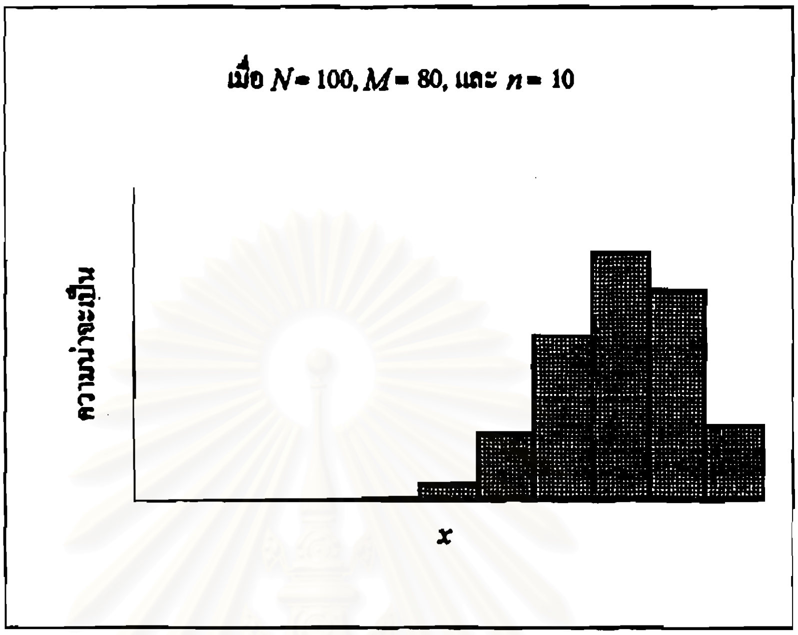
$$
\begin{aligned}
\sigma^2 &= E[X(X-1)]+E[X]-[E[X]]^2 \\
&= \frac{M(M-1)n(n-1)}{N(N-1)}+\frac{nM}{N}-\left(\frac{nM}{N}\right)^2 \\
&= \frac{nM(N-M)(N-n)}{N^2(N-1)} \\
&= \frac{n(N-n)}{(N-1)}\cdot\left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right)
\end{aligned}
$$

ฟังก์ชันความน่าจะเป็นของแจกแจงไฮเปอร์จีออเมตริกจะขึ้นกับขนาดของประชากรทั้งหมด N, ขนาดประชากรย่อยที่สนใจ M, และ ขนาดตัวอย่าง n ซึ่งเป็นพารามิเตอร์ของการแจกแจง กล่าวคือ เมื่ออัตราส่วนระหว่างขนาดประชากรย่อยที่สนใจกับขนาดประชากรทั้งหมด มีค่าน้อย กราฟแสดงฟังก์ชันความน่าจะเป็นจะมีลักษณะเบ้ขวา ในทางกลับกัน เมื่ออัตราส่วนระหว่างขนาดประชากรย่อยที่สนใจกับขนาดประชากรทั้งหมด มีค่ามาก กราฟแสดงฟังก์ชันความน่าจะเป็นจะมีลักษณะเบ้ซ้าย และ เมื่ออัตราส่วนดังกล่าว มีค่าเข้าใกล้ 0.5 กราฟแสดงฟังก์ชันความน่าจะเป็นจะเข้าใกล้สมมาตร ดังแสดงในรูปที่ 2.3 อย่างไรก็ตามถึงแม้ว่าอัตราส่วนระหว่างขนาดประชากรย่อยที่สนใจกับขนาดประชากรทั้งหมด จะไม่ใช่ 0.5 แต่เมื่อเพิ่มขนาดตัวอย่าง n มากขึ้น กราฟแสดงลักษณะการแจกแจงก็จะเข้าใกล้สมมาตรมากขึ้น ดังจะกล่าวไว้ในหัวข้อการประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงปกติ





รูปที่ 2.3 แสดงฟังก์ชันความน่าจะเป็นของการแจกแจงไฮเปอร์จีออเมตริก



รูปที่ 2.3 (ต่อ)

2.5.5 การแจกแจงแบบปกติ (Normal Distribution : $N(\mu,\sigma^2)$)

การแจกแจงความน่าจะเป็นแบบนี้นับว่าเป็นการแจกแจงที่สำคัญ และมีการใช้มากที่สุดในวิชาสถิติ ลักษณะที่เห็นเด่นชัด คือ มีลักษณะเป็นรูปโค้งสมมาตรคล้ายระฆัง (Bell-Shape) ซึ่งเรียกว่า โค้งปกติ (Normal Curve) ในปี 1733 เดอมัวร์ (De Mivre) ได้สร้างสมการของโค้งปกติซึ่งได้กลายมาเป็นรากฐานของทฤษฎีต่าง ๆ ในวิชาสถิติ และต่อมา เกาส์ (Gauss) ก็ได้สมการนี้จากการศึกษาเรื่องความผิดพลาดในการวัดปริมาณเดียวกันหลาย ๆ ครั้ง การแจกแจงปกติบางครั้งจึงถูกเรียกอีกชื่อว่า "การแจกแจงความน่าจะเป็นแบบเกาซ์" (Gaussian Distribution)

ตัวแปรสุ่ม X มีการแจกแจงความน่าจะเป็นแบบปกติ ด้วยค่าเฉลี่ย μ และค่าแปรปรวน σ^2 จะมีฟังก์ชันความหนาแน่นในรูป

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad ; -\infty < x < \infty$$

ฟังก์ชันที่ให้โมเมนต์ของการแจกแจงปกติ

(Moment generating function of the Normal Distribution)

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{xt}\cdot\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{-\infty}^{\infty} e^{-\frac{\left[-2xt\sigma^2+(x-\mu)^2\right]}{2\sigma^2}}\,dx \\
&= e^{\mu t+\frac{1}{2}t^2\sigma^2}\left\{\frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{-\infty}^{\infty} e^{-\frac{\left[x-(\mu+t\sigma^2)\right]^2}{2\sigma^2}}\,dx\right\} \\
&= e^{\mu t+\frac{1}{2}t^2\sigma^2}, \qquad -\infty < t < \infty
\end{aligned}$$

สามารถแสดงได้ว่า

$$E(X) = \mu$$

$$E(X^2) = \mu^2 + \sigma^2$$

ค่าคาดหวังของการแจกแจงปกติ

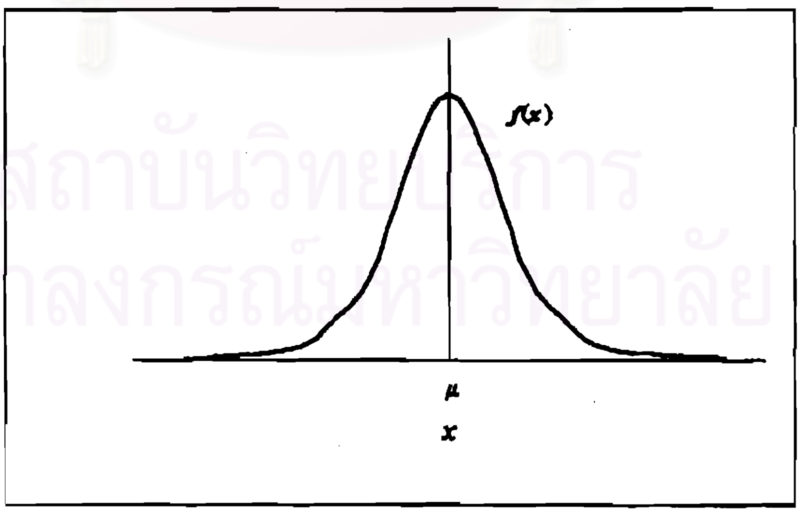
$$\mu = E[X]$$

ค่าความแปรปรวนของการแจกแจงปกติ

$$\sigma^2 = E[X^2]-[E[X]]^2$$

การแจกแจงปกติมีลักษณะและคุณสมบัติดังนี้

1. ลักษณะของโค้งเป็นรูประฆังคว่ำ (Bell Shaped)
2. เส้นแบ่งครึ่งโค้งอยู่ที่จุดที่เป็นค่าเฉลี่ยของข้อมูล และเส้นนี้ทำให้เป็นเส้นโค้งที่อยู่สองข้าง มีลักษณะสมมาตร (Symmetry)
3. จุดที่เป็นค่าเฉลี่ย มัธยฐาน และฐานนิยม เป็นจุดเดียวกันหรือมีค่าเท่ากัน
4. มีความโด่ง (kurtosis) ของเส้นโค้ง เท่ากับ 3 ซึ่งเรียกว่า เมโซเคอร์ติค (Mesokurtic) และจุดเปลี่ยนโค้งทั้งสองข้างจะอยู่ ณ ตรง 1 ส่วนเบี่ยงเบนมาตราฐาน
5. ค่าความเบ้ (Skewness) เท่ากับศูนย์
6. ปลายทั้งสองข้างของเส้นโค้งจะค่อย ๆ ลดต่ำลง แต่ไม่จรดกับฐานของโค้งหรือแกน



รูปที่ 2.4 แสดงฟังก์ชันความหนาแน่นของการแจกแจงปกติ

<u>การแจกแจงแบบปกติในรูปมาตรฐาน</u> (Standard Normal Distribution or The Unit Normal)

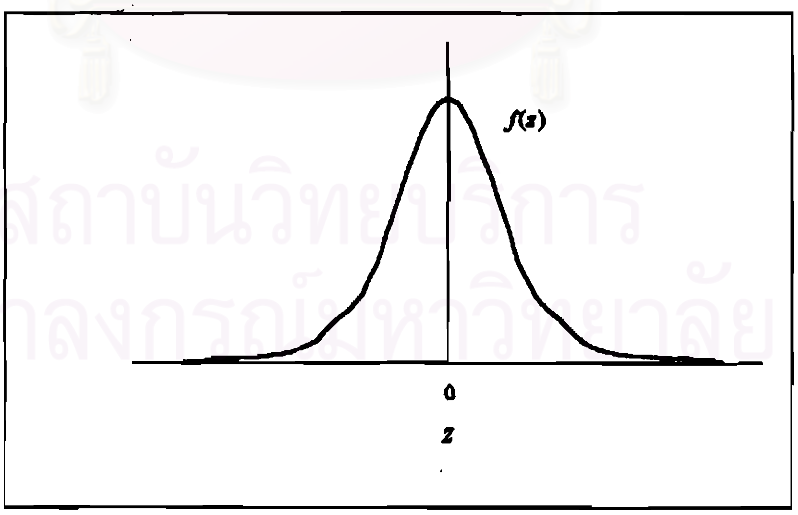
ถ้าการแจกแจงปกติมี $\mu = 0$ และ $\sigma = 1$ เราจะเรียกการแจกแจงนี้ว่าการแจกแจงปกติมาตรฐาน (Standard normal distribution) ใช้สัญลักษณ์ดังนี้ $N(0,1)$ ประโยชน์ของการแจกแจงปกติมาตรฐานจะใช้ในการเปิดตารางเพื่อหาพื้นที่ เพราะตารางสถิติทั่วไปจะแสดงค่าพื้นที่ของเส้นโค้งปกติมาตรฐานเท่านั้น วิธีการเปลี่ยนการแจกแจงปกติทั่วไปให้เป็นในรูปการแจกแจงปกติมาตรฐานสามารถกระทำได้โดยสูตรแปลงค่านี้

$$Z = \frac{X - \mu}{\sigma}$$

และฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติมาตรฐานคือ

$$f(z) = \frac{1}{\sqrt{2\pi}} . e^{-z^2/2} \quad ; \quad -\infty < z < \infty$$

และเนื่องจากการแจกแจงนี้ เป็นกรณีเฉพาะของการแจกแจงปกติ ดังนั้นคุณสมบัติต่าง ๆ ก็เป็นเช่นเดียวกัน



<u>รูปที่ 2.5</u> แสดงฟังก์ชันความหนาแน่นของการแจกแจงปกติมาตรฐาน

2.6 การประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยแจกแจงทวินาม

(Approximation of the Hypergeometric Distribution by the Binomial Distribution)

ถ้า n มีค่าน้อยเมื่อเทียบกับ N การสุ่มตัวอย่างแบบไม่ใส่คืนก็มีผลใกล้เคียงกับแบบใส่คืน ค่า $\frac{M}{N}$ ใกล้เคียงความน่าจะเป็น p และการทดลองแบบไฮเปอร์จีออเมตริกจะเข้าใกล้การแจกแจงแบบทวินาม เราจึงอาจประมาณค่าความน่าจะเป็นแบบไฮเปอร์จีออ-เมตริกด้วยแบบทวินามได้ โดยใช้พารามิเตอร์ n และ p ซึ่ง $p=\frac{M}{N}$

สมมติว่าขนาดตัวอย่างสุ่ม n เป็นแบบไม่แทนที่จากประชากร N หน่วย มีสัดส่วนของความสำเร็จ $= p$ และสัดส่วนของความไม่สำเร็จ $= (1-p)$ ให้ X เป็นจำนวนครั้งของความสำเร็จที่เกิดขึ้นในตัวอย่าง, M แทน จำนวนความสำเร็จของประชากร จะได้ว่า $M = Np$ และ $N-M = N(1-p)$

ดังนั้นฟังก์ชันความน่าจะเป็นของการแจกแจงไฮเปอร์จีออเมตริกจะเขียนได้ดังนี้

$$
\begin{aligned}
P(X=x) &= \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \\
&= \binom{n}{x}\frac{(M)_x(N-M)_{(n-x)}}{(N)_n} \\
&= \binom{n}{x}\frac{(Np)_x(N(1-p))_{(n-x)}}{(N)_n} \\
&= \binom{n}{x}\frac{(Np)_x(N(1-p))_{(n-x)}}{(N)_x(N-x)_{(n-x)}}
\end{aligned}
$$

ทั้งนี้

$$
\begin{aligned}
(N)_n &= [N(N-1)(N-2)\ldots(N-x+1)].[(N-x)\ldots(N-x-(n-x)+1)] \\
&= (\mathrm{N})_x(N-x)_{(n-x)}
\end{aligned}
$$

เมื่อ $N \to \infty$,

$$\frac{(Np)_x}{(N)_x} = \left[\frac{Np}{N}\right]\left[\frac{Np-1}{N-1}\right]..\left[\frac{Np-x+1}{N-x+1}\right]$$
$$= p\left(p-\frac{1-p}{N-1}\right)..\left(p-\frac{(x-1)(1-p)}{N-x+1}\right)$$
$$\to (p)(p)(p)...(p) = p^x$$

และ

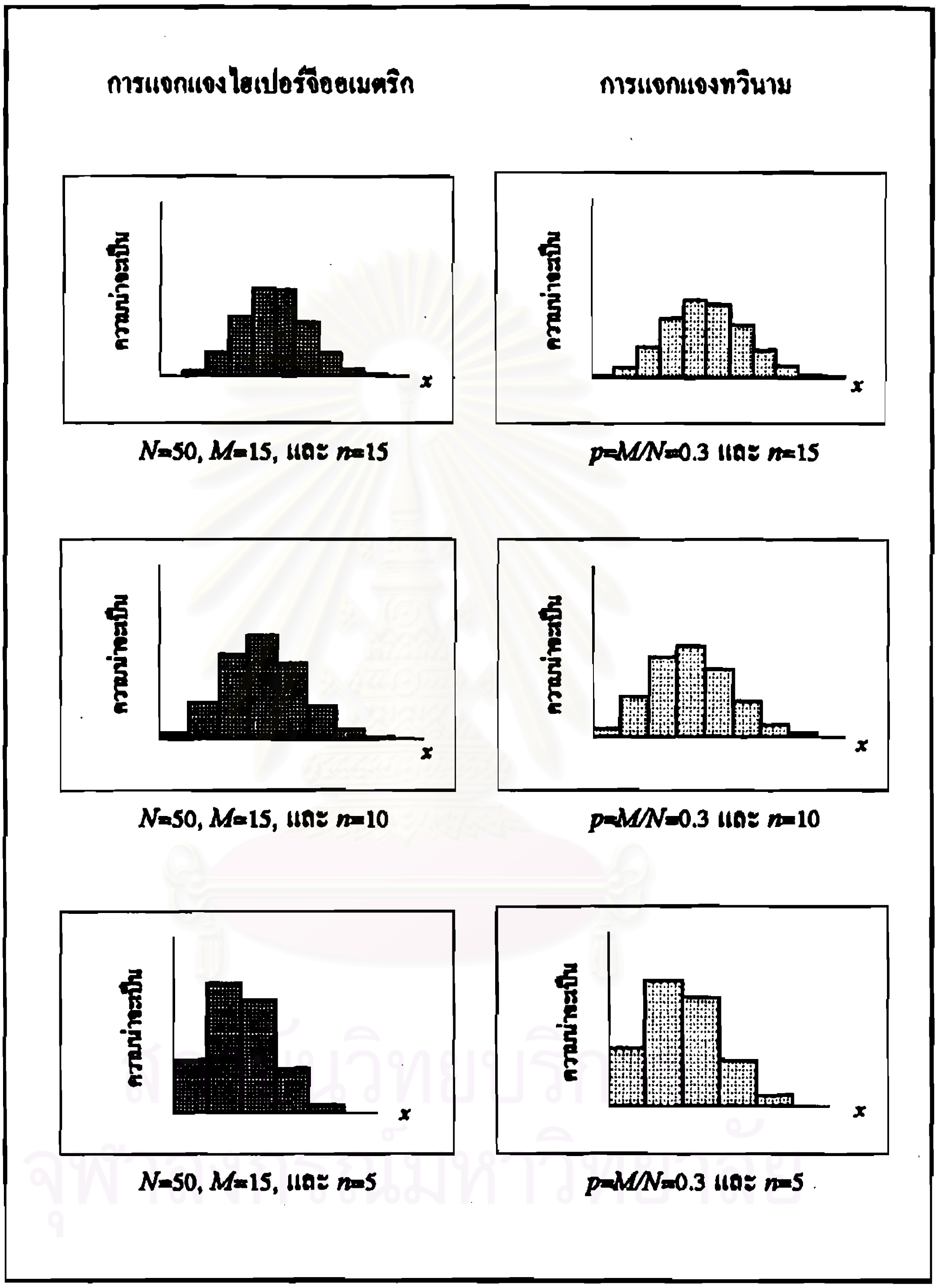
$$\frac{\left(N(1-p)\right)_{(n-x)}}{(N-x)_{(n-x)}} = \left[\frac{N(1-p)}{N-x}\right]\left[\frac{N(1-p)-1}{N-x-1}\right]..\left[\frac{N(1-p)-n+x+1}{N-n+1}\right]$$
$$= \left[(1-p)+\frac{x(1-p)}{N-x}\right]\left[(1-p)+\frac{x(1-p)-p}{N-x-1}\right]$$
$$..\left[(1-p)+\frac{x-p(n-1)}{N-n+1}\right]$$
$$\to (1-p)(1-p)...(1-p) = (1-p)^{n-x}$$

ดังนั้น

$$\lim_{\substack{N\to\infty \\ M\to Np}} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x(1-p)^{n-x} \qquad ; x = 0,1,2,3,\ldots,n$$

โดยทั่วไปแล้วผลลัพธ์ของการใช้ความน่าจะเป็นทวินามประมาณความน่าจะเป็นไฮเปอร์จีออเมตริกจะยอมรับว่ามีความถูกต้องเมื่อ $\frac{n}{N} \le 0.1$

การประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงทวินามนั้นขึ้นกับขนาดประชากรทั้งหมด (N) และ ขนาดตัวอย่าง (n) ซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงไฮเปอร์จีออเมตริก กล่าวคือ ณ ขนาดประชากร (N) ค่าหนึ่ง เมื่อขนาดตัวอย่าง (n) ลดลง การประมาณการแจกแจงดังกล่าวจะทำให้ความคลาดเคลื่อนระหว่างการแจกแจงไฮเปอร์จีออเมตริกกับการแจกแจงทวินาม (ε) มีค่าน้อยลง ดังแสดงในรูปที่ 2.6



<u>รูปที่ 2.6</u> แสดงการเปรียบเทียบฟังก์ชันความน่าจะเป็นของการแจกแจงไฮเปอร์ฯและทวินาม

2.7 การประมาณการแจกแจงทวินามด้วยการแจกแจงปัวส์ซอง

(Approximation of the Binomial Distribution by the Poisson Distribution)

การแจกแจงทวินาม เมื่อ $n \to \infty$ และ $p \to 0$ การหาค่าความน่าจะเป็นอาจมีปัญหาด้านการคำนวณ ดังนั้นในกรณีดังกล่าวเราจะทำการประมาณการแจกแจงทวินาม ด้วยการแจกแจงอื่น ๆ การแจกแจงที่ใช้ประมาณการแจกแจงทวินาม คือ การแจกแจงปัวส์ซอง กล่าวได้ว่าการแจกแจงปัวส์ซอง เป็นการแจกแจงของเหตุการณ์ที่เกิดขึ้นยาก การประมาณการแจกแจงความน่าจะเป็นทวินามด้วยการแจกแจงปัวส์ซองได้ โดยใช้พารามิเตอร์ $\lambda = np$

แสดงได้ดังนี้

$$p = \frac{\lambda}{n} \quad ;$$

$$P(X=x) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{1.\left(1-\frac{1}{n}\right).\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)}{x!}\,(\lambda)^x\left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-x}$$

เมื่อ $n \to \infty$;

$$1.\left(1-\frac{1}{n}\right).\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right) \to 1$$

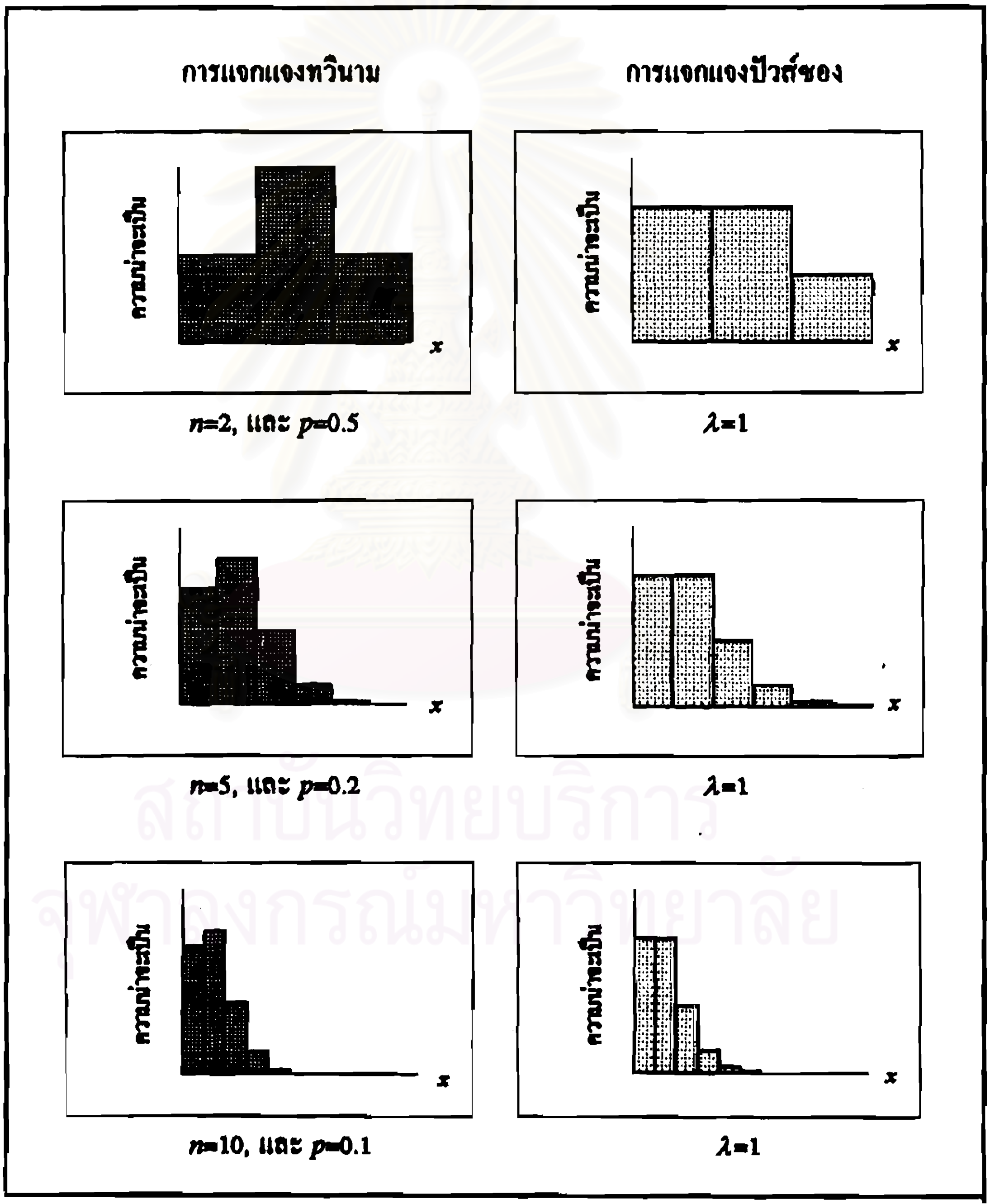
$$\left(1-\frac{\lambda}{n}\right)^{-x} \to 1$$

$$\left(1-\frac{\lambda}{n}\right)^{-n/\lambda} \to e$$

ดังนั้น

$$\lim_{\substack{n\to\infty \\ p\to 0}} \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!} \qquad ; \; x = 0,1,2,\ldots$$

การประมาณการแจกแจงทวินามด้วยการแจกแจงปัวส์ซองนั้น ขึ้นกับขนาดตัวอย่าง (n) ซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงทวินาม กล่าวคือ ณ λ ค่าหนึ่ง เมื่อเพิ่มขนาดตัวอย่าง n มากขึ้น (p ค่าลดลง ทั้งนี้เท่ากับว่าเป็นการลดค่าความน่าจะเป็นของการเกิดความสำเร็จลง) การประมาณการแจกแจงดังกล่าวจะทำให้ความคลาดเคลื่อน (ε) ระหว่างการแจกแจงทวินามกับการแจกแจงปัวส์ซอง มีค่าน้อยลง ดังแสดงในรูปที่ 2.7



<u>รูปที่ 2.7</u> แสดงการเปรียบเทียบฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามและปัวส์ซอง

2.8 <u>ทฤษฎีบทลิมิตสู่ส่วนกลาง</u> (Central Limit Theorem)

การลู่เข้าในเชิงการแจกแจงนั้นกล่าวถึงการแจกแจงขีดจำกัด (limiting distribution) ของตัวแปรสุ่ม เมื่อขนาดตัวอย่างเข้าใกล้อนันต์ โดยมีทฤษฎีที่สำคัญ คือ ทฤษฎีบทลิมิตสู่ส่วนกลาง ซึ่งมีประโยชน์มากในการอนุมานเชิงสถิติ

ตัวสถิติตัวหนึ่งที่ใช้มากในการอนุมานเชิงสถิติจากตัวอย่างสุ่ม คือ ค่าเฉลี่ย ซึ่งถ้าทราบว่า ตัวอย่างสุ่ม $(X_1, X_2, \ldots, X_n)$ มาจากประชากรที่มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 แล้ว ตัวแปรสุ่ม $\overline{X}$ ก็จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และ ค่าความแปรปรวนเป็น $\frac{\sigma^2}{n}$ แต่ในทางปฏิบัติ ตัวอย่างสุ่มที่ใช้อาจไม่ได้มาจากการแจกแจงแบบปกติ ในกรณีเช่นนี้การหาการแจกแจงความน่าจะเป็นของ $\overline{X}$ อาจทำได้ยากโดยเฉพาะเมื่อ n มีค่ามาก จึงต้องใช้การแจกแจงโดยประมาณแทน ดังนั้นถ้าสามารถหาได้ว่า การแจกแจงความน่าจะเป็นของ $\overline{X}$ เข้าใกล้การแจกแจงใดเมื่อ $n \to \infty$ ก็ย่อมทำให้การอนุมานทำได้สะดวกดียิ่งขึ้น ทฤษฎีบทลิมิตสู่ส่วนกลางเป็นทฤษฎีที่ว่าด้วยการแจกแจงขีดจำกัด (limiting distribution) ของตัวแปรสุ่ม $\overline{X}$ หรือผลบวก $\sum X$ ทฤษฎีบทนี้อาจเขียนได้หลายรูปแบบตามกรณีที่กำลังพิจารณา

<u>ทฤษฎีบทลิมิตสู่ส่วนกลาง</u> (Central Limit Theorem)

ให้ $X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มที่มีการแจกแจงเดียวกันและเป็นอิสระต่อกัน โดยที่มีค่าเฉลี่ย μ และ ค่าความแปรปรวน $\sigma^2 < \infty$ และกำหนดให้ $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n}$ เมื่อ $S_n = \sum_{i=1}^{n} X_i$ จะได้ว่า $Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ ลู่เข้าในเชิงการแจกแจงสู่ตัวแปรสุ่ม Z ซึ่งเป็นตัวแปรสุ่มที่มีการแจกแจงปกติที่มีค่า $\mu = 0$ และ $\sigma^2 = 1$ ขณะที่ $n \to \infty$

2.9 การประมาณการแจกแจงด้วยการแจกแจงแบบปกติ

(Approximation by the Normal Distribution)

การแจกแจงแบบปกติ อาจกล่าวได้ว่า เป็นการแจกแจงที่สำคัญที่สุดในวิชาสถิติ และเป็นการแจกแจงซึ่งใช้อธิบายข้อมูลที่เกิดขึ้นโดยทั่วไปทางธรรมชาติ การประมาณการแจกแจงต่าง ๆ โดยใช้การแจกแจงแบบปกตินั้นมีจุดประสงค์สำคัญก็เพื่อผลในการอนุมานใช้กับงานต่าง ๆ ในการประมาณการแจกแจงต่าง ๆ โดยใช้การแจกแจงแบบปกติจะใช้หลักการจากทฤษฎีบทลิมิตสู่ส่วนกลางดังที่กล่าวไว้ในหัวข้อข้างต้น สำหรับงานวิจัยนี้จะพิจารณา 3 กรณี ดังนี้

2.9.1 การประมาณการแจกแจงทวินามด้วยการแจกแจงปกติ

(Approximation of the Binomial Distribution by the Normal Distribution)

ในหัวข้อ 2.7 ได้กล่าวถึง การประมาณการแจกแจงทวินามด้วยการแจกแจงปัวส์ซองเมื่อขนาดตัวอย่างใหญ่ ความน่าจะเป็นของผลสำเร็จมีค่าน้อย และผลคูณระหว่างขนาดตัวอย่างกับความน่าจะเป็นของผลสำเร็จมีค่าคงที่ สำหรับในหัวข้อนี้จะกล่าวถึงการประมาณการแจกแจงทวินามด้วยการแจกแจงปกติมาตรฐานเมื่อผลคูณระหว่างขนาดตัวอย่างกับความน่าจะเป็นของผลสำเร็จและความน่าจะเป็นของความล้มเหลวมีค่ามากพอสมควร โดยที่การประมาณจะได้ผลดี เมื่อความน่าจะเป็นของผลสำเร็จลู่ไปสู่ 0.5

เป็นที่ทราบกันดีว่าตัวแปรสุ่มทวินามได้มาจากผลรวมของตัวแปรสุ่มแบร์นูลลี่ที่มีค่าเฉลี่ยเป็น $\mu = p$ และความแปรปรวนเป็น $\sigma^2 = pq$ เมื่อ $0 \le p \le 1$ แล้ว S_n จะมีการแจกแจงทวินามที่มีค่าเฉลี่ยเป็น np และ ความแปรปรวนเป็น npq

โดยทฤษฎีบทลิมิตสู่ส่วนกลางได้ว่า การแจกแจงของ

$$Z_n = \frac{S_n - np}{\sqrt{npq}}$$

จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับหนึ่งในขณะที่ $n \to \infty$

สำหรับกรณีนี้นอกจากจะใช้ทฤษฎีบทลิมิตสู่ส่วนกลางพิสูจน์แล้วยังสามารถแสดงให้เห็นได้อีกกรณีหนึ่งดังนี้

จากทฤษฎีบท De Moivre-Laplace

$X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระกันในการทดลอง n ครั้ง ซึ่งการทดลองในแต่ละครั้งจะมีค่าเป็น 1 ด้วยความน่าจะเป็นเท่ากับ p และ มีค่าเป็น 0 ด้วยความน่าจะเป็นมีค่าเท่ากับ $q = 1 - p$ ให้

$$S_n = \sum_{k=1}^{n} X_k, \qquad S_n^* = \frac{S_n - E[S_n]}{\sqrt{V[S_n]}}$$

ดังนั้น

$$\lim_{n\to\infty} P\{x' \le S_n^* \le x''\} = \frac{1}{\sqrt{2\pi}}\int_{x'}^{x''} e^{-x^2/2}\,dx$$

พิสูจน์ได้ดังนี้

จะเห็นได้ว่า S_n เป็นตัวแปรสุ่มทวินาม มีค่าเฉลี่ย $E[S_n] = np$ และความแปรปรวน $V[S_n] = npq$ จากนี้แปลงเป็นตัวแปรมาตรฐาน S_n^* เป็นตัวแปรสุ่มมีค่า

$$x = \frac{k - np}{\sqrt{npq}} \qquad ; k = 0,1,2\ldots,n$$

และมีความน่าจะเป็น

$$P(S_n^* = x) = P_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} \qquad ; k = 0,1,2,\ldots,n$$

เมื่อ $n \to \infty$;

$$k = np + \sqrt{npq}\,x \to \infty, \qquad n - k = nq - \sqrt{npq}\,x \to \infty \qquad \ldots\ldots(*)$$

จากสูตรของสเตอร์ลิง (Stirling's Formula)

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n} \qquad ; n \to \infty$$

จะได้ว่า

$$P_n(k) \sim \frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi k}\, k^k e^{-k} \sqrt{2\pi(n-k)}\ (n-k)^{n-k} e^{-(n-k)}}\, p^k q^{n-k}$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

จากสมการ (*)

$$\frac{k}{np} = 1 + \sqrt{\frac{q}{np}}x, \qquad \frac{n-k}{nq} = 1 - \sqrt{\frac{p}{nq}}x$$

และจากความเป็นจริงที่ว่า

$$\ln(1+\alpha_n) \sim \alpha_n - \frac{\alpha_n^2}{2}$$

ทำให้ได้ว่า

$$\begin{aligned} \ln\left(\frac{k}{np}\right)^{-k} &= -k \ln\left(1 + \sqrt{\frac{q}{np}}x\right) \\ &\sim -(np + \sqrt{npq}x)\left(\sqrt{\frac{q}{np}}x - \frac{1}{2}\frac{q}{np}x^2\right) \end{aligned}$$

$$\begin{aligned} \ln\left(\frac{n-k}{nq}\right)^{-(n-k)} &= -(n-k) \ln\left(1 - \sqrt{\frac{p}{nq}}x\right) \\ &\sim -(nq - \sqrt{npq}x)\left(-\sqrt{\frac{p}{nq}}x - \frac{1}{2}\frac{p}{nq}x^2\right) \end{aligned}$$

เมื่อ $n \to \infty$;

$$\lim_{n\to\infty} \ln\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} = -\frac{x^2}{2},$$

$$\lim_{n\to\infty}\left(\frac{np}{k}\right)^{k}\left(\frac{nq}{n-k}\right)^{n-k} = e^{-x^2/2}$$

สำหรับค่าในช่วง $x' \le x \le x''$;

$$\sqrt{\frac{n}{n(n-k)}} \sim \sqrt{\frac{k}{np.nq}} = \frac{1}{\sqrt{npq}},$$

จะได้ว่า

$$\lim_{n\to\infty} P\{S_n^* = x\} = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\Delta x, \quad \Delta x = \frac{1}{\sqrt{npq}},$$

ดังนั้น

$$\lim_{n\to\infty} P\{x' \le S_n^* \le x''\} = \lim_{n\to\infty}\sum_{x'\le x\le x''} P\{S_n^* = x\}$$

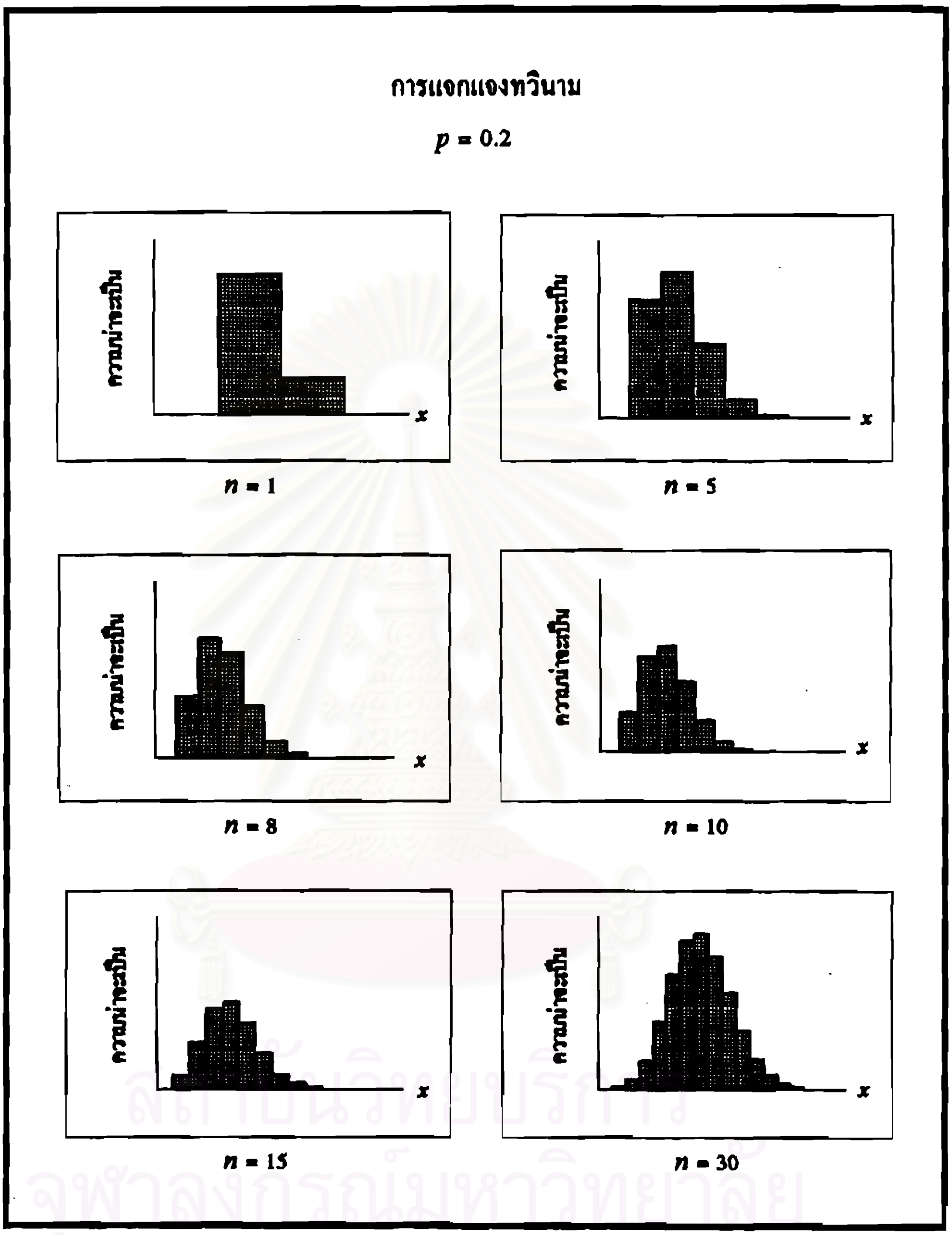
$$= \lim_{n\to\infty}\sum_{x'\le x\le x''}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\Delta x,$$

ผลบวกค่า x ทั้งหมดในช่วง $x' \le x \le x''$ จะสามารถเขียนได้ใหม่ว่า

$$\lim_{n\to\infty}\sum_{x'\le x\le x''}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\Delta x = \frac{1}{\sqrt{2\pi}}\int_{x'}^{x''} e^{-x^2/2}dx$$

จากที่แสดงให้เห็นจะเห็นว่าการแจกแจงแบบปกติสามารถใช้ประมาณการแจกแจงทวินามได้ เมื่อลักษณะข้อมูลเป็นไปตามเงื่อนไข

การประมาณการแจกแจงทวินามด้วยการแจกแจงปกตินั้นขึ้นกับขนาดตัวอย่าง ซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงทวินาม กล่าวคือ ณ ความน่าจะเป็นของการเกิดความสำเร็จค่าหนึ่ง เมื่อเพิ่มขนาดตัวอย่างมากขึ้น การแจกแจงทวินามจะลู่เข้าสู่การแจกแจงปกติมากขึ้น ดังแสดงในรูปที่ 2.8



รูปที่ 2.8 แสดงการเปรียบเทียบฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามเมื่อขนาดตัวอย่างเพิ่มขึ้น

2.9.2 <u>การประมาณการแจกแจงปัวส์ซองด้วยการแจกแจงปกติ</u>

(Approximation of the Poisson Distribution by the Normal Distribution)

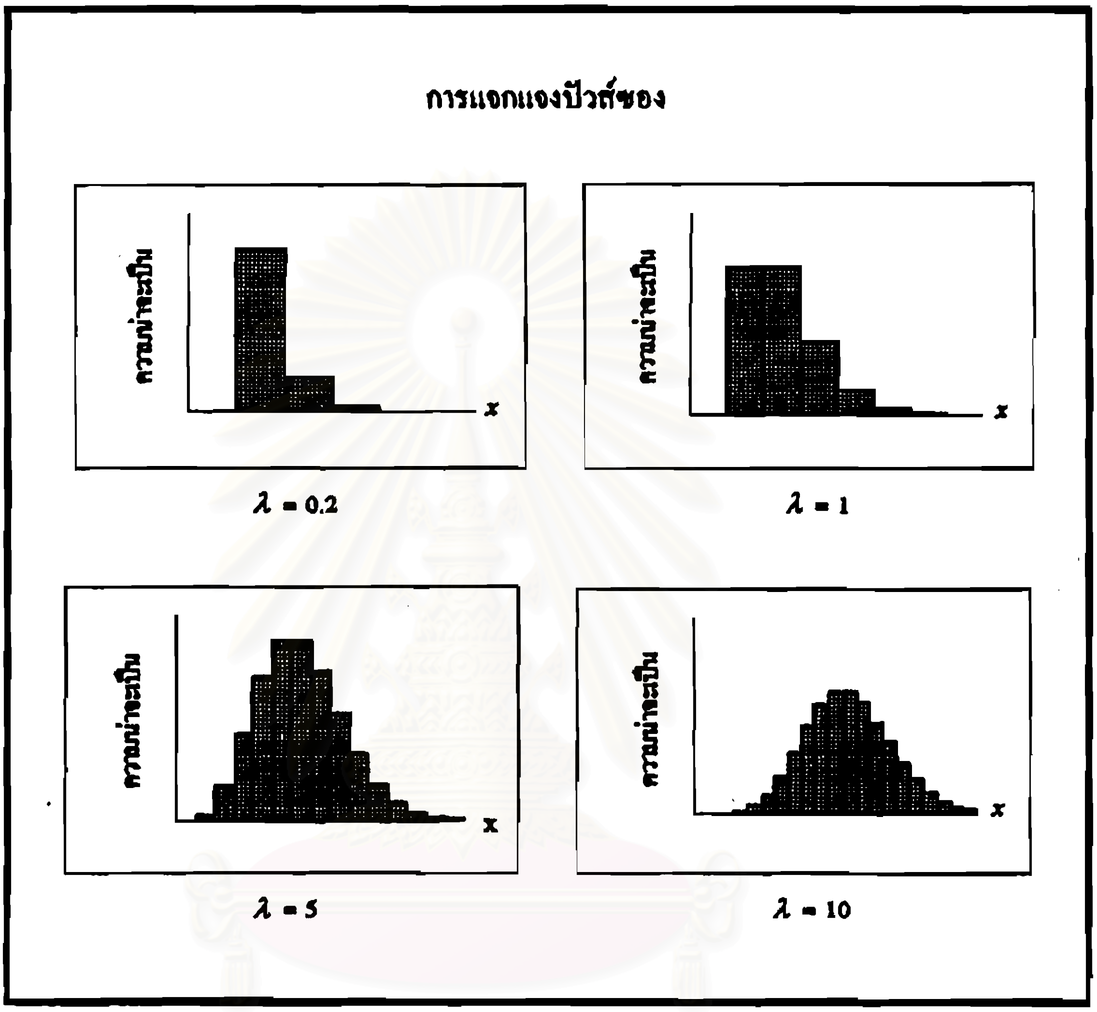
การแจกแจงปัวส์ซองสามารถประมาณได้ด้วยการแจกแจงปกติมาตรฐาน เมื่อ λ มีขนาดใหญ่ ดังนั้น ถ้าให้ S_n เป็นตัวแปรสุ่มที่มีการแจกแจงปัวส์ซอง มีค่าเฉลี่ย และความแปรปรวนเท่ากับ λ

โดยทฤษฎีบทลิมิตสู่ส่วนกลางได้ว่า การแจกแจงของ

$$Z_n = \frac{S_n - \lambda}{\sqrt{\lambda}}$$

จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับหนึ่ง ในขณะที่ $n \to \infty$ การประมาณค่านี้จะได้ผลดี ถ้าค่า λ มีขนาดใหญ่พอสมควร

การประมาณการแจกแจงปัวส์ซองด้วยการแจกแจงปกตินั้น ขึ้นกับขนาด λ ซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงปัวส์ซอง กล่าวคือ เมื่อ λ มีค่ามากขึ้น การแจกแจงปัวส์ซองจะลู่เข้าสู่การแจกแจงปกติมากขึ้น ดังแสดงในรูปที่ 2.9



<u>รูปที่ 2.9</u> แสดงการเปรียบเทียบฟังก์ชันความน่าจะเป็นของการแจกแจงปัวส์ซอง เมื่อ λ เพิ่มขึ้น

2.9.3 <u>การประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงปกติ</u>

(Approximation of the Hypergeometric Distribution by the Binomial Distribution)

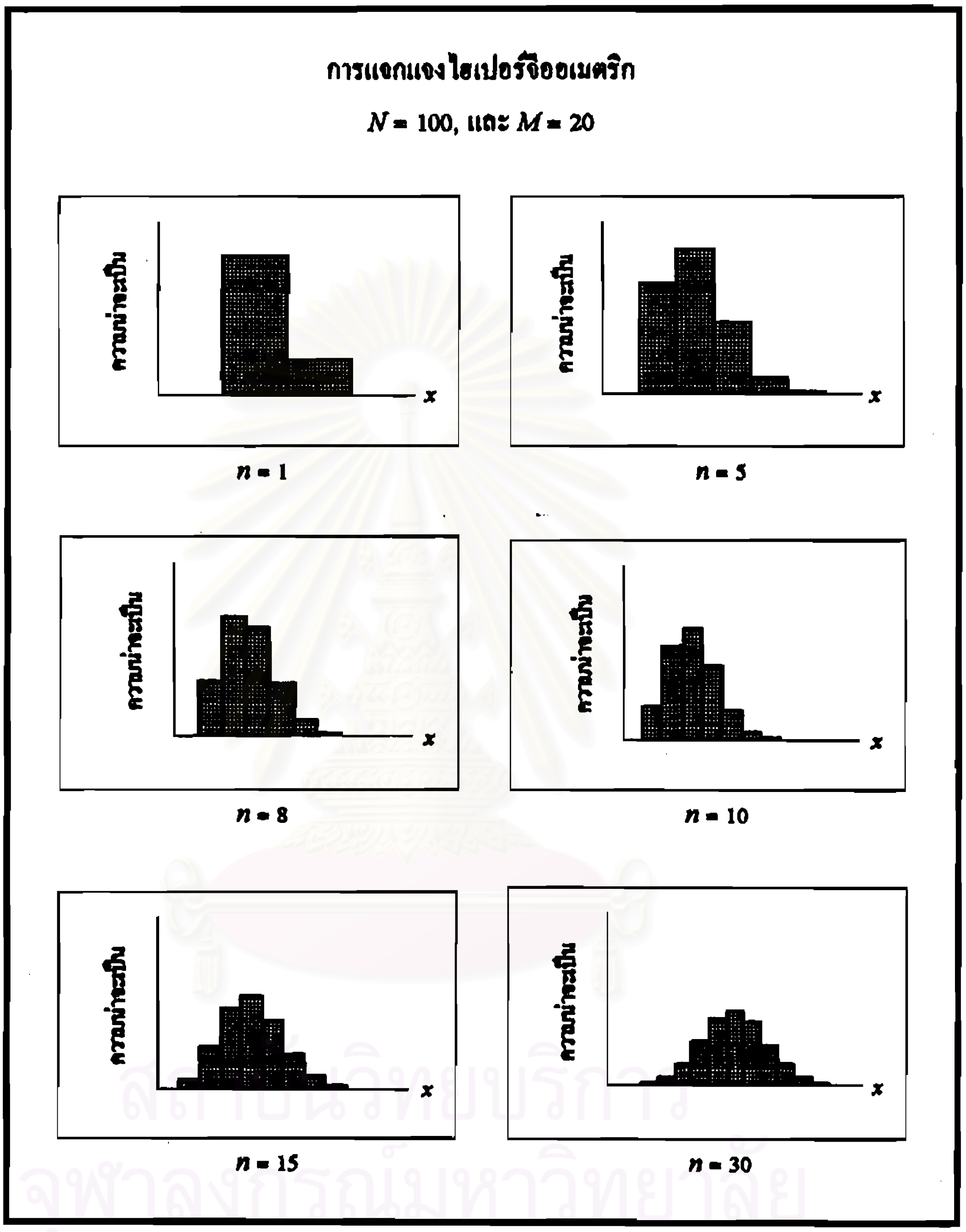
การแจกแจงไฮเปอร์จีออเมตริกสามารถประมาณได้ด้วยการแจกแจงปกติมาตรฐาน เมื่อ n มีขนาดใหญ่ และ $\frac{M}{N}$ มีค่ามาก ดังนั้น ถ้าให้ S_n เป็นตัวแปรสุ่มที่มีการแจกแจงไฮเปอร์จีออเมตริก มีค่าเฉลี่ยเท่ากับ $\frac{nM}{N}$ และ ความแปรปรวนเท่ากับ $\frac{n(N-n)}{(N-1)}\left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right)$

โดยทฤษฎีบทลิมิตสู่ส่วนกลางได้ว่า การแจกแจงของ

$$Z_n = \frac{S_n - nM/N}{\sqrt{n(N-n).(N-1)^{-1}(M/N).(1-M/N)}}$$

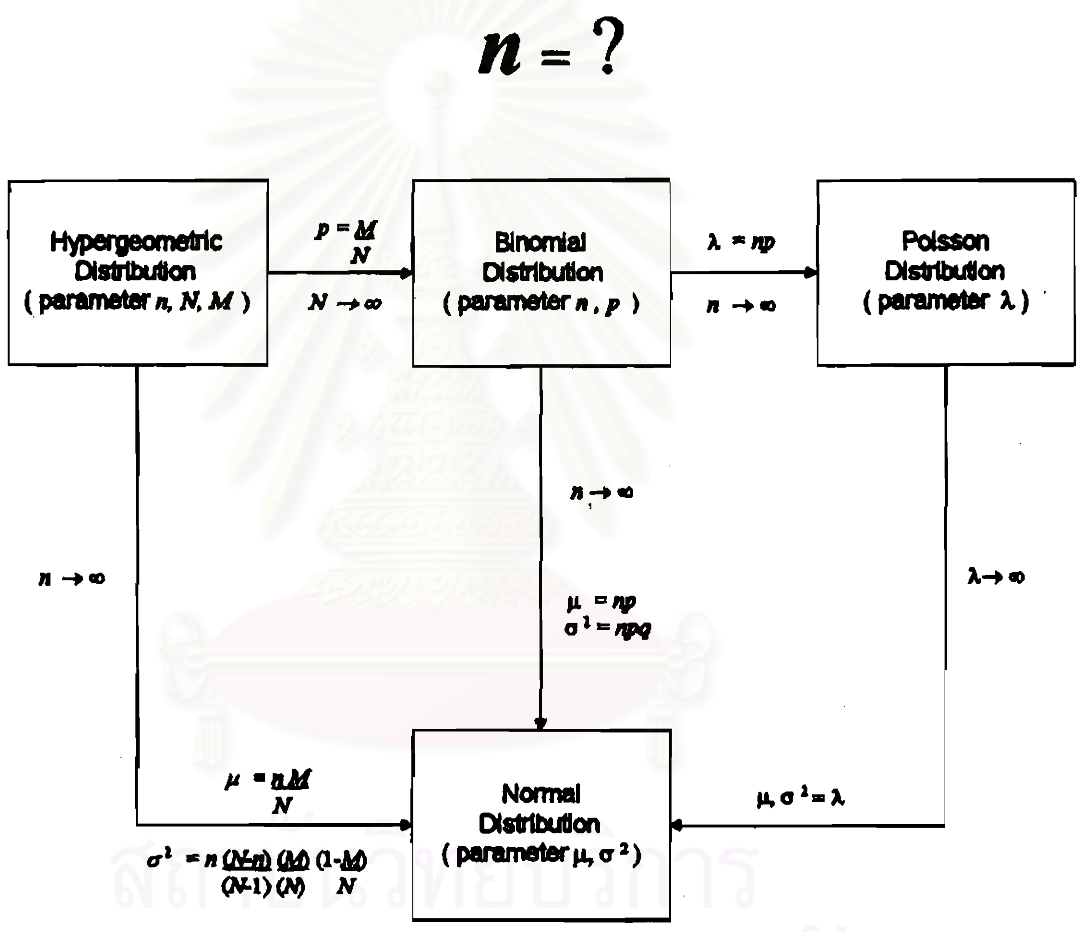
จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนเท่ากับหนึ่ง ขณะที่ $n \to \infty$

การประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงปกตินั้น ขึ้นกับขนาดตัวอย่าง n ซึ่งเป็นค่าพารามิเตอร์ของการแจกแจงไฮเปอร์จีออเมตริก กล่าวคือ ณ อัตราส่วนระหว่างขนาดประชากรที่เราสนใจและขนาดประชากรทั้งหมด (M/N) ค่าหนึ่ง เมื่อเพิ่มขนาดตัวอย่างมากขึ้น การแจกแจงไฮเปอร์จีออเมตริกจะลู่เข้าสู่การแจกแจงปกติมากขึ้น ดังแสดงในรูปที่ 2.10



__รูปที่ 2.10__ แสดงการเปรียบเทียบฟังก์ชันความน่าจะเป็นของการแจกแจงไฮเปอร์จีออเมตริก เมื่อขนาดตัวอย่างเพิ่มขึ้น

จากหัวข้อ 2.6, 2.7, และ 2.9 ได้กล่าวถึง การประมาณการแจกแจงไฮเปอร์จีออเมตริกด้วยการแจกแจงทวินาม, การประมาณการแจกแจงทวินามด้วยการแจกแจงปัวส์ซอง, และ การประมาณการแจกแจงทวินาม, การแจกแจงปัวส์ซอง, และ การแจกแจงไฮเปอร์จีออเมตริก ด้วยการแจกแจงปกติ ตามลำดับ สามารถสรุปเป็นแผนผังสำหรับงานวิจัยได้ดังนี้



รูปที่ 2.11 แสดงแผนผังการวิจัย

2.10. การทดสอบเทียบความกลมกลืนกับโดยใช้ไคสแควร์

(Chi-Square Goodness-of-Fit Test)

เนื้อหาส่วนสำคัญที่จัดเป็นหัวข้อใหญ่ในเรื่องการอนุมานเชิงสถิติอีกหัวข้อหนึ่ง นอกจากการประมาณค่าพารามิเตอร์ ก็คือ การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ของประชากร ซึ่งเป็นวิธีการเปรียบเทียบความเชื่อ ที่นักวิจัยมีต่อค่าพารามิเตอร์ของประชากรกับข้อมูลที่เก็บรวมรวมมาได้จากการสุ่มเลือกตัวอย่าง โดยผ่านขั้นตอนการวิเคราะห์เชิงสถิติ แล้วหาข้อสรุปจากการเปรียบเทียบ

ข้อความเกี่ยวกับการทดสอบสมมติฐานจะถูกต้องหรือไม่นั้น ขึ้นอยู่กับข้อมูลที่ใช้ในการตัดสินใจเป็นสำคัญ กล่าวคือ จะสามารถบอกได้แต่เพียงว่า จากข้อมูลที่มีอยู่ ข้อสรุปเกี่ยวกับพารามิเตอร์ของประชากร ควรเป็นเช่นนี้เท่านั้น

กระบวนการของการทดสอบสมมติฐานแบ่งออกเป็น 4 ขั้นตอนใหญ่ ๆ คือ

1. กำหนดข้อความที่เป็นสมมติฐาน

ประกอบไปด้วย สมมติฐานว่าง (null hypothesis) และสมมติฐานแย้ง (alternative hypothesis)

2. กำหนดระดับนัยสำคัญของการทดสอบสมมติฐานทางสถิติ (statistical significance level)

ค่าระดับนัยสำคัญ เป็นค่าความน่าจะเป็นที่จะเกิดความผิดพลาดชนิดที่ 1 (Type I error probability) ซึ่งเป็นโอกาสที่ผู้ทำการตัดสินใจจะทำการตัดสินใจผิดพลาด โดยการปฏิเสธสมมติฐานว่างที่ตั้งขึ้น เมื่อในความเป็นจริงสมมติฐานว่างนั้นถูกต้อง ค่าระดับนัยสำคัญที่ใช้กันเสมอ ๆ ได้แก่ 0.05, 0.01, และ 0.10

3. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test statistic) และกฎเกณฑ์การตัดสินใจ (decision rules) ของการทดสอบ

ในการกำหนดตัวสถิติที่ใช้ในการทดสอบ มีความจำเป็นที่จะต้องทราบการแจกแจงของตัวสถิตินั้น เพื่อที่จะได้สามารถหาค่าสถิติ จากฟังก์ชันการแจกแจงของตัวสถิติ ณ ระดับนัยสำคัญที่กำหนด แล้วจะใช้ค่านั้น เป็นค่าวิกฤต (critical value) ซึ่งจะใช้ในการกำหนด กฎเกณฑ์การตัดสินใจของการทดสอบสมมติฐาน

ค่าวิกฤตดังกล่าว สามารถมีค่าเป็นค่าเดียวหรือสองค่าก็ได้ ขึ้นอยู่กับลักษณะของการทดสอบว่าเป็นการทดสอบแบบด้านเดียว (one-sided test) หรือ เป็นการทดสอบสองด้าน (two-sided test) ค่าวิกฤตจะเป็นค่ากำหนดบริเวณวิกฤต ซึ่งอาจจะมีด้านเดียวหรือสองด้านขึ้นอยู่กับการทดสอบว่าเป็นแบบด้านเดียวหรือสองด้าน

กฎเกณฑ์ของการตัดสินใจ คือการกำหนดบริเวณวิกฤต (critical region) ซึ่งเป็นบริเวณที่เราจะตัดสินใจปฏิเสธสมมติฐานว่าง (reject H_0) ถ้าค่าสถิติที่คำนวณได้จากตัวอย่าง ตกอยู่ในบริเวณนี้ ในกรณีที่ค่าสถิติที่คำนวณได้จากตัวอย่างไม่ได้มีค่าอยู่ในบริเวณวิกฤตจะตัดสินใจไม่ปฏิเสธสมมติฐานว่าง (do not reject H_0) ในความหมายที่ว่า ข้อมูลที่รวบรวมมาได้ไม่ได้ปรากฏหลักฐานที่จะให้ปฏิเสธสมมติฐานว่าง

4. เก็บรวบรวมข้อมูล

ขั้นตอนการเก็บรวบรวมข้อมูลนับเป็นขั้นตอนที่สำคัญมากขั้นตอนหนึ่ง ข้อมูลที่นำมาใช้ทดสอบสมมติฐานจะมีค่าถูกต้องเชื่อถือได้มากน้อยเพียงใด ขึ้นอยู่กับว่าตัวอย่างที่เลือกมานั้นมีความเป็นตัวแทนที่ดีของประชากรที่ต้องการทดสอบสมมติฐานหรือไม่ ภายหลังจากที่ได้เก็บรวมรวมข้อมูลเสร็จสิ้นแล้ว ก็จะสามารถคำนวณค่าสถิติได้ โดยการแทนค่าสังเกตที่ได้จากตัวอย่าง ซึ่งเป็นค่าของตัวแปรสุ่มลงในสูตรของตัวสถิติ แล้วทำการตัดสินใจเกี่ยวกับข้อความในสมมติฐานว่าง ตามกฎเกณฑ์การตัดสินใจที่กำหนดไว้ในข้อ 3

จากขั้นตอนของกระบวนการทดสอบสมมติฐาน จะพบว่าในการหาการแจกแจงของตัวสถิติ จะต้องตั้งข้อกำหนดเกี่ยวกับรูปแบบการแจกแจงของข้อมูลเสียก่อน หลังจากนั้น จึงจะหาการแจกแจงของตัวสถิติ

ภายหลังจากรวบรวมค่าสังเกต $X_1, X_2, \ldots, X_n$ มาได้แล้ว จะทำการตรวจสอบว่า ข้อมูลที่ได้มามีความสอดคล้องกับข้อตกลงเบื้องต้นของการแจกแจงที่กำหนดหรือไม่ โดยการตั้งสมมติฐานว่างว่าชุดตัวอย่างสุ่ม $X_1, X_2, \ldots, X_n$ มีรูปแบบการแจกแจงเป็นไปตามข้อกำหนดที่ตั้งขึ้น

ปัญหาการทดสอบสมมติฐานในลักษณะดังกล่าวเรียกว่า การทดสอบเทียบความกลมกลืนกัน (Goodness-of-Fits Test) ซึ่งเป็นการทดสอบหรือตรวจสอบว่าลักษณะการแจกแจงที่กำหนดไว้บนสมมติฐานว่าง จะสามารถเข้ากันได้ กับการแจกแจงของข้อมูล $X_1, X_2, \ldots, X_n$ ซึ่งเป็นค่าสังเกต ที่เราสังเกตมาได้ดีเพียงใด

ในทางปฏิบัติ เราจะไม่พยายามทดลองไปเรื่อย ๆ เพื่อหาว่า ข้อมูลมีความสอดคล้องกับการแจกแจงใดมากที่สุด กล่าวคือ จะไม่ทำการทดสอบข้อมูลกับการแจกแจงทุกการแจกแจง แต่จะเลือกทดสอบกับการแจกแจงที่มีเหตุผล หรือมีความเชื่อที่ได้มาจากประสบการณ์มากพอที่จะกล่าวว่า ค่าสังเกต $X_1, X_2, ..., X_n$ น่าจะถูกสุ่มเลือกมาจากประชากรที่มีการแจกแจงนั้น ๆ หรือใช้วิธีพิจารณากราฟแสดงการแจกแจงของข้อมูล เช่น กราฟฮิสโทแกรม (Histogram) เทียบเคียงกับกราฟการแจกแจงเชิงทฤษฎีของการแจกแจงที่รู้จักกันทั่ว ๆ ไป

ดังนั้นอาจกล่าวได้ว่า การทดสอบเทียบความกลมกลืนกัน คือ การหาคำตอบของคำถามซึ่งมีใจความว่า "เป็นไปได้หรือไม่ ที่จะกล่าวว่าค่าสังเกต $X_1, X_2, ..., X_n$ ได้ถูกสุ่มเลือกมาจากประชากร ที่มีรูปแบบการแจกแจงตามที่ระบุไว้ในสมมติฐานว่าง"

ข้อได้เปรียบของการนำเอารูปแบบการแจกแจงทางทฤษฎีมาใช้ที่เห็นได้ชัด คือ ในกรณีที่สมมติฐานว่างไม่ถูกปฏิเสธ จะสามารถดึงเอาคุณสมบัติต่าง ๆ ของการแจกแจงที่ระบุไว้ในสมมติฐานว่าง มาใช้เป็นข้อสรุปเกี่ยวกับประชากร ทำให้ได้ข้อสรุปที่กว้างกว่าการใช้ข้อสรุปที่ได้รับจากข้อมูลตัวอย่าง

วิธีการทดสอบเทียบความกลมกลืนกันมีหลายวิธี ในงานวิจัยนี้จะใช้วิธีไคสแควร์ ซึ่งสามารถใช้ได้กับการแจกแจงแบบต่อเนื่อง และการแจกแจงแบบไม่ต่อเนื่อง โดยมีหลักการ ดังนี้

1. ทำการแบ่งช่วงของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม หรือของข้อมูลที่นำมาศึกษา ออกเป็น k ช่วง โดยที่แต่ละช่วงจะต้องไม่มีสมาชิกซ้ำกัน

สมมติว่าได้เป็น $(a_0, a_1), [a_1, a_2), ..., [a_{k-1}, a_k)$

โดยที่ a_0 สามารถมีค่าเป็น $-\infty$ และ a_k สามารถมีค่าเป็น $+\infty$ ได้

กำหนดให้

O_i = จำนวนค่าสังเกต x ที่ตกอยู่ในช่วงที่ $i : [a_{i-1}, a_i)$

สำหรับทุกค่า $i = 1,2,...,k$ และจะต้องได้ว่า $O_1 + O_2 + ... + O_k$ เท่ากับ n^* เมื่อ n^* คือจำนวนทั้งหมดของค่าสังเกต X

2. คำนวณค่าสัดส่วนที่คาดหมายของจำนวนค่าสังเกต X ที่ตกอยู่ในช่วงที่ i ในการหาค่าสัดส่วนที่คาดหมาย จะต้องกระทำภายใต้เงื่อนไขว่าค่าสังเกต $X_1, X_2, \ldots, X_n$ ได้ถูกสุ่มเลือกมาจากประชากร ที่มีรูปแบบการแจกแจงเป็นไปตามที่ระบุในสมมติฐานว่าง

กำหนดให้

p_i เป็นค่าสัดส่วนที่คาดหมายของจำนวนค่าสังเกต X ที่ตกอยู่ในช่วงที่ i

$n^* p_i$ เป็นจำนวนที่คาดหมายของค่าสังเกต X ที่ตกอยู่ในช่วงที่ i แทนด้วยสัญลักษณ์ E_i

3. คำนวณค่าสถิติ χ_c^2 จากสูตร

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

3.1 กรณีทราบค่าพารามิเตอร์ทุกตัวของการแจกแจงในสมมติฐานว่าง

ภายใต้เงื่อนไขสมมติฐานว่างเป็นจริง และทราบค่าพารามิเตอร์ทุกตัวของการแจกแจงในสมมติฐานว่าง ตัวสถิติ χ_c^2 จะมีการแจกแจงสู่เข้าสู่การแจกแจงไคสแควร์ ที่มีค่าองศาอิสระ $k-1$ เมื่อขนาดตัวอย่าง n^* ค่าใหญ่

ในการทดสอบสมมติฐานว่าง (H_0) ที่ระดับนัยสำคัญ α มีกฎการตัดสินใจดังนี้

1. ตัดสินใจปฏิเสธสมมติฐานว่าง (H_0) เมื่อค่าสถิติ χ_c^2 ที่คำนวณได้จากตัวอย่างมีค่ามากกว่าค่าวิกฤต $\chi^2_{\alpha,k-1}$ ซึ่งเป็นเปอร์เซ็นไทล์ที่ 100(1-α) ของตัวแปรสุ่มไคสแควร์องศาอิสระ $k-1$

2. ตัดสินใจไม่ปฏิเสธสมมติฐานว่าง (H_0) เมื่อค่าสถิติ χ_c^2 มีค่าน้อยกว่าหรือเท่ากับ $\chi^2_{\alpha,k-1}$

3.2 กรณีไม่ทราบค่าพารามิเตอร์ทุกตัวของการแจกแจงในสมมติฐานว่าง

ในกรณีที่การกำหนดสมมติฐานว่างมีความจำเป็นต้องใช้ค่าประมาณแบบจุดที่คำนวณได้จากข้อมูลตัวอย่างแทนเป็นค่าของพารามิเตอร์ในสมมติฐาน จำนวนทั้งสิ้น m ตัว $(m \geq 1)$

ในการทดสอบสมมติฐานว่าง (H_0) ที่ระดับนัยสำคัญ α มีกฎการตัดสินใจดังนี้

1. ตัดสินใจปฏิเสธสมมติฐานว่าง (H_0) เมื่อค่าสถิติ χ_c^2 ที่คำนวณได้จากตัวอย่าง มีค่ามากกว่า ค่าวิกฤต $\chi^2_{\alpha,k-m-1}$ ซึ่งเป็นเปอร์เซ็นไทล์ที่ 100(1-α) ของตัวแปรสุ่มไคสแควร์ องศาอิสระ $k-m-1$

2. ตัดสินใจปฏิเสธสมมติฐานว่าง (H_0) เมื่อค่าสถิติ χ_c^2 มีค่าน้อยกว่า หรือเท่ากับ $\chi^2_{\alpha,k-m-1}$