บทที่ 3

การวิเคราะห์ความถดถอยและสหสัมพันธ์อย่างง่าย

(Simple Regression Analysis and Correlation)

3.1 การวิเคราะห์ความถดถอย (Regression Analysis)

เป็นการศึกษาถึงความสัมพันธ์ของตัวแปรตั้งแต่ 2 ตัวขึ้นไป โดยมีวัตถุประสงค์ที่จะประมาณหรือพยากรณ์ค่าของตัวแปรตัวหนึ่งจากตัวแปรตัวอื่น ๆ ที่มีความสัมพันธ์กับตัวแปรที่ต้องการพยากรณ์ โดยจะต้องมีการกำหนดหรือทราบค่าตัวแปรอื่น ๆ ล่วงหน้า เช่น ถ้าทราบความสัมพันธ์ระหว่างยอดขายกับค่าโฆษณาแล้ว จะทำให้สามารถประมาณ / พยากรณ์ยอดขายเมื่อกำหนดหรือทราบงบประมาณในการโฆษณา และจะศึกษาถึงการเปลี่ยนแปลงของยอดขายเมื่องบประมาณในการโฆษณาเปลี่ยนแปลงไป โดยอาศัยหลักการของการวิเคราะห์ความถดถอย การวิเคราะห์ความถดถอยมี 2 ประเภท คือ

1. การวิเคราะห์ความถดถอยอย่างง่าย
2. การวิเคราะห์ความถดถอยเชิงพหุ

เป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปร 2 ตัว หรือลักษณะที่สนใจศึกษา 2 ลักษณะ โดยที่ต้องทราบค่าของตัวแปรตัวหนึ่งหรือต้องกำหนดค่าของตัวแปรตัวหนึ่งไว้ล่วงหน้า เช่น ถ้าศึกษาถึงความสัมพันธ์ระหว่างรายจ่ายกับรายได้ ยอดขายกับค่าโฆษณา ฯลฯ ซึ่งจะต้องทราบหรือกำหนดรายได้ และค่าโฆษณาไว้ล่วงหน้า เช่น ทราบว่าเงินเดือนพนักงานทำความสะอาดของบริษัทแห่งหนึ่งเป็น 2,000, 2,500, 3,000, 3,500, และ 4,000 บาท ผู้วิเคราะห์จะต้องสอบถามพนักงานทำความสะอาดที่มีเงินเดือนดังกล่าวถึงรายจ่ายต่อเดือน จึงจะสามารถหาความสัมพันธ์ระหว่างรายได้กับรายจ่ายได้ หรือในการหาความสัมพันธ์ระหว่างยอดขาย กับค่าโฆษณาจะต้องทราบถึงงบประมาณในการโฆษณาที่บริษัทกำหนดไว้หรือใช้ไปจริง แล้วจึงจะทราบถึงยอดขาย โดยจะเรียกว่ารายได้ และค่าโฆษณา ซึ่งเป็นตัวแปรที่ต้องกำหนดค่าไว้ล่วงหน้าว่า ตัวแปรอิสระ (Independent Variable) และมักจะใช้สัญลักษณ์ x ส่วนยอดขายกับรายจ่ายจะเรียกว่า ตัวแปรตาม (Dependent Variable) และใช้สัญลักษณ์ y ซึ่งหมายถึง ยอดขายเป็นตัวแปรที่ขึ้นอยู่กับค่าโฆษณาและรายจ่ายเป็นตัวแปรที่ขึ้นกับรายได้

วัตถุประสงค์ของการวิเคราะห์ความถดถอยและสหสัมพันธ์

การวิเคราะห์ความสัมพันธ์ของตัวแปรต่าง ๆ มีวัตถุประสงค์ดังนี้

1. เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรว่ามีความสัมพันธ์กันมากน้อยเพียงใด ถ้า x และ y มีความสัมพันธ์กันมาก แสดงว่าถ้า x มีค่าเปลี่ยนแปลงไปจะมีผลกระทบต่อค่าของ y เป็นอย่างมาก

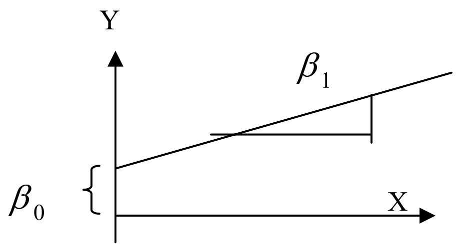
2. ใช้ความสัมพันธ์ที่วิเคราะห์ได้มาประมาณค่าหรือพยากรณ์ค่า y ในอนาคต เมื่อกำหนดค่า x

สำหรับการหารูปแบบความสัมพันธ์ระหว่างตัวแปร y และ x นั้นในขั้นแรกจะนำเอาข้อมูลของตัวแปรทั้งสองมาเขียนกราฟแสดงความสัมพันธ์ ซึ่งจะเรียกกราฟนี้ว่า แผนภาพการกระจาย (Scatter Diagram) ผู้วิเคราะห์จะต้องพิจารณาจากแผนภาพการกระจายว่าความสัมพันธ์ของตัวแปรทั้งสองจะอยู่ในรูปแบบใด เช่น เส้นตรง พาราโบลา เส้นโค้ง ฯลฯ โดยที่จะต้องสามารถเขียนความสัมพันธ์อยู่ในรูปแบบทางคณิตศาสตร์ได้

3.2 การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis)

เป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปร 2 ตัว ที่ความสัมพันธ์อยู่ในรูปเชิงเส้น ซึ่งสามารถแสดงความสัมพันธ์ในรูปสมการเชิงเส้นได้ดังนี้

$$y_i = \beta_i + \beta_i \chi_i + e_i \qquad ; i=1, 2, \ldots,$$



รูปที่ 3.1 ความสัมพันธ์ 2 ตัวแปร ในรูปเชิงเส้น

โดยที่ y คือตัวแปรตาม (Dependent Variable) เนื่องจากค่าของ y ขึ้นอยู่กับค่าของ x

x = ตัวแปรอิสระ (Independent Variable)

β_0 = ส่วนตัดแกน y หรือ คือค่าของ y เมื่อ x มีค่าเป็นศูนย์

e = ความคลาดเคลื่อนอย่างสุ่ม (random error)

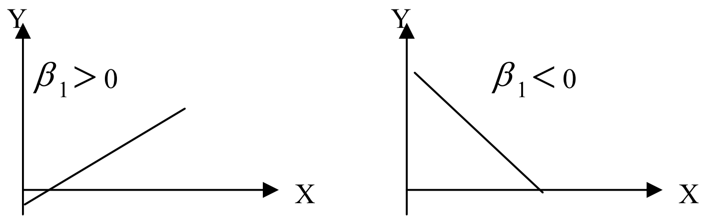
β_1 = ความชัน (slope) ของเส้นตรง ซึ่งเป็นค่าที่แสดงถึงอัตราการเปลี่ยนแปลงของ y เมื่อ x เปลี่ยนไป 1 หน่วย และจะเรียก β_1 ว่าสัมประสิทธิ์ความถดถอย (Regression Coefficient) ค่าของ β_1 อาจจะเป็น

1. $\beta_1 > 0$ แสดงว่า x และ y มีความสัมพันธ์ในทางเดียวกันคือถ้า x เพิ่ม y จะเพิ่มด้วย แต่ถ้า X ลดลง y จะลดลงด้วย

2. $\beta_1 < 0$ แสดงว่า x และ y มีความสัมพันธ์ในทางตรงข้ามกันคือถ้า x เพิ่ม y จะลดลง แต่ถ้า x ลดลง y จะเพิ่มขึ้น

3. β_1 มีค่าเข้าใกล้ศูนย์ แสดงว่าค่า x และ y มีความสัมพันธ์กันน้อย

4. $\beta_1 = 0$ แสดงว่า x และ y ไม่มีความสัมพันธ์กัน



รูปที่ 3.2 แสดงค่า β_1 เมื่อ x และ y มีความสัมพันธ์ในรูปเส้นตรง

หน่วยของ β_0 และ β_1 จะมีหน่วยเหมือนกับหน่วยของ Y ดังตัวอย่างเช่น

1. ให้ y = ยอดขาย มีหน่วยเป็นล้านบาท และ x เป็นค่าโฆษณาที่มีหน่วยเป็นแสนบาท จะทำให้ β_0 และ β_1 มีหน่วยเป็นล้านบาท

2. ให้ y เป็นจำนวน (หน่วย) สินค้าที่ขายได้ และ x เป็นราคาขายต่อหน่วย (บาท) จะได้ β_0 และ β_1 มีหน่วยเป็นจำนวนสินค้าที่ขายได้ เช่น x เป็นราคาขายทีวีต่อเครื่อง (บาท) y เป็นจำนวนทีวีที่ขายได้ (เครื่อง) β_0 และ β_1 จะมีหน่วยเป็นเครื่อง

3.3 สมมติฐานหรือเงื่อนไขของการวิเคราะห์ความถดถอย

1. ค่า x จะต้องเป็นค่าที่กำหนดไว้ล่วงหน้าหรือทราบค่า
2. ความคลาดเคลื่อน e_i เป็นตัวแปรที่มี ค่าเฉลี่ย = 0 หรือ E $(e_i) = 0$ ค่าแปรปรวนของ e_i มีค่าเท่ากันทุกค่าของ i และมีค่าเท่ากับค่าแปรปรวนของ y
   $V(e_i) = V(y) = \sigma^2_{y.x} = \sigma^2$
3. e_i และ $\theta_{j\,j}$ เป็นอิสระกัน นั่นคือ Cov $(e_i, \theta_j) = E(e_i, \theta_j) = 0;\ i \neq j$
4. θ_j มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์และค่าแปรปรวน σ^2 นั่นเอง

$$\theta_j \sim \text{normal}\ (0\ ,\ \sigma^2\ )$$

จากข้อสมมติข้างต้น จะได้ว่า

$$y_i \sim \text{normal}\ (E(Y_i)\ ,\ \sigma^2)$$

โดยที่ $E(y_i) = E(\beta_0 + \beta_1 x_i + e_i)$

$= \beta_0 + \beta_1 x_i + E(e_i)$

$= \beta_0 + \beta_1 \chi_i$ เนื่องจาก $E(e_i) = 0$

3.4 การประมาณค่าพารามิเตอร์ของสมการความถดถอย

เมื่อพิจารณาจากแผนภาพการกระจาย ซึ่งแสดงความสัมพันธ์ระหว่าง x และ y แล้วพบว่า x และ y สัมพันธ์กันในรูปเส้นตรง จะต้องคำนวณหาค่า β_0 และ β_1 ซึ่งจะทำให้ทราบถึงความสัมพันธ์ระหว่าง x และ y ว่ามีความสัมพันธ์ตามกันหรือตรงข้ามกันและความสัมพันธ์นั้นมากหรือน้อยเพียงใด ถ้า β_1 มีค่ามากกว่าแสดงว่า y มีความสัมพันธ์กับ X มากด้วย

การที่จะหาค่า β_0 และ β_1 ได้จำเป็นต้องทราบค่า x และ y ทุกค่าที่ได้เกิดขึ้นแล้วในอดีตเช่น ถ้า x = รายได้ของคนกรุงเทพมหานคร (กทม.) y = รายจ่ายของคนกทม. การหาค่า β_0 และ β_1 จะต้องทราบถึงรายได้และรายจ่ายของคนกทม. ทุกคน ซึ่งเป็นไปได้ยากในทางปฏิบัติเราจึงใช้ข้อมูลตัวอย่างขนาด n ในการประมาณค่า β_0 และ β_1 ดังนั้นค่าประมาณของ y คือ

$$y_i = \beta_0 + \beta_1 x_i$$

หรือ $\hat{y}_i = a + bx_i$

โดยที่ $\hat{\beta}_0 = a \quad , \quad \hat{\beta}_1 = b$

3.4.1 การประมาณค่า β_0 และ β_1 โดยวิธีกำลังสองน้อยที่สุด

การประมาณค่า β_0 และ β_1 ด้วย a และ b ตามลำดับขั้น มีเป้าหมายเพื่อให้ความคลาดเคลื่อนในการประมาณค่า y_i ด้วย $\hat{y}_i$ มีค่าต่ำสุด โดยใช้วิธีที่เรียกว่า วิธีกำลังสองน้อยที่สุด (Least Square Method) ซึ่งเป็นวิธีที่ต้องการหาค่า a และ b ที่ทำให้ผลบวกของค่าคลาดเคลื่อนยกกำลังสองมีค่าน้อยที่สุด

เนื่องจาก $y_i = \beta_0 + \beta_1 x_i + e_i$

หรือ $\hat{y}_i = a + bx_i$

$$\therefore y_i - \hat{y}_i = e_i$$

ผลบวกของค่าคลาดเคลื่อนยกกำลังสอง $= \sum_{i=1}^{n} e_i^2 = \sum \left(y_i - \hat{y}_i \right)^2$

ดังนั้นวิธีกำลังสองน้อยที่สุดคือการหาค่า a และ b ที่ทำให้ $\sum_{i=1}^{n} e_i^2$ มีค่าต่ำสุด การที่ประมาณค่า β_0 และ β_1 ด้วยค่า a และ b โดยใช้วิธีกำลังสองน้อยที่สุดจะทำให้

1. ผลรวมของค่าคลาดเคลื่อนในการประมาณค่า y_i ด้วย $\hat{y}_i$ เป็นศูนย์ คือ

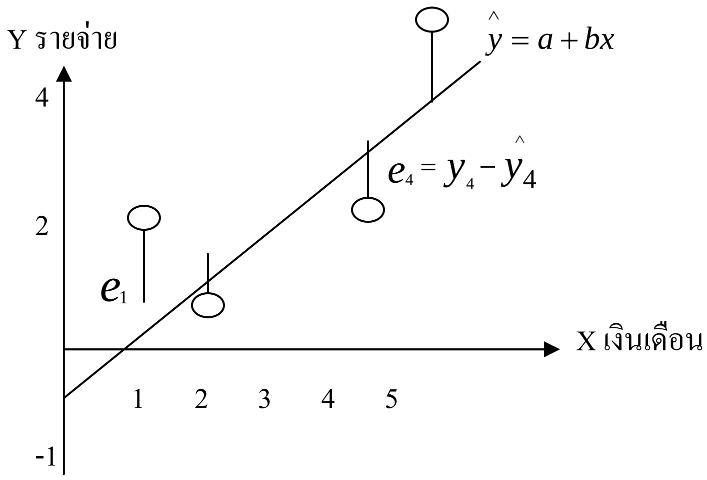
$$\sum \left(y_i - \hat{y}_i \right) = \sum e_i = 0$$

2. จุด $(\bar{X}, \bar{Y})$ เป็นจุดที่อยู่บนเส้นความถดถอย

3. $\sum\left(y_i - \hat{y}_i\right)^2$ มีค่าต่ำสุด

3.5 การประมาณค่าแปรปรวนของความคลาดเคลื่อน

ในการประมาณค่า $y = \beta_0 + \beta_1 x + e$ ด้วย $\hat{y} = a + bx$ ซึ่งจะเกิดค่าคลาดเคลื่อนในการประมาณ y_i ด้วย $\hat{y}_i$ เป็น $e = y_i - \hat{y}_i$ ดังแสดงในรูป



รูปที่ 3.3 แสดงค่าคลาดเคลื่อน

จากหัวข้อที่ 3.3 ได้ว่า $e_j \sim$ normal $(0 , \sigma^2)$

และ $Y_i \sim$ normal $(E(Y_i) , \sigma^2)$

$$\sigma^2 = V(y_i) = V(e_i) = E\left(e_i - E(e_i)\right)^2$$

$$= E(e_i - 0)^2 = E(e_i)^2 = E\left(y_i - \hat{y}_i\right)^2 = \sigma^2_{\bar{y},x}$$

การที่ใช้สัญลักษณ์ $\sigma^2_{\bar{y},x}$ หมายถึงค่าแปรปรวนของ Y ที่เกิดขึ้นเนื่องจากอิทธิพลของ X โดยทั่วไปจะใช้สัญลักษณ์ σ^2 แทน $\sigma^2_{y,x}$

ค่าประมาณของ $\sigma^2_{y,x}$ คือ $S^2_{y,x}$ หรือ S^2 โดยที่ $S^2 = \dfrac{\left[\sum\left(y_i - \hat{y}_i\right)^2\right]}{n-2}$ หรือ

$$S^2 = \frac{SSE}{n-2}$$

3.6 สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

จากหัวข้อที่ 3.1-3.5 เป็นการวิเคราะห์เพื่อหาความสัมพันธ์ของตัวแปร x และ y และเป็นการทดสอบว่า x และ y มีความสัมพันธ์กันหรือไม่ โดยการทดสอบสัมประสิทธิ์ความถดถอย (β_1) โดยใช้ตัวประมาณ (b) ซึ่งไม่สามารถระบุได้ว่า x และ y มีความสัมพันธ์กันมากหรือน้อย เนื่องจากค่า b ที่ได้มีหน่วยตามค่าของ y เช่น y มีหน่วยเป็นบาท ค่า b จะมีหน่วยเป็นบาทด้วย แต่ถ้าค่า y มีหน่วยเป็นล้านบาท ค่า b จะมีหน่วยเป็นล้านบาทด้วย การที่ค่า b ที่มีหน่วยเป็นบาทมีค่ามากกว่าค่า b ที่มีหน่วยเป็นล้านบาท ไม่ได้หมายความว่าความสัมพันธ์ x และ y ที่ y มีหน่วยเป็นบาทจะมีความสัมพันธ์กันมากกว่ากรณีที่ y มีหน่วยเป็นล้านบาท นั่นคือการที่กำหนดให้หน่วยของ y แตกต่างกัน จะทำให้ค่า b แตกต่างกันด้วย (x และ y ไม่จำเป็นต้องมีหน่วยเหมือนกัน)

สำหรับสถิติที่ใช้วัดความสัมพันธ์ระหว่าง x และ y ว่ามากหรือน้อยนั้นจะเรียกว่า สัมประสิทธิ์สหสัมพันธ์ (ρ) ซึ่งในกรณีที่ค่าของ y ขึ้นกับ x เพียงตัวเดียวจะเรียกว่า สัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple Correlation Coefficient) โดยที่ ρ จะไม่มีหน่วย จึงสามารถใช้วัดความสัมพันธ์ระหว่าง y และ x ได้ว่ามีความสัมพันธ์มากหรือน้อยเพียงใด เนื่องจากค่า ρ จะมีค่าสูงสุดเป็น 1 และต่ำสุดเป็น -1

นอกจากนั้นความถดถอยและสหสัมพันธ์ยังมีวิธีการในการเก็บข้อมูลที่แตกต่างกัน คือ สำหรับเรื่องความถดถอยตัวแปรอิสระ x จะต้องถูกกำหนดไว้ล่วงหน้า ในขณะที่เป็นตัวแปรตามหรือ y เป็นตัวแปรสุ่ม แต่สหสัมพันธ์นั้นทั้ง x และ y จะเป็นตัวแปรสุ่มทั้งคู่

เนื่องจากเราใช้ข้อมูลตัวอย่าง จึงประมาณค่า ρ ด้วยค่า r โดยที่ r คือ สัมประสิทธิ์สหสัมพันธ์ตัวอย่าง

ความหมายของค่า r

1. ค่า r เป็นลบ แสดงว่า x และ y มีความสัมพันธ์ในทิศทางตรงข้าม คือถ้า x เพิ่ม y จะลด แต่ถ้า x ลด y จะเพิ่ม

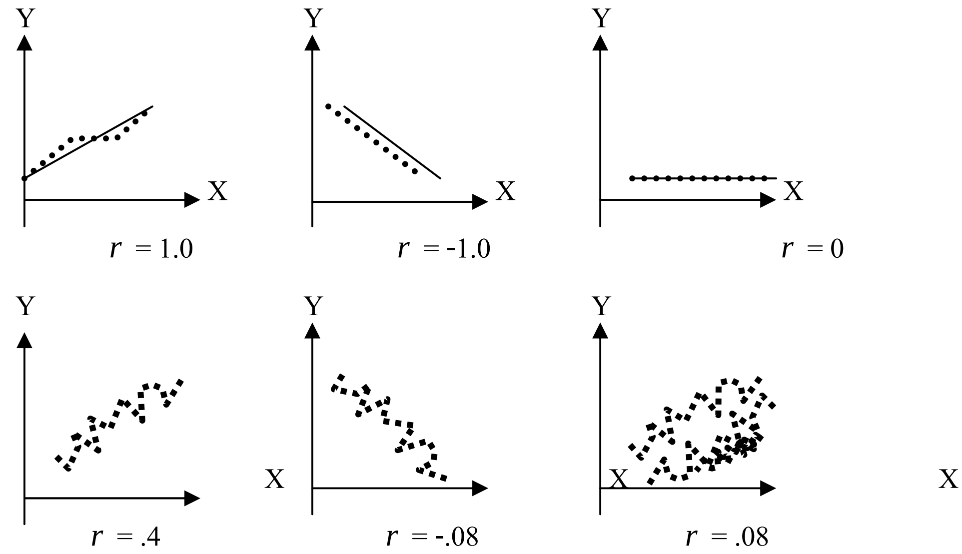
2. ค่า r เป็นบวกแสดงว่า x และ y มีความสัมพันธ์ในทางเดียวกัน คือ ถ้า x เพิ่ม y จะเพิ่มด้วย แต่ถ้า x ลด y จะลดลงด้วย

3. ถ้า r มีค่าเข้าใกล้ 1 หมายถึง x และ y สัมพันธ์ในทิศทางเดียวกันและมีความสัมพันธ์กันมาก

4. ถ้า r มีค่าเข้าใกล้ -1 หมายถึง x และ y สัมพันธ์ในทิศทางตรงกันข้ามและมีความสัมพันธ์กันมาก

5. ถ้า r =0 แสดงว่า x และ y ไม่มีความสัมพันธ์กัน

6. ถ้า r เข้าใกล้ 0 แสดงว่า x และ y มีความสัมพันธ์กันน้อย



รูปที่ 3.4 แสดงค่าของ r ที่มีค่า $-1 < r < 1$

ในทางปฏิบัติการที่จะพิจารณาว่าตัวแปร x และ y มีความสัมพันธ์กันมากหรือน้อย จะพิจารณาจากค่า r (ไม่ใช่จากค่า b) เนื่องจาก r ไม่มีหน่วย และมีขอบเขตคือมีต่ำสุด $= -1$ ค่าสูงสุด $= 1$ นอกจากนั้น r และ b จะมีเครื่องหมายเดียวกัน คือ เป็นบวกเหมือนกันหรือเป็นลบเหมือนกัน เนื่องจากทั้ง r และ b เป็นค่าแสดงถึงความสัมพันธ์ระหว่าง x และ y

ความสัมพันธ์ระหว่างสัมประสิทธิ์การตัดสินใจ และสัมประสิทธิ์สหสัมพันธ์ สำหรับเมื่อมี 2 ตัวแปร

$$\text{สัมประสิทธิ์สหสัมพันธ์} = \sqrt{\text{สัมประสิทธิ์การตัดสินใจ}}$$

หรือ

$$r = \sqrt{r^2}$$

3.7 แนวทางการประมาณค่าพารามิเตอร์

เมื่อต้องการศึกษาธรรมชาติของประชากรหนึ่ง ผู้ศึกษาต้องทราบว่าจะใช้ตัวแปรสุ่มใดแสดงธรรมชาติของประชากรนั้น สมมติว่าผู้ศึกษาจะใช้ตัวแปรสุ่ม x ในการแสดงคุณสมบัติของประชากร x อาจจะเป็นตัวแปรสุ่มมิติเดียว หรือหลายมิติ (เวคเตอร์) ก็ได้ ในการนี้การแจกแจงความน่าจะเป็นของ x มีความสำคัญ นำไปสู่ธรรมชาติหรือคุณสมบัติของประชากรได้ เช่น ทำให้ทราบค่าเฉลี่ยและความแปรปรวนของประชากรได้ โดยทั่วไปแล้วฟังก์ชันการแจกแจง (ความน่าจะเป็น) หรือฟังก์ชันความหนาแน่น (ของความน่าจะเป็น) ของตัวแปรสุ่ม มักขึ้นอยู่กับค่าของพารามิเตอร์ตัวหนึ่งหรือมากกว่า พารามิเตอร์ดังกล่าวมักเป็นตัวไม่ทราบค่า ดังนั้นการประมาณ

ค่าพารามิเตอร์จึงเป็นสิ่งที่ควรศึกษา เมื่อประมาณค่าพารามิเตอร์ได้แล้ว จะทำให้ผู้ศึกษาสามารถประมาณพารามิเตอร์อื่น ๆ ที่สนใจได้อีก เช่น ถ้าตัวแปรสุ่ม x มีฟังก์ชันการแจกแจงขึ้นอยู่กับพารามิเตอร์ θ สิ่งที่แสดงคุณสมบัติบางประการของประชากรมักเป็นฟังก์ชันของ θ ด้วย เช่น ค่าเฉลี่ยของประชากรหรือค่าคาดหมาย และความแปรปรวนของประชากร $\mu = E(X)$ ก็จะเป็นฟังก์ชันของ $\sigma^2 = E(X-\mu)^2$ การประมาณค่าของ θ จึงทำให้ผู้ศึกษาสามารถประมาณ μ และ σ^2 ได้อีกด้วย

ในการประมาณพารามิเตอร์นั้น ต้องใช้ข้อมูลหรือค่าสังเกตจากตัวอย่าง กล่าวคือ จะใช้สถิติบางตัว มาเป็นตัวประมาณพารามิเตอร์ที่สนใจ ในการนี้จะต้องใช้ตัวอย่างสุ่ม เนื่องจากจะทำให้เราได้ข้อมูลโดยไม่ใช้ความรู้สึก ความเห็น ความชอบ ความเชื่อมั่น (ที่อาจผิด) หรือความเอนเอียงที่อาจมีของผู้ศึกษา หรือผู้ทำหน้าที่เก็บรวบรวมข้อมูล สามารถวัดระดับความคลาดเคลื่อนของค่าประมาณได้ นอกจากนี้ทฤษฎีสถิติที่ใช้ในการวิเคราะห์ข้อมูลล้วนมีพื้นฐานมาจากการใช้ตัวอย่างสุ่มทั้งสิ้น ดังนั้นหากใช้ข้อมูลจากตัวอย่างที่มิใช่ตัวอย่างสุ่มแล้ว อาจทำให้ผู้ศึกษาไม่สามารถใช้วิธีการทางสถิติที่มีผู้พัฒนาขึ้นมาแล้วอย่างถูกต้องได้ ไม่สามารถวัดระดับความคลาดเคลื่อนของการประมาณ ผลการศึกษาอาจไม่ใกล้เคียงความเป็นจริง ที่เราจะศึกษาต่อไปจึงอาศัยตัวอย่างสุ่มเป็นพื้นฐานสำคัญ

นิยาม 1. ให้ x เป็นตัวแปรสุ่มที่มีฟังก์ชันความหนาแน่นขึ้นอยู่กับพารามิเตอร์ θ ที่ไม่ทราบค่า เซตของ θ (ซึ่งอาจเป็นเวคเตอร์ $\theta = (\theta_1, \theta_2, \ldots, \theta, \theta_n)$ ในกรณีทั่วไป) ที่เป็นไปได้ ทั้งหมดเรียกว่า **ปริภูมิพารามิเตอร์** (Parameter space) และมักแทนด้วย Ω

ถ้า X มีฟังก์ชันความหนาแน่น f(x) ที่ขึ้นอยู่กับพารามิเตอร์ θ เรามักเขียนฟังก์ชันนั้นในรูป $f(x;\theta)$ หรือ $f(x|\theta)$ ดังนั้น เซต $\{f(x;\theta) : \theta \in \Omega\}$ จึงหมายถึง **วงศ์** (Family) ของฟังก์ชันความหนาแน่นของ x ที่เป็นไปได้ทั้งหมด

ถ้า f(x) มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ย μ และความแปรปรวน σ^2 ฟังก์ชันความหนาแน่นของ X ได้แก่

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , -\infty < X < \infty,\ -\infty < \mu < \infty,\ \sigma^2 > 0$$

ในกรณีนี้ปริภูมิพารามิเตอร์ คือ

$$\Omega = \{(\mu,\sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$$

หรือส่วนบนของระนาบนั่นเอง

ถ้า x มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน $\sigma^2 = \sigma_0^2 =$ ค่าคงที่ที่ทราบค่า

$$f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2}, -\infty < x < \infty, -\infty < \mu < \infty$$

ในกรณีนี้ปริภูมิพารามิเตอร์ คือ

$$\Omega \quad = \{(\mu, \sigma_0^2) : -\infty < \mu < \infty\}$$
$$= \{\mu : -\infty < \mu < \infty\}$$

หรือเส้นจำนวนจริงนั่นเอง

ถ้า x มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย $\mu = 0$ และความแปรปรวน σ^2

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

ในกรณีนี้ปริภูมิพารามิเตอร์ คือ

$$\Omega \quad = \{(0, \sigma^2) : \sigma^2 > 0\}$$
$$= \{\sigma^2 : \sigma^2 > 0\}$$

หรือครึ่งขวาของเส้นจำนวนจริงนั่นเอง

การประมาณค่า (Estimation) ในที่นี้ก็คือ การใช้ข้อมูลจากตัวอย่างสุ่ม ในรูปของ**ค่าสถิติ**เพื่อกะประมาณ หรือคาดหมายว่าพารามิเตอร์ θ ควร**มีค่าเท่าใด**หรือ**อยู่ในช่วงใด** มีการสร้างตัวประมาณ หรือสูตร (ฟังก์ชัน) ที่ใช้ในการหาค่าประมาณ การดำเนินการประมาณพารามิเตอร์จึงมี 2 แนวทางด้วยกัน คือ

1) การประมาณด้วยจุด (Point estimation) หมายถึง การประมาณพารามิเตอร์ออกมาเป็นค่าเดียวหรือจุดเดียว

นิยาม 2. ให้ $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta)$, $\theta \in \Omega$

ตัวประมาณ (Estimator) ของ θ คือ ฟังก์ชันของตัวอย่างสุ่ม $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$

ค่าประมาณ (Estimate) ของ θ คือ ค่าหนึ่งของตัวประมาณ $\hat{\theta}$ ของ θ นั่นคือ ถ้า $X_1 = x_1, \ldots, x_n = x_n$ เป็นข้อมูลจากตัวอย่างสุ่ม $x_1, \ldots, x_n$ ค่าประมาณของ θ คือ $\hat{\theta}(x_1, \ldots, x_n)$

คำว่าตัวประมาณและค่าประมาณพารามิเตอร์ θ มีความแตกต่างกัน เช่นเดียวกับคำว่าฟังก์ชันและค่าของฟังก์ชัน ตัวประมาณเป็นฟังก์ชันของตัวอย่างสุ่ม แต่ค่าประมาณเป็นค่าของฟังก์ชันหรือค่าของตัวประมาณนั่นเอง

ตัวประมาณฟังก์ชันที่บอกหลักเกณฑ์หรือสูตรที่ใช้ในการหาค่าประมาณ

การประมาณด้วยจุด เป็นการค้นหาตัวประมาณที่เหมาะสม เพื่อใช้ประมาณพารามิเตอร์ การหาค่าประมาณกระทำเมื่อได้มีการเก็บรวบรวมข้อมูลมาแล้ว

อย่างไรก็ตาม มีผู้ใช้คำว่า**ค่าประมาณ**ในความหมายของ**ตัวประมาณ**ด้วย แต่ไม่มีใครใช้ตัวประมาณในความหมายของค่าประมาณเลย

2) การประมาณด้วยช่วง (Interval estimation) หมายถึง การประมาณพารามิเตอร์ θ ออกมาเป็นหลายค่า มักเป็นช่วง(ของจำนวนจริง) หรือเซตของจุดหลายจุด เรียกว่า **ช่วงความเชื่อมั่น** (Confidence interval) หรือ**เซตความเชื่อมั่น** (Confidence set) ช่วงหรือเซตดังกล่าวได้มาโดยอาศัยตัวประมาณ $\hat{\theta} = \hat{\theta}(X_1,..., X_n)$ ของ θ มีการใช้การแจกแจงความน่าจะเป็นของ $\hat{\theta}(X_1,..., X_n)$ ในการสร้างช่วงความเชื่อมั่นหรือเซตความเชื่อมั่น และใช้ข้อความที่เกี่ยวกับความน่าจะเป็นประกอบด้วย

3.8 คุณสมบัติที่พึงปรารถนาของตัวประมาณ

ในการประมาณพารามิเตอร์ด้วยจุด เป็นการใช้ฟังก์ชัน $\hat{\theta} = \hat{\theta}(x_1,...,x_n)$ ของตัวอย่างสุ่ม $X_1,...X_n$ ซึ่งอาจใช้ฟังก์ชันที่มีคุณสมบัติที่พึงปรารถนาบางประการเป็นตัวประมาณของ θ การพิจารณาคุณสมบัติของ $\hat{\theta}$ ส่วนใหญ่อาศัยการแจกแจงของการสุ่มตัวอย่าง ในหลายกรณีจะมี $\hat{\theta}$ ที่อาจใช้ได้หลายตัวหรือหลายฟังก์ชัน จึงควรมีหลักเกณฑ์ที่ใช้พิจารณาคุณสมบัติที่พึงปรารถนาของตัวประมาณ เพื่อให้เลือกใช้ตัวประมาณให้เหมาะสม แต่หลักเกณฑ์ที่ใช้เลือกตัวประมาณที่เหมาะสมมีอยู่หลายหลักเกณฑ์ ตัวประมาณตัวหนึ่งอาจมีคุณสมบัติที่พึงปรารถนาได้หลายประการด้วย

นักสถิติมีวิธีการหาตัวประมาณที่เหมาะสมอยู่หลายวิธี ดังจะพิจารณาต่อไป บางครั้งการใช้สามัญสำนึกอาจนำไปสู่ตัวประมาณที่มีคุณสมบัติที่พึงปรารถนาบางประการด้วยซ้ำ เช่น ในการแจกแจงแทบทุกชนิดมักใช้ค่าเฉลี่ยของตัวอย่างเป็นตัวประมาณของค่าเฉลี่ยของประชากร และค่าเฉลี่ยของตัวอย่างมักมีคุณสมบัติที่พึงปรารถนาหลายประการ หลายครั้งวิธีการหา ตัวประมาณของค่าเฉลี่ยของประชากรก็ให้ผลเป็นค่าเฉลี่ยของตัวอย่างด้วย อย่างไรก็ตามยังมีตัวประมาณอื่น ๆ ของค่าเฉลี่ยของประชากรอีก เช่น อาจใช้มัธยฐานตัวอย่าง ค่ากึ่งกลางพิสัยตัวอย่าง ค่าฐานนิยมตัวอย่าง ฯลฯ ในการประมาณค่าเฉลี่ยของประชากร จึงควรเลือกใช้ตัวประมาณให้เหมาะสมกับสถานการณ์ และมีคุณสมบัติที่พึงปรารถนาบางประการ

คุณสมบัติที่พึงปรารถนาของตัวประมาณที่อาจใช้เป็นเกณฑ์ในการประเมินการเลือกใช้ให้เหมาะสมมีหลายประการ ที่จะพิจารณาต่อไปได้แก่

1) ความไม่เอนเอียง (Unbiased ness)
2) ความคงเส้นคงวา หรือความไม่ขัดแย้งกัน (Consistency)
3) ความพอเพียง (Sufficiency)
4) ความมีความแปรปรวนต่ำสุด (Minimum Variance)
5) ความมีประสิทธิภาพ (Efficiency)
6) ความมีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (Minimum mean squared error)
7) ความยืนยง หรือความไม่เปลี่ยนแปลง (Invariance)

3.9 ตัวประมาณไม่เอนเอียง

นิยาม 3. ตัวประมาณ $\hat{\theta} = \hat{\theta}(X_1,..., X_n)$ เป็น**ตัวประมาณไม่เอนเอียง** (Unbiased estimator) ของพารามิเตอร์ θ ก็ต่อเมื่อ

$$\mathrm{E}\left(\hat{\theta}\right) = \theta$$

$\hat{\theta}$ เป็น**ตัวประมาณเอนเอียง**(Biased estimator) ของ θ ก็ต่อเมื่อ $\mathrm{E}\left(\hat{\theta}\right) \neq \theta$ และเรียกผลต่าง $\mathrm{E}\left(\hat{\theta}\right) - \theta$ ว่า**ค่าเอนเอียง** (Bias) ของ $\hat{\theta}$ ในการประมาณ θ

วิธีการหาตัวประมาณที่ไม่เอนเอียงของพารามิเตอร์ทำได้หลายวิธี เช่น

1) ใช้สามัญสำนึก โดยเลือกตัวประมาณ $\hat{\theta}$ ของ θ มาตัวหนึ่ง แล้วหาค่าคาดหมายของตัวประมาณนั้น ถ้า $\mathrm{E}\left(\hat{\theta}\right) = \mathrm{a}\theta + \mathrm{b}$ เมื่อ a , b เป็นค่าคงที่และ $a \neq 0$ เราอาจปรับแก้ได้ $\hat{\theta}_1 = \frac{1}{2}\left(\hat{\theta} - \mathrm{b}\right)$ เป็นตัวประมาณไม่เอนเอียงของ θ

2) ใช้คุณสมบัติของตัวประมาณของ θ ที่ได้มาโดยวิธีการหนึ่ง (ตามรูปแบบ) ที่เป็นตัวประมาณไม่เอนเอียงของ θ เช่น ตัวแบบเชิงเส้นอย่างง่ายของ Y บน X ที่ $\mathrm{Y} = \alpha + \beta X + \varepsilon$ ภายใต้ข้อสมมติฐาน ตัวประมาณแบบกำลังสองต่ำสุด (Least squares estimators) ของ α และ β เป็นตัวประมาณไม่เอนเอียง

3) เลือกตัวประมาณ $\hat{\theta}$ ของ θ มาตัวหนึ่ง แล้วหา $\mathrm{E}\left(\hat{\theta}\right)$ ถ้าได้ $\mathrm{E}\left(\hat{\theta}\right) = \theta$ แสดงว่า $\hat{\theta}$ เป็นตัวประมาณไม่เอนเอียงของ θ แต่ถ้า $\mathrm{E}\left(\hat{\theta}\right) \neq \theta$ พยายามหาตัวประมาณไม่เอนเอียงของค่าเอนเอียง $b(\theta) = \mathrm{E}\left(\hat{\theta}\right) - \theta$ นำมาปรับแก้ $\hat{\theta}$ จะได้ตัวประมาณไม่เอนเอียงของ θ

4) หาวิธีการลดค่าเอนเอียงของ $\hat{\theta}$ ลง เช่นใช้วิธีแจคไนฟ์ (Jackknife) ตามแบบของเควนูลล์ (Quenouille) ที่จะกล่าวต่อไป

ตัวประมาณไม่เอนเอียงของ θ ถ้ามี อาจมีได้หลายตัวก็ได้

ทฤษฎีบทที่ 1. ให้ $X_1,...X_n$ เป็นตัวอย่างสุ่มจากประชากรหนึ่งที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 แล้ว

1) $\hat{\mu} = \sum_{i=1}^{n} a_i X_i$ ที่ $\sum_{i=1}^{n} a_i = 1$ เป็นตัวประมาณไม่เอนเอียงของ μ ที่มีความแปรปรวน $\sigma^2 \sum_{i=1}^{n} a_i^2$

2) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ เป็นตัวประมาณไม่เอนเอียงของ μ ที่มีความแปรปรวน $\frac{\sigma^2}{n}$

3) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ เป็นตัวประมาณไม่เอนเอียงของ σ^2 ที่มีความแปรปรวน $\frac{1}{n}\left[\mu_4 - \frac{n-3}{n-1}\sigma^4\right]$ เมื่อ $\mu_4 = \mathrm{E}(X-\mu)^4$ คือ โมเมนต์ที่ 4 รอบ μ ของประชากร

ทฤษฎีบทที่ 2 ถ้า $\hat{\theta}$ เป็นตัวประมาณไม่เอนเอียงของ θ และ $V\left(\hat{\theta}\right)$ มากกว่า 0 แล้ว $\sqrt{\hat{\theta}}$ เป็นตัวประมาณเอนเอียงของ $\sqrt{\theta}$

พิสูจน์ จากอสมการโคชี-ชวอร์ท [1] (Cauchy-Schwatz inequality)

$$\mathrm{E}\left(\sqrt{\hat{\theta}}\right) \cdot \mathrm{E}\left(\sqrt{\hat{\theta}}\right) \leq \mathrm{E}\left(\sqrt{\hat{\theta}} \cdot \sqrt{\hat{\theta}}\right) = 0$$

เราได้ $\mathrm{E}\left(\sqrt{\hat{\theta}}\right) \leq \sqrt{\theta}$ ซึ่งจะเป็นสมการเมื่อ $V\left(\hat{\theta}\right) = 0$

---

[1] อสมการนี้กล่าวว่าถ้า เป็นตัวแปรสุ่มใด ๆ ซึ่งอสมการจะเป็นสมการ ก็ต่อเมื่อ สำหรับค่าคงที่บางตัว

ดังนั้น $E\sqrt{\hat{\theta}} \neq \sqrt{\theta}$ หรือ $\sqrt{\hat{\theta}}$ เป็นตัวประมาณที่เอนเอียงของ $\sqrt{\theta}$

ในบางกรณีอาจไม่มีตัวประมาณไม่เอนเอียงของพารามิเตอร์เลยก็ได้ เช่นเมื่อประชากรมีการแจกแจงแบบปัวส์ซองที่มีพารามิเตอร์ θ เราสามารถหาตัวประมาณไม่เอนเอียงของ θ ได้ (คือ $\overline{X}$) แต่จะไม่มีตัวประมาณไม่เอนเอียงของ $\frac{1}{\theta}$

ถ้า $\hat{\theta}$ เป็นตัวประมาณเอนเอียงของ θ โดยใช้ตัวอย่างขนาด n เควนูลล์ ได้เสนอวิธีการหาตัวประมาณของพารามิเตอร์ θ จาก $\hat{\theta}$ ที่มีค่าเอนเอียงต่ำลงวิธีหนึ่ง วิธีการของเควนูลล์เป็นดังนี้

ให้ $\hat{\theta}_{n-1}$ เป็นค่าประมาณที่คิดจากค่า n-1 ค่าของตัวอย่างสุ่ม $X_1, \ldots, X_n$ โดยใช้ตัวประมาณ $\hat{\theta}_n$ (รูปเดียวกัน)

ให้ $\hat{\theta}_{n-1}$ เป็นค่าเฉลี่ย (เลขคณิต) ของ $\hat{\theta}_{n-1}$ ที่เป็นไปได้ทั้งหมด (ซึ่ง $\hat{\theta}_{n-1}$ จะมีทั้งสิ้น n ค่า) และให้

$$\hat{\theta}'_n = n\hat{\theta}_n - (n-1)\overline{\hat{\theta}}_{n-1}$$

ผลปรากฏว่า $\hat{\theta}'_n$ มีค่าเอนเอียงน้อยกว่า $\hat{\theta}_n$
ในทำนองเดียวกัน

$$\hat{\theta}''_n = \frac{n^2\,\hat{\theta}'_n + (n-1)^2\,\hat{\theta}'_{n-1}}{n^2 - (n-1)^2}$$

จะมีค่าเอนเอียงน้อยกว่า $\hat{\theta}'_n$ ในการประมาณ θ

3.10 ตัวประมาณคงเส้นคงวา

นิยาม 4 ให้ $\hat{\theta}_n$ เป็นตัวประมาณของ θ ที่มาจากตัวอย่างขนาด n เราถือว่า $\hat{\theta}_n$ เป็น**ตัวประมาณคงเส้นคงวา** หรือ**ตัวประมาณไม่ขัดแย้งกัน** (Consistent estimator) ของ θ ก็ต่อเมื่อ $\hat{\theta}_n$ ลู่เข้าในความน่าจะเป็น (Converge in probability or converge stochastically) ไปหา θ นั่นคือ

เมื่อ $\in$ เป็นจำนวนจริงบวกใด ๆ

$$\lim_{n\to\infty} P\left[\left|\hat{\theta}_n - \theta\right| \geq \in\right] = 0$$

หรือเมื่อ $\in$ และ δ เป็นจำนวนจริงบวกใด ๆ จะมีจำนวนเต็มบวกที่ถ้า n มากกว่า N แล้ว

$$P\left[\left|\hat{\theta}_n - \theta\right| \geq \in\right] \leq \delta$$

ทฤษฎีบทที่ 3 ถ้า $\hat{\theta}_n$ เป็นตัวประมาณของพารามิเตอร์ θ จากตัวอย่างขนาด n และ

$$\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta$$

$$\text{และ} \quad \lim_{n\to\infty} V\left(\hat{\theta}_n\right) = 0$$

แล้ว $\hat{\theta}_n$ เป็นตัวประมาณคงเส้นคงวาของ θ

ทฤษฎีบทที่ 4 ถ้าประชากรมีโมเมนต์ที่ K คือ $\mu_n^{'} = E\left(X^k\right)$ และความแปรปรวนของ X^k เป็นจำนวนจำกัดแล้วโมเมนต์ที่ k ของตัวอย่าง $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ เป็นตัวประมาณคงเส้นคงวาของ $\mu_k^{'}$

ทฤษฎีบทที่ 5 ถ้า $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณคงเส้นคงวาของ θ_1 และ θ_2 ตามลำดับ แล้ว $\theta_1 + \theta_2$ เป็นตัวประมาณคงเส้นคงวาของ θ และ เป็นตัวประมาณคงเส้นคงวาของ $\theta_1.\theta_2$

ทฤษฎีบทที่ 6 ผลบวกและผลคูณของตัวประมาณคงเส้นคงวาจำนวนหนึ่งเป็นตัวประมาณคงเส้นคงวาของผลบวกและผลคูณของพารามิเตอร์ที่สมนัยกัน

ทฤษฎีบทที่ 7 พหุนาม (Polynomial) ของโมเมนต์ตัวอย่างเป็นตัวประมาณคงเส้นคงวาของพหุนามเดียวกันของโมเมนต์ประชากร

นิยาม 5 ตัวประมาณ $\hat{\theta}_n$ ของพารามิเตอร์ θ ที่ได้จากตัวอย่างสุ่มขนาด n เป็น**ตัวประมาณคงเส้นคงวาในความคลาดเคลื่อนกำลังสองเฉลี่ย** (Mean squared error consistent estimator) หรือ**ตัวประมาณคงเส้นคงวาในค่าเฉลี่ยกำลังสอง** (Consistent estimator in quadratic mean) ก็ต่อเมื่อ

$$\lim_{n\to\infty} E\left(\hat{\theta}_n - \theta\right)^2 = 0$$

ทฤษฎีบทที่ 8 ถ้า $\hat{\theta}_n$ เป็นตัวประมาณคงเส้นคงวาในความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์ θ แล้ว $\hat{\theta}_n$ เป็นตัวประมาณคงเส้นคงวาของ θ

นิยาม 6 ตัวประมาณ $\hat{\theta}_n$ ของพารามิเตอร์ θ ที่ได้จากตัวอย่างสุ่มขนาด n เป็น**ตัวประมาณดีที่สุดและแจกแจงเชิงกำกับแบบปกติ** (Best asymptotically normal-BAN estimator) ก็ต่อเมื่อ $\hat{\theta}_n$ เป็นตัวประมาณคงเส้นคงวาของ θ ที่มีการแจกแจงเชิงกำกับแบบปกติและมีความแปรปรวนต่ำที่สุด

ตัวประมาณ $\hat{\theta}_n$ ของ θ ที่มีคุณสมบัตินี้ จะต้องเป็นไปตามเงื่อนไข 3 ประการ คือ

1) $\lim\limits_{n \to \infty} P\left[\left|\hat{\theta}_n - \theta\right| > \epsilon\right] = 0$ เมื่อ $\epsilon > 0$

2) เมื่อ $n \to \infty$ แล้ว $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ มีการแจกแจงใกล้ปกติที่มีค่าเฉลี่ย 0 และความแปรปรวน σ^2 (ที่ขึ้นอยู่กับ θ)

3) ถ้า $\hat{\theta}_n^*$ เป็นตัวประมาณคงเส้นคงวาของ θ ที่ $\sqrt{n}\left(\hat{\theta}_n^* - \theta\right)$ มีการแจกแจงใกล้ปกติที่มีค่าเฉลี่ย 0 และความแปรปรวน σ^{*2} เมื่อ $n \to \infty$ แล้ว $\sigma^2 \le \sigma^{*2}$

3.11 ตัวประมาณที่มีความแปรปรวนต่ำสุด

นิยาม 7 ตัวประมาณ $\hat{\theta}$ ของพารามิเตอร์ θ เป็น**ตัวประมาณที่มีความแปรปรวนต่ำสุด** (Minimum variance estimator - MVE) หรือ**ตัวประมาณที่ดีที่สุด** (Best estimator) ของ θ ก็ต่อเมื่อ $\hat{\theta}$ มีความแปรปรวนมากกว่าความแปรปรวนของตัวประมาณอื่นใดของ θ

ในการเปรียบเทียบความแปรปรวนของตัวประมาณของพารามิเตอร์ θ ควรเปรียบเทียบกับความแปรปรวนของตัวประมาณของ θ ที่มีคุณสมบัติอื่น ๆ เหมือน ๆ กัน ที่นิยมพิจารณากันมาก ได้แก่ ความแปรปรวนของตัวประมาณที่ไม่เอนเอียงของ θ หรือตัวประมาณที่เป็นฟังก์ชันเชิงเส้นของค่าสังเกตและไม่เอนเอียงของ θ

นิยาม 8 ตัวประมาณ $\hat{\theta}$ ของพารามิเตอร์ θ เป็น**ตัวประมาณไม่เอนเอียงที่มีความแปรปรวนต่ำสุด** หรือเป็น**ตัวประมาณไม่เอนเอียงดีที่สุด** (Minimum variance unbiased estimator - MVUE) หรือเป็น**ตัวประมาณไม่เอนเอียงดีที่สุด** (Best unbiased estimator) ของ θ ก็ต่อเมื่อ $\hat{\theta}$ เป็นตัวประมาณไม่เอนเอียงของ θ ที่มีความแปรปรวนต่ำสุดในบรรดาตัวประมาณไม่เอนเอียงของ θ ด้วยกัน

นิยาม 9 ตัวประมาณ $\hat{\theta}$ ของพารามิเตอร์ θ เป็น**ตัวประมาณเชิงเส้นดีที่สุดไม่เอนเอียง** (Best linear unbiased estimator-BLUE) ของ θ ก็ต่อเมื่อ $\hat{\theta}$ เป็นฟังก์ชันกำลังที่หนึ่งของตัวอย่างสุ่มที่ไม่เอนเอียงของ θ และมีความแปรปรวนต่ำสุดในบรรดาตัวประมาณที่เป็นฟังก์ชันกำลังที่หนึ่งของตัวอย่างสุ่มและไม่เอนเอียงของ θ ด้วยกัน

ความแปรปรวนของตัวประมาณบางตัวของพารามิเตอร์ θ จะมีค่าขอบเขตต่ำ (Lower bound) ภายใต้เงื่อนไขบางประการที่จะพิจารณาภายหลัง ค่าขอบเขตล่างเป็นค่าต่ำสุดของความแปรปรวนของตัวประมาณใด ๆ ของ θ

นิยาม 10 ตัวประมาณ $\hat{\theta}$ ของ θ เป็น**ตัวประมาณที่มีความแปรปรวนถึงค่าขอบเขตต่ำสุด** (Minimum variance bound estimator) ของ θ ก็ต่อเมื่อ $\hat{\theta}$ มีความแปรปรวนต่ำที่สุดและเท่ากับค่าขอบเขตล่างของความแปรปรวนของตัวประมาณใด ๆ ของ θ

ทฤษฎีบทที่ 9 ถ้ามีตัวประมาณไม่เอนเอียงที่มีความแปรปรวนต่ำสุดของพารามิเตอร์ θ แล้ว จะต้องมีเพียงตัวเดียว

3.12 วิธีหาตัวประมาณพารามิเตอร์

ให้ $X_1,...X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta),\theta \in \Omega$ ในที่นี้ให้ถือว่าทราบรูปของฟังก์ชัน f แต่ไม่ทราบค่าพารามิเตอร์ θ โดยที่ Ω เป็นปริภูมิพารามิเตอร์ บางครั้งอาจถือว่า $\theta = (\theta_1,...,\theta_k)$ เป็นเวคเตอร์ของพารามิเตอร์ k ตัวที่ θ_i แต่ละตัวเป็นจำนวนจริง นั่นคือถือว่าปริภูมิพารามิเตอร์ □ เป็นปริภูมิยูคลิเดียนที่มีมิติ k

นักสถิติต้องการหาสถิติ (ฟังก์ชัน) $\hat{\theta} = \hat{\theta}(X_1,..., X_n)$ ของตัวอย่างสุ่ม $X_1,...X_n$ ที่ใช้ประมาณ θ หรือฟังก์ชันของ θ เช่น $g_1(\theta), g_2(\theta),..., g_r(\theta)$ เป็นต้น

มีผู้เสนอวิธีการที่ใช้หา $\hat{\theta}$ ไว้หลายวิธี แต่ละวิธีมักใช้สามัญสำนึกหรือมีเหตุผลประกอบ ในที่นี้จะพิจารณาวิธีหาตัวประกอบ $\hat{\theta}$ ต่อไปนี้

1. วิธีโมเมนต์ (Method of moment)
2. วิธีภาวะน่าจะเป็นสูงสุด (Method of maximum likelihood)
3. วิธีไคกำลังสองต่ำสุด (Method of minimum chi-square)
4. วิธีระยะห่างต่ำสุด (Method of minimum distance)
5. วิธีของเบส์ (Method of Bayes)
6. วิธีมินิแมกซ์ (Method of minimax)
7. วิธีกำลังสองต่ำสุด (Method of least squares)
8. วิธีของพิทมัน (Method of Pitman)

3.13 วิธีโมเมนต์

วิธีการ การหาตัวประมาณด้วยวิธีโมเมนต์ ทำได้ง่ายมาก เป็นวิธีการหาตัวประมาณของพารามิเตอร์เก่าแก่ที่สุด นำเสนอครั้งแรกโดยคาร์ลเพียร์สัน (Karl Pearson) เมื่อประมาณปี ค.ศ. 1894 วิธีการนี้มักให้ตัวประมาณได้เสมอในหลายกรณี และมักจะทำให้ได้ตัวประมาณที่ยังปรับปรุงให้ดีขึ้นได้อีก แต่เป็นการเริ่มต้นที่ดีในกรณีที่วิธีอื่นทำให้หาตัวประมาณได้ยาก

แนวความคิดที่นำมาใช้หาตัวประมาณด้วยวิธีโมเมนต์ ได้แก่ การถือว่าโมเมนต์ตัวอย่างเป็นตัวประมาณที่ไม่เอนเอียงของโมเมนต์ของประชากรที่สมนัยกัน และฟังก์ชันของโมเมนต์ตัวอย่างก็อาจใช้เป็นตัวประมาณฟังก์ชันที่สมนัยกันของพารามิเตอร์ได้

ให้ $X_1,...X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\underline{\theta}),\underline{\theta}\in\Omega$ ซึ่งมีโมเมนต์ที่ k คือ อาจหา $\mu'_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x;\underline{\theta})dx$ ได้ อันจะทำให้มีโมเมนต์ที่ 1, 2, …, k-1 ด้วย

นิยาม 11 เมื่อประชากรมีฟังก์ชันความหนาแน่นขึ้นอยู่กับพารามิเตอร์ $\theta_1,\theta_2,...,\theta_k$ และมีโมเมนต์ที่ k คือ μ'_k และให้โมเมนต์ที่ k ของตัวอย่างเป็น $M'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ **ตัวประมาณแบบโมเมนต์** ของ $\theta_1,...,\theta_k$ คือค่าของพารามิเตอร์เหล่านี้เทอมของค่าสังเกต $X_1,...X_n$ ที่ได้จากการแก้สมการ

$$\mu'_j = M'_j \qquad j = 1, 2, \ldots, k$$

3.14 ปัญหาในการหาตัวประมาณแบบโมเมนต์

การใช้วิธีโมเมนต์หาตัวประมาณพารามิเตอร์ หรือฟังก์ชันของพารามิเตอร์เป็นจุดเริ่มต้นที่ดี แต่ผลที่ได้อาจไม่เป็นตัวประมาณที่มีคุณภาพพึงปรารถนาเสมอไปในการหาตัวประมาณอาจประสบปัญหาเกี่ยวกับการแก้สมการได้ ซึ่งอาจได้ใช้การประมาณด้วยตัวเลขให้ได้ค่าใกล้เคียงในบางกรณี

ในการประมาณฟังก์ชันของพารามิเตอร์ เช่น $g_1(\theta_1,...,\theta_k), g_2(\theta_1,...,\theta_k),...,g_r(\theta_1,...,\theta_k)$ โดยวิธีโมเมนต์ อาจดำเนินการได้หลายทาง

1. ใช้วิธีโมเมนต์หาตัวประมาณของ $\theta_1,...,\theta_k$ คือให้ $\mu'_j = M'_j, j=1,...,k$ แล้วแก้สมการหาค่าของ θ_i ในเทอมของค่าสังเกตเป็นตัวประมาณ $\hat{\theta}_j$ ของ $\theta_j, j=1,...,k$ แล้วใช้ $\hat{\theta}_j$ ในการประมาณ $g_1,...,g_r$ คือ $\hat{g}_1 = g_1\left(\hat{\theta}_1,...,\hat{\theta}_k\right),...,\hat{g}_r = g_r\left(\hat{\theta}_1,...,\hat{\theta}_k\right)$

2. ใช้โมเมนต์ $\mu_j^{'} = E(X^j), j = 1,2,...,k$ ที่อยู่ในรูปฟังก์ชันของ $g_1,...,g_r$ คือ $\mu_j^{'} = \mu_j^{'}(g_1,...,g_r), j = 1,...,k$ แล้วหา $g_1,...,g_r$ จากสมการ $\mu_j^{'}(g_1,...,g_r) = M_J^{'}$ ในเทอมของค่าสังเกตเป็นตัวประมาณแบบโมเมนต์ของ $g_1,...,g_r$

วิธีการทั้งสองที่กล่าวนี้ ไม่จำเป็นต้องให้ผลลัพธ์เดียวกัน

3.15 คุณสมบัติของตัวประมาณแบบโมเมนต์

ตัวประมาณพารามิเตอร์ที่ได้มาโดยวิธีโมเมนต์มีคุณสมบัติที่น่าสนใจหลายประการ

1. โดยทั่วไปตัวประมาณแบบโมเมนต์เป็นตัวประมาณคงเส้นคงวา เนื่องจากตัวประมาณแบบนี้เป็นฟังก์ชันของโมเมนต์ตัวอย่าง และโมเมนต์ตัวอย่างเป็นตัวประมาณคงเส้นคงวาของโมเมนต์ของประชากร

2. ภายใต้เงื่อนไขทั่วไป ตัวประมาณแบบโมเมนต์มีการแจกแจงเชิงกำกับแบบปกติ (Asymptotically normal)

3. ตัวประมาณแบบโมเมนต์อาจเอนเอียงหรือไม่เอนเอียงก็ได้ หากเป็นตัวประมาณเอนเอียงอาจลดค่าเอนเอียงลงได้ เช่นใช้วิธีแจคไนฟ์ตามแบบของเควนูลล์ที่กล่าวมาแล้ว

4. ค่าเอนเอียงของตัวประมาณแบบโมเมนต์มักมีค่าน้อย (ในอันดับหรือขนาด $n-1$) และส่วนเบี่ยงเบนมาตรฐานมักอยู่ในรูป $c/\sqrt{n}$

5. ตัวประมาณแบบโมเมนต์มักมีประสิทธิภาพต่ำ บางครั้งอาจได้ตัวประมาณแบบโมเมนต์ที่ขัดแย้งกับความรู้สึก หรือสามัญสำนึก

3.16 วิธีภาวะน่าจะเป็นสูงสุด

วิธีการ การประมาณพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (เป็นวิธีการที่ใช้กันแพร่หลายมากที่สุด) มีแนวคิดมานานตั้งแต่คริสศตวรรษที่ 18 คาร์ล ฟรีดริค เกาส์ (Karl Friedrich Gauss 1777-1855) และแดเนียล เบอร์นูลลี (Daniel Bernoulli) ได้ใช้วิธีการนี้มาแล้ว ต่อมาในศตวรรษที่ 20 โรนัลด์ ไอล์เมอร์ ฟิเชอร์ (Ronald Aylmer Fisher 1890-1962) ได้ทำการศึกษาคุณสมบัติของวิธีการนี้ ทำให้มีผู้ใช้กว้างขวางขึ้น และถือว่าวิธีการนี้เป็นผลงานของ ฟิเชอร์ โดยเขาได้นำเสนองานเกี่ยวกับวิธีการนี้ในปี ค.ศ. 1912 พร้อมทั้งมีการปรับปรุงแก้ไขส่วนที่เกี่ยวข้องให้เหมาะสมขึ้นอีกด้วย นักสถิติคนอื่น ๆ ก็มีส่วนทำให้วิธีการนี้เป็นที่นิยมแพร่หลายยิ่งขึ้นด้วย

นิยาม 12. ให้ $X_1,\ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta), \theta \in \Omega$ **ฟังก์ชันภาวะน่าจะเป็น** (Likelihood function) ของตัวอย่างสุ่ม คือ ฟังก์ชันความหนาแน่นรวม $L = L(\theta; x_1,..., x_n) = L(\theta)$ ของตัวอย่างสุ่มนั้นที่ถือว่าเป็นฟังก์ชันของพารามิเตอร์ นั่นเอง

$$L(\theta; x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

นิยาม 13. ค่าของพารามิเตอร์ θ ในเทอมของค่าสังเกตของตัวอย่างสุ่ม $X_1,\ldots,X_n$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่าสูงสุด เรียกว่า ตัวประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum likelihood estimator-MLE) ของ θ นั่นคือค่าของ θ คือ $\hat{\theta} = \hat{\theta}(X_1,\ldots,X_n)$ เป็น MLE ของ θ ก็ต่อเมื่อ $L\left(\hat{\theta}\right) = L\left(\hat{\theta}(X_1,\ldots,X_n)\right)$ มีค่าสูงสุด

วิธีหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด เป็นวิธีการหาค่าของพารามิเตอร์ θ ที่ทำให้ฟังก์ชัน $L(\theta)$ สูงสุด ในการนี้มีข้อควรสังเกตดังต่อไปนี้

1. เป้าหมายในการหา MLE ของ θ คือการหาค่า θ เรียกว่า $\hat{\theta} = \hat{\theta}(X_1,\ldots,X_n)$ ที่ทำให้

$$L\left(\hat{\theta}; x_1,\ldots,X_n\right) \geq L(\theta; x_1,\ldots,x_n)$$

เมื่อ $\theta = \Omega$ และ $X_1 = x_1,\ldots,x_n = x_n$

2. ถ้าฟังก์ชันภาวะน่าจะเป็น $L(\theta)$ เป็นฟังก์ชันที่หาอนุพันธ์ได้ (Differentiable function) เมื่อเทียบกับ θ อาจใช้อนุพันธ์หา MLE ของ θ ได้ เมื่อเรนจ์ของ $f(x;\theta)$ ไม่ขึ้นอยู่กับ θ และ θ อยู่ในช่วงจำนวนจริงช่วงหนึ่ง ในกรณีดังกล่าว $\hat{\theta}$ คือรากของสมการ

$$\frac{\partial L}{\partial \theta} = 0$$

เงื่อนไขพอเพียง(Sufficient condition) ที่ $\hat{\theta}$ ทำให้ $L(\hat{\theta}) \geq L(\theta)$, $\theta \in \Omega$ คือ

$$\frac{\partial^2 L}{\partial \theta^2} < 0 \qquad \text{เมื่อ } \theta = \hat{\theta}$$

3. การใช้อนุพันธ์หา MLEในหลายกรณีใช้ lnL จะสะดวกกว่าที่จะใช้ L ควรสังเกตว่า

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta}$$

และ $L > 0$ ดังนั้น เมื่อให้

$$\frac{\partial \ell n L}{\partial \theta} = 0$$

เราจะได้ $\frac{\partial L}{\partial \theta} = 0$ ด้วย นอกจากนั้น เมื่อ

$$\frac{\partial^2 \ln L}{\partial \theta^2} < 0$$

ก็จะทำให้ $\frac{\partial^2 L}{\partial \theta^2} < 0$ ด้วย

นิยาม 14. สมการที่ใช้หา MLE คือ $\frac{\partial L}{\partial \theta} = 0$ หรือ $\frac{\partial \ln L}{\partial \theta} = 0$ เราเรียกว่า **สมการภาวะน่าจะเป็น** (Likelihood equation)

4. ในบางกรณี อาจไม่สามารถใช้อนุพันธ์ในการหา MLE เช่นเมื่อเรนจ์ของ $f(x;\theta)$ ขึ้นอยู่กับ θ

คุณสมบัติของตัวประมาณแบบภาวะน่าจะเป็นสูงสุด ตัวประมาณแบบภาวะน่าจะเป็นสูงสุดมีคุณสมบัติที่น่าสนใจหลายประการ ซึ่งจะรวบรวมนำเสนอไว้บางประการต่อไปนี้

1. ตัวประมาณแบบภาวะน่าจะเป็นของพารามิเตอร์ θ เป็นฟังก์ชันของสถิติพอเพียงของ θ (ซึ่งจะพิสูจน์ให้เห็นจริงภายหลัง)

2. ในกรณีที่มีตัวประมาณซึ่งมีความแปรปรวนต่ำสุดเท่าค่าขอบเขตล่างของพารามิเตอร์ θ ตัวประมาณนั้นอาจหาได้โดยวิธีภาวะน่าจะเป็นสูงสุด (ซึ่งจะพิสูจน์ให้เห็นจริงภายหลัง)

3. ตัวประมาณแบบภาวะน่าจะเป็นสูงสุดอาจมีได้มากกว่า 1 ตัว

ทฤษฎีบทที่ 10 ถ้า $\hat{\theta}_n$ เป็นตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ θ แล้ว $\lim_{n\to\infty} P\left[\hat{\theta}_n = \theta\right] = 1$

ทฤษฎีบทที่ 11 ภายใต้เงื่อนไขธรรมดา ถ้า $\hat{\theta}$ เป็นตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของ θ ที่คงเส้นคงวาแล้ว มีได้เพียงตัวเดียว $\hat{\theta}$

ทฤษฎีบทที่ 12 ถ้า $\hat{\theta}$ เป็นตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ θ และมีอนุพันธ์ของฟังก์ชันภาวะน่าจะเป็นอันดับที่ 1 และ 2 เทียบกับ θ ในช่วงที่คลุมค่าแท้จริง θ_0 ของ θ และ $R^2(\theta) = \mathrm{E}\left(\frac{\partial \ell nL}{\partial \theta}\right)^2$ ต่างจาก 0 เมื่อ θ อยู่ในช่วงดังกล่าว แล้ว $\hat{\theta}$ มีการแจกแจงเชิงกำกับแบบปกติที่มีค่าเฉลี่ย θ_0 และความแปรปรวน $1/R^2(\theta_0)$

ปัญหาในการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด ในการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ θ เราจะพยายามหาค่าของ θ ในเทอมของค่าสังเกต $X_1, \ldots, X_n$

ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น L(θ) มีค่าสูงสุด เมื่อใช้อนุพันธ์ นักสถิติจะหาค่าของ θ ดังกล่าวจากสมการภาวะน่าจะเป็น $\frac{\partial L}{\partial \theta} = 0$ หรือ $\frac{\partial \ell nL}{\partial \theta} = 0$ คือหาจุดวิกฤต (Critical point) ที่จะทำให้ L สูงสุดนั่นเอง บางครั้งการแก้สมการดังกล่าวอาจทำได้ยากเช่นเมื่อสมการภาวะน่าจะเป็นเป็นสมการระดับขั้นสูง ๆ หรือเป็นสมการซับซ้อน

ในบางกรณี ไม่อาจใช้อนุพันธ์เพื่อหาค่าของพารามิเตอร์ θ ในเทอมของค่าสังเกต $X_1,\ldots, X_n$ ได้เช่นเมื่อ L หรือ ℓnL เป็นฟังก์ชันของ θ ที่หาอนุพันธ์ไม่ได้ หรือเมื่อเรนจ์ของฟังก์ชันขึ้นอยู่กับ θ หรืออนุพันธ์ไม่มี θ อยู่ด้วย จึงต้องใช้วิธีการอื่น ๆ เช่นการสังเกตว่า L(θ) จะสูงสุดได้เมื่อไร หรือการเปรียบเทียบค่าของ L(θ) เมื่อค่าของ θ เปลี่ยนไป

แม้ในกรณีที่ใช้อนุพันธ์ในการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด ควรมีการตรวจสอบเงื่อนไขพอเพียงที่ค่าของ θ นั้นจะทำให้ L(θ) สูงสุด มิใช่ให้ค่าต่ำที่สุด

ในกรณีที่ใช้อนุพันธ์ในการหา θ ที่ทำให้ L(θ) มีค่าสูงสุด แต่ไม่อาจแก้สมการภาวะน่าจะเป็น หรือแก้ได้ยาก อาจใช้การประมาณ (Approximation) ค่าของ $\hat{\theta} = \hat{\theta}(x_1,\ldots,x_n)$ โดยอาศัยค่าสังเกต $X_1 = x_1,\ldots, X_n = x_n$ ตามวิธีการของนิวตัน (Newton's approximation) ได้เมื่อ n มีค่ามาก

ทฤษฎีบทที่ 13 ค่าประมาณของรากของสมการภาวะน่าจะเป็น $\frac{\partial \ell nL}{\partial \theta} = 0$ ได้แก่

$$\hat{\theta} = \theta_0 + \left(\frac{\partial \ell nL}{\partial \theta}\right)_{\theta=\theta_0} V\left(\hat{\theta}\right)$$

โดยที่ $V\left(\hat{\theta}\right) = -1 \Big/ E\left(\frac{\partial^2 \ell nL}{\partial \theta^2}\right)$

พิสูจน์ จากอนุกรมเทย์เลอร์ จะได้ คือ

$$0 = \left(\frac{\partial \ell nL}{\partial \theta}\right)_{\theta=\hat{\theta}} = \left(\frac{\partial \ell nL}{\partial \theta}\right)_{\theta=\theta_0} + \left(\hat{\theta}+\theta_0\right)\left(\frac{\partial^2 \ell nL}{\partial \theta^2}\right)_{\theta=\theta^*}$$

เมื่อ θ^* อยู่ระหว่าง $\hat{\theta}$ กับ θ_0 ซึ่งเป็นค่าที่แท้จริงของ θ

$$\lim_{n\to\infty} P\left[\theta^* = \theta_0\right] = 1$$

และ $$\ell im_{n\to\infty}\ P\left[\left.\frac{\partial^2 \ell nL}{\partial \theta^2}\right|_{\theta=\theta_0} = E\left(\frac{\partial^2 \ell nL}{\partial \theta^2}\right)\right] = 1$$

หรือ $$\lim_{n\to\infty} P\left[\frac{\partial^2 \ell nL}{\partial\theta^2} = -\frac{1}{V\left(\hat{\theta}\right)}\right] = 1$$

ดังนั้น เมื่อ n มีค่ามาก จะได้

$$0 = \left(\frac{\partial \ell nL}{\partial\theta}\right)_{\theta=\theta_0} - \left(\hat{\theta} - \theta_0\right)\frac{1}{V\left(\hat{\theta}\right)}$$

หรือ $$0 = \theta_0\left(\frac{\partial \ell nL}{\partial\theta}\right)_{\theta=\theta_0} \cdot V\left(\hat{\theta}\right)$$

โดยที่ $$V\left(\hat{\theta}\right) = -1 \Big/ E\left(\frac{\partial^2 \ell nL}{\partial\theta^2}\right) = \frac{1}{E\left(\frac{\partial \ell nL}{\partial\theta}\right)^2}$$

ให้ θ_0 เป็นค่าประมาณเริ่มต้นของ $\hat{\theta}$ ค่าของ $\hat{\theta}$ ที่หาได้ในรอบแรกจะใช้เป็น θ_0 ในรอบที่ 2 เช่นนี้ซ้ำ ๆ ได้หลายรอบ จนกระทั่งได้ค่าของ $\hat{\theta}$ ที่ค่อนข้างนิ่ง คือ มีค่าต่างจากค่าที่หาได้ในรอบก่อนไม่มากนัก

ตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของพารามิเตอร์หลายตัว ในกรณีที่พารามิเตอร์ θ เป็นเวคเตอร์ $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ อาจหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของ $\theta_1, ..., \theta_k$ ที่ปรากฏอยู่ในฟังก์ชันความหนาแน่น $f(x;\theta_1, ..., \theta_k)$ ได้ โดยใช้หลักเกณฑ์เดิม นอกจากนั้นตัวแปรสุ่มแต่ละตัวของตัวอย่างสุ่ม ยังอาจเป็นเวคเตอร์ได้ด้วย

นิยาม 15. เมื่อ $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta_1, ..., \theta_k)$ ค่าของ $\theta_1, ..., \theta_k$ ในเทอมของค่าสังเกตของตัวอย่างสุ่มที่ทำให้ฟังก์ชันภาวะน่าจะเป็น $L(\theta_1, ..., \theta_k)$ มีค่าสูงสุด เรียกว่า **ตัวประมาณแบบภาวะน่าจะเป็นสูงสุด** ของ $\theta_1, ..., \theta_k$ นั่นคือ $\hat{\theta}_1(x_1, ..., x_n), ..., \hat{\theta}_k(x_1, ..., x_n)$ เป็น MLE ของ $\theta_1, ..., \theta_k$ ตามลำดับ ก็ต่อเมื่อ $L\left(\hat{\theta}_1, ..., \hat{\theta}_k\right)$ มีค่าสูงสุด

เงื่อนไขจำเป็นที่ทำให้ L($\theta_1,...,\theta_k$) สูงสุด ได้แก่ $\dfrac{\partial L(\theta_1,...,\theta_k)}{\partial\theta_j}=0, j=1,...,k$ หรือที่สมมูลย์กันก็คือ $\dfrac{\partial \ell n L(\theta_1,...,\theta_k)}{\partial\theta_j}=0, j=1,...,k$ และเงื่อนไขพอเพียงที่จะได้ L สูงสุดได้แก่ การที่เมตริกซ์ $\left(\dfrac{\partial^2 \ell n L}{\partial\theta_i\partial\theta_j}\right)$ เป็นเมตริกซ์นิเสธแน่ (Negative definite matrix) ขนาด k x k

เมื่อสามารถใช้อนุพันธ์ในการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ $\theta_1,...,\theta_k$ ได้ เราเรียกระบบสมการ $\dfrac{\partial L}{\partial\theta_j}=0$ หรือ $\dfrac{\partial \ell nL}{\partial\theta_j}=0$ ว่า **ระบบสมการภาวะน่าจะเป็น**

3.17 วิธีไคกำลังสองต่ำสุด

นิยาม 16. ให้ $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $fax;\theta f$ ให้ S เป็นปริภูมิตัวอย่างของตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นดังกล่าว มีการแบ่งกั้น (Partition) S ออกเป็น $S_1\cup S_2,...\cup S_k$ โดยที่ $S_i\cap S_j=\Phi$ เมื่อ $i\neq j$ ให้ $P_j=P_j(\theta)$ เป็นความน่าจะเป็นที่ X_i มีค่าอยู่ใน $S_i, \sum_{j=1}^{k}P_j=1$ ให้ n_j เป็นจำนวน X_i ในตัวอย่างมีค่าตกอยู่ใน $S_i, \sum_{j=1}^{k}n_j=n$ ค่าของ θ ในเทอมของค่าสังเกตที่ทำให้

$$X^2=n\sum_{j=1}^{k}\left[\frac{n_j}{n}-P_j(\theta)\right]^2 \Big/ P_j(\theta) \qquad (1)$$

มีค่าต่ำสุด ที่เรียกว่า **ตัวประมาณแบบไคกำลังสองต่ำสุด** (Minimum chi-squares estimator) ของ θ รูปอื่นของ X^2 ใน (1) ได้แก่

$$X^2=\sum_{j=1}^{k}\left[n_j-np_j(\theta)\right]^2 \Big/ np_j(\theta) \qquad (2)$$

และฟังก์ชันการแจกแจงของตัวอย่าง คือ

$$F_n(x)=\begin{cases}0 & \text{เมื่อ } x<0\\ \dfrac{n_0}{n} & \text{เมื่อ } 0\le x<1\\ 1 & \text{เมื่อ } x\ge 1\end{cases}$$

ฟังก์ชันระยะห่างระหว่าง $F(x;\theta)$ กับ $F_n(x)$ คือ

$$d(F(x;\theta),F_n(x)) = \sup_x |F(x;\theta) - F_n(x)|$$

ดังนั้น $d(F(x;\theta),F_n(x))$ มีค่าต่ำสุดเมื่อ $1-\theta = \frac{n_0}{n}$ หรือ $\theta = 1 - \frac{n_0}{n} = \frac{n_1}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

นั่นคือ ตัวประมาณแบบระยะห่างต่ำสุด คือ $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}$

3.18 วิธีของเบส์

แนวความคิด สถิติตามแนวของเบส์ (Bayesian approach) แตกต่างจากสถิติตามแนวเดิม (Chassical approach) ในแนวเดิม การประมาณพารามิเตอร์ θ มักถือว่าเริ่มจากการสุ่ม ตัวอย่างจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta)$ และถือว่าพารามิเตอร์ θ เป็นค่าคงที่ แต่ไม่ทราบค่า แต่ในแนวของเบส์จะมีความพยายามใช้ความรู้เดิมหรือข้อมูลเดิมเกี่ยวกับ θ ให้เป็นประโยชน์ในการประมาณ θ ให้ได้ดียิ่งขึ้น ดังนั้นจึงถือว่า θ เป็นค่าของตัวแปรสุ่ม Θ มีการแจกแจงความน่าจะเป็นรูปใดรูปหนึ่ง

ให้ $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta) = f(x|\theta)$ โดยที่ θ เป็นค่าของตัวแปรสุ่ม Θ

ในที่นี้ เราถือว่า $f(x|\theta)$ เป็นฟังก์ชันความหนาแน่นเงื่อนไข (Conditional probability density function) ของตัวแปรสุ่ม เมื่อกำหนดให้ว่า $\Theta = \theta$

นิยาม 17. ให้ตัวแปรสุ่ม Θ มีฟังก์ชันความหนาแน่น g (θ) ซึ่งเรียกว่า **ฟังก์ชันความหนาแน่นเริ่ม** (Prior or initial probability density function) ของ Θ

ให้ $h(\theta|X_1,\ldots,X_n)$ เป็นฟังก์ชันความหนาแน่น แบบมีเงื่อนไขของตัวแปรสุ่ม Θ เมื่อกำหนดให้ว่า $X_1 = x_1,\ldots,X_n = x_n$ มักเรียกฟังก์ชันนี้ว่า **ฟังก์ชันความหนาแน่นภายหลัง** (Posterior probability density function) หรือ **ฟังก์ชันความหนาแน่นปรับแก้** (Revised probability density function) ของ Θ

ในที่นี้ ฟังก์ชันความหนาแน่นร่วมของ $X_1,\ldots,X_n$ เมื่อกำหนดให้ว่า Θ =0 ได้แก่ $f(x_1,\ldots,x_n;\theta) = f(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f(x_1|\theta)$

อาจหาฟังก์ชันความหนาแน่นภายหลังของ Θ ได้จากฟังก์ชันความหนาแน่นเริ่มต้น และฟังก์ชันความหนาแน่นแบบมีเงื่อนไขของตัวอย่างสุ่ม โดยใช้ความรู้เกี่ยวกับฟังก์ชันความหนาแน่นแบบมีเงื่อนไข

ให้ Θ มีฟังก์ชันความหนาแน่นเริ่มต้นเป็น $g(\theta)$
ดังนั้น ฟังก์ชันความหนาแน่นร่วมของ $X_1,...,X_n$ และ Θ ได้แก่

$$f\left(x_1,...,x_n|\theta\right)g|\theta) = \prod_{i=1}^{n} f\left(x_1|\theta\right)g(\theta))$$

ฟังก์ชันความหนาแน่นร่วมของ $X_1,...,X_n$ (ไม่มีเงื่อนไข คือ $f\left(x_1,...,x_n|\theta\right)g(\theta))d\theta$
ดังนั้น ฟังก์ชันความหนาแน่นภายหลังของ Θ ได้แก่

$$h(\theta|x_1,...,x_m)) = \frac{f\left(x_1,...,x_n|\theta\right)g(\theta))}{\int f\left(x_1,...,x_n|\theta\right)g(\theta))d\theta} \qquad (3.33)$$

หรือ
$$h(\theta|x_1,...,x_m)) = \frac{\prod_{i=1}^{n} f\left(x_i|\theta\right)g(\theta))}{\int \prod_{i=1}^{n} f\left(x_i|\theta\right)g(\theta))d\theta} \qquad (3.34)$$

ตัวประมาณแบบเบส์ภายหลัง

นิยาม 18 ให้ $X_1,...,X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f\left(x|\theta\right)$ โดยที่ θ เป็นค่าของตัวแปรสุ่ม Θ ที่มีฟังก์ชันความหนาแน่นเริ่มต้น $g(\theta)$

ตัวประมาณแบบเบส์ภายหลัง (Posterior Bayes estimator) ของฟังก์ชัน $t(\theta)$ ของ θ เทียบกับฟังก์ชันเริ่มต้น $g(\theta)$ ได้แก่ $\hat{t}(\theta) = \mathrm{E}\left(t(\Theta)|X_1,...,X_n\right)$ นั่นคือ

$$\hat{t}(\theta) = \int t(\theta)h\left(\theta|x_1,...,x_n\right)d\theta$$

หรือ
$$\hat{t}(\theta) = \frac{\int t(\theta)\prod_{i=1}^{n} f\left(X_i|\theta\right)g(\theta)d\theta}{\int \prod_{i=1}^{n} f\left(X_i|\theta\right)g(\theta)d\theta}$$

3.19 วิธีมินิแมกซ์

ในการประมาณค่าของพารามิเตอร์ โดยการพิจารณาฟังก์ชันสูญเสีย มักต้องการตัวประมาณที่มีความสูญเสียต่ำ สิ่งที่ควรพิจารณาจึงได้แก่ ค่าคาดหมายของความสูญเสีย หรือการเสี่ยง นั่นเอง

นิยาม 19 ตัวประมาณ $T_1\left(X_1,...,X_n\right)$ ของ $g(\theta)$ **ดีกว่า** ตัวประมาณ $T_2\left(X_1,...,X_n\right)$ ก็ต่อเมื่อ $R(\theta,T_1) \le R(\theta,T_2)$ ทุกค่า θ และ $R(\theta,T_1) < R(\theta,T_2)$ อย่างน้อย θ ค่าหนึ่ง

ตัวประมาณ $T_1\left(X_1,...,X_n\right)$ ของ $g(\theta)$ **เป็นตัวประมาณที่ยอมรับได้** ก็ต่อเมื่อไม่มีตัวประมาณใดของ $g(\theta)$ ที่ดีกว่า $T_1\left(X_1,...,X_n\right)$

นิยาม 20 ตัวประมาณ $T_0(X_1,...,X_n)$ ของ g(θ) เป็นตัวประมาณแบบมินิแมกซ์ (Minimax Estimator) ก็ต่อเมื่อ

$$\sup_{\theta\in\Omega} R(\theta,T_0) \le \sup_{\theta\in\Omega} R(\theta,T) \tag{3.43}$$

ไม่ว่า $T_1(X_1,...,X_n)$ จะเป็นตัวประมาณใด ๆ ของ g(θ)

ตัวประมาณแบบมินิแมกซ์ จึงเป็นตัวประมาณที่มีการเสี่ยงสูงสุด ต่ำกว่าหรือเท่ากับการเสี่ยงสูงสุดของตัวประมาณใด ๆ ของฟังก์ชันของพารามิเตอร์ที่สนใจ

ทฤษฎีบทที่ 14 ถ้า $T_0(X_1,...,X_n)$ เป็นตัวประมาณแบบเบส์ของ t(θ) ที่มีการเสี่ยง คงที่ แล้ว T_0 เป็นตัวประมาณแบบมินิแมกซ์ของ t(θ)

3.20 วิธีกำลังสองต่ำสุด

แนวความคิดและวิธีการหาตัวประมาณแบบกำลังต่ำสุด วิธีการนี้เป็นวิธีการสำคัญมีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of linear estimation) โดยไม่จำเป็นต้องทราบรูปการแจกแจงความน่าจะเป็น แต่อาศัยผลต่างระหว่างค่าสังเกตและค่าคาดหมายเป็นสำคัญ วิธีการนี้คิดขึ้นโดย คาร์ล ฟรีดริค เกาส์ (Karl Friedrich Gauss 1777-1855) และ อังเดร อังเดรเยวิช มาร์โคฟ (Andrei Andreevich Markov 1856-1922)

นิยาม 21 ให้ $X_1,...,X_n$ เป็นตัวอย่างสุ่มจากประชากรหนึ่งที่มีค่าคาดหมาย E(X) ค่าของพารามิเตอร์ในเทอมของค่าสังเกตที่ทำให้ผลบวกของกำลังที่ 2 ของผลต่างระหว่างค่าสังเกตกับค่าคาดหมายต่ำสุด จะเรียกว่า **ตัวประมาณแบบกำลังสองต่ำสุด** (Least-squares estimator) ของพารามิเตอร์ที่เกี่ยวข้อง

นั่นคือ ตัวประมาณแบบกำลังสองต่ำสุดของพารามิเตอร์ θ (อาจเป็นเวคเตอร์) ที่ปรากฏอยู่ในค่าคาดหมาย E(X) ได้แก่ ค่าของ θ ในเทอมของค่าสังเกต ที่ทำให้

$$Q = \sum_{i=1}^{n} \left(x_i - E(X_i)\right)^2$$

ต่ำที่สุด

มักหาตัวประมาณดังกล่าว โดยการใช้อนุพันธ์กล่าวคือ ใช้

$$\frac{\partial Q}{\partial \theta} = 0$$

และเรียกสมการ ว่า **สมการปกติ** (Normal equation(s))