บทที่ 2

ทฤษฎี และแบบจำลองเบื้องต้นที่ใช้ในการศึกษา

ขบวนการทางอุทกวิทยา (Hydrologic Process) มีความยุ่งยากสลับซับซ้อนมากมีปรากฏการณ์ธรรมชาติทางอุทกวิทยาเป็นจำนวนมากที่ไม่สามารถจะหาคำตอบที่แน่นอนได้โดยตรง เพียงแต่สามารถที่จะแปลความหมายหรืออธิบายเกี่ยวกับปรากฏการณ์นั้น ๆ ในแนวความคิดของความน่าจะเป็นเท่านั้น ทั้งนี้เพราะสาเหตุของการเกิดเหตุการณ์ทางอุทกวิทยานั้นมีความไม่แน่นอนขึ้นกับตัวแปรต่าง ๆ ที่เป็นปรากฏการณ์ธรรมชาติที่ไม่สามารถจะกำหนดหรือควบคุมได้

การเกิดฝน (Rainfall) เป็นขบวนการทางอุทกวิทยาอย่างหนึ่งที่ต้องอาศัยการแปลความ หรืออธิบายเกี่ยวกับขนาดของเหตุการณ์ (Magnitude of Event) ที่เกิดขึ้นในลักษณะของความน่าจะเป็น จึงได้มีผู้พยายามคิดหาวิธีการทางสถิติ (Statistical Method) เพื่อใช้ในการวิเคราะห์ข้อมูลฝนและนำทฤษฎีของความน่าจะเป็นแบบต่าง ๆ มาประยุกต์ใช้สำหรับการวิเคราะห์ข้อมูลทางอุทกวิทยา สำหรับหลักการของความน่าจะเป็นและสถิติที่เกี่ยวข้องนั้นจะแสดงไว้โดยสังเขปดังต่อไปนี้

2.1 ทฤษฎีความน่าจะเป็น และสถิติเบื้องต้น

ปรากฏการณ์ทางอุทกวิทยาที่เกิดขึ้นมักมีความไม่แน่นอน การที่จะคาดหมายปรากฏการณ์เหล่านี้อย่างเป็นวิทยาศาสตร์นั้น สามารถใช้ทฤษฎีความน่าจะเป็นและสถิติมาวิเคราะห์พฤติกรรมต่าง ๆ ซึ่งเริ่มตั้งแต่การเก็บรวบรวมข้อมูล การสรุปข้อมูลให้อยู่ในรูปความน่าจะเป็นและสถิติการแปลความหมายของข้อมูลที่สรุปนั้นเพื่อให้เข้าใจลักษณะของปรากฏการณ์ที่เกิดขึ้น ตลอดจนทำนายพฤติกรรมที่มีโอกาสจะเกิดขึ้นในอนาคต สำหรับทฤษฎีที่จะกล่าวในหัวข้อนี้เป็นการทบทวนโดยสังเขปประกอบด้วย กฎความน่าจะเป็น การกระจายความน่าจะเป็น พารามิเตอร์ทางสถิติที่ใช้บอกคุณสมบัติการกระจายความน่าจะเป็น และชนิดของการกระจายความน่าจะเป็น

2.1.1 กฎความน่าจะเป็น

ความน่าจะเป็น (Probability) ขึ้นกับสถิติของการทดลอง/สำรวจ เช่น เหตุการณ์

E_1 เป็นเหตุการณ์ที่เกิดขึ้นหลังจากการทดลอง/สำรวจ N ครั้ง ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 ระหว่างการทดลอง/สำรวจ กับจำนวนครั้งที่ทดลอง/สำรวจทั้งหมด ดังสมการ (2-1)

$$P(E_1) = \frac{n_1}{N} \qquad (2\text{-}1)$$

โดย $P(E_1)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1

n_1 เป็นจำนวนครั้งที่เกิดเหตุการณ์ E_1 ระหว่างการทดลอง/สำรวจ

N เป็นจำนวนครั้งที่ทดลอง/สำรวจทั้งหมด

คุณสมบัติของความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ สรุปเป็นกฎของความน่าจะเป็นดังนี้

1) ความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ต้องไม่เป็นลบและไม่เกินหนึ่ง

$$0 \leqslant P(E_i) \leqslant 1 \qquad (2\text{-}2)$$

2) ผลรวมของความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ทุกเหตุการณ์คือหนึ่ง

$$\Sigma_i P(E_i) = 1 \qquad (2\text{-}3)$$

3) ในกรณีที่เหตุการณ์ 2 เหตุการณ์ E_1 และ E_2 เป็นเหตุการณ์ที่ไม่เป็นอิสระต่อกัน (Dependent Events)

ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2 เท่ากับผลรวมของความน่าจะเป็นที่จะเกิดแต่ละเหตุการณ์ ลบความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2

$$P(E_1 \cup E_2) = P(E_1)+P(E_2)-P(E_1 \cap E_2) \qquad (2\text{-}4)$$

โดย $P(E_1 \cup E_2)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 หรือ E_2

$P(E_1 \cap E_2)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2

4) ในกรณีที่เหตุการณ์ 2 เหตุการณ์ E_1 และ E_2 เป็นเหตุการณ์ที่เป็นอิสระต่อกัน (Independent Events)

ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 หรือ E_2 เท่ากับผลรวมของความน่าจะเป็น ที่จะเกิดแต่ละเหตุการณ์

$$P(E_1 \cup E_2) = P(E_1)+P(E_2) \qquad (2\text{-}5)$$

5) ในกรณีที่เหตุการณ์ E_1 เกิดขึ้นก่อนเหตุการณ์ E_2

ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 เมื่อกำหนดให้เกิดเหตุการณ์ E_1 (Conditional probability) เท่ากับ อัตราส่วนของความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2 กับความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1

$$P(E_2 \mid E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \qquad (2\text{-}6)$$

โดย $P(E_2 \cap E_1)$ เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 หลังจากเกิดเหตุการณ์ E_1

ดังนั้น
$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2 \mid E_1) \qquad (2\text{-}7)$$

ถ้าเหตุการณ์ E_1 และ E_2 เป็นอิสระต่อกัน (Independent Events) ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 เมื่อกำหนดให้เกิดเหตุการณ์ E_1 เท่ากับ ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_2 ดังนั้น ความน่าจะเป็นที่จะเกิดเหตุการณ์ E_1 และ E_2 เท่ากับผลคูณของความน่าจะเป็นที่จะเกิดเหตุการณ์ทั้ง 2 เหตุการณ์

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) \qquad (2\text{-}8)$$

สำหรับปรากฏการณ์ทางอุทกวิทยานั้น มักแสดงความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ด้วยค่ารอบปี (Return Period) ซึ่งเป็นส่วนกลับของความน่าจะเป็นที่จะเกิดเหตุการณ์นั้น ๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ดังสมการ (2-9)

$$T = \frac{1}{P(E_1)} \qquad (2\text{-}9)$$

โดย T เป็น ค่ารอบปี

ความหมายของค่ารอบปีนั้น มักเข้าใจสับสน เช่น ค่ารอบปีเท่ากับ 10 ไม่ได้ หมายถึง ในระยะเวลา 10 ปี จะเกิดเหตุการณ์นี้ครั้งหนึ่ง แต่มีความหมาย ถ้าพิจารณาระยะยาวแล้วจะได้ว่าเหตุการณ์นั้นเกิดขึ้นโดยเฉลี่ย 1 ครั้ง ในระยะเวลา 10 ปี หรือกล่าวอีกนัยหนึ่งว่าเหตุการณ์นั้นมีความน่าจะเป็นที่จะเกิดในหนึ่งปีเท่ากับ 1/10 โดยทั่วไปสามารถสรุปความสัมพันธ์ของความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ กับค่ารอบปี ได้ดังนี้

1) ความน่าจะเป็นที่จะเกิดเหตุการณ์ E ในแต่ละปี

$$P(E) = \frac{1}{T} \qquad (2\text{-}10)$$

2) ความน่าจะเป็นที่ไม่เกิดเหตุการณ์ E ในแต่ละปี

$$P(\overline{E}) = 1 - P(E) = 1 - \frac{1}{T} \qquad (2\text{-}11)$$

3) ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ E ในเวลา n ปีติดกัน

$$[P(\overline{E})]^n = (1 - \frac{1}{T})^n \qquad (2\text{-}12)$$

4) ความน่าจะเป็นที่จะเกิดเหตุการณ์ E อย่างน้อย 1 ครั้ง ในเวลา n ปี ซึ่งเรียกอีกอย่างหนึ่งว่าความเสี่ยง (Risk)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$R_n(E) = 1 - [P(\overline{E})]^n = 1 - (1 - \frac{1}{T})^n \qquad (2\text{-}13)$$

ตารางที่ 2-1 แสดงค่ารอบปีสัมพันธ์กับความเสี่ยงและระยะเวลาที่พิจารณา ตัวอย่างเช่น ต้องการหาค่ารอบปีที่ควรใช้ในการออกแบบระบบระบายน้ำ โดยยอมให้ความเสี่ยง 20% ที่จะเกิดน้ำท่วม 5 ปีต่อมา จากตารางนี้จะได้ว่าค่ารอบปีที่ควรใช้ในการออกแบบคือ 22.9 ปี เป็นต้น [2]

2.1.2 การกระจายความน่าจะเป็น

การกระจายความน่าจะเป็น (Probability Distribution) เป็นความสัมพันธ์ระหว่างความน่าจะเป็นของตัวแปร กับค่าของตัวแปรนั้น ๆ ใช้สัญญลักษณ์ P(X=x) โดย X แทนตัวแปรสุ่ม และ x แทนค่าของตัวแปร P(X=x) นี้ อาจเขียนแทนด้วย P(x) เช่น ตัวแปรมีค่าที่เป็นไปได้คือ 0 และ 1 โดยมีความน่าจะเป็นที่จะมีค่าเป็น 0 เท่ากับ 0.3 และความน่าจะเป็นที่จะมีค่าเป็น 1 เท่ากับ 0.7 สามารถเขียนแทนด้วย P(X=0) = 0.3 และ P(X=1) = 0.7 หรือแทนด้วย P(0) = 0.3 และ P(1) = 0.7 เป็นต้น

ในการนำเสนอการกระจายความน่าจะเป็นนี้มักนิยมใช้กราฟ ทั้งในรูปของความน่าจะเป็น (Probability Density) P(x) หรือ f(x) และความน่าจะเป็นสะสม (Cummulative Probability) P(X(x)) หรือ F(X) ซึ่งมีความสัมพันธ์ ดังนี้

$$P(X < x') = F(x') = \Sigma_I p(x_i) \quad \text{โดย} \quad x_i < x' \qquad (2\text{-}14)$$

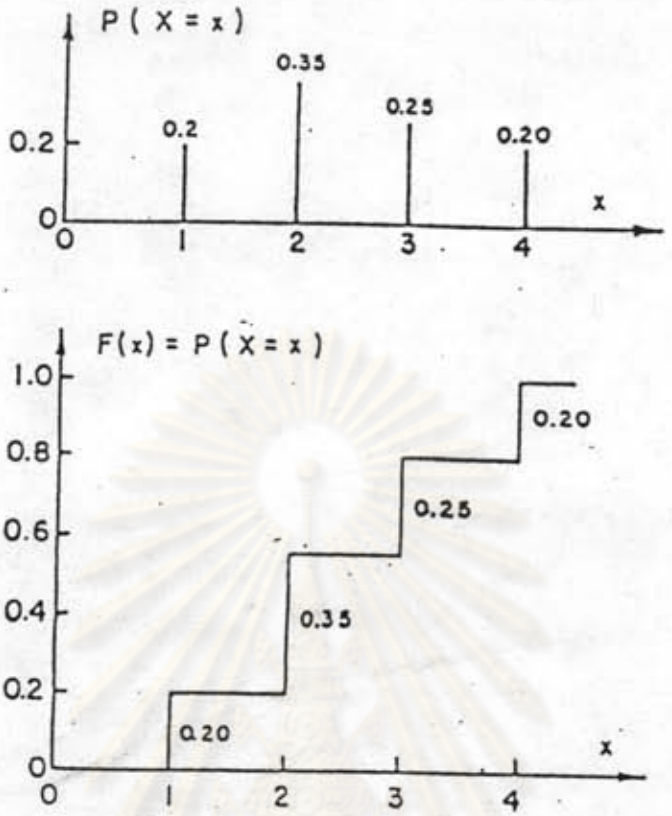
หรือ
$$P(X < x') = F(x') = \int_{-\infty}^{\infty} p(x)dx = \int_{-\infty}^{\infty} f(x)dx \qquad (2\text{-}15)$$

ในรูปที่ 2-1 ก แสดงการกระจายความน่าจะเป็นของตัวแปรที่ไม่ต่อเนื่องและรูปที่ 2-1 ข แสดงการกระจายความน่าจะเป็นของตัวแปรที่ต่อเนื่อง

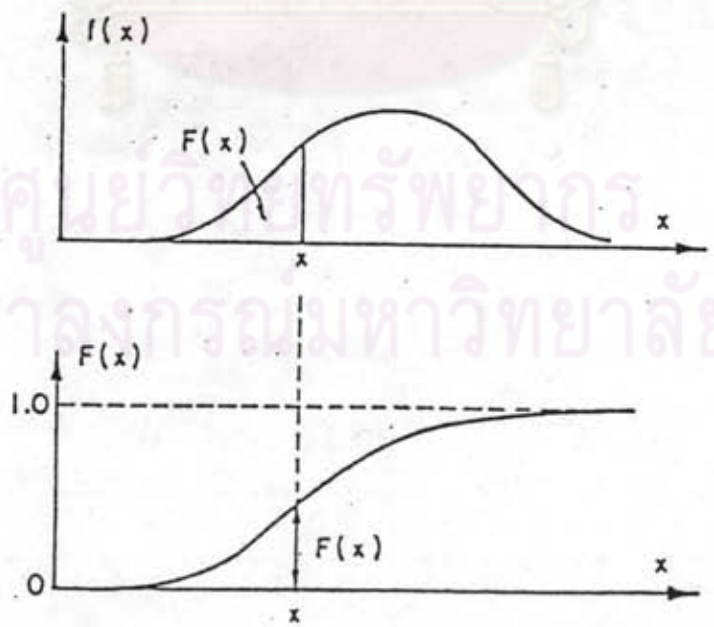
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2-1 ค่ารอบปีสัมพันธ์กับค่าความเสี่ยง และระยะเวลาที่พิจารณา [2]

Risk	Expected Design Life (years)							
(%)	2	5	10	15	20	25	50	100
75	2.0	4.1	7.7	11.3	14.9	18.5	36.6	72.6
50	3.4	7.7	14.9	22.1	29.4	36.6	72.6	144.8
40	4.4	10.3	20.1	29.9	39.7	49.4	98.4	196.3
30	6.1	14.5	28.5	42.6	56.6	70.6	140.7	280.9
25	7.5	17.9	35.3	52.6	70.0	87.4	174.3	348.1
20	9.5	22.9	45.3	67.7	90.1	112.5	224.6	448.6
15	12.8	31.3	62.0	92.8	123.6	154.3	308.2	615.8
10	19.5	48.0	95.4	142.9	190.3	237.8	475.1	949.6
5	39.5	98.0	195.5	292.9	390.4	487.9	975.3	1950.1
2	99.5	248.0	495.5	743.0	990.5	1238.0	2475.4	4950.3
1	199.5	498.0	995.5	1493.0	1990.5	2488.0	4975.5	9950.4



ก) การกระจายความน่าจะเป็นของตัวแปรไม่ต่อเนื่อง



ข) การกระจายความน่าจะเป็นตัวแปรต่อเนื่อง

รูปที่ 2-1 การกระจายความน่าจะเป็นของตัวแปร [17]

2.1.3 พารามิเตอร์ทางสถิติที่ใช้บอกคุณสมบัติการกระจายความน่าจะเป็น

คุณสมบัติการกระจายความน่าจะเป็นสามารถอธิบายได้ด้วยพารามิเตอร์ทางสถิติ ซึ่งใช้การหาค่าพารามิเตอร์เหล่านี้มักใช้ค่าพารามิเตอร์ของตัวแปรสุ่มตัวอย่างเพื่อประมาณค่าพารามิเตอร์ของตัวแปรทั้งหมด พารามิเตอร์ที่สำคัญมีดังนี้

1) พารามิเตอร์ที่ใช้วัดศูนย์กลางของการกระจาย โดยทั่วไปใช้ค่าเฉลี่ยเลขคณิต (Arithmatic Mean) เขียนแทนด้วย μ สำหรับตัวแปรทั้งหมด และ $\bar{X}$ สำหรับตัวแปรสุ่มตัวอย่าง สามารถหาค่าได้จากสมการ (2-16) และ (2-17)

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad (2\text{-}16)$$

โดย N เป็นจำนวนตัวแปรทั้งหมด

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (2\text{-}17)$$

โดย n เป็นจำนวนตัวแปรสุ่มตัวอย่าง

ส่วนพารามิเตอร์อื่นที่ใช้วัดศูนย์กลางนี้ได้แก่ ค่ามัธยฐาน (Median) เป็นค่าของตัวแปรที่แบ่งการกระจายออกเป็น 2 ส่วน แต่ละส่วนมีความน่าจะเป็นเท่ากับ 0.5 ค่าฐานนิยม (Mode) เป็นค่าของตัวแปรที่มีความน่าจะเป็นสูงสุด

2) พารามิเตอร์ที่ใช้วัดความผันแปรของการกระจาย โดยทั่วไปใช้ค่าความแปรปรวน (Variance) เขียนแทนด้วย σ^2 สำหรับตัวแปรทั้งหมด และ S^2 สำหรับตัวแปรสุ่มตัวอย่าง สามารถหาค่าได้จากสมการ (2-18) และ (2-19)

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \qquad (2\text{-}18)$$

โดย N เป็นจำนวนตัวแปรทั้งหมด

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (2\text{-}19)$$

n เป็นจำนวนตัวแปรสุ่มตัวอย่าง

ส่วนพารามิเตอร์อื่นที่ใช้วัดความผันแปรได้แก่ ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็นค่ารากที่ 2 ของความแปรปรวนเขียนแทนด้วย σ สำหรับตัวแปรทั้งหมดและ S สำหรับตัวแปรสุ่มตัวอย่าง ค่าพิสัย (Range) เป็นค่าความแตกต่างระหว่างค่าสูงสุดและต่ำสุดของของตัวแปร ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variation) เป็นอัตราส่วนระหว่าง σ กับ μ หรือ S กับ $\bar{X}$ เขียนแทนด้วย C_V ใช้สำหรับเปรียบเทียบความผันแปรของการกระจายหลาย ๆ ชุด

3) พารามิเตอร์ที่ใช้วัดความสมมาตรของการกระจาย โดยทั่วไปใช้ค่าความเบ้ (Skewness) เขียนแทนด้วย α สำหรับตัวแปรทั้งหมด และ a สำหรับตัวแปรสุ่มตัวอย่าง สามารถหาค่าได้จากสมการ (2-20) และ (2-21)

$$\alpha = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^3 \qquad (2\text{-}20)$$

โดย N เป็นจำนวนตัวแปรทั้งหมด

$$a = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (x_i - \bar{x})^3 \qquad (2\text{-}21)$$

โดย n เป็นจำนวนตัวแปรสุ่มตัวอย่าง

ส่วนพารามิเตอร์อื่นที่ใช้วัดความสมมาตรได้แก่ สัมประสิทธิ์ความเบ้ (Coefficient of Skewness) เป็นอัตราส่วนระหว่าง α กับ σ^3 หรือ a กับ S^3 เขียนแทนด้วย C_s ถ้า $C_s = 0$ แสดงว่า การกระจายความน่าจะเป็นสมมาตรรอบค่าเฉลี่ย ถ้าค่า C_s เป็นบวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

018498

แสดงว่าการกระจายนี้จะเบ้ซ้าย และถ้า C_s เป็นลบ แสดงว่าการกระจายนี้จะเบ้ขวา

2.1.4 ชนิดของการกระจายความน่าจะเป็นมาตรฐาน

ในการอธิบายปรากฏการณ์ทางอุทกวิทยา สามารถใช้การกระจายความน่าจะเป็นมาตรฐาน (Standard Probability Distribation) ซึ่งมีอยู่หลายทฤษฎีดังตารางที่ 2-2 แสดงทฤษฎีการกระจายความน่าจะเป็นมาตรฐานชนิดต่าง ๆ ส่วนหนึ่งที่นิยมใช้กันมาก โดยแสดงฟังค์ชั่นการกระจายความน่าจะเป็นค่าเฉลี่ย μ และความแปรปรวนของฟังค์ชั่น σ^2 การเลือกใช้ทฤษฎีเหล่านี้ให้เหมาะสมนั้นจะต้องมีการทดสอบความเหมาะสมในเชิงสถิติก่อน ซึ่งรายละเอียดของทฤษฎีและการทดสอบนั้นสามารถศึกษาได้จากหนังสือสถิติทั่วไป

2.2 ลักษณะของกระบวนการทางสโตแคสติก

ทฤษฎีความน่าจะเป็นและสถิติ ซึ่งใช้ในการหาโอกาสเกิดของเหตุการณ์ที่ไม่แน่นอน โดยตัวแปร X ของเหตุการณ์นี้เรียกว่า ตัวแปรที่ไม่แน่นอน (Random Variable) อนุกรม $X, X_1, X_2, X_3, \ldots$ ที่เกิดจากตัวแปร X นี้ เรียกว่า อนุกรมทางความน่าจะเป็น (Probabilistic Series) กลไกที่ทำให้เกิดการเปลี่ยนแปลงค่าตัวแปร X นี้ เรียกว่า กระบวนการทางความน่าจะเป็น (Probabilistic Process)

ในทางตรงข้าม ถ้าตัวแปร X นั้นสามารถทำนายผล X อย่างแน่นอนได้ ตัวแปรนี้เรียกว่าตัวแปรดีเทอร์มินิสติค (Deterministic Variable) อนุกรม $X_1, X_2, X_3 \ldots$ ที่เกิดจากตัวแปร X นี้เรียกว่า อนุกรมทางดีเทอร์มินีสติก (Deterministic Series) กลไกที่ทำให้เกิดการเปลี่ยนแปลงค่าตัวแปร X นี้ เรียกว่า กระบวนการทางดีเทอร์มินีสติค (Deterministic Process)

สำหรับตัวแปร X ที่ไม่สามารถทำนายอย่างแน่นอน แต่จะมีส่วนที่สัมพันธ์กับเหตุการณ์ที่เกิดขึ้นแล้ว ตัวแปร X นี้เรียกว่า ตัวแปรสโตแคสติก (Stochastic Variable) อนุกรม $X_1, X_2, X_3 \ldots$ ที่เกิดจากตัวแปร X นี้ เรียกว่า อนุกรมทางสโตแคสติค (Stochastic Series) กลไกที่ทำให้เกิดการเปลี่ยนแปลงค่าตัวแปร X นี้ เรียกว่าขบวนการทางสโตแคสติก

ตารางที่ 2-3 ทฤษฎีการกระจายความน่าจะเป็นมาตรฐานชนิดต่าง ๆ [16]

Distribution of Random Variable X	Probability Distribution Function	Range	Mean $\bar{x}$ or μ	Variance s^2 or σ^2
Binomial	$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$	$0 \leq x \leq n$	np	$np(1-p)$
Poisson	$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$0 \leq x \leq \cdots$	λ	λ
Uniform	$f(x) = \frac{1}{b-a}$	$a \leq x \leq b$	$(b+a)/2$	$\frac{(b-a)^2}{12}$
Exponential	$f(x) = \frac{1}{a} e^{-x/a}$	$0 \leq x \leq \infty$	a	a^2
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$	$-\infty \leq x \leq \infty$	μ	σ^2
Lognormal ($y = \ln x$)	$f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-(y-\mu_y)^2/2\sigma_y^2\right]$	$-\infty \leq y \leq \infty$ ($0 \leq x \leq \infty$)	μ_y	σ_y^2
Gamma	$f(x) = \frac{x^\alpha e^{-x/\beta}}{\beta^{\alpha+1}\Gamma(\alpha+1)}$	$0 \leq x \leq \infty$	$\beta(\alpha+1)$	$\beta^2(\alpha+1)$
Extreme value	$f(x) = a \exp\left[-a(x-u) - e^{-a(x-u)}\right]$	$-\infty \leq x \leq \infty$	$u + \frac{0.5772}{a}$	$\frac{\pi^2}{6a^2}$

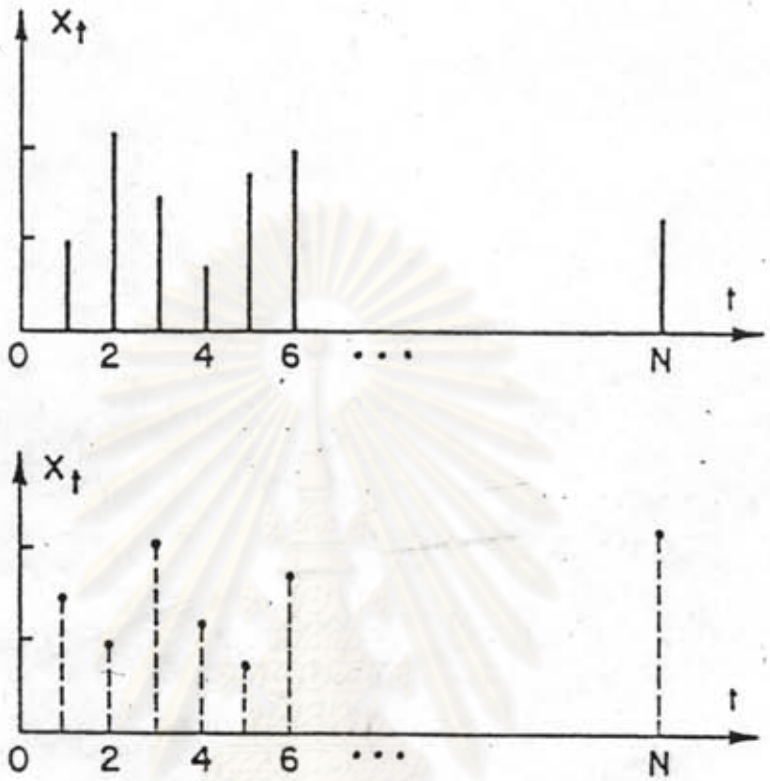
(Stochastic Process) ในรูปที่ 2-2 แสดงตัวอย่างขบวนการทางสโตแคสติก 2 กระบวนการเขียนในรูปกราฟแท่ง ซึ่งสามารถเขียนแทนให้อยู่ในรูปกราฟเส้นแบบต่อเนื่องทั้ง ๆ ที่ข้อมูลไม่ต่อเนื่อง เพื่อให้ง่ายต่อการพิจารณาลำดับเหตุการณ์มากขึ้น ดังในรูปที่ 2-2 ข

คุณสมบัติของขบวนการทางสโตแคสติกนี้ ขึ้นกับความสัมพันธ์ของค่าตัวแปร X_1, X_2, X_3... สามารถอธิบายด้วยพารามิเตอร์ทางสถิติ ซึ่งได้แก่ ค่าเฉลี่ย (Mean), ความแปรปรวน (Variance), ค่าฟังค์ชั่นความสัมพันธ์ในตัวเอง (Autocorrelation Function) และค่าความหนาแน่นของสเป็คตั้ม (Spectrum Density) ซึ่ง 2 ค่าหลังนี้จะกล่าวรายละเอียดการวิเคราะห์ในหัวข้อ 2.4

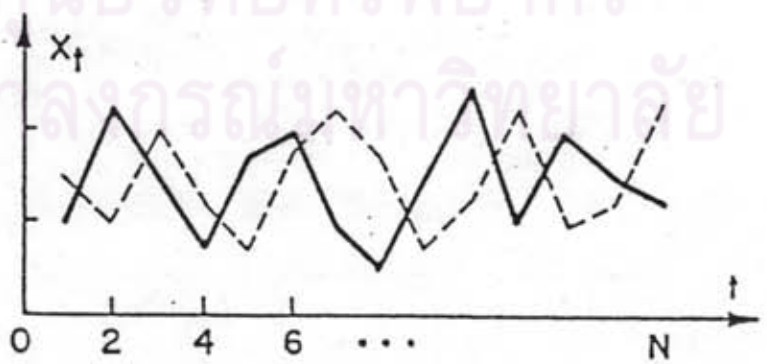
กระบวนการทางสโตแคสติก สามารถเรียกชื่อตามความสมดุลย์ (Stationary) ให้หลายแบบคือ กระบวนการทางสโตแคสติกที่มีค่าเฉลี่ยคงที่ไม่ขึ้นกับเวลานั้น เรียกว่า กระบวนการทางสโตแคสติกที่สมดุลย์สำหรับค่าเฉลี่ย (Stationary in Mean) ในทำนองเดียวกัน กระบวนการทางสโตแคสติกที่มีค่าความแปรปรวนคงที่ไม่ขึ้นกับเวลานั้น เรียกว่า ขบวนการทางสโตแคสติกที่สมดุลย์สำหรับค่าความแปรปรวน (Stationary in Variance) สำหรับกระบวนการทางสโตแคสติกที่สมดุลย์ทั้ง 2 อย่าง เรียกว่า สมดุลย์อย่างอ่อน (Weak Stationary) คุณสมบัติทางสถิติอย่างอื่นรวมทั้ง 2 อย่างนี้สมดุลย์ไม่ขึ้นกับเวลา เรียกว่า สมดุลย์อย่างมาก (Strong Stationary) แต่ถ้าคุณสมบัติทางสถิติทุกตัวขึ้นกับเวลาจะเรียกว่า ไม่สมดุลย์ (Nonstationary)

2.3 อนุกรมทางอุทกวิทยา

อนุกรมทางอุทกวิทยา (Hydrologic Series) เป็นอนุกรมที่เกิดจากตัวแปรทางอุทกวิทยา อนุกรมนี้สามารถแบ่งตามชนิดของตัวแปรได้เป็นอนุกรมที่ตัวแปรขึ้นกับเวลา ซึ่งเรียกว่าอนุกรมเวลา (Time Series) เช่น ปริมาณน้ำฝนรายปีที่สถานีวัดน้ำฝนแห่งหนึ่ง อนุกรมที่ตัวแปรขึ้นกับระยะทาง ซึ่งเรียกว่าอนุกรมระยะทาง (Line Series) เช่น ความลึกของแม่น้ำตามความยาวของแม่น้ำ และอนุกรมที่ขึ้นกับจำนวนครั้งที่เกิดปรากฏการณ์ที่สนใจในระยะเวลาหรือพื้นที่ที่พิจารณา ซึ่งเรียกว่าอนุกรมจำนวนครั้ง (Counting Series) เช่น จำนวนวันฝนตกในแต่ละปีของสถานีวัดน้ำฝนแห่งหนึ่งที่กล่าวมาทั้ง 3 ชนิดนี้เป็นอนุกรมที่พิจารณาตัวแปรใน 1 มิติ ยังมีอนุกรมทางอุทกวิทยาที่อาจพิจารณาตัวแปรในหลายมิติอีก แต่อนุกรมทางอุทกวิทยาที่เกิดขึ้นส่วน



ก) ตัวอย่างขบวนการสโตแคสติค 2 ขบวนการในรูปกราฟแท่ง



ข) ตัวอย่างขบวนการสโตแคสติค 2 ขบวนการในรูปกราฟเส้น

รูปที่ 2-2 ตัวอย่างขบวนการสโตแคสติก 2 ขบวนการ [17]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ใหญ่จะเป็นอนุกรมเวลา

อนุกรมเวลาโดยทั่วไปประกอบด้วยส่วนที่เป็นดีเทอร์มินีสติก และส่วนที่เป็นสโตแคสติก ในส่วนที่เป็นดีเทอร์มินีสติกอาจประกอบไปด้วยแนวโน้ม หรือ กระโดด หรือ เพอร์ริโอดิก และในส่วนที่เป็นสโตแคสติกอาจจะประกอบไปด้วยกระบวนการสโตแคสติกที่เป็นอิสระ (การสุ่มล้วน ๆ) หรือกระบวนการสโตแคสติกที่ขึ้นอยู่กับการสุ่ม แบบจำลองทางอุทกวิทยาที่เป็นสโตแคสติกใด ๆ จะต้องมีความเหมาะสมทั้งส่วนที่เป็นดีเทอร์มินีสติก และส่วนที่เป็นสโตแคสติกในอนุกรมเวลา

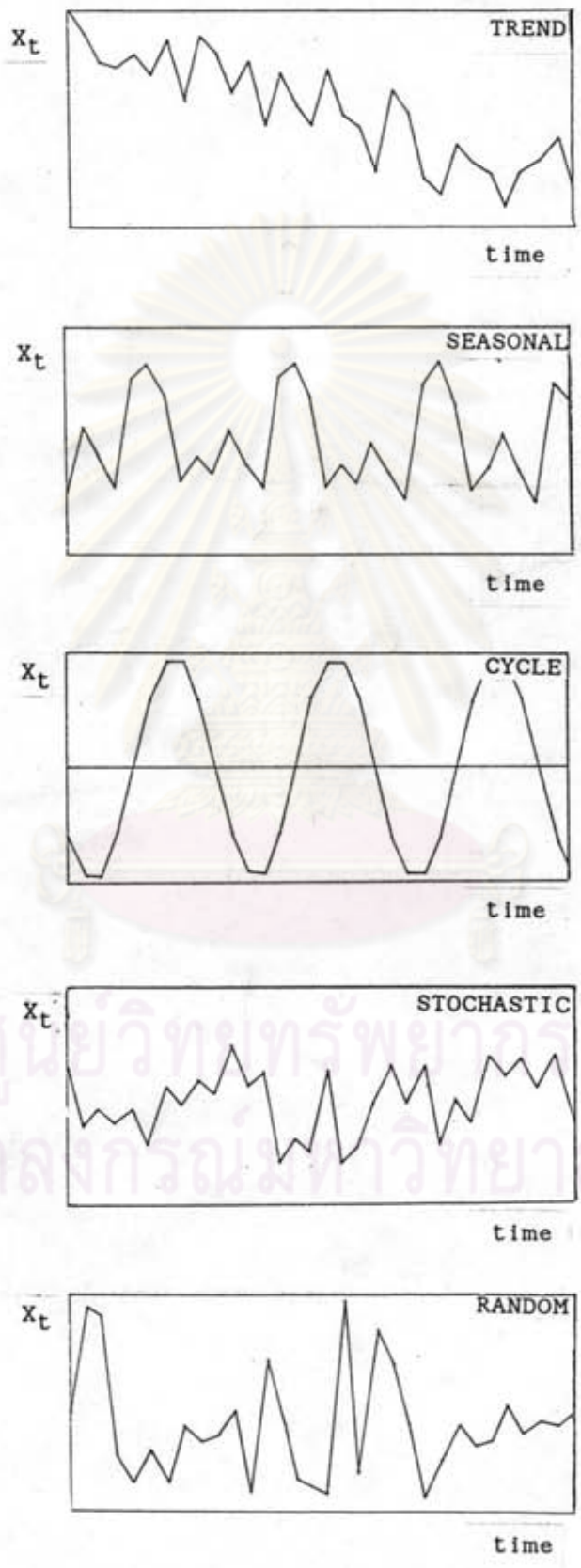
ในการพิจารณาลักษณะของอนุกรมทางอุทกวิทยา ซึ่งเป็นอนุกรมเวลานี้ สามารถแยกส่วนประกอบของอนุกรมออกเป็นดังนี้

อนุกรมเวลาหนึ่ง = ส่วนแนวโน้ม + ส่วนฤดูกาล + ส่วนวงจร + ส่วนสโตแคสติก + ส่วนที่ไม่แน่นอน

ในรูปที่ 2-3 แสดงตัวอย่างของอนุกรมเวลาที่มีส่วนประกอบต่าง ๆ กัน รายละเอียดของส่วนประกอบต่าง ๆ มีดังต่อไปนี้

1) ส่วนที่เปลี่ยนแปลงอย่างมีแนวโน้ม (Trend Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่เกิดการเปลี่ยนแปลงค่าของตัวแปรตามลำดับเหตุการณ์ โดยไม่คำนึงถึงความผันแปรระยะสั้น มักเกิดจากสาเหตุขององค์ประกอบอื่นเปลี่ยนแปลง เช่น ปริมาณน้ำใช้ของคนในกรุงเทพมหานครมากขึ้นเรื่อย ๆ เนื่องจากคนในกรุงเทพมหานครมีมากขึ้น สำหรับการวิเคราะห์การเปลี่ยนแปลงทำได้โดยเขียนกราฟของข้อมูลในอดีตตามลำดับเหตุการณ์ แล้วสังเกตุการเปลี่ยนแปลงของข้อมูล ถ้าข้อมูลมีความผันแปรมาก ก็ต้องทำการปรับข้อมูลให้เรียบมากขึ้นเพื่อหาทางกำจัดออกไปเสียจากอนุกรม นอกจากการสังเกตุแล้ว ยังสามารถใช้สมการทางคณิตศาสตร์แบบต่างๆ เช่น สมการเส้นตรง สมการเลขยกกำลัง เพื่อศึกษาการ เปลี่ยนแปลงอย่างมีแนวโน้มของของอนุกรมได้ ในการศึกษาครั้งนี้ได้วิเคราะห์การเปลี่ยนแปลงโดยการวิเคราะห์การถดถอย ซึ่งกล่าวรายละเอียดในหัวข้อ 2.4.1

2) ส่วนที่เป็นฤดูกาล (Seasonal Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่ เกิดขึ้นในฤดูกาลประจำในแต่ละปี มักเป็นปรากฏการณ์ที่มีผลมาจากการโคจรของ โลกรอบดวงอาทิตย์ในหนึ่งปี ตลอดจากเหตุอื่น ๆ ที่เกิดขึ้นในลักษณะฤดูกาล เช่น ปริมาณน้ำฝนมากในช่วงฤดูฝน คือ เดือน พ.ค.-ต.ค. และปริมาณน้ำฝนน้อยในฤดูแล้ง คือ เดือน พ.ย.-เม.ย. สำหรับการ



รูปที่ 2-3 อนุกรมของการเปลี่ยนแปลงในรูปแบบต่าง ๆ [2]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิเคราะห์ผลของฤดูกาล อาจทำได้โดยการหา ค่าเฉลี่ยแต่ละฤดูกาล เช่น มีข้อมูลน้ำฝนรายเดือน 30 ปี ก็เอาค่าเฉลี่ยของน้ำฝน เดือน ม.ค. จนถึง ธ.ค. ใน 30 ปีนั้น มาเป็นตัวแทนผลของฤดูกาลเพื่อขจัดออกจากอนุกรมที่พิจารณาเป็นต้น แต่ถ้าข้อมูลที่สนใจใช้ช่วงเวลาเป็นปี เช่นน้ำฝนรายปี ผลของฤดูกาลก็ไม่จำเป็นต้องนำมาพิจารณา

3) ส่วนที่เป็นวงจร (Cyclicity Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่มีความสำคัญมาก หมายถึงรอบของเหตุการณ์ที่เกิดขึ้น โดยมากวงจรนี้มีผลจากการ โคจรของดวงดาว เช่น โลก ดวงจันทร์ และดวงอาทิตย์ เป็นต้น ซึ่งการโคจรเป็นวง ดังนั้นวงจรพื้นฐานที่สำคัญทางอุทกวิทยามักประกอบด้วย 1 วัน, 7 วัน, 15 วัน, 30 วัน, 1 ปี และหลาย ๆ ปี แล้วแต่กระบวนการที่เกิดขึ้นมีองค์ประกอบใดที่มีอิทธิพลต่อขบวนการ สำหรับการวิเคราะห์วงจร วงจรอาจทำได้โดยใช้การวิเคราะห์ทางสเป็คทรอล (Spectral Analysis) ซึ่งกล่าวรายละเอียดในหัวข้อ 2.4.3

4) ส่วนที่เป็นสโตแคสติก (Stochastic Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่ตัวแปรไม่สามารถทำนายหรือคาดการณ์อย่างแน่นอนได้ แต่มีส่วนที่ สัมพันธ์กับเหตุการณ์ที่เคยเกิดขึ้น มักจะพิจารณาหลังจากวิเคราะห์ส่วนเปลี่ยนแปลงที่สามารถทำนายได้อย่างแน่นอน เช่น แนวโน้ม ฤดูกาล และวงจร สำหรับการวิเคราะห์ส่วนที่เป็นสโตแคสติก ทำได้โดยใช้แบบจำลองทางสโตแคสติก เช่น AR, MA, ARIMA ซึ่งในการเลือกแบบจำลองนี้ทำโดยเปรียบเทียบค่าความสัมพันธ์ในตัวเอง (Autocorrelation Function) ของอนุกรมนี้กับแบบจำลอง จากนั้นก็ ประมาณค่าพารามิเตอร์ที่จะใช้ในแบบจำลองตามความเหมาะสม ซึ่งการวิเคราะห์ ความสัมพันธ์ในตัวเองนี้กล่าวรายละเอียดในหัวข้อ 2.4.2

5) ส่วนที่ไม่แน่นอน (Random Component) เป็นส่วนของอนุกรมทางอุทกวิทยาที่ไม่สามารถทำนายได้ด้วยองค์ประกอบข้างต้น สามารถวิเคราะห์โดยใช้ทฤษฎีความน่าจะเป็น เพื่อหาฟังก์ชั่นการกระจายความน่าจะเป็นที่เหมาะสมกับข้อมูลที่พิจารณา จากนั้นจึงหาความน่าจะเป็นที่จะเกิดเหตุการณ์ต่าง ๆ ดังกล่าวในหัวข้อ 2.1

2.4 เทคนิคที่ใช้วิเคราะห์โครงสร้างของอนุกรม

การวิเคราะห์โครงสร้างของอนุกรมใช้สำหรับศึกษาพฤติกรรม และส่วนประกอบของอนุกรม ซึ่งเป็นประโยชน์ต่อการอธิบายคุณสมบัติของอนุกรมนั้น ๆ และการเข้าใจลักษณะทาง

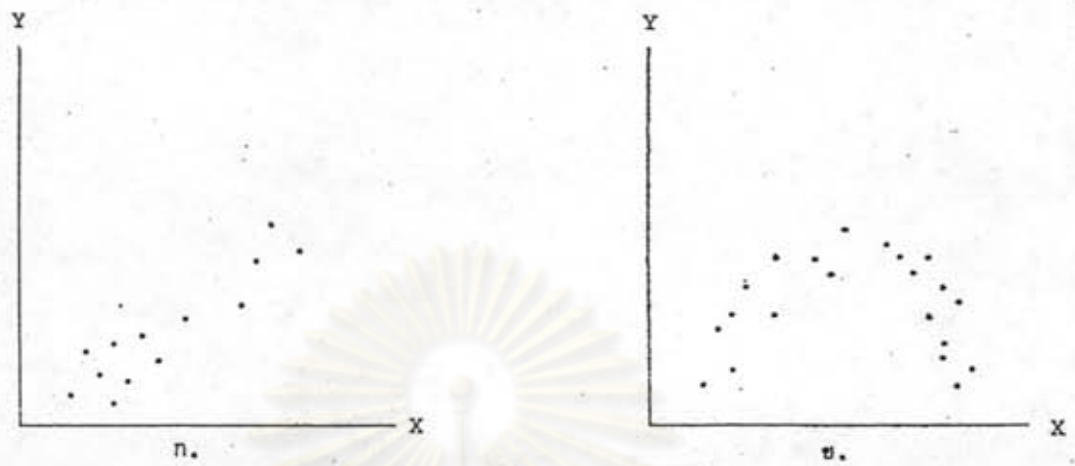
กายภาพที่มีผลต่อโครงสร้างของอนุกรมนั้น ๆ โดยเทคนิคที่ใช้วิเคราะห์ที่จะกล่าวถึงมีการวิเคราะห์การถดถอย (Regression Analysis) การวิเคราะห์ความสัมพันธ์ในตัวเอง (Autocorrelation Analysis) และการวิเคราะห์ทางสเป็กทรอล (Spectral Analysis)

2.4.1 การวิเคราะห์การถดถอย (Regression Analysis)

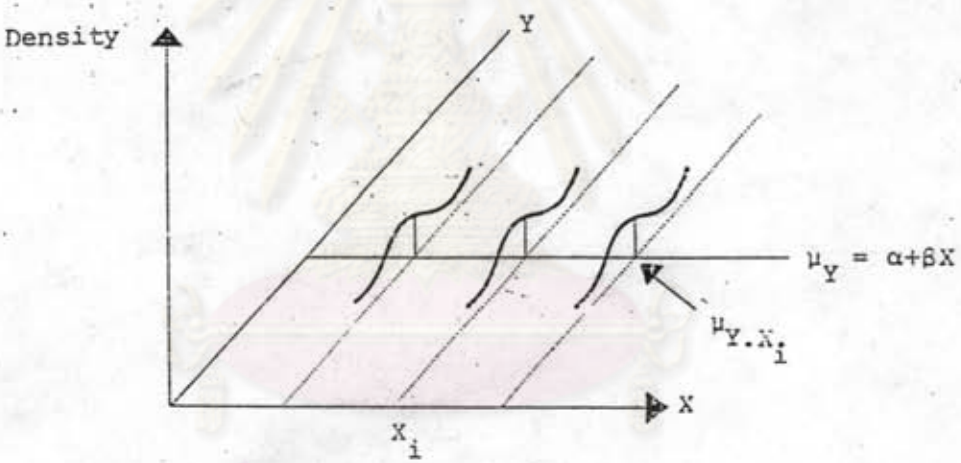
การวิเคราะห์การถดถอยเป็นการพยากรณ์ค่าของตัวแปรตัวหนึ่งซึ่งมีชื่อเรียกว่าตัวแปรตาม (Dependent Variable) มักจะแทนด้วย Y โดยใช้ค่าของตัวแปรอิสระ (Independent Variable) ซึ่งแทนด้วย X ตัวแปรอิสระนี้อาจใช้เพียงตัวเดียวหรืออาจใช้ตั้งแต่ 2 ตัวขึ้นไป ถ้าใช้ตัวแปรอิสระเพียงตัวเดียววิธีการพยากรณ์จะเรียกว่าการวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis) แต่ถ้าใช้ตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป เรียกว่าการวิเคราะห์การถดถอยเชิงหลายมิติ (Multiple Regression Analysis) โดยทั่ว ๆ ไปการใช้ตัวแปรอิสระมากกว่าหนึ่งตัวย่อมให้ผลในการพยากรณ์ถูกต้องมากกว่าการใช้ตัวแปรอิสระเพียงตัวเดียว อย่างไรก็ตาม ในหัวข้อนี้จะได้แนะนำการวิเคราะห์การถดถอยอย่างง่ายเพื่อให้เข้าใจวิธีการทางทฤษฎี

แนวความคิดที่ใช้ค่าของตัวแปรอิสระมาพยากรณ์ค่าของตัวแปรตามก็เนื่องจากว่าตัวแปรทั้งสองประเภทนี้มีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่ง ซึ่งอาจจะเป็นที่แน่ชัดหรือไม่เป็นที่แน่ชัดก็ได้ว่า ตัวแปรใดคือเหตุ และตัวแปรใดคือผล ลักษณะความสัมพันธ์อาจเป็นแบบเชิงเส้น (Linear) หรือแบบไม่ใช่เชิงเส้น (Non-linear) ซึ่งพอจะทราบได้โดยการทำแผนภูมิการกระจาย (Scatter Diagram) ดังตัวอย่างในรูปที่ 2-4

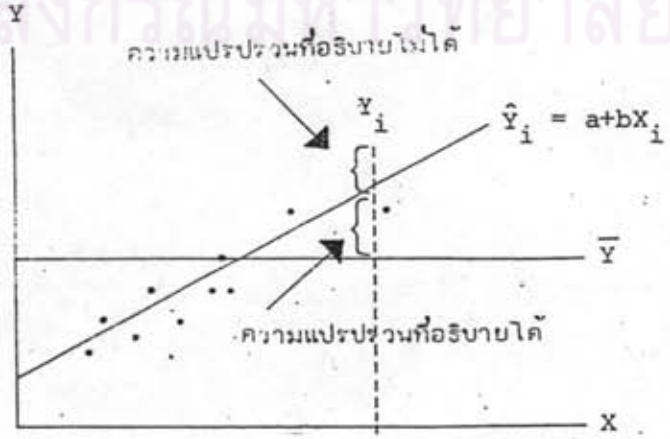
ในรูปที่ 2-4ก ความสัมพันธ์ระหว่าง X และ Y มีแนวโน้มเป็นลักษณะเชิงเส้นซึ่งควรมีการทดสอบทางสถิติก่อนว่าความสัมพันธ์เป็นลักษณะเชิงเส้นหรือไม่ ส่วนรูปที่ 2-4ข ความสัมพันธ์มีลักษณะไม่ใช่เชิงเส้น ความสัมพันธ์ในลักษณะเชิงเส้นเป็นลักษณะที่พบบ่อย ๆ ในการวิเคราะห์การถดถอย ในกรณีที่ความสัมพันธ์เป็นแบบไม่ใช่เชิงเส้นก็สามารถแปลง (Transform) ตัวแปรอิสระให้เป็นรูปใดรูปหนึ่งซึ่งเป็นฟังก์ชั่นของตัวแปรเดิมและให้ลักษณะความสัมพันธ์ที่มีต่อตัวแปรตามเป็นในแบบเชิงเส้น ดังนั้น จะพิจารณาการวิเคราะห์การถดถอยแบบเชิงเส้นเป็นสำคัญ



รูปที่ 2-4 แผนภูมิการกระจายของตัวแปรอิสระ X และตัวแปรตาม Y [1]



รูปที่ 2-5 เส้นตรงการถดถอย [1]



รูปที่ 2-6 ความแปรปรวนของตัวแปรตาม [1]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.4.1.1 การวิเคราะห์การถดถอยแบบเชิงเส้น (Linear Regression Analysis)

การวิเคราะห์การถดถอยแบบเชิงเส้นเป็นการพยากรณ์ค่าของตัวแปรตาม Y โดยอาศัยค่าของตัวแปรอิสระ X เพียงตัวเดียว Y และ X มีความสัมพันธ์ในลักษณะเชิงเส้นซึ่งเขียนเป็นสมการการถดถอยได้ดังนี้

$$\mu_Y = \alpha + \beta X \qquad (2\text{-}22)$$

เมื่อ μ_Y คือค่าเฉลี่ยของ Y สำหรับค่าหนึ่งของ X

α เรียกว่า intercept คือค่า μ_Y เมื่อ X = 0 ในข้อมูลบางประเภทเราสามารถอธิบายความหมายของค่า intercept นี้ได้ แต่สำหรับข้อมูลบางประเภทค่านี้ไม่ได้ให้ความหมายอะไรที่น่าสนใจ

β เรียกว่าสัมประสิทธิ์การถดถอย (Regression Coefficient) คืออัตราการเปลี่ยนแปลงของค่าเฉลี่ยของ Y เมื่อ X เปลี่ยนแปลงไป 1 หน่วย

α และ β เป็นพารามิเตอร์ (Parameter) ของสมการการถดถอย

สมการ (2-22) มาจากความจริงที่ว่าแต่ละค่าของ X นั้น Y เป็นตัวแปรเชิงสุ่ม (Random Variable) ซึ่งมีได้หลายค่า การวิเคราะห์การถดถอยซึ่งสนใจความสัมพันธ์ระหว่างหว่างค่าเฉลี่ย Y และค่าใดค่าหนึ่งของ X สมมติว่า Y มีการกระจายเป็นแบบปกติ (Normal Distributed) ในแต่ละค่าของ X เส้นตรงที่แสดงความสัมพันธ์ระหว่างค่าเฉลี่ยของ Y และ X ตามสมการ (2-22) จะแสดงได้ตามรูปที่ 2-5

นอกเหนือจากค่าเฉลี่ยของ Y ค่า Y แต่ละค่าก็เป็นสิ่งที่น่าสนใจ ค่า Y แต่ละค่านี้จะแทนด้วย Y_i ซึ่งสามารถแสดงเป็นตัวแบบการถดถอย (Regression Model) ได้ดังนี้

$$Y_i = \alpha + \beta X_i + e_i \qquad (2\text{-}23)$$

เมื่อ e_i คือความคลาดเคลื่อนที่เกิดจากการที่ Y_i แตกต่างไปจากค่าเฉลี่ย μ_Y

นั่นคือ $e_i = Y_i + \mu_Y$

ข้อสมมุติเกี่ยวกับ e_i ได้แก่

1. e_i มีการกระจายแบบปกติ
2. e_i มีค่าคาดหมาย (expected value) เป็น 0 นั่นคือ $E(e_i) = 0$
3. e_i มีความแปรปรวนคงที่ นั่นคือ $E(e_i^2) = \sigma^2$ ไม่ว่า X จะมีค่าเท่าใด
4. e_i และ e_j มีความแปรปรวนร่วม (covariance) เป็น 0 นั่นคือ $E(e_i e_j) = 0$

ข้อสมมุติข้างต้นบางข้อจำเป็นสำหรับการหาค่าประมาณของ α และ β และบางข้อจำเป็นสำหรับการทดสอบสมมติฐานเกี่ยวกับค่าของ α และ β รวมทั้งการหาช่วงความเชื่อมั่นของค่าประมาณของ α และ β กล่าวคือในการหาค่าประมาณของ α และ β โดยวิธีกำลังสองน้อยสุด (Method of Least Squares) ซึ่งจะได้กล่าวต่อไป ข้อสมมติที่จำเป็นคือข้อ 2 และ 4 แต่ถ้าใช้วิธี Maximum Likelihood ข้อสมมติที่ 1 ก็จำเป็นด้วย นอกจากนี้ถ้าข้อสมมติที่ 2,3,4 และข้อสมมติที่ว่าตัวแปร e_i ไม่ขึ้นอยู่กับค่าของ X เป็นจริง จะได้ว่าค่าประมาณของ α, β และ μ_Y ที่ได้จากวิธีกำลังสองน้อยสุดจะมีคุณสมบัติเป็น Best Linear Unbiased Estimators (BLUE) ส่วนในการทดสอบสมมุติฐานและการหาช่วงความเชื่อมั่นนั้นข้อสมมติที่จำเป็นคือ 1 และ 3

เนื่องจากค่าที่แท้จริงของ α และ β เป็นค่าที่ไม่สามารถทราบได้ จึงจำเป็นที่จะต้องประมาณโดยใช้ข้อมูลตัวอย่าง สมการการถดถอยตัวอย่าง(Sample Regression Equation) จะเป็น

$$\hat{Y} = a + bX \qquad (2\text{-}24)$$

เมื่อ $\hat{Y}$, a, b คือค่าประมาณของ μ_Y, a, b ตามลำดับ

ที่ค่า X ค่าใดค่าหนึ่ง สมมุติเป็น X_i จะพยากรณ์ค่าเฉลี่ยของ Y ได้จากสมการ (2-24) นั่นคือ

$$\hat{Y} = a + bX_i$$

ให้ e_i เป็นผลต่างระหว่างค่าที่สังเกตได้ Y_i (observed Y_i) และค่าพยากรณ์ $\hat{Y}_i$ ซึ่งได้มาจาก X_i

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$e_i = Y_i - \hat{Y}_i$$

$$Y_i = \hat{Y}_i + e_i$$

$$Y_i = a + bX_i + e_i \qquad (2\text{-}25)$$

สมการ (2-25) เรียกว่าตัวแบบการถดถอยตัวอย่าง (Sample Regression Model) และ e_i เรียกว่า residual

เมื่อทราบลักษณะของสมการการถดถอยแล้ว ขั้นต่อไปคือการหาค่า a และ b ในสมการ (2-24) การหาค่าประมาณ a และ b สามารถทำได้หลายวิธี แต่วิธีที่เป็นที่นิยมใช้กันมากได้แก่วิธีกำลังสองน้อยสุด (Method of Least Squares)คือการหาค่า a และ b ในเทอมของ X และ Y เพื่อให้ผลรวมของความคลาดเคลื่อน (ผลต่างระหว่างค่าที่พยากรณ์ได้และค่าที่สังเกตุได้) กำลังสองมีค่าน้อยที่สุด นั่นคือทำให้ $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i-\hat{Y}_i)^2$ มีค่าน้อยที่สุด

2.4.1.2 วิธีหาค่าประมาณของสัมประสิทธิ์การถดถอยแบบกำลังสองน้อยสุด

หาค่า a และ b ซึ่งทำให้ $E = \sum_{i=1}^{n} (Y_i-\hat{Y}_i)^2$ มีค่าน้อยที่สุด

$$E = \sum_{i=1}^{n} (Y_i - a - bX_i)^2$$

$$= \sum_{i=1}^{n} (Y_i^2 - 2aY_i - 2bX_iY_i + a^2 + 2abX_i + b^2X^2_i)$$

$$= \sum_{i=1}^{n} Y^2_i - 2a \sum_{i=1}^{n} Y_i - 2b \sum_{i=1}^{n} X_iY_i + na^2 + 2ab \sum_{i=1}^{n} X_i + b^2 \sum_{i=1}^{n} X^2_i$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\frac{\partial E}{\partial a} = -2\sum_{i=1}^{n} Y_i + 2an + 2b\sum_{i=1}^{n} X_i$$

$$\frac{\partial E}{\partial b} = -2\sum_{i=1}^{n} X_iY_i + 2a\sum_{i=1}^{n} X_i + 2b\sum_{i=1}^{n} X^2_i$$

ให้ $\frac{\partial E}{\partial a}$ และ $\frac{\partial E}{\partial b} = 0$ จะได้สมการปกติ (normal equation)

$$an + b\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \qquad (2\text{-}26)$$

$$a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X^2_i = \sum_{i=1}^{n} X_iY_i \qquad (2\text{-}27)$$

สมการ (2-27) x n - (2-26) x $\sum_{i=1}^{n} X_i$ จะได้

$$bn\sum_{i=1}^{n} X^2_i - b\left(\sum_{i=1}^{n} X_i\right)^2 = n\sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i$$

$$b = \frac{n\sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X^2_i - \left(\sum_{i=1}^{n} X_i\right)^2}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากสมการ(2-26) $$a = \frac{\sum_{i=1}^{n} Y_i}{n} - b\frac{\sum_{i=1}^{n} X_i}{n}$$

$$= \bar{Y} - b\bar{X}$$

2.4.1.3 ความแปรปรวนของตัวแปรตาม

จากข้อมูลตัวอย่าง เราสามารถพิจารณาลักษณะความแปรปรวนของ Y_i จากค่าเฉลี่ย $\bar{Y}$ ได้จากรูปที่ 2-6 ซึ่งแสดงกราฟของการถดถอยอย่างง่าย จะเห็นว่าความแปรปรวนทั้งหมดประกอบด้วยความแปรปรวน 2 ส่วน ส่วนแรกคือส่วนที่ Y_i แตกต่างจากค่าประมาณที่ได้จากเส้นตรงการถดถอย $(\hat{Y}_i)$ ซึ่งเรียกว่าความแปรปรวนที่อธิบายไม่ได้ (Unexplained Variation) ส่วนที่สองคือส่วนที่ $\hat{Y}_i$ แตกต่างจาก $\bar{Y}$ ซึ่งเรียกว่าความแปรปรวนที่อธิบายได้ (Explained Variation) นั่นคือ

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$$

จากสมการข้างต้นหาผลบวกกำลังสองจะได้

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

ให้ $$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\text{และ} \qquad SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

$$\text{จะได้} \qquad SST = SSE + SSR$$

$$\text{ให้} \qquad R^2 = \frac{\text{ความแปรปรวนที่อธิบายได้}}{\text{ความแปรปรวนทั้งหมด}}$$

$$= \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

R^2 นี้เรียกว่า Coefficient of Determination ซึ่งมีค่าอยู่ระหว่าง 0 กับ 1 $R^2 \times 100$ จะหมายถึงเปอร์เซ็นต์ของความแปรปรวนของตัวแปรตามที่สามารถอธิบายได้ด้วยตัวแปรอิสระ หรืออีกนัยหนึ่งคือเปอร์เซ็นต์ของบทบาทของตัวแปรอิสระที่มีส่วนในการพยากรณ์ตัวแปรตาม

ในการวิเคราะห์ความถดถอย ค่า R^2 ที่สูง เป็นสิ่งที่ต้องการเพราะนั่นหมายถึงตัวแปรอิสระสามารถใช้ในการพยากรณ์ตัวแปรตามได้ดี หรือกล่าวอีกนัยหนึ่งคือสมการถดถอยสามารถอธิบายใช้ในการพยากรณ์ตัวแปรไม่อิสระได้ ส่วนค่า R ซึ่งเรียกว่าสัมประสิทธิ์ความสัมพันธ์หลายมิติ (Multiple Correlation Coefficient) ใช้วัดความสัมพันธ์ระหว่างตัวแปรไม่อิสระกับตัวแปรอิสระทั้งหมดโดยทั่ว ๆ ไปมักจะพิจารณา R^2 มากกว่า R เพราะ R^2 ให้ความหมายที่ชัดเจนกว่า และใช้ในการพิจารณาว่าตัวแปรอิสระตัวใดควรจะเข้าในสมการการถดถอย หรือไม่โดยดูว่าเปอร์เซนต์ของความแปรปรวนที่อธิบายได้เพิ่มขึ้นเท่าใดจากการเพิ่มตัวแปรอิสระนั้นในสมการ

เมื่อกล่าวถึง R^2 และ R ควรจะได้กล่าวถึงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระแต่ละตัว ซึ่งวัดด้วยสัมประสิทธิ์ความสัมพันธ์บางส่วน (Partial Correlation Coefficient) ซึ่งใช้สัญญลักษณ์ r_{YX}

$$r_{YX} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})\ (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}\ \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

r_{YX} มีค่าระหว่าง -1 และ 1 ที่ค่าใกล้ -1 หรือ 1 แสดงว่า X และ Y มีความสัมพันธ์กันมากในทางตรงกันข้าม (-) หรือในทางเดียวกัน(+) ถ้า r_{YX} มีค่าเป็น 0 หมายถึงตัวแปรที่สองไม่มีความสัมพันธ์กันและในเชิงเส้นตรง

2.4.2 การวิเคราะห์ความสัมพันธ์ในตัวเอง (Autocorrelation Analysis)

การวิเคราะห์ความสัมพันธ์ในตัวเองเป็นเทคนิคที่ใช้สำหรับหาความสัมพันธ์และรูปแบบของอนุกรม เพื่อเป็นแนวทางในการเลือกแบบจำลองทางสโตแคสติค โดยค่าฟังก์ชั่นความสัมพันธ์ในตัวเองเป็นค่าความสัมพันธ์ระหว่างสัมประสิทธิ์ของความสัมพันธ์ในตัวเอง(Autocorrelation Coefficient) กับช่วงเวลาถัดไป k (lag time) ซึ่งค่าสัมประสิทธิ์ความสัมพันธ์ในตัวเองใช้วัดความสัมพันธ์เชิงเส้นของตัวแปร X_t กับ X_{t+k} เขียนแทนด้วย ρ_k สำหรับตัวแปรทั้งหมด และ r_k สำหรับตัวแปรตัวอย่าง หาค่าได้โดย

$$\rho_k = \frac{\sum_{t=1}^{N-k} (X_t-\mu)(X_{t+k}-\mu)}{\sum_{t=1}^{N} (X_t-\mu)^2} \qquad (2\text{-}28)$$

$$r_k = \frac{\sum_{t=1}^{N-k} (X_t-\bar{X})(X_{t+k}-\bar{X})}{\sum_{t=1}^{N} (X_t-\bar{X})^2} \qquad (2\text{-}29)$$

ในกรณีที่อนุกรมไม่สมดุลย์ ค่า r_k จากสมการ (2-29) ใช้ประมาณ ρ_k ได้ไม่แม่นยำ โดยเฉพาะอนุกรมทางอุทกวิทยาที่เป็นอนุกรมเปิด ไม่มีจุดเริ่มต้นและจุดสิ้นสุด (Open Series) มักใช้สมการ (2-30) ในการประมาณ ρ_k แทน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$r_k = \frac{\sum_{t=1}^{N-k} (X_t-\bar{X}_t)(X_{t+k}-\bar{X}_{t+k})}{\left(\sum_{t=1}^{N-k} (X_t-\bar{X}_t)^2 \sum_{t=1}^{N-k} (X_{t+k}-\bar{X}_{t+k})^2\right)^{1/2}} \qquad (2\text{-}30)$$

โดย $\bar{X}_t$ เป็นค่าเฉลี่ย N-k ค่าแรกคือ $X_1, X_2, \ldots, X_{N-k}$

$\bar{X}_{t+k}$ เป็นค่าเฉลี่ย N-k ค่าหลังคือ $X_{k+1}, X_{k+2}, \ldots, X_N$

ทั้งสมการ (2-29) และ (2-30) ถ้า k = 0 จะได้ $r_k = 1$ และ r_k อยู่ระหว่าง -1 ถึง 1 สำหรับการเลือก k นั้น ไม่ควรให้ k มีค่ามากนัก เพราะถ้า k มากจะทำให้ N-k น้อยลง ดังนั้นความแม่นยำของ r_k ก็น้อยลง

เมื่อนำค่าสัมประสิทธิ์ของความสัมพันธ์ในตัวเองไปเขียนกราฟสัมพันธ์กับ k จะได้กราฟซึ่งเรียกว่ากราฟความสัมพันธ์ (Correlogram) หรือฟังก์ชั่นความสัมพันธ์ในตัวเอง (Autocorrelation Function) ดังในรูปที่ 2-7 แสดงตัวอย่างกราฟความสัมพันธ์ โดยทั่วไปนิยมแสดงกราฟความสัมพันธ์ด้วยกราฟเส้นมากกว่ากราฟแท่ง เพราะสามารถแสดงลำดับเหตุการณ์ได้ชัดเจนกว่า

จากกราฟความสัมพันธ์ ถ้าตัวแปรไม่เป็นอิสระต่อกัน (Dependent Variables) ค่า r_k จะมีค่าเข้าใกล้ 1 ถ้าตัวแปรเป็นอิสระต่อกัน (Independent Variables) ค่า r_k จะมีค่าเป็น 0 หรือมีค่าใกล้ 0 โดยอยู่ในขอบเขตที่ยอมรับได้ (Confidence Interval) ดังสมการ (2-31 ก และ ข)

$$r_k(95\%) = \frac{-1 + 1.96\sqrt{N-k-1}}{N-k} \qquad (2\text{-}31\ ก)$$

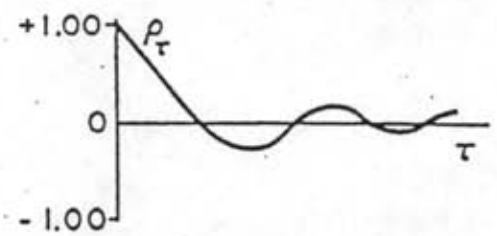
โดย $r_k(95\%)$ เป็นค่าขอบเขตของ r_k ที่ 95% ของข้อมูล N ตัว

$$r_k(99\%) = \frac{-1 + 2.326\sqrt{}}{N-k} \qquad (2\text{-}31\ ข)$$

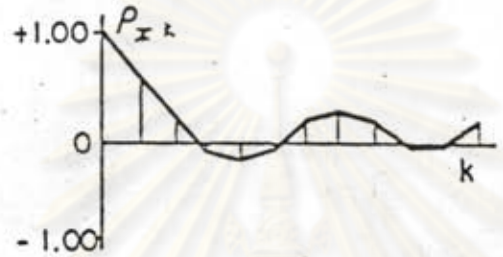
โดย $r_k(99\%)$ เป็นค่าขอบเขตของ r_k ที่ 99% ของข้อมูล N ตัว

รูปที่ 2-8 แสดงกราฟความสัมพันธ์ r_k ซึ่งหาจากสมการ (2-30) ของน้ำท่า

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

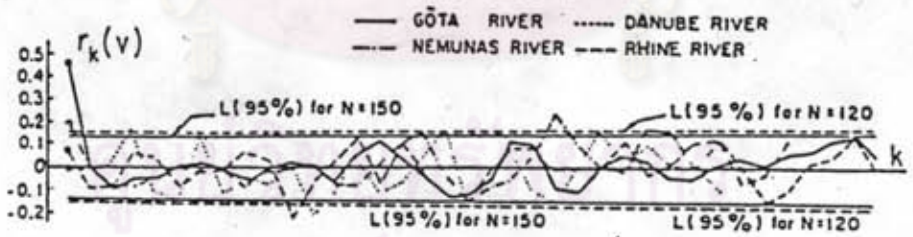


A continuous correlogram, or the autocorrelation function of a variable



A discrete correlogram, or the autocorrelation function of a variable

รูปที่ 2-7 ลักษณะของกราฟความสัมพันธ์ (Correlogram) [17]



Correlograms of annual runoff series of four European rivers: the Göta River, Sweden (N = 150), the Nemunas River, Lithuania (N = 132), the Rhine River at Basle, Switzerland (N = 150), and the Danube River at Orsova, Romania (N = 120). Probability limits at the 95 percent level are given for the normal independent variables for two lengths; N = 150 (max) and N = 120 (min), (Yevjevich, 1964).

รูปที่ 2-8 กราฟความสัมพันธ์ (Correlogram) ของน้ำท่ารายปีที่แม่น้ำ 4 สาย ในยุโรป [14]

รายปีสำหรับแม่น้ำ 4 สายในยุโรป และขอบเขตหาจากสมการ (2-31 ก) สำหรับ N = 150 และ N = 120 ค่า r_1 ของแม่น้ำไรน์ (Rhine) และดานูป (Danube) อยู่ในขอบเขต 95% ส่วนค่า r_1 ของแม่น้ำโกต้า (Gota) และเนมูนัส (Nemunies) อยู่นอกขอบเขตนี้ [12]

2.4.3 การวิเคราะห์ทางสเป็คทรอล (Spectral Analysis)

การวิเคราะห์ทางสเป็กทรอล เป็นเทคนิคที่ใช้สำหรับศึกษาโครงสร้างของอนุกรมอีกวิธีหนึ่ง โดยเฉพาะใช้หาคุณสมบัติของวงจร (Cyclicity) การวิเคราะห์นี้มี 2 แบบคือ การวิเคราะห์ค่าสเป็คทรอลเชิงเส้น (Line Spectral Analysis) และการวิเคราะห์ค่าความหนาแน่นของสเป็คทรอล (Variance Density Spectral Analysis)

2.4.3.1 การวิเคราะห์ค่าสเป็คทรอลเชิงเส้น

การวิเคราะห์ค่าสเป็คทรอลเชิงเส้น ใช้หาคุณสมบัติอนุกรมที่มีวงจรโดยสมมุติค่าความถี่เชิงมุม (Angular Frequency) ทั้งหมดที่เป็นไปได้ จากนั้นประมาณค่าแอมปิจูดและเฟสของแต่ละความถี่เชิงมุมนั้นจากอนุกรมแล้วแทนในสมการ (2-32)

$$\mu_t = \mu_x + \sum_{j=1}^{m} [A_j \cos \lambda_j t + B_j \sin \lambda_j t] \qquad (2-32)$$

$$A_j = \frac{2}{N} \sum_{t=1}^{N} X_t \cos \lambda_j t \ , \ j = 1,2,\ldots,m \qquad (2-33)$$

$$B_j = \frac{2}{N} \sum_{t=1}^{N} X_t \sin \lambda_j t \ , \ j = 1,2,\ldots,m \qquad (2-34)$$

$$C_j = \frac{1}{2} [A_j^2 + B_j^2]^{1/2} \qquad (2-35)$$

$$\phi_j = \tan^{-1} [B_j/A_j] \qquad (2-36)$$

โดย μ_t เป็นอนุกรมของส่วนที่เป็นวงจร

μ_x เป็นค่าเฉลี่ยของอนุกรมประมาณได้ด้วย X

A_j, B_j เป็นค่าสัมประสิทธิ์ฟูเรียร์ (Fourier Coefficient)

C_j เป็นค่าแอมปิจูด ซึ่งค่า $C_j^2/2$ เป็นค่าความแปรปรวนที่ความถี่นั้น ๆ หรือเรียกว่าค่าสเป็กทรอลเชิงเส้น

ϕ_j เป็นค่าเฟส

N เป็นจำนวนข้อมูล

m เป็นจำนวนความถี่ที่เป็นไปได้ซึ่งขึ้นกับ N คือ ถ้า N เป็นเลขคู่ m มีค่า $N/2-1$ ถ้า N เป็นเลขคี่ m มีค่า $(N-1)/2$ เป็นความถี่เชิงมุมมีค่า

λ_j เป็นความถี่เชิงมุม มีค่า $2\pi f_j$

f_j เป็นความถี่ของวงจร มีค่า j/N

ในทางปฏิบัติ ไม่นิยมใช้ j ทั้ง m ค่า แต่จะเลือกใช้ j ที่มีความสำคัญเพียงบางตัวเท่านั้น ดังในรูปที่ 2-9 แสดงการวิเคราะห์สเป็กทรอลเชิงเส้นของข้อมูล 36 ค่า โดยรูปที่ 2-7ก เป็นข้อมูล รูปที่ 2-9ข เป็นค่าสเป็คทรอลเชิงเส้นที่ความถี่ต่าง ๆ หรือเรียกกราฟนี้ว่ากราฟวงจร (Periodogram) รูปที่ 2-7ค แสดงเปรียบเทียบข้อมูลกับอนุกรมที่สร้างขึ้นจากความถี่ 5 ค่า [10]

สำหรับการเลือกความถี่ที่สำคัญสามารถพิจารณาจากกราฟวงจร แล้วเลือกค่าความถี่ที่ให้ค่าสเป็กทรอลเชิงเส้นสูง ๆ หรือสามารถพิจารณาจากกราฟวงจรสะสม (Cummulative Periodogram) ซึ่งเกิดจากการเขียนกราฟระหว่าง P_i กับ i ตามสมการ

$$P_i = \frac{\sum_{j=1}^{i} C_j^2/2}{S^2} \qquad (2\text{-}37)$$

โดย P_i เป็นค่ากราฟวงจรสะสม

$C_j^2/2$ เป็นค่าสเป็กทรอลเชิงเส้น

S^2 เป็นค่าความแปรปรวนของข้อมูลที่พิจารณา

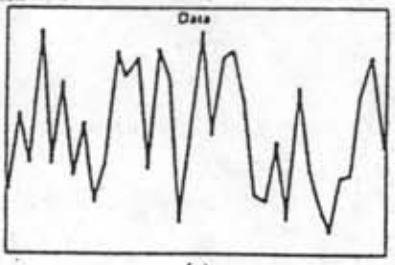
i เป็นจำนวนเต็ม มีค่า $1, 2, \ldots, m$

โดยทั่วไปกราฟนี้จะประกอบด้วย 2 ส่วน ส่วนแรกเป็นส่วนที่เป็นวงจร ค่า P_i จะเห็นอย่างรวดเร็วเมื่อเพิ่ม i อีกส่วนเป็นส่วนที่ไม่แน่นอน ค่า P_i จะเพิ่มขึ้นอย่างช้า ๆ เมื่อเพิ่ม i

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

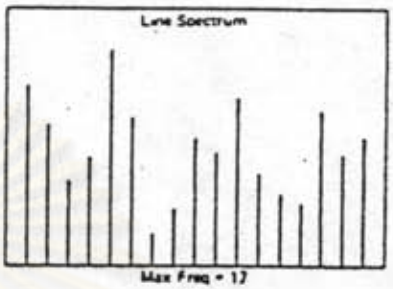
ท.ว.194660

(a) A PLOT OF THE RAW DATA



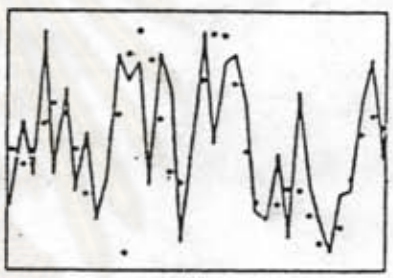
(a)

(b) THE LINE SPECTRUM RESULTS
(also called the Periodogram)



(b)

(c) THE RECONSTRUCTED SERIES
using the first 5 waves
(The dots are the reconstructed series.
The solid line is the original data.)

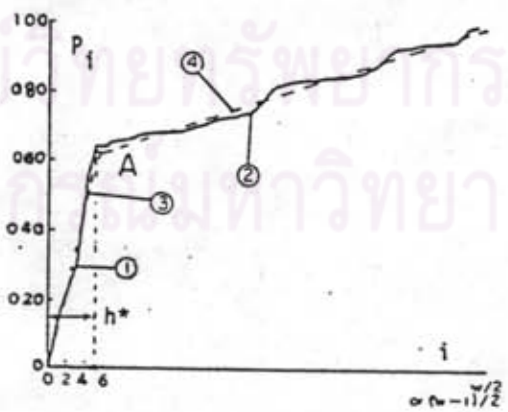


(c)

THE LINE SPECTRUM ANALYSIS OF THE 36 DATA POINTS

F. 2-9

รูปที่ 2-9 การวิเคราะห์สเป็คทรอลเชิงเส้นของข้อมูลตัวอย่าง [11]



Separation of the cumulative periodogram into the periodic part, for both the observed (1) and the fitted (3), and the sampling variation part, also for both the observed (2) and the fitted (4), in case of a periodic series with an independent stochastic component.

F. 2-10

รูปที่ 2-10 กราฟวงจรสะสม (Cumulative Periodogram) ของอนุกรมที่มีวงจร [14]

จุดที่เกิดจากการปรับกราฟทั้ง 2 ส่วนให้เรียบตัดกันจะเป็นจำนวนความถี่ของวงจรที่สำคัญตามต้องการ ดังแสดงในรูปที่ 2-10 เป็นกราฟวงจรสะสมที่มีความถี่ของวงจรเท่ากับ h* ถ้าอนุกรมมีตัวแปรที่เป็นอิสระต่อกันกราฟนี้จะเป็นเส้นตรงเชื่อมโยงระหว่างจุด (0,0) กับ (m,1) หรืออาจจะขึ้นลงรอบเส้นตรงนี้ แต่ยังอยู่ในขอบเขตที่เชื่อถือได้ ซึ่งตามการทดสอบของโคโมโกรอฟสไมร์นอฟ (Kolmogorov-Smirnov) [6] จะได้

$$\Delta P_i(95\%) = \frac{+1.36}{\sqrt{m}} \qquad (2\text{-}38ก)$$

$$\Delta P_i(75\%) = \frac{+1.02}{\sqrt{m}} \qquad (2\text{-}38ข)$$

ซึ่ง $P_i(95\%)$ และ $P_i(75\%)$ เป็นค่าขอบเขตที่ผลต่างของ P_i ของข้อมูลกับ P_i ตามทฤษฎี มีความเชื่อมั่น 95% และ 75% ตามลำดับ

2.4.3.2 การวิเคราะห์ค่าความหนาแน่นของสเป็คทรอล

การวิเคราะห์สเป็คทรอลเชิงเส้นที่กล่าวมานั้นใช้ความถี่เชิงมุมที่ไม่ต่อเนื่อง ส่วนการวิเคราะห์ค่าความหนาแน่นของสเป็คทรอล จะใช้ความถี่เชิงมุมอย่างต่อเนื่อง ซึ่งสามารถใช้กับอนุกรมทางอุทกวิทยาที่ต่อเนื่องได้ และยังประยุกต์ใช้กับอนุกรมทางอุทกวิทยาที่ไม่ต่อเนื่องได้ดีกว่าการวิเคราะห์สเป็คทรอลเชิงเส้น เพราะสามารถประมาณค่าความถี่พื้นฐานและแอมปิจูดของส่วนที่เป็นวงจรได้

จากสมการ (2-32) สามารถแปลงให้อยู่ในรูปเลขยกกำลังด้วยจำนวนเชิงซ้อนได้ คือ

$$f(t) = \mu_\tau - \mu_x = \sum_{j=1}^{\infty} C_j e^{i\lambda_j t} \qquad (2\text{-}39)$$

$$C_j = \frac{2}{T}\int_0^T f(t)e^{-i\lambda_j t}\,dt \qquad (2\text{-}40)$$

โดย i เป็นจำนวนเชิงซ้อน = $\sqrt{-1}$ และ T เป็นความอนุกรมแทนค่า C_j จากสมการ (2-40) ในสมการ (2-39) จะได้

$$f(t) = \mu_\tau - \mu_x = \sum_{j=1}^{\alpha}\left[\frac{2}{T}\int_0^T f(t)e^{-i\lambda_j t}\,dt\right] e^{i\lambda_j t} \qquad (2\text{-}41)$$

เขียนให้อยู่ในรูปฟูเรียร์อินทิกรัล (Fourier Integral) และฟูเรียร์ทรานส์ฟอร์ม (Fourier Transform) จะได้สมการ (2-42) และ (2-43) ตามลำดับ

$$f(t) = \frac{1}{\pi}\int_0^{\infty} \phi(\lambda)e^{i\lambda t} \qquad (2\text{-}42)$$

$$\phi(\lambda) = \int_0^{\infty} f(t)e^{i\lambda t}\,dt \qquad (2\text{-}43)$$

ซึ่ง $\phi(\lambda)$ เป็นฟังชั่นความหนาแน่นของสเป็คทรอลสามารถอธิบายในรูปจำนวนเชิงซ้อนได้เป็น

$$\phi(\lambda) = \alpha(\lambda) + i\beta(\lambda) \qquad (2\text{-}44)$$

$$\alpha(\lambda) = \int_0^{\infty} f(t)\cos \lambda t\,dt \qquad (2\text{-}45)$$

$$\beta(\lambda) = \int_0^{\infty} f(t)\sin \lambda t\,dt \qquad (2\text{-}46)$$

โดย $\alpha(\lambda)$, $\beta(\lambda)$ เป็นความหนาแน่นของ A_j และ B_j ที่แต่ละ λ

โดยทั่วไปนิยมพิจารณาความหนาแน่นของสเป็คทรอลสำหรับความแปรปรวนหนึ่งหน่วย $\upsilon(\lambda) = \phi(\lambda)/\sigma^2_x$ ซึ่งฟังชั่นนี้มีความสัมพันธ์กับฟังชั่นความสัมพันธ์ในตัวเอง สามารถเขียนสมการทำนองเดียวกับสมการ (2-42) และ (2-43) ได้เป็น

$$\upsilon(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\rho(\tau)e^{i\lambda\tau}d\tau = \frac{1}{2\pi}\int_{-\infty}^{\infty}\rho(\lambda)\cos \lambda\tau\, d\tau \qquad (2\text{-}47)$$

$$\rho(\tau) = \int_{-\infty}^{\infty}\upsilon(\lambda)e^{i\tau\lambda}d\lambda = \int_{-\infty}^{\infty}\upsilon(\lambda)\cos \tau\lambda\, d\lambda \qquad (2\text{-}48)$$

เนื่องจากฟังชั่นความสัมพันธ์ในตัวเองสมมาตรรอบแกน y และใช้ λ อยู่ระหว่าง o ถึง ∞ τ อยู่ระหว่าง 0 ถึง ∞ สามารถแทนในสมการ (2-47) และ (2-48) ได้

$$\gamma(\lambda) = \frac{2}{\pi} \int_0^{\infty} \rho(\tau)\cos \lambda\tau \, d\tau \qquad (2\text{-}49)$$

$$\rho(\tau) = \int_0^{\infty} \gamma(\lambda)\cos \tau\lambda \, d\lambda \qquad (2\text{-}50)$$

โดย $2v(\lambda)$ มีค่าเท่ากับ $\gamma(\lambda)$

ดังนั้นสามารถเขียนสมการ (2-32) เป็นฟังค์ชันต่อเนื่องได้เป็น

$$\mu_\tau = \mu_x + \sigma_x^2 \int_0^{\infty} \alpha(\lambda)\cos \tau\lambda \, d\lambda + \sigma_x^2 \int_0^{\infty} \beta(\lambda)\sin \tau\lambda \, d\lambda \qquad (2\text{-}51)$$

ซึ่ง $\alpha(\lambda)$ และ $\beta(\lambda)$ เป็นฟังค์ชั่นต่อเนื่องประมาณได้จาก $a(\lambda)$ และ $b(\lambda)$ ดังสมการ (2-44) และ (2-45) โดยใช้ความยาวของอนุกรม T ค่า $\alpha(\lambda)$ และ $\beta(\lambda)$ นั้นสามารถเขียนให้สัมพันธ์กับ $\gamma(\lambda)$ ได้เป็น

$$\gamma(\lambda) = \frac{1}{2}[\alpha^2(\lambda) + \beta^2(\lambda)]^{1/2} \qquad (2\text{-}52)$$

จากสมการ (2-49) และ (2-50) แปลงให้อยู่ในรูปของความถี่ได้เป็น

$$\gamma(f) = 4\int_0^{\infty} \rho(\tau)\cos 2\pi f\tau \, d\tau \qquad (2\text{-}53)$$

$$\rho(\tau) = \int_0^{\infty} \gamma(f)\cos 2\pi\tau f \, df \qquad (2\text{-}54)$$

ถ้าใช้กับอนุกรมที่ไม่ต่อเนื่อง สมการ (2-49) และ (2-53) จะกลายเป็น

$$\gamma(\lambda) = \frac{1}{\pi}[1+2\sum_{k=1}^{\infty} \rho(k)\cos \lambda k] \qquad (2\text{-}55)$$

$$\gamma(f) = 2[1+2\sum_{k=1}^{\infty} \rho(k)\cos 2\pi fk] \qquad (2\text{-}56)$$

ซึ่งค่าผลรวมทางขวามือของ (2-55) และ (2-56) เป็น 0 สำหรับตัวแปรที่เป็นอิสระต่อกัน จะ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ได้ $\gamma(\lambda) = 1/\pi$ และ $\gamma(f) = 2$ โดย $0 < y <$ และ $0 < f < 0.5$

ในการประมาณค่า $\gamma(f)$ จากตัวอย่างซึ่งค่อนข้างจะผันแปรสูง จึงควรมีการปรับให้เรียบ (Smoothing) โดยสามารถคำนวณหาจาก

$$g(f) = 2[1 + 2 \sum_{k=1}^{m} D(k)\, \rho(k) \cos 2\pi f k] \qquad (2\text{-}57)$$

ซึ่งค่า m เป็น lag time มากที่สุดที่เป็นไปได้ของข้อมูลตัวอย่างมี N ข้อมูลจะได้ m มีค่า N-1 แต่โดยที่ไปใช้ $m < N/4$ เพื่อให้เกิดความแม่นยำในการประมาณ และค่า D(K) เป็นตัวปรับค่าฟังชั่นตามเวลาให้เรียบมากขึ้น โดยทั่วไปนิยมปรับค่าฟังชั่นตามความถี่ให้เรียบมากขึ้น โดยใช้ S(f) ตามสมการ

$$S(f_j) = \frac{1}{4} g(f_{j-1}) + \frac{1}{2} g(f_j) + \frac{1}{4} g(f_{j+1}) \qquad (2\text{-}58)$$

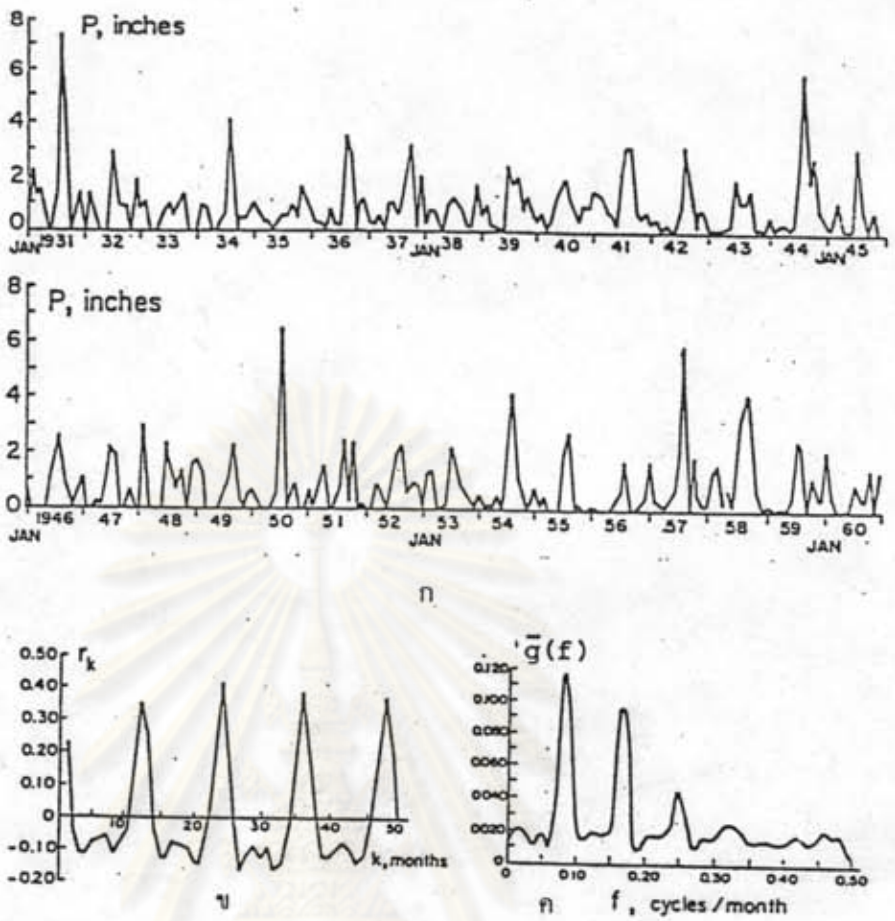
ซึ่ง g(f) หาจากสมการ (2-56) ค่า $S(f_j)$ ในสมการนี้แปลงกลับเป็นตัวปรับฟังชั่นตามเวลาได้เป็น

$$D(k) = (1 + \cos 2\pi f k)/2 \qquad (2\text{-}59)$$

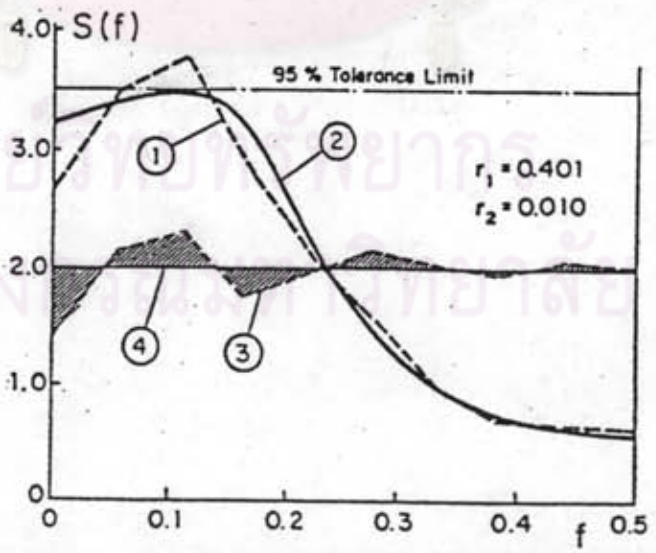
ปกติ S(f) ในสมการ (2-55) จะไม่ติดลบ อย่างไรก็ตาม ในการคำนวณอาจมีความคลาดเคลื่อนได้บ้าง ทำให้บางครั้งค่า S(f) ติดลบ จึงมักแทนค่าติดลบนี้ด้วยศูนย์

เมื่อนำค่า S(f) นี้ไปเขียนกราฟสัมพันธ์กับความถี่จะได้ฟังชั่นความหนาแน่นของสเป็คทรอล ดังในรูปที่ 2-11ค แสดงค่าความหนาแน่นของสเป็คทรอล สำหรับข้อมูลน้ำฝนรายเดือนที่ Hachita, New Mexico ซึ่งหาจากกราฟความสัมพันธ์ในรูป 2-11ข จากกราฟนี้สามารถหาความถี่พื้นฐานของวงจรต่าง ๆ ได้ โดยพิจารณาความถี่ที่ให้ค่าความหนาแน่นของสเป็คทรอลสูง ๆ [15]

ถ้าข้อมูลมีความไม่แน่นอนมาก (Random Variable) กราฟความหนาแน่นของสเป็คทรอลนี้ จะมีค่า S(f) = 2 หรือมีค่าขึ้นลงรอบ ๆ ค่านี้ ดังในรูป 2-12 โดยอยู่ในขอบเขตที่เชื่อถือได้ ซึ่งตามการทดสอบแบบไคส์แควร์ (Chi-square) [15] จะได้ ดังในรูป 2-12



รูปที่ 2-11 ชุดข้อมูลฝนรายเดือนที่ Hachita, New Mereico (1931-1960) พร้อมทั้งกราฟความสัมพันธ์ และความหนาแน่นทางสเป็คทรอล [17]



Spectra of annual flow series of the Fox River, Wisconsin, for 60 years: (1) estimated spectrum, (2) fitted spectrum of the second-order model (r_1 = 0.401, r_2 = 0.010), (3) spectrum of residuals, and (4) expected spectrum of the independent series.

รูปที่ 2-12 ความหนาแน่นทางสเป็คทรอลของปริมาณน้ำท่ารายปีที่แม่น้ำ Fox , Winsonsinnsin [17]

$$P[S(f) > \frac{1-\chi^2_{\alpha}(\upsilon)}{\upsilon}\gamma(f)] = \frac{\alpha}{2} \qquad (2\text{-}60\text{ก})$$

$$P[S(f) < \frac{\chi^2_{\alpha}(\upsilon)}{\upsilon}\gamma(f)] = 1 - \frac{\alpha}{2} \qquad (2\text{-}60\text{ข})$$

โดย υ เป็นระดับของความเป็นอิสระมีค่า $\frac{8N}{3m}$ ถึง $\frac{2N}{m}$

N เป็นจำนวนข้อมูลทั้งหมด

n เป็น lag time มากที่สุดที่พิจารณา

$\chi^2_{\alpha}(\upsilon)$ เป็นค่าจากตารางการแจกแจงไคส์แควร์

α เป็นค่าระดับความน่าจะเป็น

$\gamma(f)$ มีค่าเป็น 2

$\frac{1-\chi^2_{\alpha}(\upsilon)}{\upsilon}\gamma(f)$ เป็นขอบเขตระดับล่าง

$\frac{\chi^2_{\alpha}(\upsilon)}{\upsilon}\gamma(f)$ เป็นขอบเขตระดับบน

สำหรับค่าความหนาแน่นของสเป็คทรอลที่ f 0 และ $f = 0.5$ จะมีค่ามากกว่าค่าความหนาแน่นของสเป็คทรอลที่ f อื่น ๆ ดังนั้นค่าขอบเขตต้องเพิ่มขึ้น 2 เท่าของค่าปกติซึ่งในทางปฏิบัติมีความสำคัญมากเพราะค่า $g(f_0)$ และ $g(f_{0.5})$นี้มักมีค่าสูงกว่าขอบเขตของ f อื่น ๆ [15]

2.5 แบบจำลองเบื้องต้นที่ศึกษา

การพยากรณ์ปริมาณฝนที่ศึกษาอาศัย การวิเคราะห์อนุกรมเวลา ศึกษาคุณสมบัติของแบบจำลองต่าง ๆ ในอนุกรมเวลาเพื่อประโยชน์ในการเลือกแบบจำลอง ส่วนวิธีการเลือกแบบจำลอง การคาดคะเนค่าของพารามิเตอร์ การตรวจสอบความถูกต้องของแบบจำลองและการพยากรณ์นั้น ก่อนที่จะกล่าวถึงแบบจำลองใคร่ขอทำความเข้าใจโดยทั่วไปเกี่ยวกับการวิเคราะห์อนุกรมเวลา (Time Series Analysis) เกี่ยวกับคำจำกัดความพื้นฐานบางตัวเสียก่อน

ตัวกรองเชิงเส้น (Linear Filter) เป็นตัวควบคุมเชิงเส้นที่ทำให้ส่วนกลับของลำดับของตัวแปรสุ่มที่ไม่มีความสัมพันธ์กัน a_t ไปเป็นลำดับของตัวแปรที่มีความสัมพันธ์กัน Z_t ดังแสดงในรูปที่ 2-11 ตัวกรอง (Filter) $\psi(B)$ กรองตัวแปรสุ่ม a_t ออกมาเป็นชุดอนุกรม

ตัวควบคุมเชิงเส้น (Linear Operator) $\psi(B)$ เป็นโพลีโนเมี่ยล (Polynomial) ในตัวควบคุมย้อนกลับ (Backward Shift Operator) B ถ้า B ควบคุมตัวแปรสุ่ม Z_t จะได้ว่าตัวแปรสุ่มเป็น Z_{t-1} โดยกระบวนการได้ล่าช้าไปหนึ่งหน่วยของเวลา (One discrete time unit) แสดงในรูปทางคณิตศาสตร์คือ

$$BZ_t = Z_{t-1}$$

ตัวยกกำลังของ B คือ

$$B^2Z_t = BBZ_t = BZ_{t-1} = Z_{t-2}$$

โดยทั่วไป

$$B^mZ_t = Z_{t-m} \qquad (2\text{-}61)$$

โพลีโนเมียล ตัวควบคุม B ในรูปแบบเดียวกัน สำหรับตัวอย่าง

$$(a_0 + a_1B + a_2B^2)Z_t = a_0Z_t + a_1Z_{t-1} + a_2Z_{t-2} \qquad (2\text{-}62)$$

ตัวควบคุมเลื่อนไปข้างหน้า F เลื่อนกระบวนการสุ่มหนึ่งหน่วยเวลาไปในอนาคต

$$FZ_t = Z_{t+1} \qquad (2\text{-}63)$$

หรือ

$$F^mZ_t = Z_{t+m} \qquad (2\text{-}64)$$

คูณ B ในสมการ (2-63)

$$BFZ_t = BZ_{t+1} = Z_t$$

ซึ่งได้ว่า $F = B^{-1}$

ตัวควบคุมแตกความต่างย้อนกลับ ∇ กำหนดได้ตาม

$$\nabla Z_t = Z_t - Z_{t-1}$$

$$= (1 - B) Z_t \qquad (2\text{-}65)$$

ด้วยกกำลังของ ∇ กำหนดได้ด้วย

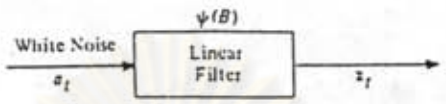
$$\begin{aligned}\nabla^2 Z_t &= \nabla \nabla Z_t = \nabla (Z_t - Z_{t-1}) \\ &= \nabla Z_t - \nabla Z_{t-1} \\ &= Z_t - Z_{t-1} - Z_{t-1} + Z_{t-2} \\ &= Z_t - 2Z_{t-1} + Z_{t-2} \\ &= (1 - B)^2 Z_t \qquad (2\text{-}66)\end{aligned}$$

ส่วนกลับของตัวควบคุมความแตกต่างย้อนกลับ ∇^{-1} เป็นตัวควบคุมรวมย้อนกลับ S กำหนดได้ตาม

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j}$$

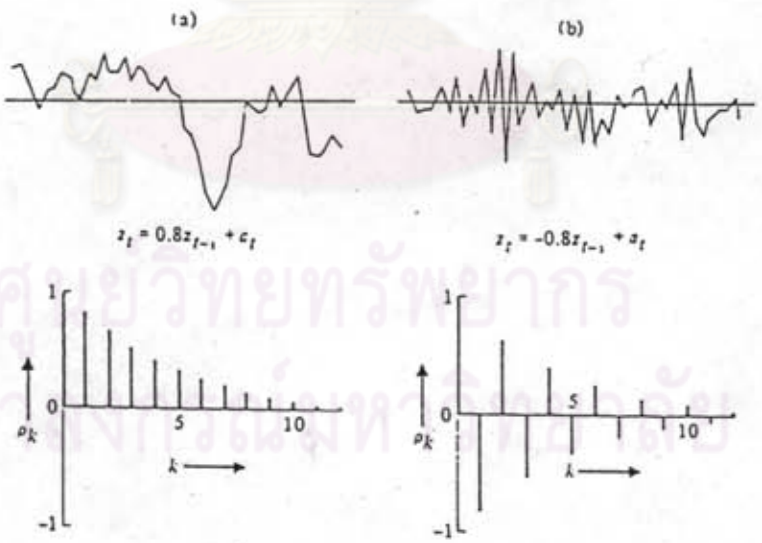
$$= (1 + B + B^2 + \ldots) Z_t \qquad (2\text{-}67)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



A linear filter acting on uncorrelated noise to create a time series

รูปที่ 2-13 การสร้างอนุกรมเวลาโดย Linear Filter [12]



Theoretical Autocorrelation Functions

Autocorrelation functions of an AR(1)

รูปที่ 2-14 ฟังก์ชั่นความสัมพันธ์ในตัวเองของ AR(1) [12]

2.5.1 แบบจำลองตัวกรองเชิงเส้นสภาวะสมดุลย์เชิงสถิติโดยทั่วไป

รูปที่ 2-13 ได้แทนแบบจำลองเชิงเส้นทั่วไปด้วยตัวควบคุม $\psi(B)$ สมมติว่าได้ผลออกมา Z_t นั้นเป็นขบวนการสภาวะสมดุลย์เชิงสถิติ (Stationary Process) ที่ปกติโดยการลดค่าเฉลี่ยของมันจากค่าไม่ต่อเนื่องทั้งหมด Z_t เป็นขบวนการค่าเฉลี่ยเป็นศูนย์ ดังนั้น Z_t ได้สัมพันธ์กันกับ white noise a_t โดย convolution ที่ไม่ต่อเนื่อง

$$\begin{aligned} Z_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \cdots \\ &= a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \\ &= \sum_{j=0}^{\infty} \psi_j a_{t-j} \\ &= \psi(B) a_t \end{aligned} \qquad (2\text{-}68)$$

ที่ $\psi_0 = 1$ และ

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \ldots$$

ลำดับของ White Noise a_t มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวน σ_a^2 ฟังชั่นความแปรปรวนในตัวเอง (Autocovariance Function) คือ

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \qquad (2\text{-}69)$$

ซึ่งให้ว่าฟังค์ชั่นความสัมพันธ์ในตัวเอง ρ_k มีค่าเป็น 1 ที่ k = 0 และเป็น 0 ในที่อื่น ๆ

การใช้สมการ (2-60) และคุณสมบัติของ a_t ฟังชั่น ความแปรปรวน ของ Z_t กลายเป็น

$$\begin{aligned} \gamma_k &= E[Z_t Z_{t-k}] = E\left[\sum_{j=0}^{\infty}\sum_{h=0}^{\infty} \psi_j \psi_h a_{t-j} a_{t-k-h}\right] \\ &= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \end{aligned} \qquad (2\text{-}70)$$

สำหรับ k = 0 สมการข้างบนจะให้ความแปรปรวนของกระบวนการ

$$\gamma_0 = \sigma_z^2 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 \qquad (2\text{-}71)$$

อย่างน้อยที่สุดผลรวมที่ไม่จำกัด ของสมการ (2-71) จะลู่เข้า (Converge) ความแปรปรวนของตัวกรองเชิงเส้นโดยทั่วไปจะไม่จำกัด (Infinity) ความแปรปรวนที่จำกัด และความสมดุลย์ต้องการการลู่เข้า (Convergence) ของไพลีโนเมียล $\Psi(B)$ สำหรับ $|B| \leq 1$ ตัวควบคุม B ในสมการ (2-68) ได้แปลความหมายตามตัวแปรดัมมี่ ซึ่งอาจจะเป็นค่าสลับซับซ้อน (Complex Value) สภาพความสมดุลย์ในสถานนั้น ไพลีโนเมียลไม่จำกัด $\Psi(B)$ จะต้องลู่เข้า (Converge) สำหรับทุกค่าของ B ภายใน หรือบนวงกลมที่มีรัศมีหนึ่งหน่วย (Unite Cycle) $|B| \leq 1$

ตัวกรองความสมดุลย์เชิงเส้นกำหนดไว้ในสมการ (2-68) เป็นค่าเฉลี่ยเคลื่อนที่ของลำดับ White Noise ที่ผ่านขั้นตอนมาแล้ว อาจเป็นไปได้ที่แสดงกระบวนการที่เหมือนกันตามลำดับการถดถอยในตัวเองที่ไม่จำกัด (Infinite Autoregressive)

$$\begin{aligned} Z_t &= \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \cdots + a_t \\ &= \sum_{j=1}^{\infty} \pi_j Z_{t-j} + a_t \end{aligned} \qquad (2\text{-}72)$$

ที่ $\pi(B)$ คือตัวควบคุมที่ถือตามรูปแบบของไพลีโนเมียล B

ตัวอย่างการเท่ากันระหว่าง ค่าเฉลี่ยเคลื่อนที่ (Moving Average) และ แบบจำลองการถดถอยในตัวเอง (Autoregressive Model) สมมติว่าเป็นแบบจำลองดังต่อไปนี้

$$\begin{aligned} Z_t &= (1 - \theta B) a_t \\ &= \psi(B) a_t \end{aligned} \qquad (2\text{-}73)$$

ซึ่งเป็นสภาวะความสมดุลย์ที่ค่าใด ๆ ของ θ

โดยการแทนค่า a_t สมการ (2-73) สามารถแสดงได้ตาม

$$a_t = Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2} + \cdots + \theta^n a_{t-n} \qquad (2\text{-}74)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ถ้า $|\theta| > 1$, z_t ขึ้นอยู่กับค่าในอดีตด้วยน้ำหนักที่เพิ่มขึ้นด้วยระยะทางในอดีตให้หลีกเลี่ยงความเหมาะสมนี้ สภาพส่วนกลับ $|\theta| < 1$ กำหนดให้มีด้วย สภาพนี้ยอมให้รวมกันระหว่างค่าปัจจุบันกับค่าในอดีต ในรูปแบบที่เห็นชัด ๆ เช่น สภาวะที่เท่ากันในสถานะนั้น $(1-\Theta B)^{-1}$ สามารถแสดงได้ตามอนุกรมเรขาคณิตไม่สู่เข้า (Convergent Infinite Peometric Serier) $(1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + ...)$

โดยการเปรียบเทียบ (Analogy) สมการ (2-72) the autoregressive version ของแบบจำลองในสมการ (2-73) มีสัมประสิทธิ์กำหนดให้โดย

$$\pi_j = \theta^j \quad \text{for } j=1,\ldots,\infty \tag{2-75}$$

โดยทั่วไปแบบจำลองเป็นส่วนกลับในโพลีโนเมียล B

$$\pi(B) = \psi^{-1}(B) \tag{2-76}$$

จะลู่เข้าสำหรับ $|B| < 1$ ทุกค่า(ภายในหรือบนวงกลมที่มีรัศมีหนึ่งหน่วย)

2.5.2 แบบจำลองการถดถอยในตัวเองอันดับ p, AR(p)

แบบจำลองการถดถอยในตัวเองของอันดับจำกัด (infinity) ด้วยกำหนดในสมการ (2-72) นั้นใช้น้อยมากในทางปฏิบัติ แบบจำลองอันดับจำกัดที่ใช้มากคือ

$$\begin{aligned}(1-\phi_1 B - \cdots - \phi_p B^p) Z_t &= a_t \\ \phi(B) Z_t &= a_t\end{aligned} \tag{2-77}$$

ตัวอย่างสำหรับแบบจำลอง lag-1 คือ

$$(1-\phi_1 B) Z_t = a_t \tag{2-78}$$

เป็นเพราะว่าโพลีโนเมียล $\phi(B)$ ได้จำกัด (Finite) สมการ (2-78) เป็นส่วน

กลับปราศจากเงื่อนไข (Unconditionally Invertible) คุณสมบัติได้ถือตามแบบจำลอง AR อันดับจำกัด p

ความสมดุลย์สามารถหาใหม่ได้โดยการจัดสูตรใหม่แบบจำลองในรูปแบบของสมการ (2-68)

$$Z_t = (1-\phi_1 B)^{-1} a_t = \sum_{j=0}^{\infty} \phi_1^j a_{t-j}$$
$$= \left(1 + \phi_1 B + \phi_1^2 B^2 + \cdots\right) a_t = \psi(B) a_t \qquad (2-71)$$

ซึ่งแสดงว่า แบบจำลอง AR จำกัด เท่ากับ ค่าเฉลี่ยเคลื่อนที่ไม่จำกัด มีการค้นพบว่า แบบจำลองค่าเฉลี่ยเคลื่อนที่ไม่จำกัดอยู่ในสภาวะความสมดุลย์ $\psi(B)$ จะต้องลู่เข้า สำหรับ $|B| \leq 1$ สำหรับแบบจำลอง AR(1) ที่กำหนดข้างบนจะได้ว่า $|\phi_1| < 1$ อยู่ในภาวะความสมดุลย์ความสมดุลย์ สังเกตว่ารากของ $1 - \phi_1 B = 0$ คือ $B = \phi_1^{-1}$ โดยที่ $|\phi_1| < 1$ ดังนั้นสภาวะความสมดุลย์คือรากของโพลีโนเมี่ยล B, $\phi(B)$ จะต้องอยู่ด้านนอกวงกลมรัศมีหนึ่งหน่วย การสรุปข้างบนใช้ได้สำหรับแบบจำลองการถดถอยในตัวเอง (Autoregressive) ของอันดับใด ๆ

ฟังก์ชั่นความแปรปรวนในตัวเอง (Autorcovariance Function) ของ AR(p) กำหนดให้โดยการคูณ $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi p Z_{t-p} + a_t$ โดย Z_{t-k} และถือว่าเป็นค่าที่ได้คาดหมาย

$$E[Z_{t-k} Z_t] = E\left[Z_{t-k}(\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t)\right]$$
$$= E\left[\phi_1 Z_{t-k} Z_{t-1} + \phi_2 Z_{t-k} Z_{t-2} + \cdots + \phi_p Z_{t-k} Z_{t-p} + Z_{t-k} a_t\right] \qquad (2-80)$$

การใช้ความล่าช้าของความสัมพันธ์ระหว่าง Z_{t-k} และ a_t และมีความสมดุลย์ความแปรปรวนในตัวเองที่ช่วงถัดไป k γ_k กลายเป็น

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} \qquad (2-81)$$

ที่ $r_{k-q} = E[Z_{t-k} Z_{t-q}]$

การหาร γ_0 ด้วยความแปรปรวนกระบวนการความสมดุลย์ (Stationary Process Variance) จะได้ว่า

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \qquad (2\text{-}82)$$

ดังนั้น ความสัมพันธ์ในตัวเองของ AR (p) ปฏิบัติตามสมการ $\phi(B)\rho_k = 0$

สมการ (2-82) เขียนในรูประบบสมการ Yule-Walker สำหรับ k = 1,2, ..., P ได้ว่า

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p
\end{aligned} \qquad (2\text{-}83)$$

สังเกตุว่าในสูตรระบบข้างบน ชี้ให้เห็นว่า $\rho_0 = 1$ และ $\rho_k = \rho_{-k}$ ด้วยสมการ Yule-Walker สามารถใช้ได้ในการประมาณพารามิเตอร์ของ AP(p) โดยการแทนค่าความสัมพันธ์ตัวอย่าง r_k ในสมการ (2-83) และแก้สมการหาค่า ϕs ต่าง ๆ กำหนดไปว่าความสัมพันธ์ตัวอย่างไม่คงที่อย่างมากในการประมาณสำหรับค่า lag สูงๆ จะต้องใช้กรรมวิธีการประมาณพารามิเตอร์ ในรูปแบบของเมตริก (Matrix Form)

$$\Phi = \Sigma^{-1}R, \qquad (2\text{-}84)$$

ที่

$$\Phi = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}, \quad R = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_p \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}.$$

ความแปรปรวนของกระบวนการ AR(p) กำหนดได้จากสมการ (2-80) สำหรับ k = 0 ตอนนี้ การคาดหมาย $E[Z_t a_t]$ จะได้ว่า σ_a^2 และความแปรปรวนกำหนดได้โดย

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_a^2. \qquad (2\text{-}85)$$

หารตลอดด้วย $\gamma_0 = \sigma_Z^2$ และแก้สมการหา σ_Z^2 ได้ผลดังนี้

$$\sigma_z^2 = \frac{\sigma_a^2}{1-\phi_1\rho_1-\phi_2\rho_2-\cdots-\phi_p\rho_p} \qquad (2\text{-}86)$$

ได้มีการศึกษารูปแบบของฟังค์ชั่นความสัมพันธ์ ρ_k ซึ่งได้พบรูปแบบที่เหมาะสม

$$\phi(B)\rho_k = 0 \qquad (2\text{-}87)$$

ตัวควบคุมโพลีโนเมี่ยล $\phi(B)$ สามารถแสดงได้ตาม

$$\phi(B) = \prod_{i=1}^{p}(1-G_iB) \qquad (2\text{-}88)$$

กำหนดให้สมการ (2-88) การแก้ปัญหาโดยทั่วไปของ ρ_k คือ

$$\rho_k = A_1G_1^k + A_2G_2^k + \cdots + A_pG_p^k \qquad (2\text{-}89)$$

ที่ $G_1^{-1} \ldots G_p^{-1}$ เป็นรากของโพลีโนเมียล $\phi(B)$ สำหรับความสมดุลย์ในขั้นแรกพบว่า $|G_i| < 1$ ถ้า G_i เป็นจริงและแน่ชัด ดังนั้นสมการ (2-89) โดยเชิงเรขาคณิตกำหนดให้เป็นศูนย์ตาม k ที่เพิ่มขึ้น ถ้าคู่ลำดับของราก G_i, G_j สลับซับซ้อนสามารถกระจายเป็นเทอมเทอมหนึ่ง $d^k \sin(2\pi fk + F)$ ถึง ρ_k ซึ่งเป็นคลื่น sine ที่ลดลงตาม k ที่เพิ่มขึ้น

แบบจำลอง AR(p) มี p+2 พารามิเตอร์ที่ต้องประมาณ ($\phi_1, \ldots, \phi_p$, μ และ σ_a^2) ดังนั้นเป็นการยากที่จะหาค่าพารามิเตอร์ต่าง ๆ ในเชิงสถิติได้ และในการศึกษาครั้งนี้ได้พิจารณาเฉพาะแบบจำลองที่มีอันดับต่ำ ๆ คืออันดับหนึ่งเท่านั้น

2.5.3 แบบจำลองการถดถอยในตัวเองอันดับหนึ่ง AR(1)

แบบจำลองถดถอยในตัวเองช่วงถัดไปเป็นหนึ่ง หรือ แบบจำลอง Markov คือ

$$Z_t = \phi_1 Z_{t-1} + a_t$$

หรือ

$$(1-B)\, Z_t = a_t$$

$$\phi(B) Z_t = a_t \qquad (2\text{-}90)$$

จากสมการ Yule - Walker

$$\begin{aligned} \rho_1 &= \phi_1 \rho_0 \\ &= \phi_1 \\ \rho_2 &= \phi_1 \rho_1 \\ &= \rho_1^2 \end{aligned} \qquad (2\text{-}91)$$

หรือโดยทั่วไป

$$\rho_k = \rho_1^k = \phi_1^k \qquad (2\text{-}92)$$

จากสมการ (2-86) ความแปรปรวน (Variance) คือ

$$\sigma_z^2 = \frac{\sigma_a^2}{1-\phi_1\rho_1} = \frac{\sigma_a^2}{1-\rho_1^2} \qquad (2\text{-}93)$$

สภาวะความสมดุลย์ต้องการ $|\phi_1| < 1$ ซึ่งเป็นกรณีโดยที่ $\phi_1 = \rho_1$ เสมอ สำหรับ ϕ_1 เป็นบวก แน่นอนที่ว่าความสัมพันธ์ในตัวเองของ AR (1) เป็นชนิดศักยภาพที่ลดลง (Decarying Exponential Type) สำหรับ ϕ_1 เป็นลบ ความสัมพันธ์ในตัวเองจะเป็นอย่างอื่นตามเครื่องหมายและลดลงตาม k ที่เพิ่มขึ้น ดังแสดงในรูป (2-14a) และ (2-14b)

แบบจำลอง AR(1) เป็นแบบจำลองที่นิยมของการจำลองสภาพอนุกรมเวลาและการคาดหมายในทางอุทกวิทยาและทางอื่น ๆ ในทางอุทกวิทยา โดยแบบธรรมดาเรียกว่า Thomas-Fiering Model โดยปกติได้แสดงในเทอมของโมเมนต์ของกระบวนการสุ่ม X_t

$$(X_i - \mu) = \rho_1 (X_{i-1} - \mu) + \sigma_x (1-\rho_1^2)^{1/2} W_i \qquad (2\text{-}94)$$

ที่ $X_t - \mu$ คือ Z_t ในย่อหน้าก่อนหน้านี้

ในการแสดงข้างบน μ คือ ค่าเฉลี่ยความสมดุลย์ของกระบวนการ X_t เทอมสุ่ม a_t แทนโดย $\sigma_x(1-\rho_1^2)^{1/2}\, w_t$ ที่ w_t คือค่าเฉลี่ยเป็นศูนย์ ความแปรปรวนเป็น 1 โดยปกติตัวแปรสุ่มได้กระจาย การแสดง $\sigma_x(1-\rho_1^2)^{1/2}$ เป็นไปตามสมการ (2-93) ความเบี่ยงเบนมาตรฐาน (Standard Deviation) ของ a_t

แบบจำลอง AR(1) โดยความถี่ได้ใช้กับการแก้ปัญหาของการคงรักษาโมเมนต์อันดับ 1 และ 2 ของอนุกรมเวลา บางครั้งเหมาะสมสำหรับการจำลองสภาพ และโดยทั่วไปสามารถจะทำได้สำหรับการคาดหมายเทอมสั้น ๆ การประมาณพารามิเตอร์ได้ใช้วิธีการของโมเมนต์อย่างง่ายที่ต้องแทนด้วยโมเมนต์ตัวอย่าง $\bar{X}$, r_1 และ s_x^2 สำหรับโมเมนต์ประชากร μ, ρ, และ σ^2_x ตามลำดับ ในสมการ (2-94) สมการธรรมดาสำหรับสถิติเหล่านี้คือ

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2-95}$$

$$\begin{aligned} S_x^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{X})^2 \\ &= \frac{1}{N-1}\sum_{i=1}^{N} x_i^2 - \frac{N}{N-1}\bar{X}^2 \end{aligned} \tag{2-96}$$

$$r_1 = \frac{\frac{1}{N}\sum_{i=1}^{N-1}(x_i-\bar{X})(x_{i+1}-\bar{X})}{S_x^2} \tag{2-97}$$

ที่ ตัวแปร x_i เล็ก แทนค่าตัวอย่างของตัวแปรสุ่ม X

2.5.4 แบบจำลองค่าเฉลี่ยเคลื่อนที่อันดับ q, MA(q)

ตามที่ในแบบจำลองการถดถอยในตัวเอง สูตรค่าเฉลี่ยเคลื่อนที่ของอันดับไม่จำกัดไม่ได้ใช้กันในทางปฏิบัติ แบบจำลองอันดับจำกัดเขียนแทนได้โดย

$$\begin{aligned} Z_t &= a_t - \theta_1 a_{t-1} - \ldots, - \theta_q a_{t-q} \\ &= (\,1 - \theta_1 B - \ldots, - \theta_q B^q\,)\, a_t \\ &= \theta_i(B)\, a_t \end{aligned} \tag{2-98}$$

โดยที่โพลีโนเมียล $\theta(B)$ มีจำกัด และลู่เข้า (Cnverges) เสมอ กล่าวได้ว่าไม่อยู่ในสภาวะสมดุลย์ (Unconditional Stationarity)

ส่วนกลับของสมการ (2-98) ต้องการให้ $\theta^{-1}(B)$ จะลู่เข้า (Converge) สำหรับ $|B| < 1$ [5] แสดงว่าเท่ากันกับ รากของสมการลักษณะเฉพาะ $\theta(B)$ จะต้องวางอยู่ภายนอกวงกลมที่มีรัศมีหนึ่งหน่วย (unit circle) ความสำคัญอยู่ที่แบบจำลองค่าเฉลี่ยเคลื่อนที่จำกัดเท่ากันกับแบบจำลองการถดถอยในตัวเองไม่จำกัด

ฟังค์ชั่นความแปรปรวนในตัวเอง (Autocovariance Function) ของกระบวนการ MA(q) กำหนดได้โดยการกระทำการคาดหมายดังต่อไปนี้

$$\gamma_k = E\left[(a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \cdots - \theta_q a_{t-k-q})\right]$$

สำหรับ k = 0 และการใช้ "Whiteness" คุณสมบัติของอนุกรม γ_k กระบวนการความแปรปรวน (Variance) กลายเป็น

$$\gamma_0 = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma_a^2 \qquad (2\text{-}99)$$

สำหรับที่ $k \neq 0$ จะได้ว่า ความแปรปรวนในตัวเอง (Autocovariance) คือ

$$\gamma_k = \begin{cases} \left(-\theta_k + \theta_1\theta_{k-1} + \cdots + \theta_{q-k}\theta_q\right)\sigma_a^2 & k = 1,2,\ldots q \\ 0 & k > q \end{cases}$$

โดยปกติขึ้นอยู่กับความแปรปรวนในตัวเอง (Autocovariance) หารด้วย γ_0 ฟังค์ชั่นความสัมพันธ์ในตัวเอง (Autocorrelation Function) กลายเป็น

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & k = 1,2,\ldots,q \\ 0 & k > q \end{cases} \qquad (2\text{-}100)$$

ดังนั้นความสัมพันธ์ในตัวเองของแบบจำลองค่าเฉลี่ยเคลื่อนที่เป็นศูนย์เกินพอที่เป็นลำดับของแบบจำลองที่ตรงกันข้ามกับจำนวนที่ไม่จำกัด (Infinite) จำนวนของ ρ_k ในแบบจำลอง AR(p)

โดยไม่เป็นระบบเชิงเส้นของการแสดงความสัมพันธ์ในตัวเองทำให้การประมาณพารามิเตอร์ สำหรับแบบจำลอง MA(q) ใช้วิธีการของไมเมนต์สูง ๆ นั้นที่ไม่มั่นคงโดยปกติต้องใช้กรรมวิธีกำลังสองน้อยที่สุดทำซ้ำๆกัน [5]

2.5.5 แบบจำลองค่าเฉลี่ยเคลื่อนที่อันดับหนึ่ง MA(1)

รูปแบบของ MA(1) คือ

$$\begin{aligned} Z_t &= a_t - \theta_1 a_{t-1} \\ &= (1-\theta_1 B)a_t \end{aligned} \qquad (2\text{-}101)$$

ซึ่งเป็นความสมดุลย์ (Stationary) สำหรับทุกค่าของ θ_1 แต่เป็นส่วนกลับเพียงอย่างเดียวสำหรับ $\left|B\right| > 1$ ซึ่งได้ว่า (โดยที่ $B = \theta^{-1}$) $\left|\theta_1\right| < 1$

ความแปรปรวนของกระบวนการคือ

$$\gamma_0 = (1+\theta_1^2)\sigma_a^2 \qquad (2\text{-}102)$$

และฟังก์ชันความสัมพันธ์ในตัวเอง

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 \end{cases} \qquad (2\text{-}103)$$

ซึ่ง (สำหรับ k = 1) นำไปสู่ความสัมพันธ์ดังต่อไปนี้ระหว่าง θ_1 และ ρ_1

$$\theta_1^2 + \frac{\theta_1}{\rho_1} + 1 = 0 \qquad (2\text{-}104)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สังเกตว่าความสัมพันธ์ในตัวเองไม่ปรากฏหลังจากช่วงเวลาถัดไปหนึ่งหน่วย (lag one) รากที่ 2 ของสมการ (2-104) จะต้องเป็น และ ถ้า $\left|\theta_1\right| < 1$ ดังนั้น $\left|\theta_1^{-1}\right| > 1$ และสภาวะส่วนกลับจะเป็นตัวทำลายได้ ดังแสดงในตารางที่ 2-4 แสดงการหาส่วนกลับของสมการ (2-104) สำหรับความสัมพันธ์ระหว่าง -0.5 และ 0.5 [5]

2.5.6 แบบจำลองค่าเฉลี่ยการถดถอยในตัวเอง ARMA(p,q)

แบบจำลองการถดถอยในตัวเอง (Autoregressive) และแบบจำลองค่าเฉลี่ยเคลื่อนที่สามารถรวมกันได้ในกระบวนการแบบจำลองนั้น อีกนัยหนึ่งควรจะเป็นการควบคุมที่เป็นไปไม่ได้ที่แทนด้วยแบบจำลอง AR หรือ MA จำกัดอย่างเดียว แบบจำลอง ARMA(p,q) แสดงได้ตามรูปแบบ

$$\begin{aligned}\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)Z_t &= \left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)a_t \\ \phi(B)Z_t &= \theta(B)a_t\end{aligned} \qquad (2\text{-}105)$$

สภาวะความสมดุลย์และส่วนกลับของแบบจำลอง ARMA(p,q) ตรงกันกับองค์ประกอบของแบบจำลอง MA และ AR สำหรับความสมดุลย์และส่วนกลับรากของ $\phi(B)$ และของ $\theta(B)$ จะต้องอยู่ภายนอกวงกลมที่มีรัศมีหนึ่งหน่วย(unit circle)

ฟังค์ชั่นความแปรปรวนในตัวเองพบได้โดยการคูณ $Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$ โดย Z_{t-k} และหาค่าที่ได้คาดหมาย

$$\begin{aligned}\gamma_k = E[Z_t Z_{t-k}] &= \phi_1\gamma_{k-1}+\cdots+\phi_p\gamma_{k-p}+\gamma_{za}(k) \\ &\quad -\theta_1\gamma_{za}(k-1)-\cdots-\theta_q\gamma_{za}(k-q)\end{aligned} \qquad (2\text{-}106)$$

ที่

$$\begin{aligned}\gamma_{za}(k) &= E[Z_{t-k}a_t] \\ \gamma_{za}(k-1) &= E[Z_{t-k}a_{t-1}]\end{aligned} \qquad (2\text{-}107)$$

ค่าสำหรับ $\gamma_{za}(k)$ จะเป็นศูนย์เท่าที่ $k > 0$ โดยที่ไม่สัมพันธ์กันสำหรับค่าของ Z ก่อน t ค่าสำหรับ $\gamma_{za}(k)$ จะไม่เป็นศูนย์สำหรับ $k <= 0$

ค่าที่กล่าวข้างบนนั้นควรจะชัดเจนว่าสำหรับ $k > q$ ความแปรปรวนในตัวเอง (และความสัมพันธ์ในตัวเอง) ในสมการ (2-113) ลดลงเป็นแบบจำลอง AR(p) นั้น

$$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} \qquad \text{สำหรับ } k > q$$
$$\rho_k = \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p} \qquad \text{สำหรับ } k > q$$

สำหรับค่าของ k น้อยกว่า หรือเท่ากับ q ความแปรปรวนในตัวเองจะเป็นฟังค์ชั่นของเทอมค่าเฉลี่ยเคลื่อนที่ และจะขึ้นอยู่กับสัมประสิทธิ์ทั้งหมด คือ $\phi_1,\ldots,\phi_p$, $\theta_1,\ldots,\theta_q$ และความแปรปรวน σ_a^2 แบบจำลอง ARMA(p,q) นั้นมีคุณสมบัติที่สะดวกกว่า ความสัมพันธ์ในตัวเอง q ครั้งแรกขึ้นอยู่กับเทอมค่าเฉลี่ยเคลื่อนที่เหมือนกันเทอมการถดถอยในตัวเอง หลังจากช่วงถัดไป q พฤติกรรมการถดถอยในตัวเองถือได้จากค่าความสัมพันธ์ครั้งสุดท้าย

ความแปรวนของกระบวนการกำหนดได้โดยสมการ (2-106) สำหรับ k = 0 การประมาณค่าของความแปรปรวนี้ต้องแก้สมการของค่า $\gamma_1,\ldots,\gamma_p$

2.5.7 แบบจำลอง ARMA(1,1)

แบบจำลองที่นิยมกันและใช้ในทางอุทกวิทยาคือ

$$Z_t - \phi_1 Z_{t-1} = a_t - \theta_1 a_{t-1}$$
$$(1-\phi_1 B)Z_t = (1-\theta_1 B)a_t \qquad (2\text{-}108)$$

สภาวะความสมดุลย์ และส่วนกลับตรงกันกับแบบจำลองเฉพาะ AR(1) และ MA(1) และดังนั้นได้ว่าพิสัยของพารามิเตอร์คือ

$$-1 < \phi_1 < 1$$
$$-1 < \theta_1 < 1$$

รูปที่ 2-15 แสดงพื้นที่พารามิเตอร์ที่แรเงา

การใช้สมการ (2-106) ได้ฟังค์ชั่นความแปรปรวนในตัวเองคือ

$$\gamma_0 = \phi_1\gamma_1 + \sigma_a^2 - \theta_1\gamma_{z_a}(-1)$$
$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2$$
$$\gamma_k = \phi_1\gamma_{k-1} \qquad k \geq 2 \qquad (2\text{-}109)$$

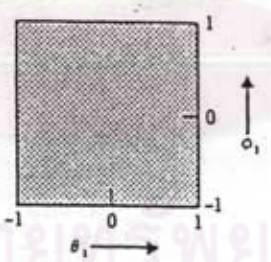
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

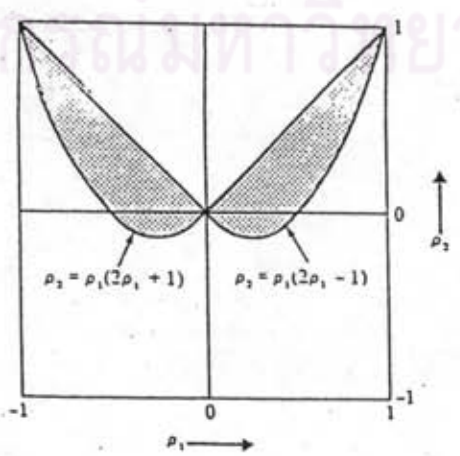
Table relating ρ_1 to θ for a first-order moving-average process

θ	ρ_1	θ	ρ_1
0.00	0.000	0.00	0.000
0.05	−0.050	−0.05	0.050
0.10	−0.099	−0.10	0.099
0.15	−0.147	−0.15	0.147
0.20	−0.192	−0.20	0.192
0.25	−0.235	−0.25	0.235
0.30	−0.275	−0.30	0.275
0.35	−0.315	−0.35	0.315
0.40	−0.349	−0.40	0.349
0.45	−0.374	−0.45	0.374
0.50	−0.400	−0.50	0.400
0.55	−0.422	−0.55	0.422
0.60	−0.441	−0.60	0.441
0.65	−0.457	−0.65	0.457
0.70	−0.468	−0.70	0.468
0.75	−0.480	−0.75	0.480
0.80	−0.488	−0.80	0.488
0.85	−0.493	−0.85	0.493
0.90	−0.497	−0.90	0.497
0.95	−0.499	−0.95	0.499
1.00	−0.500	−1.00	0.500

ตารางที่ 2-4 แสดงความสัมพันธ์ ρ_1 กับ θ

สำหรับกระบวนการค่าเฉลี่ยเคลื่อนที่อันดับหนึ่ง [12]





Valid regions for the parameters and correlations of a stationary and invertible ARMA(1,1) process

รูปที่ 2-15 แสดงพื้นที่พารามิเตอร์แรเงา [12]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กำหนดให้ $\gamma_{Za}(-1)$ สมการ (2-115) คูณโดย a_{t-1} และการคาดหมายเขียนได้คือ

$$\gamma_{za}(-1) = E[Z_t a_{t-1}] = (\phi_1 - \theta_1)\sigma_a^2 \qquad (2\text{-}110)$$

การใช้สมการ (2-110) ในสมการ (2-109) ฟังค์ชั่นความแปรปรวนในตัวเองของกระบวนการกำหนดได้ตาม

$$\gamma_0 = \frac{1 - \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_a^2$$

$$\gamma_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma_a^2$$

$$\gamma_k = \phi_1\gamma_{k-1} \qquad k \geq 2 \qquad (2\text{-}111)$$

สังเกตุว่าความแปรปรวนในตัวเองจะลดลงโดย exponential จากค่าเริ่มต้น γ_1 ซึ่งขึ้นอยู่กับ θ_1 เครื่องหมาย γ_1 (และ ρ_1) กำหนดได้โดย $\phi_1 - \theta_1$ เครื่งหมายของ ϕ_1 หาได้ ถ้าความสัมพันธ์ลดลงเป็นแบบราบเรียบ (Smooth) หรือเป็นอย่างอื่นตามเครื่องหมาย

ฟังค์ชั่นความสัมพันธ์กำหนดได้โดย

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$$

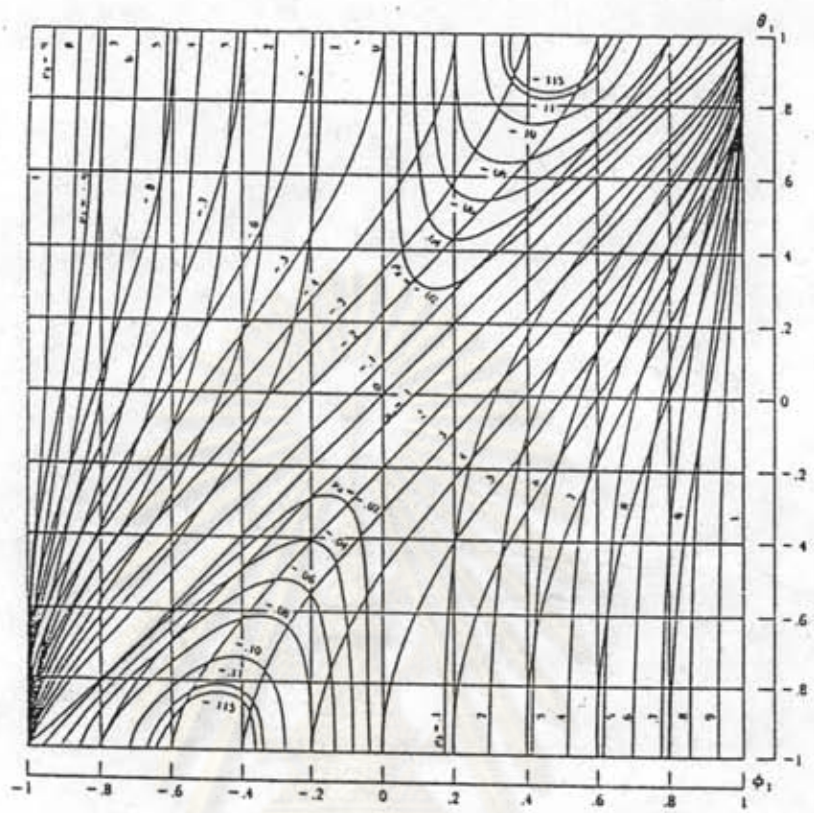
$$\rho_k = \phi_1\rho_{k-1} \qquad k \geq 2 \qquad (2\text{-}112)$$

ความสัมพันธ์ได้แสดงในสมการ (2-112) และพื้นที่พารามิเตอร์ ส่วนกลับ-ความสมดุลย์กำหนดได้เป็นบริเวณแรเงา สำหรับความสัมพันธ์ 2 ค่าในครั้งแรก

$$|\rho_2| < |\rho_1|$$
$$\rho_2 > \rho_1(2\rho_1 + 1) \qquad \rho_1 < 0$$
$$\rho_2 > \rho_1(2\rho_1 - 1) \qquad \rho_1 > 0$$

(2-113)

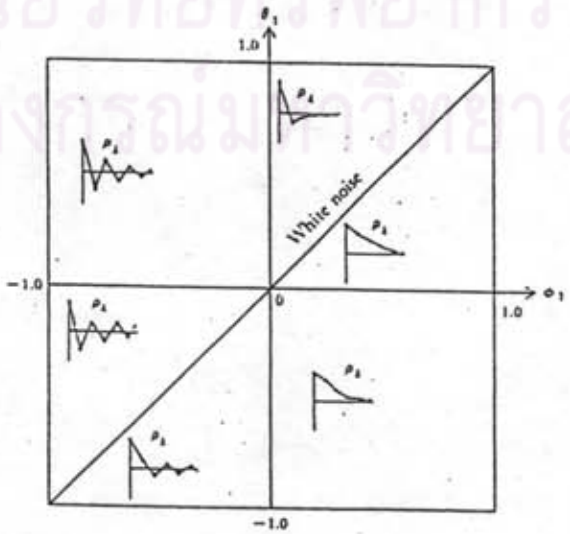
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



Relation between correlations and parameters for a stationary and invertible ARMA(1,1). Diagram may be used for parameter estimation using the method of moments.

รูปที่ 2-16 การสัมพันธ์ระหว่างความสัมพันธ์ และพารามิเตอร์สำหรับความสมดุลย์และส่วนกลับ ARMA(1,1) แผนผังอาจจะใช้ได้สำหรับการประมาณพารามิเตอร์ใช้วิธีของโมเมนต์ [12]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



Autocorrelation functions for various ARMA(1,1) models

รูปที่ 2-17 ฟังก์ชั่นความสัมพันธ์ในตัวเองสำหรับแบบจำลอง ARMA(1,1) เปลี่ยนแปลง [12]

รูปที่ 2-15 ตัวอย่างบริเวณข้างบน ความสัมพันธ์ภายนอกที่ชี้ให้เห็นว่า ARMA(1,1) ไม่เป็นแบบจำลองที่ดี รูปที่ 2-16 แผนผังการแก้สมการของพารามิเตอร์ ϕ_1 และ θ_1 ในเทอมของ ρ_1 และ ρ_2 ตามที่กำหนดในสมการ (2-112) รูปที่ 2-7 ให้รูปแบบที่เหมือนกันของความสัมพันธ์ในตัวเองที่ได้คาดหมายสำหรับบริเวณที่เปลี่ยนแปลงของพื้นที่พารามิเตอร์

2.5.8 แบบจำลองความไม่สมดุลย์ : แบบจำลองการถดถอยในตัวเองที่ได้รวม แบบจำลองค่าเฉลี่ยเคลื่อนที่ (ARIMA)

2.5.8.1 ความไม่สมดุลย์เฉพาะแห่ง

การอธิบายอย่างง่ายโดยความสัมพันธ์ของทฤษฎีสำหรับกระบวนการ ARMA ยอมให้ทำได้กับชนิดที่ได้จำกัดของแบบจำลองความไม่สมดุลย์ โดยส่วนสำคัญของความไม่สมดุลย์ในค่าเฉลี่ยของกระบวนการ

ในหัวข้อก่อนนี้ พบว่า แบบจำลอง

$$\psi(B)Z_t = \theta(B)a_t \qquad (2\text{-}114)$$

มีความสมดุลย์ถ้ารากของ $\psi(B)$ อยู่ภายนอกวงกลมที่มีรัศมีหนึ่งหน่วย สมมติว่าตอนนี้ค่าบางค่าของรากของ $\psi(B)$ ได้อยู่บนวงกลมที่มีรัศมีหนึ่งหน่วยแบบจำลองสามารถแสดงได้ตาม

$$\psi(B)Z_t = \phi(B)(1-B)^d Z_t = \theta(B)a_t \qquad (2\text{-}115)$$

หรือ

$$\phi(B)\nabla^d Z_t = \theta(B)a_t$$

ที่รากของ $\phi(B)$ อยู่ภายนอกวงกลมที่มีรัศมีหนึ่งหน่วย เพราะว่าการควบคุมความแตกต่าง $\nabla^d (d>=1)$ แบบจำลองสามารถจัดรูปแบบใหม่ได้ในเทอมของกระบวนการนี้ค่าเฉลี่ยไม่ใช่ศูนย์ X_t

โดยที่ $\nabla^d Z_t = \nabla^d X_t$ ดังนั้น ข้างบนเท่ากับ $\phi(B)\nabla^d X_t = \theta(B)a_t$ ที่รากของ $\phi(B)$ อยู่ภายนอกวงกลมที่มีรัศมีหนึ่งหน่วยกำหนดได้ว่า

$$Y_t = \nabla^d X_t \qquad (2\text{-}116)$$

สมการ (2-115) แทนกระบวนการ ARMA ความสมดุลย์-ส่วนกลับด้วยตัวแปร Y_t กำหนดให้ Y_t, X_t สามารถกำหนดได้โดยการกระทำการควบคุมรวม

$$X_t = S^d Y_t \qquad (2\text{-}117)$$

ที่

$$SY_t = \sum_{h=-\infty}^{t} Y_h \quad .$$

$$S^2 Y_t = \sum_{i=-\infty}^{t} \sum_{h=-\infty}^{i} Y_h$$

และอื่น ๆ

โดยที่ X_t เป็นตัวสำคัญในการรวมของแบบจำลอง ARMA การทำสูตรโดยเฉพาะนี้เรียกว่า ค่าเฉลี่ยเคลื่อนที่ได้รวมการถดถอยในตัวเองด้วยอันดับ p, d และ q [ARIMA(p,d,q) ที่ d คืออันดับของความแตกต่างของข้อมูลเดิมที่จำเป็นในกระบวนการความสมดุลย์ สังเกตว่า

ARMA (p,q) = ARIMA (p,0,q)

AR (p) = ARIMA (p,0,0)

MA (q) = ARIMA (0,0,q)

เป็นที่แน่ชัดว่าแบบจำลอง ARIMA ถือว่าเป็นชนิดโดยเฉพาะของความไม่สมดุลย์โดยถือว่าเป็นความแตกต่างอันดับหนึ่ง เป็นไปได้ที่กำการตัวไม่รู้ค่าคาดหมายที่มีแนวโน้มในข้อมูล ข้อมูลที่ได้แตกต่างควรจะเป็นความสมดุลย์กับพฤติกรรมต่อเนื่องอีกนัยหนึ่งถ้ากระบวนการได้แสดงการเปลี่ยนแปลงสุ่มกับความลาดเอียง และระดับ แต่เป็นเนื้อเดียวกันอีกนัยหนึ่ง ดังนั้นการแตก

ต่างครั้งที่สองจะได้กระบวนการที่มีความสมดุลย์

แนวโน้มเชิงเส้นที่หาได้แน่นอน หรือแนวโน้มไลรไนเมียลอันดับสูง ๆ สามารถรวมกันได้โดยสถานะแบบจำลอง ARIMA ตาม

$$\phi(B)\nabla^{d}X_t = \theta_0 + \theta(B)a_t \qquad (2\text{-}118)$$

แบบจำลอง ARIMA สามารถแสดงได้ในรูปแบบแตกต่าง 3 รูปแบบ อันแรกเป็นสมการความแตกต่างโดยการขยายไพลีไนเมียล

$$\psi(B) = \phi(B)(1-B)^{d} = 1 - \psi_1 B - \psi_2 B^2 - \cdots - \psi_{p-d}B^{p-d}$$

ดังนั้น

$$X_t = \psi_1 X_{t-1} + \cdots + \psi_{p-d}X_{t-p-d} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} + a_t \qquad (2\text{-}119)$$

สำหรับตัวอย่าง ARIMA(1,1,1)

$$(1-\phi B)(1-B)X_t = (1-\theta B)a_t$$

เท่ากันกับ

$$\{1-(1+\phi)B+\phi B^2\}X_t = (1-\theta B)a_t$$

แบบจำลองสามารถแสดงได้ตามฟังค์ชั่นของ past random shochs a_t

$$X_t = \Omega(B)a_t \qquad (2\text{-}120)$$

การควบคุมด้วย $\psi(B)$ ทั้งสองข้างของสมการ (2-120)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\psi(B)X_t = \psi(B)\Omega(B)a_t = \theta(B)a_t \qquad (2-121)$$

ซึ่งชี้ให้เห็นว่า

$$\psi(B)\Omega(B) = \theta(B) \qquad (2-122)$$

น้ำหนัก Ω_j กำหนดให้โดยสัมประสิทธิ์ของ B ในโพลินเมียล 2 ตัว ของสมการ (2-122)

$$(1-\psi_1 B-\cdots-\psi_{p+d}B^{p-d})(1+\Omega_1 B+\Omega_2 B^2+\cdots) = (1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q) \qquad (2-123)$$

สำหรับตัวอย่างที่ได้พบ ARIMA(1,1,1)

$$\psi(B) = 1-(1+\phi)B+\phi B^2$$
$$\theta(B) = 1-\theta B$$

การแทนค่าในสมการ (2-122) และสัมประสิทธิ์สมการของ B ได้ว่า

$$\Omega_0 = A_0 + A_1 = 1$$
$$\Omega_1 = A_0 + A_1\phi$$
$$\Omega_2 = A_0 + A_1\phi^2$$
$$\vdots$$
$$\Omega_j = A_0 + A_1\phi^j$$

ที่

$$A_0 = \frac{1-\theta}{1-\phi}$$
$$A_1 = \frac{\theta-\phi}{1-\phi}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ดังนั้นแบบจำลอง ARIMA(1,1,1,) เท่ากับ

$$X_t = \sum_{j=0}^{\infty} (A_0 + A_1\phi^j)a_{t-j} \qquad (2\text{-}124)$$

ในขั้นสุดท้ายแบบจำลอง ARIMA สามารถแสดงในเทอมของ X_3 ก่อนหน้านี้ และ current shook a_t เริ่มต้นด้วยแบบจำลองโดยทั่วไป

$$\psi(B)X_t = \theta(B)a_t \qquad (2\text{-}125)$$

เป้าหมายกำหนดให้ว่า

$$\pi(B)X_t = a_t \qquad (2\text{-}126)$$

ที่

$$\pi(B) = (1 - \sum_{j=1}^{\infty} \pi_j B^j)$$

แทนค่าสมการ(2-125) ในสมการ (2-126) ได้ว่า

$$\psi(B)X_t = \theta(B)\pi(B)X_t \qquad (2\text{-}127)$$

หรือ

$$(1 - \psi_1 B - \cdots - \psi_{p+d} B^{p-d}) = (1 - \theta_1 B - \cdots - \theta_q B^q)(1 - \pi_1 B - \pi_2 B^2 - \cdots) \qquad (2\text{-}128)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สัมประสิทธิ์ ψ_j กำหนดได้โดยสัมประสิทธิ์สมการของ B^j เป็นการง่ายในการประสานว่า สัมประสิทธิ์ ψ_j จะต้องเพิ่มขึ้นเป็น 1 สำหรับ d >= 1

การใช้ ARIMA (1,1,1) อีกครั้งหนึ่งตามตัวอย่างสมการ (2-128) ได้เป็นรูปแบบ

$$\begin{aligned}\pi(B) &= \psi(B)\theta^{-1}(B) \\ &= \left[1-(1+\phi)B+\phi B^2\right](1+\theta B+\theta^2B^2+\cdots)\end{aligned}$$

ผลลัพธ์เป็น

$$\begin{aligned}\pi_1 &= \phi+(1-\theta), \\ \pi_2 &= (\theta-\phi)(1-\theta), \\ &\vdots \qquad \vdots \\ \pi_j &= (\theta-\phi)(1-\theta)\theta^{j-2} \qquad j\geq 3\end{aligned} \tag{2-129}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย