บทที่ ๒

ทฤษฎีที่ใช้ในการวิจัย

ในการวิจัยนี้ได้นำทฤษฎีทางสถิติต่าง ๆ มาใช้ดังต่อไปนี้

๒.๑ ลูกโซ่มาคอฟ (Markov Chain)

กระบวนการลูกโซ่มาคอฟ เป็นกระบวนการที่การเกิดของเหตุการณ์ในอนาคต พิจารณาจากเหตุการณ์ในสภาวะปัจจุบันเท่านั้น ลูกโซ่มาคอฟที่ง่ายที่สุดได้แก่ First order Markov chain ซึ่งมีคุณสมบัติดังนี้ คือ

๑. สภาวะ (state) และเวลาที่ทำการวัดเป็นจำนวนนับได้ (finite)

๒. ความน่าจะเป็นของการเกิดแต่ละสภาวะขึ้นอยู่กับสภาวะสุดท้ายที่เกิดก่อนหน้านั้น

๓. ความน่าจะเป็นในข้อ ๒ มีค่าคงที่

ในการทดสอบข้อมูลฝนรายวันว่า มีคุณสมบัติดังกล่าวหรือไม่ สิ่งแรกที่จะต้องหาก็คือ ความน่าจะเป็นของการเกิดฝน ซึ่งค่าเหล่านี้สามารถประมาณได้โดยวิธีที่เรียกว่า ความน่าจะเป็นสูงสุด (Maximum likelihood)

๒.๑.๑ Maximum likelihood estimates

จากการตั้งสมมติฐานว่า ความน่าจะเป็นของการเกิดฝนขึ้นอยู่กับลักษณะอากาศของวันที่อยู่ก่อนหน้านั้นเพียง ๑ วัน ดังนั้น ความน่าจะเป็นของการเกิดฝนจึงอยู่ในรูปของความน่าจะเป็นอย่างมีเงื่อนไข (Conditional probabilities) กล่าวคือ

ถ้าให้ i, j = สภาวะต่าง ๆ ที่ประกอบด้วยการเกิดฝน และการไม่เกิดฝน ซึ่งในที่นี้จะแทนด้วย ๑ และ ๒ ตามลำดับ

และ t = เวลาที่เก็บรวบรวมข้อมูล (Times of observation)

= 0, 1, 2, ..., 30 (หรือ 31)

จะได้ว่า P_{ij} = ความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝน

= ความน่าจะเป็นของการเกิดสภาวะ j เมื่อเวลา t

โดยที่ได้เกิดสภาวะ i แล้วเมื่อเวลา $t-1$

และ $P_{ij} \geqslant 0 \;,\; \sum_{j=1}^{2} P_{ij} = 1$

จากการประมาณความน่าจะเป็นอย่างมีเงื่อนไข P_{ij} โดยวิธี ความน่าจะเป็นสูงสุด จะได้ว่า

$$\hat{P}_{ij} = n_{ij} \Big/ \sum_{j=1}^{2} n_{ij}$$

โดยที่ $n_{ij} = \sum_{t=1}^{T} n_{ij}(t)$

$n_{ij}(t)$ = จำนวนวันที่อยู่ในสภาวะ i และ j เมื่อเวลา $t-1$ และ t ตามลำดับ

และ $\sum_{j=1}^{2} n_{ij}$ = จำนวนวันที่อยู่ในสภาวะ i

$$= \sum_{t=1}^{T} n_i(t-1)$$

๒.๑.๒ <u>โมเมนต์ (moments) ของ $n_{ij}(t) - n_i(t-1)\, P_{ij}$</u>

การหาโมเมนต์ต่าง ๆ ของ $n_{ij}(t) - n_i(t-1)\, P_{ij}$ นี้ จุดประสงค์เพื่อที่จะนำค่าเหล่านี้ไปประกอบการพิจารณาการแจกแจงของตัวประมาณค่า (estimates) $\sqrt{n}\,(\hat{P}_{ij} - P_{ij})$

ให้ $n_k(0)$ = จำนวนวันที่อยู่ในสภาวะ k, k = 1, 2 เมื่อเวลา 0 (initial state)

n = จำนวนปีทั้งหมด

จะได้ว่า $\frac{n_k(0)}{n} \xrightarrow[n \to \infty]{} \eta_k$

โดยที่ $\eta_k > 0$

และ $\sum \eta_k = 1$

เนื่องจาก $n_{ij}(t)$ จะมีการแจกแจงแบบปกติ เมื่อขนาดตัวอย่างเพิ่มขึ้น

และ $n_{ij}(t) = \sum_{k=1}^{2} n_{k;\, ij}(t)$

โดย $n_{k;ij}(t)$ = จำนวนวันที่อยู่ในสภาวะ k, i และ j เมื่อเวลา 0, t-1 และ t ตามลำดับ

ให้ $P_{ij}^{(t)}$ = ความน่าจะเป็นของการเกิดสภาวะ j ที่เวลา t โดยที่สภาวะที่เวลา 0 คือ i

ฉะนั้น $P\left[n_{k;ij}(t)\right] = P_{ki}^{(t-1)} P_{ij}$

และ โมเมนต์ต่าง ๆ ของ $n_{k;ij}(t)$ คือ

$$E\left[n_{k;ij}(t)\right] = n_k(0)\, P_{ki}^{(t-1)} P_{ij}$$

$$\mathrm{Var}\left[n_{k;ij}(t)\right] = n_k(0) P_{ki}^{(t-1)} P_{ij} \left[1-P_{ki}^{(t-1)} P_{ij}\right]$$

และ $$\mathrm{Cov}\left[n_{k;ij}(t),\, n_{k;gh}(t)\right] = -n_k(0) P_{ki}^{(t-1)} P_{ij} P_{kg}^{(t-1)} P_{gh}$$

$(i,j) \neq (g,h)$

ถ้ากำหนด $n_{k;i}(t-1)$ โดยที่

$$n_{k;i}(t-1) = \sum_{j=1}^{2} n_{k;ij}(t)$$

จะได้ว่า การแจกแจงอย่างมีเงื่อนไข (Conditional distribution) ของ $n_{k;ij}(t)$ เมื่อกำหนด $n_{k;i}(t-1)$ เป็นแบบ multinomial และมีความน่าจะเป็น P_{ij} โดยที่

$$E\left[n_{k;ij}(t) \mid n_{k;i}(t-1)\right] = n_{k;i}(t-1)\, P_{ij}$$

นั่นคือ โมเมนต์ต่าง ๆ ของ $n_{k;ij}(t) - n_{k;i}(t-1)\, P_{ij}$ หาได้ดังนี้

$$E\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right] = EE\left\{\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right] \Big/ n_{k;i}(t-1)\right\}$$

$$= 0$$

$$\mathrm{Var}\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right] = E\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right]^2$$

$$= EE\left\{\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right]^2 \Big/ n_{k;i}(t-1)\right\}$$

$$= E\left[n_{k;i}(t-1)P_{ij}(1-P_{ij})\right]$$

$$= n_k(0)P_{ki}^{[t-1]}\, P_{ij}(1-P_{ij})$$

$$\mathrm{Cov}\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij},\ n_{k;ih}(t) - n_{k;i}(t-1)P_{ih}\right]$$

$$= E\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right]\left[n_{k;ih}(t) - n_{k;i}(t-1)P_{ih}\right]$$

$$= EE\left\{\left[n_{k;ij}(t) - n_{k;i}(t-1)P_{ij}\right]\left[n_{k;ih}(t) - n_{k;i}(t-1)P_{ih}\right] \Big/ n_{k;i}(t-1)\right\}$$

$$= E\left[-n_{k;i}(t-1)P_{ij}P_{ih}\right]$$

$$= -n_k(0)P_{ki}^{\{t-1\}} P_{ij}P_{ih} \qquad j \neq h$$

$$\mathrm{Cov}\left[n_{k;ij}(t)-n_{k;i}(t-1)P_{ij} \ , \ n_{k;gh}(t)-n_{k;g}(t-1)P_{gh}\right]$$

$$= E\left[n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}\right]\left[n_{k;gh}(t)-n_{k;g}(t-1)P_{gh}\right]$$

$$= EE\left\{\left[n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}\right]\left[n_{k;gh}(t)-n_{k;g}(t-1)P_{gh}\right] \Big/ n_{k;i}(t-1), n_{k;g}(t-1)\right\}$$

$$= 0 \qquad i \neq g$$

$$\mathrm{Cov}\left[n_{k;ij}(t)-n_{k;i}(t-1)P_{ij} \ , \ n_{k;gh}(t+r)-n_{k;g}(t+r-1)P_{gh}\right]$$

$$= E\left[n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}\right]\left[n_{k;gh}(t+r)-n_{k;g}(t+r-1)P_{gh}\right]$$

$$= EE\left\{\left[n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}\right]\left[n_{k;gh}(t+r)-n_{k;g}(t+r-1)P_{gh}\right] \Big/ n_{k;g}(t+r-1), n_{k;i}(t-1), n_{k;ij}(t)\right\}$$

$$= 0 \qquad r > 0$$

จากการหาโมเมนต์ต่าง ๆ ของตัวแปร $n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}$ สามารถสรุปได้ว่า

$n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}$ เป็น multinomial variable ซึ่ง mean เป็น 0 ความน่าจะเป็น P_{ij} และขนาดตัวอย่างเท่ากับ $n_k(0)P_{ki}^{[t-1]}$ และ $n_{k;ij}(t)-n_{k;i}(t-1)P_{ij}$ และ $n_{k;gh}(s)-n_{k;g}(s-1)P_{gh}$ ต่างก็เป็นอิสระต่อกัน (independent) ถ้า $t \neq s$ หรือ $i \neq g$ เนื่องจาก $n_k(0)$ คงที่ และ $n_{k;ij}(t)$, $n_{l;gh}(t)$ ต่างก็เป็นอิสระต่อกัน (เมื่อ $k \neq l$) จะได้ว่า

$$E\left[n_{ij}(t)-n_i(t-1)P_{ij}\right] = 0$$

$$\mathrm{Var}\left[n_{ij}(t)-n_i(t-1)P_{ij}\right] = \sum_{k=1}^{2} n_k(0)P_{ki}^{[t-1]}P_{ij}(1-P_{ij})$$

$$\mathrm{Cov}\left[n_{ij}(t)-n_i(t-1)P_{ij}\ ,\ n_{ih}(t)-n_i(t-1)P_{ih}\right] = -\sum_{k=1}^{2} n_k(0)P_{ki}^{[t-1]}P_{ij}P_{ih} \qquad j \neq h$$

$$\mathrm{Cov}\left[n_{ij}(t)-n_i(t-1)P_{ij}\ ,\ n_{gh}(s)-n_g(s-1)P_{gh}\right] = 0 \qquad t \neq s \text{ หรือ } i \neq g$$

๒.๑.๓ การแจกแจงแบบแอสซิมโทดของตัวประมาณค่า (Asymtotic distribution of the estimates)

การหาการแจกแจงของตัวประมาณค่า $\sqrt{n}\ (\hat{P}_{ij} - P_{ij})$ พิจารณาจาก

$$\sqrt{n}\ (\hat{P}_{ij} - P_{ij}) = \sqrt{n}\left[\frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} n_i(t-1)} - P_{ij}\right]$$

$$= \sqrt{n}\left[\frac{\sum_{t=1}^{T}\left[n_{ij}(t)-n_i(t-1)P_{ij}\right]}{\sum_{t=1}^{T} n_i(t-1)}\right]$$

$$= \frac{\sum_{t=1}^{T}\left[n_{ij}(t)-n_i(t-1)P_{ij}\right]\Big/ n^{\frac{1}{2}}}{\sum_{k=1}^{2}\sum_{t=1}^{T} \eta_k P_{ki}^{[t-1]}} \qquad (3.1)$$

$\sqrt{n}\,(\hat{P}_{ij} - P_{ij})$ จะมี limiting distribution เดียวกันกับ (3.1) ดังนั้น จาก ๒.๑.๒ จะได้ว่า

$\sqrt{n}\,(\hat{P}_{ij} - P_{ij})$ มี limiting normal distribution ที่มี

$$\begin{aligned} \text{mean} &= 0 \\ \text{variance} &= P_{ij}(1-P_{ij})/\,\emptyset_i \\ \text{covariance} &= -\delta_{ig}\, P_{ij}\, P_{gh}\,/\,\emptyset_i \end{aligned}$$

โดยที่ $$\delta_{ig} = \begin{cases} 1 & i = g \\ 0 & i \neq g \end{cases}$$

และ $$\emptyset_i = \sum_{k=1}^{2} \sum_{t=1}^{T} \eta_k\, P_{ki}^{[t-1]}$$

หรือ $(n\emptyset_i)^{\frac{1}{2}}\,(\hat{P}_{ij} - P_{ij})$ มี limiting normal distribution ที่มี

$$\begin{aligned} \text{mean} &= 0 \\ \text{variance} &= P_{ij}(1 - P_{ij}) \\ \text{covariance} &= -\delta_{ig}\, P_{ij}\, P_{gh} \end{aligned}$$

หรือกล่าวอีกนัยหนึ่งว่า

ในแต่ละค่าของ i ตัวแปร $(n\emptyset_i)^{\frac{1}{2}}\,(\hat{P}_{ij} - P_{ij})$ จะมี limit distribution เดียวกันกับตัวประมาณค่า ของ multinomial probabilities P_{ij} โดยที่ขนาดของตัวอย่างเท่ากับ $n\emptyset_i$

เมื่อ $$\begin{aligned} n\emptyset_i &= \text{จำนวนวันทั้งหมดที่อยู่ในสภาวะที่ } i \\ &= \sum_t n_i(t-1) \end{aligned}$$

และ $\left[\sum_t n_i(t-1)\right]^{\frac{1}{2}}\,(\hat{P}_{ij} - P_{ij})$ จะเป็นอิสระต่อกันในแต่ละค่าของ i เมื่อขนาดของตัวอย่างเพิ่มขึ้น

๒.๒ <u>การแจกแจงแบบจีโอเมตริก</u> (Geometric distribution)

ในการวิจัยทางสถิติ บางครั้งเราไม่ทราบว่าข้อมูลที่เก็บรวบรวมได้นั้น มีการแจกแจงในรูปใด และเมื่อไม่ทราบรูปของการแจกแจงก็อาจเป็นเหตุให้การทดสอบต่อไปเกิดการผิดพลาดได้ง่าย จึงเป็นความจำเป็นอย่างยิ่งที่จะต้องทราบว่าข้อมูลที่เก็บรวบรวมได้นั้นมีการแจกแจงแบบใด

ในการวิจัยครั้งนี้ได้นำเอาการแจกแจงแบบจีโอเมตริกมาพิจารณาดังต่อไปนี้

ให้ X เป็นตัวแปรใด ๆ ที่มีค่าเป็นเลขจำนวนเต็มบวก

p เป็นความน่าจะเป็นของความสำเร็จ (Probability of success)

$p > 0$

q เป็นความน่าจะเป็นของความล้มเหลว (Probability of failure)

และ $p + q = 1$

X จะมีการแจกแจงแบบจีโอเมตริก ถ้า

$$P\left[X = k\right] = q\,p^{k-1} \qquad k = 1, 2, 3, \ldots$$

และ

$$\mu = E(X) = \frac{1}{1-p}$$

$$\sigma^2 = Var(X) = \frac{p}{(1-p)^2}$$

๒.๒.๑ <u>ช่วงเวลา (นับเป็นวัน) ของการเกิดฝน</u> (Length of wet spell)

การแจกแจงของช่วงเวลาของการเกิดฝน สามารถจะเขียนให้อยู่ในรูปของการแจกแจงแบบจีโอเมตริกได้ดังนี้

ถ้า X = ช่วงเวลาของการเกิดฝน

ให้ P_1 = ความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝนในวันใดวันหนึ่งโดยที่กำหนดว่ามีฝนตกในวันก่อนหน้านั้น ๑ วัน

= $P(W/W)$

ฉนั้น $1-P_1$ = $P(D/W)$

= ความน่าจะเป็นอย่างมีเงื่อนไขของการไม่เกิดฝนในวันใดวันหนึ่งโดยที่กำหนดว่า มีฝนตกในวันก่อนหน้า ๑ วัน

= ความน่าจะเป็นของการหมดช่วงของการเกิดฝน

ช่วงเวลาของการเกิดฝน X สามารถเขียนให้อยู่ในรูปของการแจกแจงแบบจีโอเมตริก ดังนี้

$$P\left[X = k\right] = P\left[\text{ช่วงเวลาของการเกิดฝน} = k \text{ วัน}\right]$$

$$= P_1^{k-1}(1 - P_1)$$

๒.๒.๒ <u>ช่วงเวลา (นับเป็นวัน) ของการไม่เกิดฝน</u> (Length of dry spell)

การแจกแจงของช่วงเวลาของการไม่เกิดฝน สามารถจะเขียนให้อยู่ในรูปของการแจกแจงแบบจีโอเมตริกได้ดังนี้

ถ้า Y = ช่วงเวลาของการไม่เกิดฝน

ให้ P_o = ความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝนในวันใดวันหนึ่งโดยที่กำหนดว่า ไม่มีฝนตกในวันก่อนหน้านั้น ๑ วัน

= $P(W/D)$

= ความน่าจะเป็นของการหมดช่วงของการไม่เกิดฝน

ฉนั้น $1-P_o$ = $P(D/D)$

= ความน่าจะเป็นอย่างมีเงื่อนไขของการไม่เกิดฝนในวันใดวันหนึ่ง โดยที่วันก่อนหน้านั้น ๑ วัน ก็ไม่มีฝนตกเช่นกัน

ช่วงเวลาของการไม่เกิดฝน Y สามารถเขียนให้อยู่ในรูปของการแจกแจงแบบจีโอเมตริกได้ดังนี้

$$P\left[Y = m\right] = P\left[\text{ช่วงเวลาของการไม่เกิดฝน} = m \text{ วัน}\right]$$

$$= (1-P_o)^{m-1} P_o$$

๒.๒.๓ <u>ช่วงเวลา (นับเป็นวัน) ของวงจรลักษณะอากาศ</u> (Length of weather cycle)

ให้ X = ช่วงเวลาของการเกิดฝน

จะได้ว่า

$$P\left[X = k\right] = (1-P_1)\, P_1^{k-1}$$

$$= a_k$$

ในทำนองเดียวกันถ้าให้ Y = ช่วงเวลาของการไม่เกิดฝน

จะได้ว่า

$$P\left[Y = m\right] = (1-P_o)^{m-1} P_o$$

$$= b_m$$

ถ้า Z = ช่วงเวลาของวงจรลักษณะอากาศ

= ผลรวมของช่วงเวลาของการเกิดฝน (หรือไม่เกิดฝน) และการไม่เกิดฝน (หรือการเกิดฝน)

= n

และช่วงเวลาของการเกิดฝนและการไม่เกิดฝน เป็นอิสระต่อกัน (จะทดสอบใน ๒.๓) จะได้ว่า การแจกแจงของวงจรลักษณะอากาศยังคงเป็นแบบจีโอเมตริก และสามารถจะเขียนให้อยู่ในรูปต่อไปนี้

$$
\begin{aligned}
P\left[Z = n\right] &= a_1 b_{n-1} + a_2 b_{n-2} + \ldots\ldots + a_{n-1} b_1 \\
&= (1-P_1)P_1^{o}(1-P_o)^{n-2}P_o + (1-P_1)P_1(1-P_o)^{n-3}P_o \\
&\quad +(1-P_1)P_1^{2}(1-P_o)^{n-4}P_o + (1-P_1)P_1^{3}(1-P_o)^{n-5}P_o \\
&\quad +\ldots\ldots+(1-P_1)P_1^{n-3}(1-P_o)P_o + (1-P_1)P_1^{n-2}(1-P_o)P_o \\
&= P_o(1-P_1)\Big[(1-P_o)^{n-2} + P_1(1-P_o)^{n-3} + P_1^{2}(1-P_o)^{n-4} \\
&\quad +P_1^{3}(1-P_o)^{n-5} + \ldots\ldots + P_1^{n-3}(1-P_o) + P_1^{n-2}\Big] \\
&= P_o(1-P_1)\,\frac{(1-P_o)^{n-1} - P_1^{n-1}}{1-P_o - P_1}
\end{aligned}
$$

โดยที่

$$
\begin{aligned}
\mu &= E(Z) \\
&= E(X + Y) \\
&= E(X) + E(Y) \\
&= \frac{1}{1-P_1} + \frac{1}{P_o}
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2 &= \mathrm{Var}\ (Z) \\
&= \mathrm{Var}\ (X + Y) \\
&= \mathrm{Var}\ (X) + \mathrm{Var}\ (Y) \\
&= \frac{P_1}{(1-P_1)^2} + \frac{1-P_o}{P_o^2}
\end{aligned}
$$

๒.๓ การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

ในการทดสอบว่า ช่วงเวลาของการเกิดฝน และการไม่เกิดฝนต่างก็เป็นอิสระต่อกัน สามารถทดสอบได้โดยอาศัยการวิเคราะห์สหสัมพันธ์

ถ้า X = ช่วงเวลาของการเกิดฝน

Y = ช่วงเวลาของการไม่เกิดฝน

จะได้ว่า r_{XY} = สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ของช่วงเวลาทั้ง ๒

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

๒.๔ การหาการแจกแจงของจำนวนวันที่มีฝนตกใน ๑ สัปดาห์ (Distribution of the number of wet days in a week)

ในช่วงฤดูฝนโดยทั่ว ๆ ไป เป็นช่วงที่มีฝนตกติดต่อกันเกือบทุกวัน แต่ในบางระยะอาจจะเกิดมีเหตุการณ์ฝนทิ้งช่วง ดังนั้น การพิจารณาถึงจำนวนวันที่ฝนตกซึ่งมีความถี่สูงสุดในแต่ละหน่วยเวลาที่ตั้งไว้ จึงมีความจำเป็น ซึ่งในที่นี้จะถือเอาสัปดาห์มาตรฐานของปฏิทินประจำปีในทางอุตุนิยมวิทยาเป็นเกณฑ์ แต่ก่อนที่จะหาจำนวนวันที่มีฝนตกโดยมีความถี่สูงสุดใน ๑ สัปดาห์ จำเป็นที่จะต้องหาการแจกแจงของจำนวนวันที่มีฝนตกเสียก่อน ซึ่งการแจกแจงดังกล่าวสามารถหาได้ดังต่อไปนี้

ให้ P_1 = ความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝนในวันใดวันหนึ่งโดยที่กำหนดว่า มีฝนตกในวันก่อนหน้านั้น ๑ วัน

= $P\ (W/W)$

P_o = ความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝนในวันใดวันหนึ่งโดยที่กำหนดว่า ไม่มีฝนตกในวันก่อนหน้านั้น ๑ วัน

= P (W/D)

X_o เป็นเหตุการณ์เริ่มแรก (initial trial)

$$X_o = \begin{cases} 1 & \text{ถ้าเป็นวันที่มีฝนตก} \\ 0 & \text{ถ้าเป็นวันที่ไม่มีฝนตก} \end{cases}$$

R = ความน่าจะเป็นของการเกิดฝนใน initial trial

จากทฤษฎีของลูกโซ่มา-คอฟที่ประกอบด้วย ๒ สภาวะ

$$\text{ความน่าจะเป็นของการเกิดฝน (R)} = \frac{P_o}{1-(P_1 - P_o)}$$

ในการหาการแจกแจงของจำนวนวันที่มีฝนตกใน ๑ สัปดาห์

ให้ S = จำนวนวันที่มีฝนตก ซึ่งมีค่าได้ตั้งแต่ 0,1,2,.....7

n = จำนวนวันใน ๑ สัปดาห์ ซึ่งเท่ากับ ๗

ในแต่ละสัปดาห์ จะประกอบด้วยวันที่มีฝนตก และวันที่ไม่มีฝนตก ซึ่งการเกิดของแต่ละเหตุการณ์ อาจจะเกิดติดต่อกันเป็นช่วง หรืออาจจะเกิดสลับกันก็ได้

ถ้า C = จำนวนครั้งของการเปลี่ยนจากวันที่มีฝนตก (และวันที่ไม่มีฝนตก) ในวันใดวันหนึ่งเป็นวันที่ไม่มีฝนตก (และวันที่มีฝนตก) ในวันถัดไป

= จำนวนครั้งทั้งหมดของการเปลี่ยนเป็นเหตุการณ์ตรงข้ามในวันถัดไป

และ a, b เป็นเลขจำนวนเต็มที่น้อยที่สุด (Least Integer) โดยที่

$$a \geqslant \frac{1}{2}(C-1)$$

$$b \geqslant \frac{1}{2}C$$

ความน่าจะเป็นของการเกิดฝน S วัน ใน ๑ สัปดาห์ สามารถพิจารณาได้ ๒ กรณี คือ

๒.๔.๑. เหตุการณ์เริ่มแรกเป็นวันที่มีฝนตก

๒.๔.๒. เหตุการณ์เริ่มแรกเป็นวันที่ไม่มีฝนตก

๒.๔.๑ <u>เหตุการณ์เริ่มแรกเป็นวันที่มีฝนตก</u>

ในจำนวน C ครั้งของการเปลี่ยนเป็นเหตุการณ์ตรงข้ามในวันถัดไป ประกอบด้วย a และ b โดยที่

a = จำนวนครั้งของการเปลี่ยนจากวันที่ไม่มีฝนตก เป็นวันที่มีฝนตกในวันถัดไป

และ b = จำนวนครั้งของการเปลี่ยนจากวันที่มีฝนตกเป็นวันที่ไม่มีฝนตกในวันถัดไป

ดังนั้นจะมี S-a ครั้ง ที่มีฝนตกในวันใดวันหนึ่ง และวันถัดไปก็มีฝนตกเหมือนกัน และ 7-S-b ครั้งที่ไม่มีฝนตกในวันใดวันหนึ่ง และวันถัดไปก็ไม่มีฝนตกเช่นกัน

ความน่าจะเป็นของการเกิดฝน S วัน ใน ๑ สัปดาห์ โดยมี C ครั้งของการเปลี่ยนเป็นเหตุการณ์ตรงข้ามในวันถัดไป

$$= (1-P_1)^b P_o^a P_1^{S-a} (1-P_o)^{7-S-b}$$

$$= P_1^S (1-P_o)^{7-S} \left(\frac{1-P_1}{1-P_o}\right)^b \left(\frac{P_o}{P_1}\right)^a$$

เนื่องจากเหตุการณ์เริ่มแรกเป็นวันที่มีฝนตก ดังนั้น ความน่าจะเป็นของการเกิดฝน S วัน ใน ๑ สัปดาห์โดยที่เหตุการณ์เริ่มแรกเป็นวันที่มีฝนตก และมีการเปลี่ยนเป็นเหตุ-การณ์ตรงข้ามในวันถัดไป = C ครั้ง

$$= P\left[S,\ C\ /\ 7,\ X_o = 1\right]$$

$$= \binom{S}{a}\binom{7-S-1}{b-1} P_1^S (1-P_o)^{7-S} \left(\frac{1-P_1}{1-P_o}\right)^b \left(\frac{P_o}{P_1}\right)^a$$

และจะได้ว่า

$$P\left[S/7,\ X_o = 1\right] = P_1^S (1-P_o)^{7-S} \sum_{C=1}^{C_1} \binom{S}{a}\binom{7-S-1}{b-1}\left(\frac{1-P_1}{1-P_o}\right)^b \left(\frac{P_o}{P_1}\right)^a$$

โดยที่ $$C_1 = \begin{cases} 7\frac{1}{2} - \left|2S - 6\frac{1}{2}\right| & S < 7 \\ 0 & S = 7 \end{cases}$$

๒.๔.๒ <u>เหตุการณ์เริ่มแรกเป็นวันที่ไม่มีฝนตก</u>

ในทำนองเดียวกันกับ ๒.๔.๑ จะได้ว่า

$$P\left[S/7,\ X_o = 0\right] = P_1^S (1-P_o)^{7-S} \sum_{C=1}^{C_o} \binom{S-1}{b-1}\binom{7-S}{a}\left(\frac{1-P_1}{1-P_o}\right)^a \left(\frac{P_o}{P_1}\right)^b$$

โดยที่ $$C_o = \begin{cases} 7\frac{1}{2} - \left|2S - 7\frac{1}{2}\right| & S > 0 \\ 0 & S = 0 \end{cases}$$

จากผลที่ได้ใน ๒.๔.๑ และ ๒.๔.๒ จะได้ว่า

ความน่าจะเป็นของการเกิดฝน S วัน ใน ๑ สัปดาห์

$$= P\left[S/n\right]$$

$$= R\,P\left[S/7,\ X_o = 1\right] + (1-R)\,P\left[S/7, X_o = 0\right]$$

๒.๕ <u>วิธีการทดสอบที่ใช้ในการวิจัย</u>

จากการที่ทราบว่า ข้อมูลต่าง ๆ ที่รวบรวมได้มีการแจกแจงอยู่ในรูปลักษณะใด หรือมีคุณสมบัติอะไรบ้าง งานขั้นต่อไปก็คือ การทดสอบค่าสถิติ (Statistic) ของการแจกแจงต่าง ๆ เหล่านั้น

๒.๕.๑ การทดสอบว่าข้อมูลฝนรายวันในช่วงฤดูฝนในเขตกรุงเทพมหานคร เป็น first order Markov Chain

การทดสอบนี้จะทำการทดสอบเป็น ๒ ส่วน คือ

๒.๕.๑.๑ ทดสอบว่าความน่าจะเป็นของการเกิดฝนในวันใดวันหนึ่งขึ้นอยู่กับลักษณะอากาศของวันที่อยู่ก่อนหน้านั้น ๑ วัน ว่าฝนตกหรือไม่ตก นั่นคือ ทดสอบว่ามีคุณสมบัติของ First order chain

๒.๕.๒.๒ ทดสอบว่าความน่าจะเป็นของการเกิดฝนในแต่ละปีมีค่าคงที่

๒.๕.๑.๑ การทดสอบข้อมูลฝนรายวันว่าเป็น First order chain

ก. การทดสอบข้อมูลว่าเป็น first order chain โดยพิจารณาจาก second order chain

ให้ P_{ijk} = ความน่าจะเป็นของการเกิดสภาวะ k ที่เวลา t โดยกำหนดว่าเกิดสภาวะ j และ i เมื่อเวลา t-1 และ t-2 ตามลำดับ

และ n_{ijk} = จำนวนวันที่อยู่ในสภาวะ i, j, k ที่เวลา t-2, t-1 และ t ตามลำดับ

จากการประมาณโดยความน่าจะเป็นสูงสุด จะได้

$$\hat{P}_{ijk} = n_{ijk} \Big/ \sum_{l=1}^{2} n_{ijl}$$

จากสมมติฐาน

$$H_o : P_{ijk} = P_{jk} \quad (i, j, k = 1, 2)$$

จะได้ว่า ในแต่ละ j

$$\chi^2_j = \sum_{i,\, k} n^*_{ij} \frac{(\hat{P}_{ijk} - \hat{P}_{jk})^2}{\hat{P}_{jk}}$$

โดยที่ $\hat{P}_{jk} = \sum_{i=1}^{2} n_{ijk} / \sum_{i=1}^{2} \sum_{l=1}^{2} n_{ijl}$

และ $n^*_{ij} = \sum_{k=1}^{2} n_{ijk}$

ที่องศาแห่งความอิสระ $= (m-1)^2$

$= 1$

ดังนั้น $\chi^2 = \sum_{j=1}^{2} \chi^2_j$

ที่องศาแห่งความอิสระ $= m(m-1)^2$

$= 2$

ข. การทดสอบข้อมูลว่า เป็น order chain ที่ r-1 โดยพิจารณาจาก order chain ที่ r

ให้ $P_{ij\cdots kl}$ = ความน่าจะเป็นของการเกิดสภาวะ l ที่เวลา t โดยกำหนดว่าเกิดสภาวะ k,......j, i เมื่อเวลา t-1,.....t-r+1 และ t-r ตามลำดับ

และ $n_{ij\cdots kl}$ = จำนวนวันที่อยู่ในสภาวะ i, j.....k, l ที่เวลา t-r, t-r+1,.....t-1 และ t ตามลำดับ

จากการประมาณโดยความน่าจะเป็นสูงสุดจะได้

$$\hat{P}_{ij\ldots kl} = n_{ij\ldots kl} / n^*_{ij\ldots k}$$

จากสมมติฐาน

$$H_o : P_{ij\ldots kl} = P_{j\ldots kl} \quad (i, j\ldots k, l = 1, 2)$$

จะได้

$$\chi^2_{j\ldots k} = \sum_{i,l} n^*_{ij\ldots k} \frac{(\hat{P}_{ij\ldots kl} - \hat{P}_{j\ldots kl})^2}{\hat{P}_{j\ldots kl}}$$

โดยที่

$$\hat{P}_{j\ldots kl} = \sum_{i=1}^{2} n_{ij\ldots kl} / \sum_{i=1}^{2} n^*_{ij\ldots k}$$

และ $n^*_{ij\ldots k} = \sum_{l=1}^{2} n_{ij\ldots kl}$

ที่องศาแห่งความอิสระ = $(m-1)^2$

ฉนั้น $\chi^2 = \sum_{j,\ldots k} \chi^2_{j\ldots k}$

ที่องศาแห่งความอิสระ = $m^{r-1} (m-1)^2$

จากข้อ ข. เป็นการพิจารณาเฉพาะ order chain ที่อยู่ติดกัน นั่นคือ order ที่ r กับ r-1 เท่านั้น ซึ่งโดยวิธีเดียวกันนี้สามารถเขียนให้อยู่ในรูปทั่วไป ที่จะสามารถทดสอบข้อมูลว่าเป็น order ที่ u โดยพิจารณาจาก order ที่ r $(u < r)$ ที่องศาแห่งความอิสระ $(m^r - m^u)(m-1)$

ในการวิจัยนี้ ทำการทดสอบสมมติฐาน เมื่อ $u = 1$ และ $r = 3$ โดยที่

$$\chi^2 = \sum_{j,i,k,l} n^*_{ijk} \frac{(\hat{P}_{ijkl} - \hat{P}_{kl})^2}{\hat{P}_{kl}}$$

เมื่อ $\hat{P}_{ijkl} = n_{ijkl} / n^*_{ijk}$

$$\hat{P}_{kl} = \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ijkl} / \sum_{i=1}^{2} \sum_{j=1}^{2} n^*_{ijk}$$

และ $n^*_{ijk} = \sum_{l=1}^{2} n_{ijkl}$

ที่องศาแห่งความอิสระ $= (2^3 - 2^1)\ (2-1)$

$= 6$

๒.๔.๑.๒ <u>การทดสอบความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝนในแต่ละปี ว่ามีค่าคงที่</u>

ในการทดสอบค่าเหล่านี้ ทดสอบโดยเทียบกับค่าคงที่ ซึ่งค่าคงที่ในที่นี้ คือค่าความน่าจะเป็นอย่างมีเงื่อนไขของการเกิดฝนที่ประมาณโดยวิธี ๒.๑.๑ จากช่วงเวลาทั้งหมดรวม ๒๘ ปี

$$H_o : P_{ij}(t') = P_{ij} \qquad (t' = 1,2,\ldots 28)$$

เนื่องจาก $\hat{P}_{ij}(t') = n_{ij} / \sum_{j=1}^{2} n_{ij}$

และ $\sum_{t} n_i(t-1) = \sum_{j=1}^{2} n_{ij}$

จะได้ว่า $$\chi^2_i = \sum_{t',j} n_i(t-1) \frac{(\hat{P}_{ij}(t') - \hat{P}_{ij})^2}{\hat{P}_{ij}}$$

ที่องศาของความอิสระ $= (m-1)(t'-1)$

$= 27$

และเนื่องจากค่า $\hat{P}_{ij}(t')$ และ $\hat{P}_{ij}$ ในแต่ละ i เป็นอิสระต่อกัน

ดังนั้น $$\chi^2 = \sum_{i=1}^{2} \chi^2_i$$

ที่องศาของความอิสระ = m(m-1)(t-1)

= 54

๒.๕.๒ <u>การหาการแจกแจงของค่าสถิติไค-สแควร์</u>

จากค่าสถิติไค-สแควร์ ที่ใช้ในการทดสอบนี้ ยังไม่ได้มีการหาการแจกแจงของมันว่าอยู่ในแบบใด ซึ่งการแจกแจงของค่าสถิติดังกล่าวหาได้ดังนี้ คือ

ให้ Q_i = ค่าสถิติไค-สแควร์

$$= \sum_{j=1}^{2} \frac{\sum_t n_i(t-1)\ (\hat{P}_{ij} - P_{ij})^2}{P_{ij}}$$

$$= \sum_t n_i(t-1) \left[\frac{(\hat{P}_{i1} - P_{i1})^2}{P_{i1}} + \frac{(\hat{P}_{i2} - P_{i2})^2}{P_{i2}} \right]$$

$$= \sum_t n_i(t-1) \left[\frac{(\hat{P}_{i1} - P_{i1})^2}{P_{i1}} + \frac{\left[(1-\hat{P}_{i1}) - (1-P_{i1})\right]^2}{1-P_{i1}} \right]$$

$$= \sum_t n_i(t-1) \left[\frac{(\hat{P}_{i1} - P_{i1})^2}{P_{i1}} + \frac{(\hat{P}_{i1} - P_{i1})^2}{1-P_{i1}} \right]$$

$$= \frac{\sum_t n_i(t-1)\ (\hat{P}_{i1} - P_{i1})^2}{P_{i1}(1-P_{i1})}$$

$$= \left[\frac{\left[\sum_t n_i(t-1)\right]^{\frac{1}{2}} (\hat{P}_{i1} - P_{i1})}{\sqrt{P_{i1}\ (1-P_{i1})}} \right]^2$$

จาก ๒.๑.๓ ได้ว่า

$\left[\sum_t n_i(t-1)\right]^{\frac{1}{2}} (\hat{P}_{ij} - P_{ij})$ มีการแจกแจงแบบปกติ เมื่อขนาดตัวอย่างเพิ่มขึ้น

โดยมี mean = 0

และ variance = $P_{ij}(1-P_{ij})$

ดังนั้น $\dfrac{\left[\sum_t n_i(t-1)\right]^{\frac{1}{2}} (\hat{P}_{ij} - P_{ij})}{\sqrt{P_{ij}(1-P_{ij})}}$ จะมี limiting standard normal distribution

และ

$$\left[\frac{\left[\sum_t n_i(t-1)\right]^{\frac{1}{2}} (\hat{P}_{ij} - P_{ij})}{\sqrt{P_{ij}(1-P_{ij})}}\right]^2$$

มี limiting χ^2 distribution ที่องศาแห่งความอิสระ = 1

ฉนั้น Q_1 มีการแจกแจงแบบไค-สแควร์ เมื่อขนาดตัวอย่างเพิ่มขึ้น

๒.๕.๓ <u>การทดสอบการแจกแจงของช่วงเวลาของการเกิดฝน, การไม่เกิดฝน วงจรลักษณะอากาศ และจำนวนวันที่มีฝนตกใน ๑ สัปดาห์</u>

ในการทดสอบการแจกแจงต่าง ๆ ที่ได้จากทฤษฎีใน ๒.๒ ถึง ๒.๔ จะทำการทดสอบโดยวิธีไค-สแควร์ กล่าวคือ เปรียบเทียบความถี่ที่ได้จากข้อมูล กับค่าความถี่ที่คาดว่าควรจะเป็น

สูตรที่ใช้ในการทดสอบ

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

โดยที่ O_i = ค่าความถี่ที่ได้จากข้อมูล

E_i = ค่าความถี่ที่คาดว่าควรจะเป็น

m = จำนวนประเภทของข้อมูลที่ศึกษา

ที่องศาของความอิสระ = $m-1-P$

เมื่อ P = จำนวนพารามิเตอร์ที่ใช้ในการคำนวณ