การนำการแจกแจงความถี่ของข้อมูลทางสถิติ แบบ Beta Distribution มาประยุกต์ใช้เพื่อหาค่าความเหมาะสมภายใต้ความแปรปรวน ความเบ้ และความโด่งที่แตกต่างกันสำหรับตัวแปรเชิงกายภาพ ร่วมกับค่าสหสัมพันธ์ด้วยโปรแกรม MATLAB และการเปรียบเทียบกับการใช้ Gaussian Function

The Calculation of Suitability Score under Beta Distribution Probability Density Function Curve using Variance, Skewness, and Excess Kurtosis as Arguments for Physical Quantities with MATLAB Software and the Comparison with Gaussian Function

วิวรรษธร ฐิตสิริวิทย์

31 สิงหาคม 2563

สารบัญ

1 บทนำ (Introduction)

1.1 การแจกแจงความถี่ทางสถิติรูปแบบต่าง ๆ

การแจกแจงทางสถิติ (Statistic distribution) มีหลากหลายรูปแบบ หนึ่งในเกณฑ์การจำแนก คือ

1. **การแจกแจงแบบสมมาตร (Symmetric distribution)** เช่น Normal Distribution, Lorentzian Distribution ฯลฯ
   คือ กราฟของชุดข้อมูลแบบฟังก์ชันแบบ Probability Density Function (PDF) ที่
   กระจายตัวอย่างสมมาตรในแนว $x = x_0$

2. **การแจกแจงแบบไม่สมมาตร (Asymmetric distribution)** เช่น Beta Distribution, Gamma Distribution, Chi-Square Distribution ฯลฯ
   คือ กราฟของชุดข้อมูลแบบฟังก์ชัน PDF ที่มีค่าความเบ้ (Skewness, Skew) ในทางซ้าย/ขวา หรือ ลบ/บวก โดย x_0 จะไม่อยู่ในแนวสมมาตร

และอาจแบ่งได้เป็นแบบตามชุดข้อมูลแบบไม่ต่อเนื่อง (Discrete distribution) และแบบต่อเนื่อง (Continuous distribution) แต่ในที่นี้จะเน้นไปในหัวข้อ การแจกแจงแบบต่อเนื่อง ซึ่งอยู่ในรูปของ PDF และจะไม่กล่าวถึง Cumulative Probability Function (CDF)

2 พารามิเตอร์ของฟังก์ชัน (Function Parameters)

2.1 สัญลักษณ์ที่ใช้เขียนฟังก์ชันที่มีตัวแปรและพารามิเตอร์ควบคุม

สัญลักษณ์ของฟังก์ชัน ใช้รูป $y = f(x_1, x_2, x_3, \ldots, x_n; a_1, a_2, a_3, \ldots, x_n)$ โดยสิ่งที่อยู่ใน $f(\ldots)$ นั้นคือ arguments, x_n คือ variable, a_n คือ fixed parameter ที่มีอยู่แล้ว ไม่ใช้หาความสัมพันธ์ เป็นตัวแปรควบคุมในการหา $x_n \mapsto f(x_n)$

3 นิยามของฟังก์ชันที่เกี่ยวข้องในทางสถิติ

3.1 ค่ากลางของข้อมูล

3.1.1 ค่าเฉลี่ยเลขคณิต (Mean, μ หรือ $\bar{x}$)

ค่าเฉลี่ยเลขคณิต คือ ผลรวมของข้อมูล หารด้วยจำนวนชุดข้อมูล อาจด้วยน้ำหนักที่เท่าหรือไม่เท่า ขึ้นอยู่กับการเลือกใช้ค่าเฉลี่ยให้เหมาะสมกับชุด ว่าต้องการให้น้ำหนักกับข้อมูลใดเป็นพิเศษหรือไม่ ดังสมการที่ 1 และ 2 ตามลำดับ

$$\begin{aligned} \mu_x &= \frac{1}{N}\sum_{i=1}^{N} x_i \\ &= \frac{1}{\beta - \alpha}\int_{\alpha}^{\beta} f(x_i)dx \end{aligned} \tag{1}$$

$$\mu_{x,w} = \frac{1}{\sum w_i}\sum_{i=1}^{N} w_i x_i \tag{2}$$

ส่วนในโปรแกรม MATLAB ใช้คำสั่ง `mean(A)`

3.1.2 ค่ามัธยฐาน (Median, Me)

ค่ามัธยฐาน คือ ค่าที่อยู่ตรงกลางของช่วงข้อมูล แจกแจงตามความถี่ในช่วงต่อเนื่อง[1]

3.1.3 ค่าฐานนิยม (Mode, Mo)

ค่าฐานนิยม คือ ค่าที่มีความถี่สูงสุดของช่วงข้อมูลต่อเนื่อง[2]

3.2 ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน

3.2.1 ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation, σ หรือ s)

ส่วนเบี่ยงเบนมาตรฐาน คือ ค่าที่ใช้ดูการกระจายของข้อมูลจากค่าเฉลี่ย หาจากรากที่สองของความแปรปรวน คำนวณได้จาก
สมการที่ 3 ซึ่งในโปรแกรม MATLAB สามารถหาได้จาก `std(A)` โดย `A` เป็นเวกเตอร์ของข้อมูลประเภทหนึ่ง

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2} \tag{3}$$

3.2.2 ความแปรปรวน (Standard Deviation, σ^2 หรือ s^2)

ความแปรปรวน คือ ตัวแปรที่ใช้วัดการกระจายของข้อมูล คิดจากการหาค่าเฉลี่ยของความต่างจากค่าเฉลี่ยกำลังสอง คำนวณได้จากสมการที่ 4 โดยในโปรแกรม MATLAB สามารถหาได้จาก `var(B)` โดย `B` เป็นเวกเตอร์ของข้อมูลประเภทหนึ่ง

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2 \tag{4}$$

3.3 ค่าความเบ้ (Skewness)

ค่าความเบ้ ใช้บอกว่าข้อมูลส่วนใหญ่ (ชุดหรือช่วงมีความถี่สูง) ไปอยู่ที่ Interval ใดของกราฟการกระจาย

3.4 ค่าความโด่ง (Excess Kurtosis)

ค่า Excess Kurtosis ใช้บอกว่าชุดข้อมูลต่อเนื่องนั้น มีการกระจาย หรือการเป็นกลุ่มมากแค่ไหน เมื่อเทียบกับ Normal Distribution

[1] ในการคำนวนมัธยฐานของช่วงข้อมูลที่ไม่ต่อเนื่อง ทั้งแบบแจกแจงความถี่และไม่แจกแจงความถี่ ต่างกับการหามัธยฐานของชุดข้อมูลแบบต่อเนื่อง (PDF) ขึ้นอยู่กับรูปแบบของการกระจายของข้อมูลว่าเป็นลักษณะใด

[2] ฐานนิยมแบบไม่ต่อเนื่องกับต่อเนื่อง ต่างกันในทำนองเดียวกันกับการหามัธยฐาน

3.5 ฟังก์ชันแกมมา (Gamma Function, $\Gamma(x)$)

ฟังก์ชันแกมมา คือ ฟังก์ชันที่ให้นิยามของแฟกทอเรียลที่นอกเหนือจากสมการที่ 5

$$n! = n(n-1)(n-2)\dots(3)(2)(1)\ ; n \in \mathbb{Z}^+ \tag{5}$$

ให้ใช้นิยามและสัญลักษณ์ดังสมการที่ 6

$$\begin{aligned} \Gamma(\xi) &= (\xi-1)! \\ &= \int_0^\infty e^{-t}t^{\xi-1}dt \end{aligned} \tag{6}$$

4 ฟังก์ชันการกระจายแบบสมมาตร (Symmetric Distribution)

4.1 Normal Distribution

Normal Distribution คือ การกระจายของข้อมูลรูปแบบที่ทั่วไปที่สุด (ที่มีการสุ่มอย่างแท้จริง, true random) สามารถใช้ได้ในหลากหลายกรณี กล่าวคือเป็นการแจกแจงความถี่ของข้อมูลในอุดมคติ โดยคำนึงถึงเพียงแค่ความแปรปรวน (σ^2) และค่าเฉลี่ยเลขคณิต (μ) ของข้อมูลเท่านั้น ทำให้ง่ายต่อการศึกษาและใช้งานในระดับเบื้องต้น โดยมี PDF ดังสมการที่ 7

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ ; \sigma > 0 \tag{7}$$

โดย $\text{Skew} = 0$, $\bar{x} = \mu$, $s^2 = \sigma^2$ และ $ExcessKurtosis = 0$

4.2 Lorentzian Distribution (Cauchy Distribution/Breit-Wigner Distribution)

Lorentzian Distribution คือ การกระจายข้อมูลซึ่งขึ้นกับตัวแปร Scale (γ) และตัวแปร ตำแหน่งของฐานนิยม (m หรือ x_0) สามารถปรับการกระจายของข้อมูลได้ มาจาก Normal Distribution ที่แตกต่างกัน 2 แบบที่ไม่ขึ้นกับกันและกัน ขณะที่ $\bar{x} = 0$ และ $s^2 = 1$ โดยมี PDF ดังสมการที่ 8

$$f(x;x_0,\gamma) = \frac{1}{\pi\gamma\left(1+\left(\frac{x-k}{\gamma}\right)^2\right)}\ ; \gamma > 0 \tag{8}$$

โดย Skew หาค่าไม่ได้, $\bar{x} = 0$, s^2 หาค่าไม่ได้ และ $ExcessKurtosis$ หาค่าไม่ได้

5 ฟังก์ชันการกระจายแบบอสมมาตร (Asymmetric Distribution)

5.1 Chi-Square (χ^2) Distribution

Chi-Square Distribution คือ การกระจายข้อมูลที่มักใช้ในการทดสอบสมมติฐาน มีความคล้ายคลึงกับ Normal Distribution และ Gamma Distribution ซึ่งขึ้นกับค่า chi-square (χ^2) และค่าองศาเสรี (Degree of freedom, k) โดยมี PDF ดังสมการที่ 9

$$f(x;k) = \frac{2^{(-\frac{k}{2})}}{\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \; ; x > 0 \tag{9}$$

โดย $\text{Skew} = 2\sqrt{\frac{2}{k}}$, $\bar{x} = k$, $s^2 = 2k$ และ $ExcessKurtosis = \frac{12}{k}$

5.2 Gamma (γ) Distribution (Erlang Distribution)

Gamma Distribution มักถูกใช้ในการวิเคราะห์ปริมาณทางฟิสิกส์ที่มีการแจกแจงความถี่ ทั้งการศึกษา และการสรุปผลการทดลองที่ไม่เป็นแบบสุ่ม (non-random events) เช่น สถิติการรอคิวในช่วงเวลาใดเวลาหนึ่ง โดยมีความคล้ายคลึงกับการแจกแจงความถี่รูปแบบอื่น ๆ ด้วย และขึ้นกับตัวแปร α และ β เท่านั้น โดยมี PDF ดังสมการที่ 10

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}} \; ; \alpha,\beta > 0 \tag{10}$$

โดย $\text{Skew} = \frac{2}{\sqrt{\alpha}}$, $\bar{x} = \alpha\beta$, $s^2 = \alpha\beta^2$ และ $ExcessKurtosis = \frac{6}{\alpha}$

5.3 Beta (β) Distribution

Beta Distribution คือการแจกแจงความถี่โค้งที่ขึ้นกับตัวแปร α และ β เท่านั้น ซึ่งใช้บอกความเบ้และความโด่งได้ เหมือนกับ Gamma Distribution และมีลักษณะคล้ายคลึงกับ Bernoulli Distribution (Yes/No) ซึ่งเป็นแบบ Discrete Distribution

การแจงแจงความถี่ในลักษณะนี้มีความยืดหยุ่นต่อลักษณะตัวแปร และข้อมูลหลากหลายประเภท จึงนำมาประยุกต์ใช้ในงานวิเคราะห์ขั้นสูงได้มาก มีการใช้งานในวงกว้างเหมือนกับแบบ Normal Distribution

อีกลักษณะที่เด่น คือการแจกแจงข้อมูลแบบนี้ สามารถกำหนดช่วงของข้อมูลที่เก็บมาได้ ให้เป็นช่วง และจำกัด Scale (γ) ให้เป็นค่าที่ต้องการได้ โดย Beta Distribution มี PDF ดังสมการที่ 11

$$f(x;\alpha,\beta) = \frac{1}{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]} x^{\alpha-1}(1-x)^{\beta-1} \; ; \alpha,\beta > 0 \tag{11}$$

โดย $\text{Skew} = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$, $\bar{x} = \frac{\alpha}{\alpha+\beta}$, $s^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}$
และ $ExcessKurtosis = \frac{3(\alpha+\beta+1)[\alpha\beta(\alpha+\beta-6)+2(\alpha+\beta)^2]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} - 3$

ในกรณีที่ $0 < \alpha,\beta < 1$ ทำให้ข้อมูลกระจายไปทางซ้ายและขวามากกว่าตรงกลาง แต่ในกรณีที่ $\alpha,\beta > 1$ จะทำให้ข้อมูลเป็นรูปโค้งคว่ำ

6 การใช้โปรแกรม MATLAB วิเคราะห์ข้อมูลเบื้องต้น

6.1 ชนิดของข้อมูล

ในโปรแกรม MATLAB ตัวแปรที่ใช้เป็นหลักจะอยู่ในรูปของอาร์เรย์ (Array) ซึ่งส่วนใหญ่ มักใช้เพียง 2 รูปแบบ คือ **อาร์เรย์ 1 มิติ (เวกเตอร์หลัก/column vector)** และ**อาเรย์ 2 มิติ (เมทริกซ์/matrix)**

การเก็บข้อมูลเชิงปริมาณ (Quantitative Data) หรือในที่นี้ ข้อมูลเชิงตัวเลข (Numerical Data) จะอยู่ในรูปของค่าที่ถูกล้อมด้วยเครื่องหมายจุลภาค (Comma-separated value หรือ Comma-delimited value, CSV Format) คือเป็นลักษณะของหลัก (column) และเมื่อเก็บข้อมูลซ้ำ ๆ จะทบไปในทางแถว (row) ขณะที่มีจำนวนหลักคงที่ตลอดการเก็บ 1 ช่วง เช่น

```
253,25.49,65.36,1012.17,25.3
254,26.50,68.77,1012.95,24.1
255,26.55,69.01,1013.00,20.4
256,26.86,69.68,1013.24,15.2
```

ซึ่งเป็นตัวอย่างชุดข้อมูลที่เก็บมา[3] ในรูปแบบ CSV (สกุลไฟล์ที่เก็บข้อมูลประเภทนี้ไม่มีความเกี่ยวข้องกับรูปแบบการจัดเรียงข้างใน หากไฟล์ถูกสร้างขึ้นจากไฟล์เปล่า) สามารถเทียบเคียงได้กับอาเรย์:

	col1	col2	col3	col4	col5
row1	253	25.49	65.36	1012.17	25.3
row2	254	26.50	68.77	1012.95	24.1
row3	255	26.55	69.01	1013.00	20.4
row4	256	26.86	69.68	1013.24	15.2

Table 1: Array A

อาร์เรย์นี้มีขนาด 4 แถว x 5 หลัก ซึ่งในโปรแกรม MATLAB สามารถหาขนาดของอาร์เรย์ได้ผ่านคำสั่ง `size(A)` ซึ่งค่าที่ออกมาจะเป็นในรูปแบบของเวกเตอร์แถว (row vector) เช่นในกรณีนี้ อาร์เรย์มีขนาด `[4,5]`

การบอกตำแหน่งของค่า เรียกว่า index จะบอกเป็นพิกัด (แถว,หลัก) ของอาร์เรย์ สามารถระบุเป็นจุดได้ `A(m,n)`, เลือกเว็กเตอร์แถวทั้งช่วง `A(m,:)` หรือบางช่วง `A(m,a:b)` และเลือกเว็กเตอร์หลักทั้งช่วง `A(:,n)` หรือบางช่วง `A(a:b,n)` หรือแม้แต่ดึงอาร์เรย์มาบางส่วน `A(a:b,c:d)` ก็ได้ ซึ่งส่วนใหญ่ในการวิเคราะห์ มักดึง column vector ออกจาก matrix เพื่อวิเคราะห์แยกทีละปัจจัยมากกว่า

6.2 การสร้างฟังก์ชันคำนวณ

ให้ไปศึกษาเอาเอง (หัวข้อ `function out = calc(path arguments)`) รวมถึงการดึง file path, การทำ composite function การสร้างสมการคำนวณ การวิเคราะห์แบบเส้น-เส้น การทำ fitting และการสร้าง loop

[3] การเขียนไฟล์ในลักษณะนี้ ใช้การ rw- แบบ +a (append if existed) ได้เลย ทำให้ข้อมูลต่อท้าย และเผื่อ algorithm ในการตรวจหาไฟล์ซ้ำที่มีอยู่แล้วเพื่อป้องกันการเขียนทับด้วย

7 การนำ Normal Distribution มาประยุกต์ใช้งาน

7.1 ที่มาของการประยุกต์และ การสร้าง Gaussian Function เพื่อวิเคราะห์ข้อมูล

ฟังก์ชันการแจกแจงความถี่โค้งแบบ Normal Distribution มีความเป็นอุดมคติในทางข้อมูลและรูปแบบการกระจายตัวที่สุด ทำให้ง่ายต่อการใช้งาน และการนำมาแก้ไข ดัดแปลงให้เป็น Non-normal distribution แบบอื่น ๆ ขณะที่ยังคงความเบ้เป็นศูนย์เหมือนเดิม เดิมทีฟังก์ชัน Normal Distribution จะอยู่ในรูปของสมการที่ 12

$$N(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{12}$$

ในการดัดแปลง เราอาจนำค่าคงที่ $\frac{1}{\sqrt{2\pi\sigma}}$, ตัวคูณ $\frac{1}{2\sigma^2}$ และ X-control variable $-\mu$ ออก จะได้เป็นสมการที่ 13 ซึ่งอยู่ในรูปมาตรฐานของฟังก์ชัน Gaussian Function (สามารถหาปริพันธ์จำกัดเขตตั้งแต่ $-\infty$ ถึง $+\infty$ ได้)

$$N(x) = Ae^{-\lambda(x-x_0)^2} \tag{13}$$

โดย x_0 คือค่าสูงสุดในช่วง (Local maximum) ในกรณี $\lambda > 0$ และให้ $A = 1$ ทำให้ค่าสูงสุดเป็น 1 เสมอสำหรับ $\forall x \forall x_0 \exists \lambda$ ซึ่งหากต้องการเทียบกับค่าสูงสุดของช่วง อาจหาได้ด้วยสมการที่ 14

$$S_N = e^{-\lambda(\frac{x-x_0}{x_0})^2} \tag{14}$$

สำหรับพารามิเตอร์ควบคุมฟังก์ชัน S_N มีเพียง λ ที่ควบคุมความโด่ง สามารถปรับได้ตามความเหมาะสมของการกระจายของข้อมูลนั้น หรืออาจใช้ $\frac{1}{2\sigma^2}$ และ $\frac{1}{\sqrt{2\pi\sigma}}$ ซึ่งเป็นค่าสัมประสิทธิ์ที่มีพารามิเตอร์ส่วนเบี่ยงเบนมาตรฐานมาช่วยได้ และ x_0 ใช้เป็นค่าสูงสุดของโค้ง (เหมาะสมที่สุด ซึ่งมีค่าเป็น 1)

7.2 ข้อได้เปรียบและข้อจำกัดของ Gaussian Function ใน Symmetric Distribution

การใช้ Normal/Non-normal Distribution มาประยุกต์หาค่าความเหมาะสม (Suitability Score) เป็นการให้ข้อมูลมีการกระจายความเหมาะสมแบบเท่ากัน และสมมาตรสองฝั่งโดยปริยาย ซึ่งปริมาณทางฟิสิกส์บางตัวแปร มิได้มีค่าความเหมาะสมและอัตราการเปลี่ยนแปลง ณ จุดข้อมูล ของความเหมาะสมเท่ากันตลอดในแนวสมมาตร อาจมีความเบ้ของความเหมาะสม ว่าข้อมูลที่อยู่ฝั่งซ้าย อาจเบ้น้อยกว่าข้อมูลที่อยู่ฝั่งขวาก็ได้ ทำให้การใช้ Symmetric Distribution อาจทำให้ค่าเพี้ยนไป ซึ่งในที่นี้ การใช้ Asymmetric Distribution อาจสื่อความหมายได้ดีกว่า และคำนวณได้แม่นยำ (ใกล้เคียงสภาพความเป็นจริง) มากกว่าแบบดังกล่าว

รวมถึงความสามารถในการประมาณค่า แบ่งเกณฑ์คิด ทดลองหา หรือทำ Curve fitting เพื่อหาความเบ้และความโด่งของโค้งจากพารามิเตอร์ที่เรากำหนด ไว้เป็นตัวแปรควบคุมได้อย่างเหมาะสมและสื่อความหมายได้ดี

7.3 การสร้างฟังก์ชัน Gaussian Suitability Score, S_N

เริ่มจากการสร้างไฟล์ของฟังก์ชันขึ้นมา โดยตั้งชื่อไฟล์ว่า `gaussianscore.m` โดยเขียนฟังก์ชันขึ้นมาพร้อมกับ argument ที่ได้กล่าวไปดังนี้

```
function out = gaussianscore(t,t0,sigma,lambda)
    e = exp(1);
    out = 1/(sqrt(2*pi*sigma))* e.^(-1/(2*sigma.^2)*lambda* ((t-t0)/t0).^2)
end
```

8 การนำ Beta Distribution มาประยุกต์ใช้งาน

8.1 ที่มาของฟังก์ชันหาค่าความเหมาะสม S_β

การสร้างฟังก์ชัน Beta Distribution ที่ขึ้นกับตัวแปรที่นอกเหนือจาก α และ β สามารถทำได้โดยการนำตัวแปรที่เราต้องการให้เป็น parameter (ทำเป็นตัวแปรควบคุม) และ variable (ทำเป็นตัวแปรต้น) มาจัดรูปเพื่อแทนค่าเข้าไปใน parameter เดิม ให้เป็นฟังก์ชันใหม่ ในที่นี้คือการสร้างฟังก์ชันหาค่าความเหมาะสม S_β เมื่อเปรียบเทียบค่า หรือค่าตัวแทนของชุดข้อมูลที่เก็บมาได้ เทียบกับค่าที่เหมาะสมหรือดีที่สุดค่าหนึ่ง ซึ่งข้อจำกัดของช่วง (constraint) ที่ต้องการกำหนดคือ ค่าที่ดีที่สุด (t_0) ค่าต่ำสุดของช่วงที่เลวที่สุด (minimum, m) ค่าสูงสุดของช่วงที่เลวที่สุด (maximum, M) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation, σ) ทั้งหมดเป็น fixed parameter ส่วน variable คือค่าของข้อมูล หรือตัวแทนของชุดข้อมูลที่เก็บมาได้ (t)

กำหนดฟังก์ชันเชิงเส้น $T(t)$ ดังสมการที่ 15

$$T(t) = at + b \tag{15}$$

ข้อสังเกต 1: เมื่อเปลี่ยน b จะทำให้ Y-intercept เปลี่ยน แต่หากเปลี่ยน t เป็น ฟังก์ชันประกอบอื่น ๆ (composite function) จะทำให้ X-intercept เปลี่ยน รวมถึงการคูณค่าคงที่ไม่เจาะจง (arbitrary constant) เพื่อลด scale ของโดเมน t ลง ในที่นี้ใช้ scale เป็นพิสัยของข้อมูลที่เป็นไปได้ $\frac{1}{M-m}$ คูณเข้าไปในฟังก์ชันประกอบเป็น $T\left(\frac{t}{M-m}\right)$ จะได้ สมการที่ 16

$$T(t) = \frac{t}{M-m} + b \tag{16}$$

รวบให้ b มีค่าขยับแกน X ด้วยจะได้ สมการที่ 17

$$\begin{aligned} T(t) &= \frac{t}{M-m} + \frac{m}{m-M} \\ T(t) &= \frac{t-m}{M-m} \end{aligned} \tag{17}$$

สมการสุดท้ายที่ต้องการหาค่าความเหมาะสม S_β จะอ้างอิงจาก Beta Distribution $B(x)$ แทนค่า $T(t)$ เข้าไปเพื่อเปลี่ยน constraint จะได้ สมการที่ 18

$$S_\beta = B(T(t)) \tag{18}$$

โดย

$$B(\xi) = \frac{1}{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]} \xi^{\alpha-1}(1-\xi)^{\beta-1}$$

และค่าต่ำสุดของ $B(\xi)$ มีค่าเป็น 0 จาก

$$\lim_{\xi \to m^+} B(\xi) = 0$$

และ

$$\lim_{\xi \to M^-} B(\xi) = 0$$

หากให้ $\alpha, \beta > 1$ เมื่อหาค่าสูงสุดของ $B(\xi)$จากการหาอนุพันธ์ฟังก์ชันเทียบค่าเท่ากับศูนย์

$$\frac{\partial B(\xi)}{\partial \xi} = 0$$

แก้สมการจะได้ $B(\xi)$ มีค่าสูงสุดที่

$$\xi_0 = \frac{\alpha - 1}{\alpha + \beta - 2} \tag{19}$$

ข้อสังเกต 2: เรนจ์ของ $B(\xi)$ มีค่าสูงสุดไม่เท่ากับ 1 แต่ขึ้นกับ α และ β ทำให้จำเป็นต้องใช้วิธี Inverse scaling ผันเรนจ์จาก $[0, \xi_0]$ เป็น $[0, 1]$ ด้วยการคูณ $\frac{1}{B(\xi_0)}$ เข้าไปใน $B(\xi)$ ได้ สมการที่ 20

$$\begin{aligned} B(\xi) &= \frac{1}{B(\xi_0)} \frac{1}{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]} \xi^{\alpha-1}(1-\xi)^{\beta-1} \\ B(\xi) &= \frac{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]}{\xi_0^{\alpha-1}(1-\xi_0)^{\beta-1}} \frac{1}{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]} \xi^{\alpha-1}(1-\xi)^{\beta-1} \\ B(\xi) &= \frac{\xi^{\alpha-1}(1-\xi)^{\beta-1}}{\xi_0^{\alpha-1}(1-\xi_0)^{\beta-1}} \end{aligned} \tag{20}$$

เมื่อแทน $\xi_0 = \frac{\alpha-1}{\alpha+\beta-2}$ เข้าไปในสมการที่ 20 และจัดรูป จะได้ผลสุดท้ายเป็น สมการที่ 21 ดังนี้

$$\begin{aligned} B(\xi) &= \frac{(\alpha+\beta-2)^{\alpha+\beta-2}}{(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}} \xi^{\alpha-1}(1-\xi)^{\beta-1} \\ B(\xi) &= \frac{(\alpha+\beta-2) \uparrow^2 2}{[(\alpha-1) \uparrow^2 2]\,[(\beta-1) \uparrow^2 2]} \xi^{\alpha-1}(1-\xi)^{\beta-1} \end{aligned} \tag{21}$$

เมื่อได้สมการที่ 21 แล้ว นำฟังก์ชัน $T(t)$ เข้ามาแทนค่าใน $B(\xi)$ จะได้สมการที่ 22

$$S_\beta = \frac{(\alpha+\beta-2) \uparrow^2 2}{[(\alpha-1) \uparrow^2 2]\,[(\beta-1) \uparrow^2 2]} (T(t))^{\alpha-1}(1-T(t))^{\beta-1} \tag{22}$$

ข้อสังเกต 3: ค่าความเหมาะสม S_β ยังคงติดตัวแปร α และ β อยู่ในขณะที่ยังไม่มีตัวแปร t_0 และ σ โดยสามารถเลือกกำจัดตัวแปรใดตัวแปรหนึ่งได้ แล้วผลเฉลยของ α กับ β มีความคล้ายคลึงกันมาก เนื่องจาก 2 ตัวแปรนี้แปรผกผันกันแบบสมมาตร

เราจึงสามารถแก้สมการหาว่าจุดสูงสุดของ $S_\beta = 1$ อยู่ที่ $t = t_0$ แก้สมการหาตัวแปร α ในเทอมของ t_0 ได้จาก สมการที่ 17 กับ 19 จะได้สมการที่ 23

$$\begin{aligned} \frac{\alpha-1}{\alpha+\beta-2} &= \frac{t_0-m}{M-m} \\ \alpha &= \frac{(t_0-m)(\beta-2)+(M-m)}{M-t_0} \end{aligned} \tag{23}$$

และ β หาจากพื้นที่ใต้โค้ง Normal Distribution กับตัวแปรสมมติเพื่อปรับความโด่งของกราฟ s ให้เป็นค่าคงที่ไม่เจาะจง ซึ่งเป็นหนึ่งใน fixed parameter ของฟังก์ชันด้วย ดังสมการที่ 24

$$\beta = \frac{1}{\sqrt{2\pi\sigma}} s \tag{24}$$

8.2 โดเมนและเรนจ์ของฟังก์ชัน

8.2.1 โดเมนของตัวแปรและพารามิเตอร์คงที่

$\alpha \in (1, +\infty)$

$\beta \in (1, +\infty)$

$t_0 \in [m, M]$

$s \in (\sqrt{2\pi\sigma}, +\infty)$

8.2.2 เรนจ์ของ S_β

$B(T(t)) \in [0, 1]$

8.3 การสร้างฟังก์ชัน Beta Distribution Suitability Score

เริ่มจากการสร้างไฟล์ของฟังก์ชันขึ้นมา โดยตั้งชื่อไฟล์ว่า `betascore.m` โดยเขียนฟังก์ชันขึ้นมาพร้อมกับ argument ที่ได้กล่าวไปดังนี้

```
function out = betascore(t,t0,m,M,s,sigma)
    e = exp(1);
    beta = s/sqrt(2*pi*sigma);
    alpha = ((t0-m)*(beta-2)+(M-m))/(M-t0);
    Tex = (t-m)/(M-m);
    o1 = (alpha+beta-2).^(alpha+beta-2)/ ...
        ((alpha-1).^(alpha-1)*(beta-1).^(beta-1));
    o2 = Tex.^(alpha-1)*(1-Tex).^(beta-1);
    out = o1*o2;
end
```

9 เนื้อหาเพิ่มเติม

9.1 สัญลักษณ์ Knuth's up-arrow Notation

สัญลักษณ์ $a \uparrow^n b$ คือ Knuth's up-arrow notation ไว้บอกการเป็น Hyperoperation, Tetration, Pentation, Hexation, ... เช่น

$$a \uparrow 2 = a^2 \quad a \uparrow^2 2 = a^a \quad a \uparrow^3 2 = a^{a^a}$$
$$a \uparrow 3 = a^3 \quad a \uparrow^2 3 = a^{a^a} \quad a \uparrow^3 3 = a^{a^{a^a}}$$

สามารถไปศึกษาเพิ่มเติมได้ที่หัวข้อที่กล่าวไป

9.2 ตัวอย่างการสร้างฟังก์ชันคำนวณด้วย MATLAB (เก่า)

```
function out = calc(file)
for loop = 1:3
for i = 7:12
    cla(figure(i))
end
dataset = csvread(file);
%dataset = magic(49);
temp = dataset(:,7);
humid = dataset(:,8);
rainfall = dataset(:,9);
red = dataset(:,10);
blue = dataset(:,11);
ccn = dataset(:,12);
carbon = dataset(:,13);

%rainfall = ANNUAL RAINFALL
temp1 = dataset(:,14);
humid1 = dataset(:,15);
rainfall1 = dataset(:,16);
red1 = dataset(:,17);
blue1 = dataset(:,18);
carbon1 = dataset(:,19);

temp2 = dataset(:,20);
humid2 = dataset(:,21);
rainfall2 = dataset(:,22);
red2 = dataset(:,23);
blue2 = dataset(:,24);
carbon2 = dataset(:,25);

temp3 = dataset(:,26);
humid3 = dataset(:,27);
rainfall3 = dataset(:,28);
red3 = dataset(:,29);
blue3 = dataset(:,30);
carbon3 = dataset(:,31);

temp4 = dataset(:,32);
humid4 = dataset(:,33);
rainfall4 = dataset(:,34);
red4 = dataset(:,35);
blue4 = dataset(:,36);
carbon4 = dataset(:,37);

temp5 = dataset(:,38);
humid5 = dataset(:,39);
rainfall5 = dataset(:,40);
red5 = dataset(:,41);
blue5 = dataset(:,42);
carbon5 = dataset(:,43);

temp6 = dataset(:,44);
humid6 = dataset(:,45);
rainfall6 = dataset(:,46);
red6 = dataset(:,47);
blue6 = dataset(:,48);
carbon6 = dataset(:,49);

r(1,1) = corr(temp,temp1);
r(1,2) = corr(temp,temp2);
r(1,3) = corr(temp,temp3);
r(1,4) = corr(temp,temp4);
r(1,5) = corr(temp,temp5);
r(1,6) = corr(temp,temp6);

r(2,1) = corr(humid,humid1);
r(2,2) = corr(humid,humid2);
r(2,3) = corr(humid,humid3);
r(2,4) = corr(humid,humid4);
r(2,5) = corr(humid,humid5);
r(2,6) = corr(humid,humid6);

r(3,1) = corr(rainfall,rainfall1);
r(3,2) = corr(rainfall,rainfall2);
r(3,3) = corr(rainfall,rainfall3);
r(3,4) = corr(rainfall,rainfall4);
r(3,5) = corr(rainfall,rainfall5);
r(3,6) = corr(rainfall,rainfall6);

r(4,1) = corr(red,red1);
r(4,2) = corr(red,red2);
r(4,3) = corr(red,red3);
```

```
r(4,4) = corr(red,red4);
r(4,5) = corr(red,red5);
r(4,6) = corr(red,red6);

r(5,1) = corr(blue,blue1);
r(5,2) = corr(blue,blue2);
r(5,3) = corr(blue,blue3);
r(5,4) = corr(blue,blue4);
r(5,5) = corr(blue,blue5);
r(5,6) = corr(blue,blue6);

r(6,1) = corr(carbon,carbon1);
r(6,2) = corr(carbon,carbon2);
r(6,3) = corr(carbon,carbon3);
r(6,4) = corr(carbon,carbon4);
r(6,5) = corr(carbon,carbon5);
r(6,6) = corr(carbon,carbon6);

%temp

dat = dataset(:,7);
dat1 = dataset(:,14);
dat2 = dataset(:,20);
dat3 = dataset(:,26);
dat4 = dataset(:,32);
dat5 = dataset(:,38);
dat6 = dataset(:,44);

p = polyfit(dat1,dat,1);
f = polyval(p,dat1);

hold on

figure(7)
plot(dat1,f,'--r')
plot(dat,dat1)
scatter(dat,dat1)
plot(dat2,f,'--r')
plot(dat,dat2)
scatter(dat,dat2)
plot(dat3,f,'--r')
plot(dat,dat3)
scatter(dat,dat3)
plot(dat4,f,'--r')
plot(dat,dat4)
scatter(dat,dat4)
plot(dat5,f,'--r')
plot(dat,dat5)
scatter(dat,dat5)
plot(dat6,f,'--r')
plot(dat,dat6)
scatter(dat,dat6)
plot(dat,dat)


hold off

%humid

dat = dataset(:,8);
dat1 = dataset(:,15);
dat2 = dataset(:,21);
dat3 = dataset(:,27);
dat4 = dataset(:,33);
dat5 = dataset(:,39);
dat6 = dataset(:,45);

p = polyfit(dat1,dat,1);
f = polyval(p,dat1);

hold on

figure(8)
plot(dat1,f,'--r')
plot(dat,dat1)
scatter(dat,dat1)
plot(dat2,f,'--r')
plot(dat,dat2)
scatter(dat,dat2)
plot(dat3,f,'--r')
plot(dat,dat3)
scatter(dat,dat3)
plot(dat4,f,'--r')
plot(dat,dat4)
scatter(dat,dat4)
plot(dat5,f,'--r')
plot(dat,dat5)
scatter(dat,dat5)
plot(dat6,f,'--r')
plot(dat,dat6)
scatter(dat,dat6)
plot(dat,dat)

hold off

%rainfall

dat = dataset(:,9);
dat1 = dataset(:,16);
dat2 = dataset(:,22);
dat3 = dataset(:,28);
dat4 = dataset(:,34);
dat5 = dataset(:,40);
dat6 = dataset(:,46);

p = polyfit(dat1,dat,1);
f = polyval(p,dat1);

hold on

figure(9)
plot(dat1,f,'--r')
plot(dat,dat1)
scatter(dat,dat1)
plot(dat2,f,'--r')
plot(dat,dat2)
```

```
scatter(dat,dat2)
plot(dat3,f,'--r')
plot(dat,dat3)
scatter(dat,dat3)
plot(dat4,f,'--r')
plot(dat,dat4)
scatter(dat,dat4)
plot(dat5,f,'--r')
plot(dat,dat5)
scatter(dat,dat5)
plot(dat6,f,'--r')
plot(dat,dat6)
scatter(dat,dat6)
plot(dat,dat)

hold off

%red
dat = dataset(:,10);
dat1 = dataset(:,17);
dat2 = dataset(:,23);
dat3 = dataset(:,29);
dat4 = dataset(:,35);
dat5 = dataset(:,41);
dat6 = dataset(:,47);

p = polyfit(dat1,dat,1);
f = polyval(p,dat1);

hold on

figure(10)
plot(dat1,f,'--r')
plot(dat,dat1)
scatter(dat,dat1)
plot(dat2,f,'--r')
plot(dat,dat2)
scatter(dat,dat2)
plot(dat3,f,'--r')
plot(dat,dat3)
scatter(dat,dat3)
plot(dat4,f,'--r')
plot(dat,dat4)
scatter(dat,dat4)
plot(dat5,f,'--r')
plot(dat,dat5)
scatter(dat,dat5)
plot(dat6,f,'--r')
plot(dat,dat6)
scatter(dat,dat6)
plot(dat,dat)

hold off

%blue
dat = dataset(:,11);
dat1 = dataset(:,18);
dat2 = dataset(:,24);
dat3 = dataset(:,30);
dat4 = dataset(:,36);
dat5 = dataset(:,42);
dat6 = dataset(:,48);

p = polyfit(dat1,dat,1);
f = polyval(p,dat1);

hold on

figure(11)
plot(dat1,f,'--r')
plot(dat,dat1)
scatter(dat,dat1)
plot(dat2,f,'--r')
plot(dat,dat2)
scatter(dat,dat2)
plot(dat3,f,'--r')
plot(dat,dat3)
scatter(dat,dat3)
plot(dat4,f,'--r')
plot(dat,dat4)
scatter(dat,dat4)
plot(dat5,f,'--r')
plot(dat,dat5)
scatter(dat,dat5)
plot(dat6,f,'--r')
plot(dat,dat6)
scatter(dat,dat6)
plot(dat,dat)

hold off

%carbon
dat = dataset(:,13);
dat1 = dataset(:,19);
dat2 = dataset(:,25);
dat3 = dataset(:,31);
dat4 = dataset(:,37);
dat5 = dataset(:,43);
dat6 = dataset(:,49);

p = polyfit(dat1,dat,1);
f = polyval(p,dat1);

hold on

figure(12)
plot(dat1,f,'--r')
plot(dat,dat1)
scatter(dat,dat1)
plot(dat2,f,'--r')
plot(dat,dat2)
scatter(dat,dat2)
plot(dat3,f,'--r')
plot(dat,dat3)
scatter(dat,dat3)
plot(dat4,f,'--r')
```

```
plot(dat,dat4)
scatter(dat,dat4)
plot(dat5,f,'--r')
plot(dat,dat5)
scatter(dat,dat5)
plot(dat6,f,'--r')
plot(dat,dat6)
scatter(dat,dat6)
plot(dat,dat)

hold off

for i = 1:6
r(7,i) = (r(1,i)+r(2,i)+r(3,i)+ ...
    r(4,i)+r(5,i)+r(6,i))/6;
end

for i = 1:7
    for j = 1:6
        r(i,j) = r(i,j).^2;
    end
end

r = 100*r;

csvwrite('G:\Project CANSAT18 ...
    Final Round\MATLAB\export.csv',r)

%type('F:\Project CANSAT18 Final ...
    Round\MATLAB\export.csv')
end
r
end
```

บรรณานุกรม

[1] David Houcque. (2005). *Introduction to **MATLAB** for Engineering Students*. Northwestern University.

[2] Johnathan Mun. (2008). *Advanced Analytical Models, Understanding and Choosing the Right Probability Distributions*. New York