ความน่าจะเป็นและสัญญาณสุ่ม (Probability & Random Signal)

ตอนที่ 3 : ตัวแปรสุ่ม (2/2) (Random Variable)

ตัวแปรสุ่มแบบต่อเนื่อง (continuous random variable)

ตัวแปรสุ่มประเภทที่สองคือ**ตัวแปรสุ่มแบบต่อเนื่อง (continuous random variable) ซึ่งปืนตัวแปรที่มีค่าที่เป็นไปได้อยู่ในเซ็ตที่นับไม่ได้** ยกตัวอย่างเช่น เวลาที่รถไฟจะมาถึงสถานี หรือ ความยาวของเส้นผ่านศูนย์กลางของเพลาที่ถูกกลึง ค่าเหล่านี้ล้วนมีค่าที่เป็นไปได้ที่ต่อเนื่อง (conitnuous)

หากเรากำหนดให้ระยะเวลาที่รถไฟจะมาถึงสถานีเป็น X ถึงแม้ว่ารถไฟจะถูกจัดตารางมาให้ถึงทุกๆ 10 นาที ความน่าจะเป็นที่ระยะเวลานั้นจะเท่ากับ 10 นาที จะมีค่าเท่ากับ 0 ในทางทฤษฎี เหตุการณ์ที่ระยะเวลาจะเท่ากับ 10 นาที มีแค่ครั้งเดียวเทียบกับเหตุการณ์อื่นๆที่ระยะเวลาเท่ากับค่าอื่นเช่น 10.1, 10.01, 10.001, 10.0001, 10.00000001, 9.9, 9.99, 9.9999 **เรานั้นสามารถเขียนเหตุการณ์ที่เกิดขึ้นนั้นได้เป็นจำนวนครั้งไม่ถ้วนเนื่องจากค่าของตัวแปรเป็นค่าตอเนื่อง** ดังนั้น ความน่าจะเป็นของเหตุการณ์ที่ตัวแปรสุ่มมีค่าเท่ากับค่าคงที่ค่าหนึ่งนั้นจะเป็น 0 เสมอ

$$\mathrm{P}(X = 0) = 0$$

หรือกล่าวได้อีกอย่างคือ **เราไม่ควรใช้สมการในการอธิบายเหตุการณ์สำหรับตัวแปรสุ่มแบบต่อเนื่อง**

ตัวอย่างที่ 1 : การกลึงเพลา

โรงงานมีเครื่องกลึงอยู่ 2 เครื่อง ซึ่งแต่ละเครื่องถูกตั้งค่าไว้สำหรับกลึงเพลาให้มีเส้นผ่านศูนย์กลาง ขนาด $4\ [\mathrm{cm}]$

กำหนดให้ D_A และ D_B เป็นขนาดของเส้นผ่านศูนย์กลางของเพลาที่สร้างโดยเครื่องกลึง A และเครื่องกลึง B ตามลำดับ

หากเรากำหนดความเผื่อ (tolerance) ในการผลิตไว้ที่ $0.01\ [\mathrm{cm}]$ เราสามารถเปรียบเทียบสมรรถนะของเครื่องกลึงแต่ละเครื่องได้ดังนี้

$$\mathrm{P}(|D_A - 4| \leq 0.01) = 0.6827$$
$$\mathrm{P}(|D_B - 4| \leq 0.01) = 0.9545$$

จากตัวอย่างที่กำหนด ทำให้เราสรุปได้ว่า เครื่องกลึง B มีสมรรถนะที่ดีกว่า (จาก หนึ่งร้อยชิ้นที่ผลิต เพลาจากเครื่องกลึง A จำนวน 68 ชิ้น ผ่านตามเกณฑ์ที่กำหนด แต่เพลาจากเครื่องกลึง B จำนวน 95 ชิ้นผ่านตามเกณฑ์ที่กำหนด)

การแจกแจงของความน่าจะเป็น (probability distribution)

เหตุการณ์ที่เกี่ยวข้องกับตัวแปรสุ่มแบบต่อเนื่องจะถูกอธิบายโดยใช้อสมการ หรือ ช่วงที่เป็นไปได้ของค่า เช่น

$$\mathrm{P}(X \leq 1)\ ,\ \mathrm{P}(-0.3 \leq X \leq 0.1)\ ,\ \mathrm{P}(|x - 2.5| \leq 0.1)$$

เช่นดียวตัวแปรสุ่มแบบไม่ต่อเนื่อง การแจกแจงของความน่าจะเป็นของตัวแปรต่อเนื่องสามารถเขียนอยู่ในรูปของฟักง์ชันได้ ซึ่งต้องมีความสัมพันธ์ดังต่อไปนี้

$$\int_{-\infty}^{\infty} \{f_X(x)\} \mathrm{d}\, x = 1$$
$$0 \leq f_X(x) \leq 1, \quad 0\ \forall x$$

เรากำหนดให้ความสัมพันธ์ระหว่าง ความน่าจะเป็น และการแจกแจงของความน่าจะเป็นดังต่อไปนี้

$$\mathrm{P}(X \leq k) = \int_{-\infty}^{k} \{f_X(x)\} \mathrm{d}\, x$$

ความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าน้อยกว่าหรือเท่ากับค่า k มีค่าเท่ากับการหาพื้นที่ใต้กราฟของฟังก์ชันการแจกแจงความน่าจะเป็นจากลบอนันต์ถึงค่า k

จากความสัมพันธ์ดังกล่าว เราสามารถคำนวณหาความน่าจะเป็นของเหตุการณ์ที่ถูกเขียนในรูปของอสมการดังต่อไปนี้

$$\begin{aligned}
\mathrm{P}(X \geq k) &= 1 - \mathrm{P}(X \leq k) = 1 - \int_{-\infty}^{k} \{f_X(x)\}\mathrm{d}\,x \\
&= \int_{-\infty}^{\infty} \{f_X(x)\}\mathrm{d}\,x - \int_{-\infty}^{k} \{f_X(x)\}\mathrm{d}\,x \\
&= \int_{k}^{\infty} \{f_X(x)\}\mathrm{d}\,x
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}(|X-\mu| \leq k) = \mathrm{P}(\mu - k \leq X \leq \mu + k) = 1 - \mathrm{P}(X \leq \mu - k) - \mathrm{P}(X \geq \mu + k) \\
&= \int_{-\infty}^{\infty} \{f_X(x)\}\mathrm{d}\,x - \int_{-\infty}^{\mu-k} \{f_X(x)\}\mathrm{d}\,x - \int_{\mu+k}^{\infty} \{f_X(x)\}\mathrm{d}\,x \\
&= \int_{\mu-k}^{\mu+k} \{f_X(x)\}\mathrm{d}\,x
\end{aligned}$$

ดังนั้น เราสามารถอธิบายการแจกแจงโดยใช้ฟังก์ชันได้ตราบใดที่ฟังก์ชันนั้นมีคุณสมบัติต่อไปนี้

$$\begin{aligned}
&\int_{-\infty}^{\infty} \{f_X(x)\}\mathrm{d}\,x = 1 \\
&f_X(x) \geq 0 \;\; \forall x
\end{aligned}$$

การแจกแจงแบบยูนิฟอร์ม (Uniform Distribution)

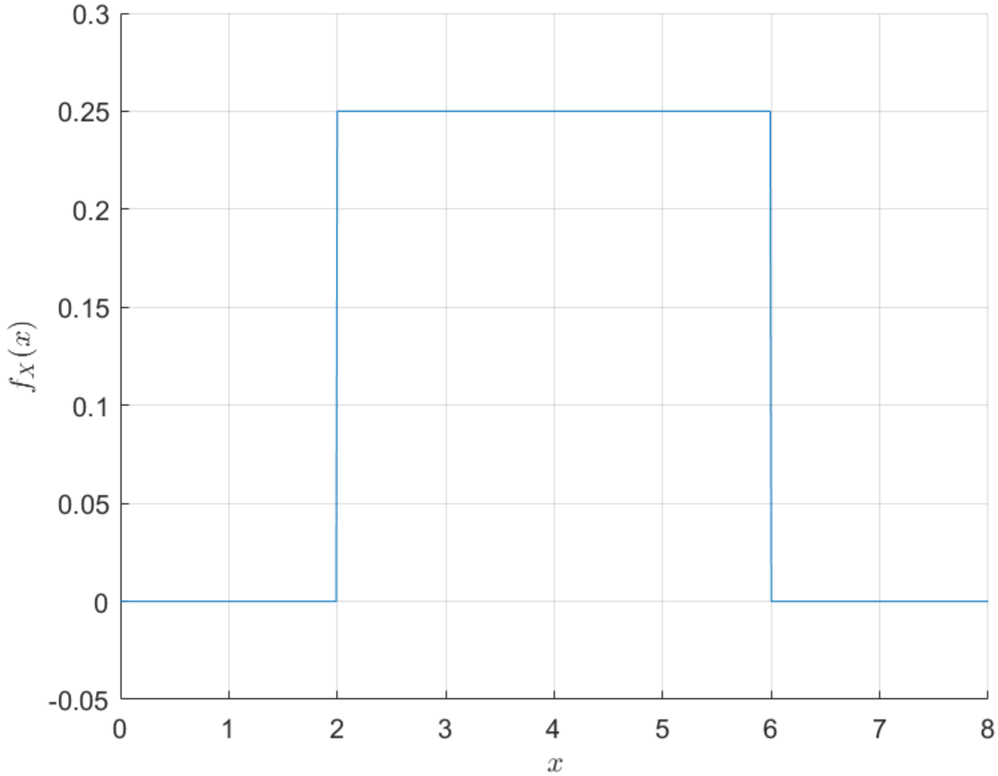
ในกรณีที่เหตุการณ์นั้นเกิดขึ้นอยู่ในช่วงที่จำกัด

$$X \in [x_{\min}, x_{\max}]$$

และมีการแจกแจงความน่าจะเป็นอย่างเท่าๆกันทุกเหตุการณ์ เราสามารถเขียนฟังก์ชันการแจกแจงความน่าจะเป็นได้ดังต่อไปนี้

$$f_X(x) = \frac{\theta(x - x_{\min}) - \theta(x - x_{\max})}{x_{\max} - x_{\min}}$$

```
clf
ax = axes;
hold(ax,'on')
grid(ax,'on')
x_max = 6;
x_min = 2;
x = 0:0.01:8;
plot(ax,x,((x>=x_min)-(x>=x_max))/(x_max-x_min))
xlabel('$x$','Interpreter',"latex");
ylabel('$f_X(x)$','Interpreter',"latex")
axis(ax,[0 8 -0.05 0.3])
```



เราสามารถนำแนวคิดนี้ไปเขียนเป็นโปรแกรมที่ใช้ในการสุ่มตัวเลขที่เป็นค่าต่อเนื่องแต่อยู่ในช่วงได้ดังต่อไปนี้

```
x_max = 6;
x_min = 2;
sz = [1,100];
X = (x_max-x_min)*rand(sz)+x_min;
```

การสุ่มแบบดังกล่าวเหมาะสำหรับตัวแปรสุ่มที่เราไม่ทราบค่าแน่ชัดแต่เราสามารถหาช่วงของค่านั้นได้และทุกค่าที่อยู่ในช่วงมีโอกาสที่จะเป็นได้เท่าๆกัน

การแจกแจงแบบปรกติ (Normal (Gaussian) Distribution)

ในหลายครั้ง การแจกจางของความน่าจะเป็นไม่ได้มีค่าเท่ากันหมดในทั้งช่วง การแจกแจงของความน่าจะเป็นนั้นอาจจะเกาะกลุ่มใกล้ค่าค่าหนึ่งมากกว่าค่าอื่นๆในช่วง **หนึ่งในการแจกแจงความน่าจะเป็นที่พบเห็นบ่อยในการิวเคราะห์เชิงสถิติและงานวิจัยได้แก่การแจกแจงแบบปรกติ** หรือ การแจกแจงแบบเกาส์เซียน (normal/Guassian distribution) ซึ่งสามารถเขียนเป็นฟังก์ชันได้ดังต่อไปนี้

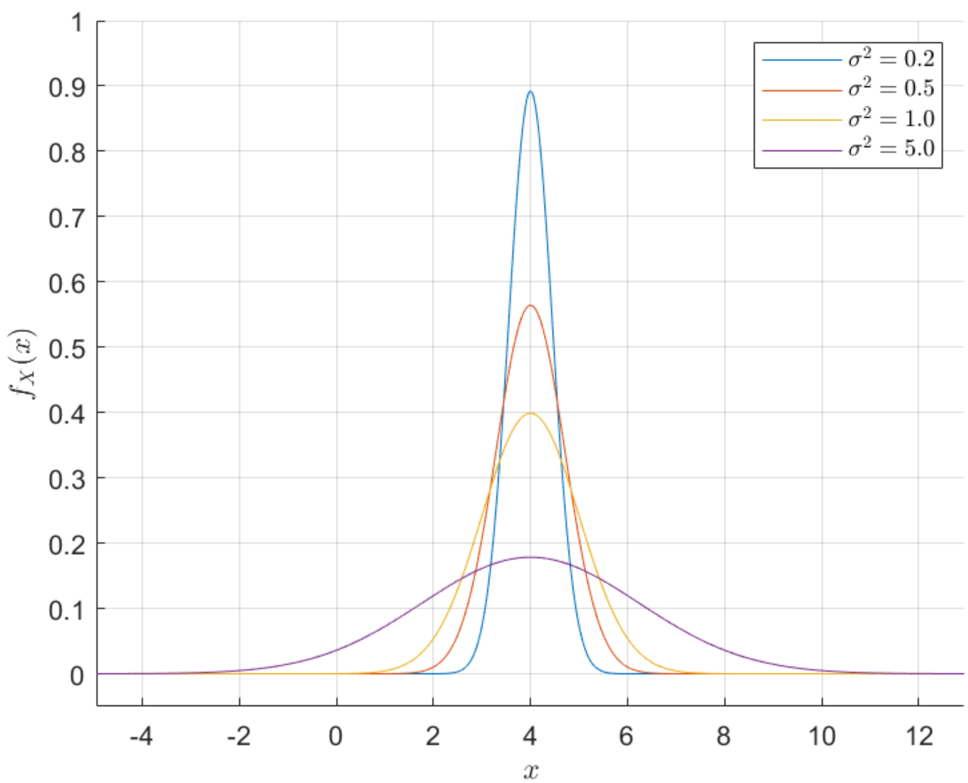
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}$$

โดยที่

μ_x และ σ_X เป็นพารามิเตอร์ของฟังก์ชันนี้ (ซึ่งมีบทบาทเกี่ยวข้องกับค่าคาดหวังและค่าแปรปรวน)

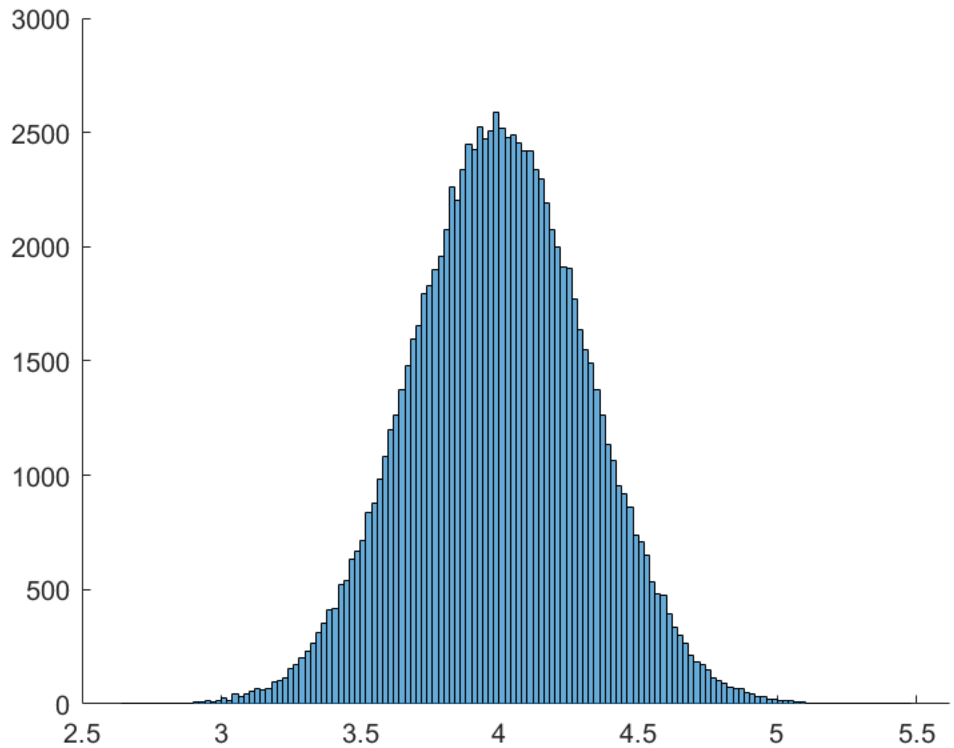
```
clf
ax = axes;
hold(ax,'on')
```

```
grid(ax,'on')
mu = 4;
sigma_sqr = [0.2 0.5 1 5];
sigma = sqrt(sigma_sqr);
lab = cell(1,numel(sigma));
 x = (-4*max(sigma):0.01:4*max(sigma))+mu;
for i = 1:numel(sigma)
    s = sigma(i);
    plot(ax,x,exp(-0.5*((x-mu)/s).^2)/sqrt(2*pi*s^2))
    xlabel('$x$','Interpreter',"latex");
    ylabel('$f_X(x)$','Interpreter',"latex")
    axis(ax,[min(x) max(x) -0.05 1]);
    lab{i} = sprintf('$\\sigma^2 = %.1f$',sigma_sqr(i));
end
legend(lab,'Interpreter',"latex")
```



ค่า σ_X ของฟังก์ชันบ่งบอกถึงความแปรปรวนของค่า X ในบริเวณรอบค่า μ_X ยิ่งค่าของ σ_X มีค่าน้อย ความน่าจะเป็นที่ X อยู่บริเวณรอบๆ μ_X จะยิ่งมีมากขึ้น **การแจกแจงความน่าจะเป็นแบบดังกล่าวมีประโยชน์ถ้าเราทราบถึงค่าคาดหวังและค่าแปรปรวนของตัวแปรสุ่ม ซึ่งเราสามารถนำมาเขียนเป็นโปรแกรมในการสุ่มได้ดังนี้**

```
sigma_sqr = 0.1;
mu = 4;
X = sqrt(sigma_sqr)*randn([1,100000])+mu;
clf;
ax = axes;
hold(ax,'on')
histogram(ax,X)
```



การแจกแจงของความน่าจะเป็นแบบปรกติเป็นการแจกแจงที่ได้รับความนิยม เนื่องจากมีความเกี่ยวข้องกับ ทฤษฎี Central Limit Theorem และพฤติกรรมหรือพลวัตต่างๆ ตามธรรมชาติล้วนแต่ให้ผลลัพธ์ทางสถิติที่มีลักษณะเป็นระฆังคว่ำหรือการแจกแจงแบบปรกตินั่นเอง

การแจกแจงแบบปรกติด้วยล็อก (Log-Normal Distribution)

สมมุติว่าเรากำหนดอัตราเร็วเชิงเส้น (linear speed) ของหุ่นยนต์ที่ค่า V แต่ในโลกความจริง อัตราเร็วของหุ่นยนต์นั้นมีความไม่แน่นอนอยู่ เราสามารถจำลองความน่าจะเป็นของอัตราเร็วด้วยการเลือกลักษณะการแจกแจง อย่างไรก็ตาม ค่าของอัตราเร็วนั้นมีค่าเป็นที่มากกว่าหรือเท่ากับ 0 ดังนั้นเราไม่สามารถใช้การแจกแจงแบบปรกติได้

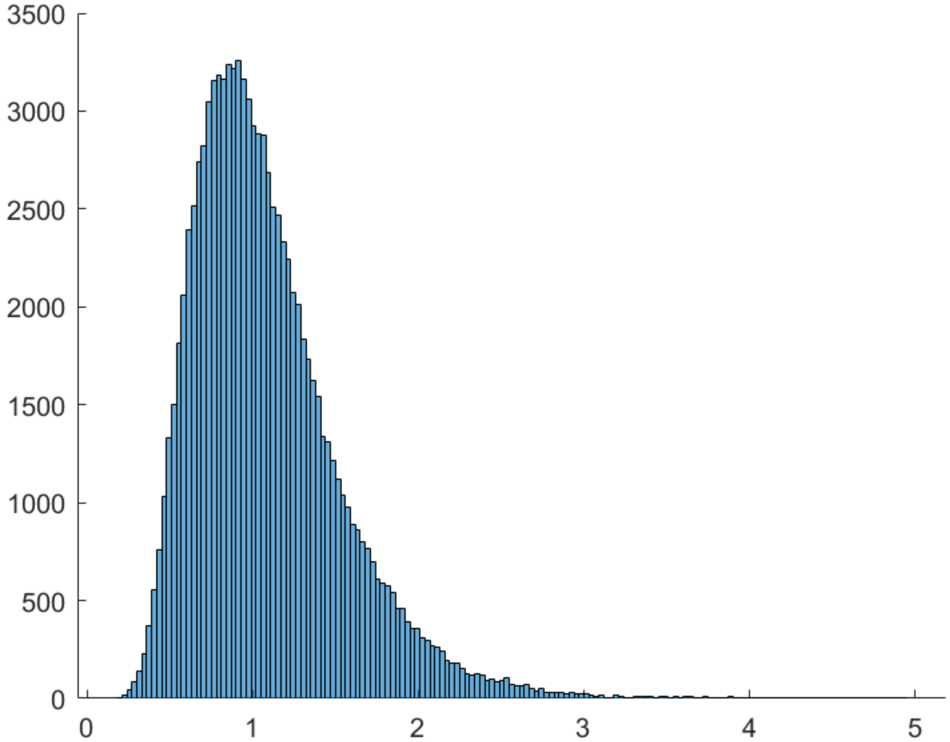
แต่ถ้าเรามาความเร็วเป็นค่าที่โดนยกกำลังดังต่อไปนี้

$$V = e^W$$

เราสามารถให้ค่าของเลขยกกำลัง W มีการแจกแจงแบบปรกติ ผลลัพทธ์ที่ได้คือความเร็วที่มีการกระจายตัวแบบปรกตด้วยล็อก (log-normal distribution) ซึ่งมีฟังก์ชันการแจกแจงของความน่าจะเป็นในรูปแบบดังต่อไปนี้

$$f_V(v) = \frac{1}{v \cdot \sigma_V \sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln(v) - \mu_V}{\sigma_V}\right)^2}$$

```
mu = log(1);
sigma = 0.4;
X = lognrnd(mu,sigma,[1,100000]);
clf;
ax = axes;
hold(ax,'on')
histogram(ax,X)
```


3500
3000
2500
2000
1500
1000
500
0
0
1
2
3
4
5