

ความน่าจะเป็นและสัญญาณสุ่ม (Probability & Random Signal)

ตอนที่ 2 : ตัวแปรสุ่ม (1/2) (Random Variable)

จะอย่างไรก็ตามการอธิบายความไม่แน่นอนในลักษณะบวกลบนี้มีข้อเสียในเรื่องของการหาแนวโน้มและความสัมพันธ์เนื่องจากการระบุความไม่แน่นอนแบบบวกลบไม่ได้ให้ข้อมูลการแจกแจงของค่ามาแต่อย่างใด ถึงแม้ค่าจะมีช่วงที่กำหนดมา แต่ค่าเฉลี่ยจริงนั้นอาจจะไม่ได้อยู่ที่กึ่งกลางของช่วงก็เป็นได้ อีกทั้งความแปรปรวนของค่าอาจจะมีมากหรือน้อยที่แตกต่างกันก็เป็นได้ ดังนั้น ในหลายครั้ง เราจะใช้ความน่าจะเป็น (probability) ในการอธิบายค่าที่เป็นไปได้ของตัวแปร

ตัวแปรที่มีความไม่แน่นอนจะถูกเรียกว่า ตัวแปรสุ่ม (random variable) ซึ่ง เราสามารถอธิบายค่าที่เป็นไปได้ของตัวแปรสุ่มนี้ผ่านความน่าจะเป็น (probability) ซึ่งตัวแปรนี้สามารถมีค่าเป็นได้ 2 ประเภทหลัก

ตัวแปรสุ่มแบบไม่ต่อเนื่อง (discrete random variable)

ประเภทแรกคือตัวแปรสุ่มแบบไม่ต่อเนื่อง (discrete random variable) ซึ่งเป็นตัวแปรที่มีค่าที่เป็นไปได้อยู่ในเซตที่นับได้ ยกตัวอย่างการโยนเหรียญหัวก้อย หากเรากำหนดให้ X เป็นตัวแปรสุ่มที่บ่งบอกว่าเหรียญเป็นหัวหรือเป็นก้อย โดยกำหนดให้ เมื่อ $X = 1$ เป็นเหตุการณ์ที่เหรียญออกเป็นหัว ส่วน $X = 0$ เป็นเหตุการณ์ที่เหรียญออกเป็นก้อย เราจะสามารถเขียนความน่าจะเป็น $P(\cdot)$ ได้ดังต่อไปนี้

$$\begin{aligned}P(X = 1) &= 0.5 \\P(X = 0) &= 0.5\end{aligned}$$

สมการการทั้งสองอธิบายถึงความน่าจะเป็นที่เหรียญจะออกหัวหรือก้อยนั้นเป็น 0.5 ทั้งคู่ (50%)

ความน่าจะเป็นนั้นจะเอาไว้ใช้อธิบายเหตุการณ์ (event) เสมอ ในการอธิบายอย่างเป็นทางการ เราจำเป็นต้องแปลความหมายของเหตุการณ์ให้อยู่ในรูปของสมการหรือสมการ ยกตัวอย่างเช่นการทอยลูกเต๋าด้าน หากต้องการหาความน่าจะเป็นที่ลูกเต๋าดูออกด้านที่มีค่าน้อยกว่า 4 เราต้องกำหนดให้ X เป็นค่าของด้านลูกเต๋าดูออก และเหตุการณ์ที่ลูกเต๋าดูออกด้านที่มีค่าน้อยกว่า 4 จะสามารถถูกเขียนเป็นสมการได้คือ $X < 4$ ดังนั้น ความน่าจะเป็นสามารถเขียนได้เป็นสมการต่อไปนี้

$$P(X < 4) = 0.5$$

เราจะไม่เขียนว่า $P(X)$ เนื่องจาก X ไม่ใช่เหตุการณ์

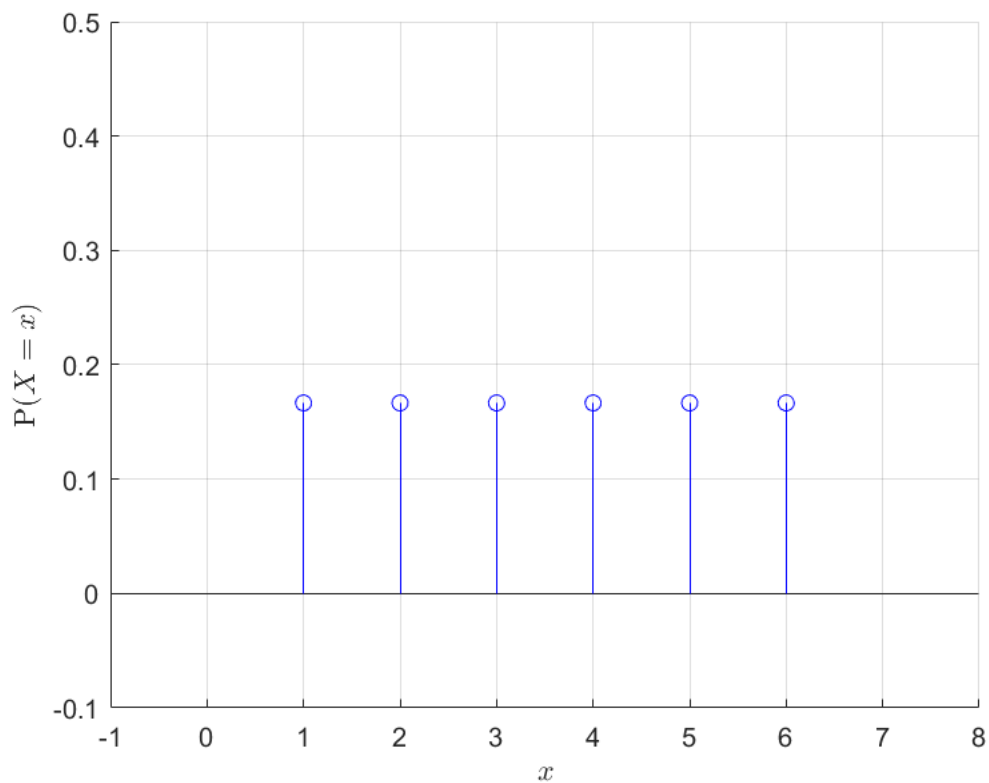
ตัวอย่างที่ 1 : การทอยลูกเต๋าด้าน

กำหนดให้ลูกเต๋าด้านมีความน่าจะเป็นที่เท่ากันที่จะออกด้านใดด้านหนึ่ง ดังนั้น เราสามารถเขียนความสัมพันธ์ของความน่าจะเป็นได้ดังต่อไปนี้

$$P(X = x) = \frac{1}{6} \quad \text{โดยที่ } x \in \{1, 2, 3, 4, 5, 6\}$$

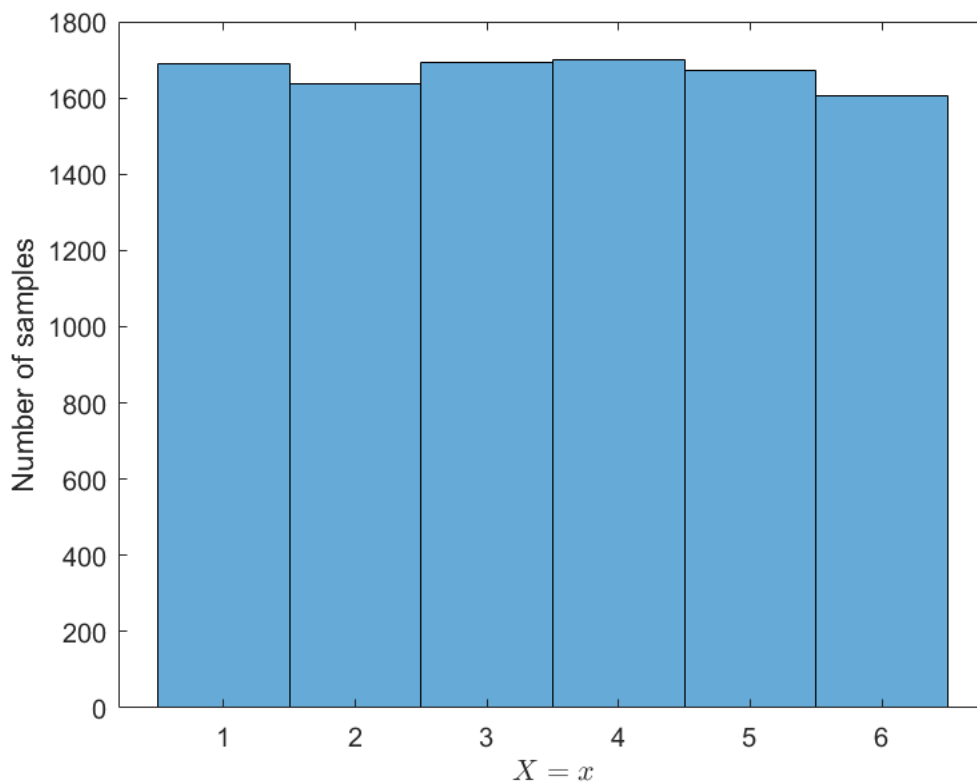
เราสามารถกราฟความสัมพันธ์ระหว่างค่าของ X และความน่าจะเป็นได้ดังต่อไปนี้

```
clf;
ax = axes;
grid(ax, 'on')
hold(ax, 'on')
x = 1:6;
stem(ax, x, (1/6)*ones(size(x)), 'b')
axis(ax, [-1 8 -0.1 0.5])
xlabel('$x$', 'Interpreter', 'latex')
ylabel('$\mathrm{P}(X=x)$', 'Interpreter', 'latex')
```



การแจกแจงที่ความน่าจะเป็นของทุกเหตุการณ์เท่ากันถูกเรียกว่า การแจกแจงแบบยูนิฟอร์ม (**uniform distribution**) ซึ่งเราสามารถเขียนเป็นโปรแกรมสุ่มโดยใช้ฟังก์ชัน **randi** ได้ดังต่อไปนี้

```
imax = 6;
imin = 1;
sz = [1 10000]; % one row 100 columns
X = randi(imax-imin+1,sz)+imin-1;
clf
ax = axes;
histogram(ax,X)
xlabel('$X=x$', 'Interpreter', "latex")
ylabel('Number of samples')
```



หากโจทย์เปลี่ยนเป็นการทอยลูกเต๋า 2 อันและนับผลรวมจากแต้มของหน้าลูกเต๋าทิ้งสอง เราจะได้ความน่าจะเป็นของแต่ละผลรวมดังต่อไปนี้

$$P(X = 2) = P(X = 12) = \frac{1}{36}$$

$$P(X = 3) = P(X = 11) = \frac{2}{36}$$

$$P(X = 4) = P(X = 10) = \frac{3}{36}$$

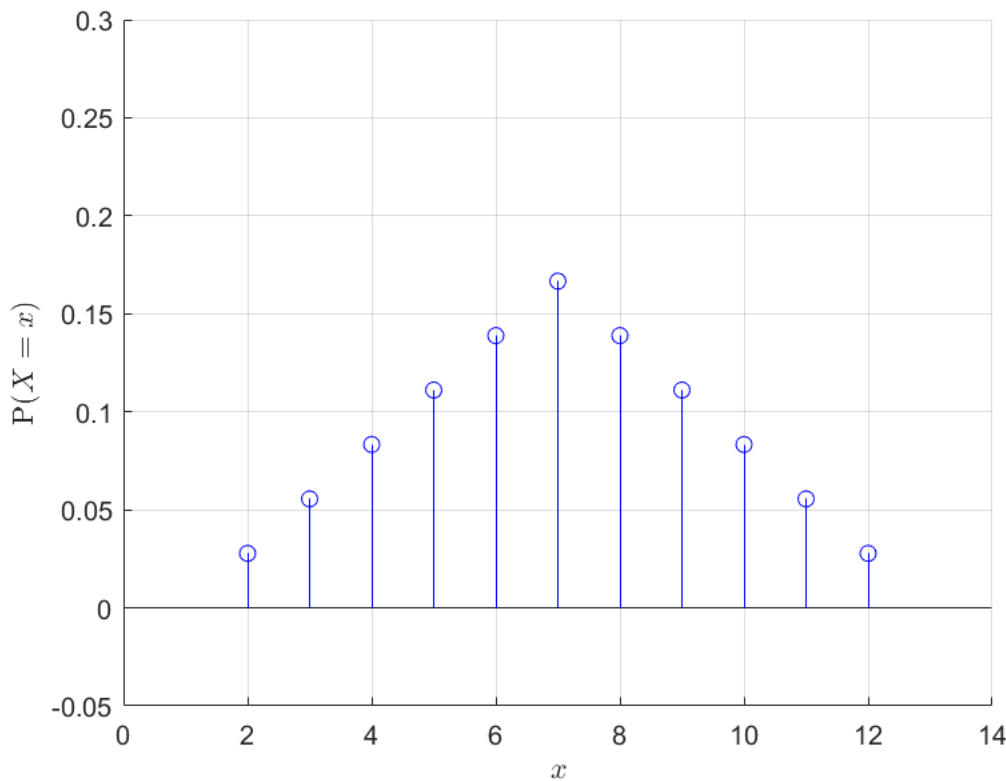
$$P(X = 5) = P(X = 9) = \frac{4}{36}$$

$$P(X = 6) = P(X = 8) = \frac{5}{36}$$

$$P(X = 7) = \frac{6}{36}$$

โดยที่ X คือผลรวมของหน้าลูกเต๋าทิ้งออก ซึ่งสามารถกราฟความสัมพันธ์ได้ดังต่อไปนี้

```
clf;
ax = axes;
grid(ax, 'on')
hold(ax, 'on')
x = 2:12;
stem(ax, x, (6-abs(x-7))/36, 'b')
axis(ax, [0 14 -0.05 0.3])
xlabel('$x$', 'Interpreter', "latex")
ylabel('$\mathrm{P}(X=x)$', 'Interpreter', "latex")
```



ในบ้อยครั้ง เหตุการณ์อาจจะมีจำนวนมากซึ่งการเขียนสมการเป็นทีละเหตุการณ์อาจไม่ใช่วิธีที่สะดวกที่สุด เราสามารถใช้ฟังก์ชันในการอธิบายการแจกแจงของความน่าจะเป็น หรือ ฟังก์ชันการแจกแจงความน่าจะเป็น $f_X(\cdot)$ (probability density function/ pdf) สำหรับตัวอย่างการทอยลูกเต๋าสองอัน เราสามารถเขียนฟังก์ชันการแจกแจงความน่าจะเป็นได้ดังต่อไปนี้

$$f_X(x) = (x-1) \cdot \theta(x-2) - 2(x-7) \cdot \theta(x-7) + (x-13) \cdot \theta(x-13)$$

โดยที่ x คือค่าของผลรวมหน้าลูกเต๋าคู่ (จำนวนเต็ม)

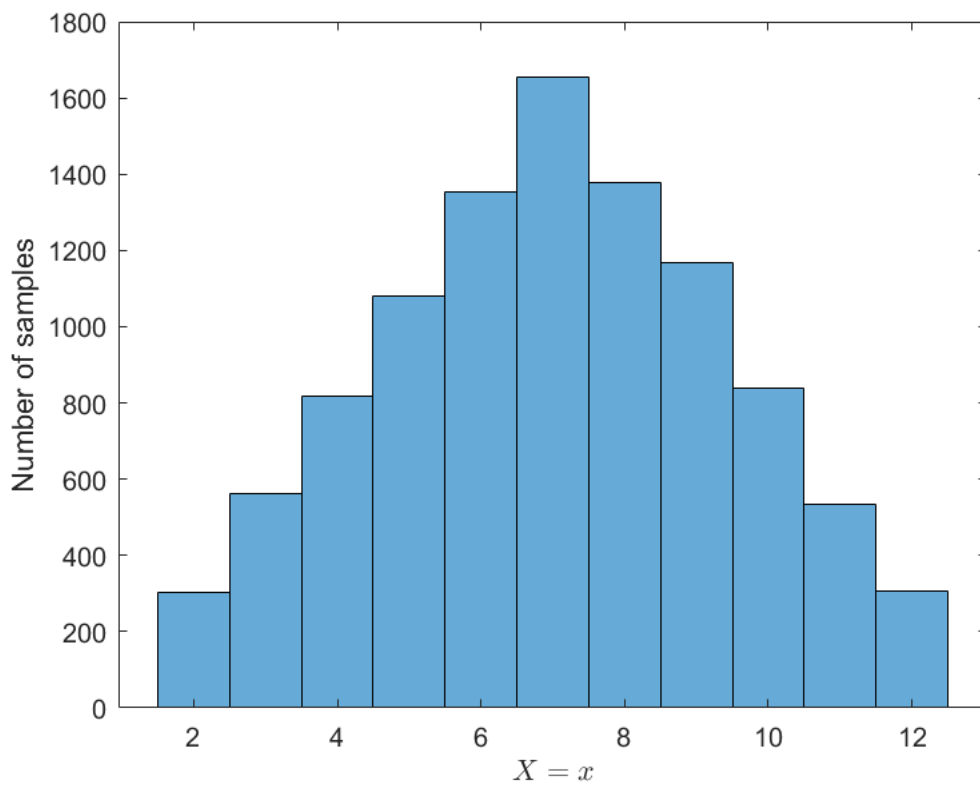
ไม่ว่าความน่าจะเป็นจะมีการแจกแจงอย่างไร เราสามารถเขียนความน่าจะเป็นของแต่ละจำนวนเต็มให้อยู่ในรูปของเวกเตอร์ที่มีจำนวนจำกัดได้ โดยการกำหนดตัวแปรที่มีจำนวนจำกัดและใช้ผ่านฟังก์ชันเช่น

$$f_X(x) = \frac{6 - |x-7|}{36} \text{ โดยที่ } x \in \{2, 3, \dots, 12\}$$

ถ้าหากเราเขียนความน่าจะเป็นในรูปแบบที่จำกัดได้ เราก็สามารถเขียนโปรแกรมในการสุ่มตามการแจกแจงใดๆได้ ซึ่งเราจะใช้ฟังก์ชัน **randsample** (จำเป็นต้องใช้ **Statistics & Machine Learning Toolbox**) ได้ดังต่อไปนี้

```
x = 2:12;
pdf = (6-abs(x-7))/36;
sz = 10000;
X = randsample(x,sz,true,pdf);

clf
ax = axes;
histogram(ax,X)
xlabel('$X=x$', 'Interpreter', 'latex')
ylabel('Number of samples')
```



หนึ่งในคุณสมบัติของฟังก์ชันการแจกแจงความน่าจะเป็นสำหรับตัวแปรที่ไม่ต่อเนื่องคือผลรวมของความน่าจะเป็นของทุกเหตุการณ์ต้องเท่ากับ **1** หรือเขียนเป็นสมการได้ดังต่อไปนี้

$$\sum_x f_X(x) = 1$$

อีกหนึ่งในคุณสมบัติของฟังก์ชันการแจกแจงคือฟังก์ชันต้องไม่เป็นลบ

$$0 \leq f_X(x) \leq 1$$