# ทฤษฎีที่อ้างอิงกับกลุ่มตัวอย่างขนาดใหญ่ (Large-Sample Theories)

รศ.ดร. วีระชาติ กิเลนทอง
มหาวิทยาลัยหอการค้าไทย

# บทนำ

บทนี้นำเสนอทฤษฎีบทที่เป็นรากฐานที่สำคัญในทางสถิติสองอัน

- กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers)
- ทฤษฎีบทลิมิตของค่ากลาง (Central Limit Theorem)

# กฎว่าด้วยจำนวนมาก (Law of Large Numbers)

- กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เป็นทฤษฎีบทที่สำคัญทางสถิติที่ทำให้**สามารถประมาณค่า (estimate) ค่าคาดหมาย (mean) ของตัวแปรที่สนใจ (ซึ่งเป็นค่าทางทฤษฎี) ด้วย ค่าเฉลี่ย (average)**
- กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เกี่ยวข้องกับการลู่เข้าเชิงความน่าจะเป็น (converge in probability)

# นิยาม: การลู่เข้าเชิงความน่าจะเป็น (converge in probability)

**Definition**

อันดับของตัวแปรสุ่ม $X_1, X_2, \ldots$ ลู่เข้าสู่ตัวแปรสุ่ม $X$ ในเชิงความน่าจะเป็น (converge to $X$ in probability) ถ้า สำหรับทุกๆ ค่า $\varepsilon > 0$

$$\lim_{n\to\infty} Pr\left(|X_n - X| < \varepsilon\right) = 1 \Leftrightarrow \lim_{n\to\infty} Pr\left(|X_n - X| \geq \varepsilon\right) = 0 \qquad (1)$$

ซึ่งอาจจะเขียนแทนด้วย $X_n \xrightarrow{p} X$ หรือ $plim\ X_n = X$

- การลู่เข้าเชิงความน่าจะเป็น (converge in probability) คือ **ความน่าจะเป็นที่** $X_n$ **จะมีค่าใกล้เคียงกับ** $c$ **นั้นมีค่าเข้าใกล้หนึ่งเมื่อ** $n$ **ลู่เข้าสู่อนันต์** หรือ เมื่อ $n$ มีค่าสูงมากๆ แล้ว โอกาสที่ค่าของตัวแปรสุ่ม $X_n$ จะแตกต่างจาก $c$ นั้นจะมีค่าเข้าใกล้ศูนย์มากๆ
- เหตุผลที่สำคัญที่ทำให้ต้องพิจารณาในรูปแบบของการลู่เข้าเชิงความน่าจะเป็น (converge in probability) เพราะว่า**ในโลกของความน่าจะเป็นหรือความไม่แน่นอน ไม่สามารถระบุค่าอะไรได้อย่างแน่นอน**
- เหตุการณ์อื่นๆ ที่มีความน่าจะเป็นลู่เข้าสู่ศูนย์ไม่ได้หมายความว่าเป็นไปไม่ได้

# ทฤษฎี: กฎว่าด้วยจำนวนมาก (Law of Large Numbers)

## Theorem

*สมมุติให้* $X_1, \ldots, X_n$ *คือกลุ่มตัวอย่างขนาด n ตัวอย่างที่สุ่มมาจากการแจกแจงที่มีค่าคาดหมาย (mean)* $E[X] = \mu$ *และค่าความแปรปรวน (variance) มีค่าจำกัด และกำหนดให้* $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ *คือค่าคาดหมายจากตัวอย่าง (sample mean) แล้ว*

$$\overline{X}_n \xrightarrow{p} E[X] = \mu \tag{2}$$

## Proof.

- จากทฤษฎีที่ว่า $E\left[\overline{X}_n\right] = \mu$ และ $Var\left[\overline{X}_n\right] = \frac{\sigma^2}{n}$ สามารถประยุกต์ **อสมการเชบิเชฟ (Chebyshev Inequality)** $\left(Pr\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2}\right)$ เพื่อแสดงว่า สำหรับทุกๆ ค่าจำนวนจริง $\varepsilon > 0$

$$Pr\left(\left|\overline{X}_n - \mu\right| \leq \varepsilon\right) = 1 - Pr\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \geq 1 - \frac{\frac{\sigma^2}{n}}{\varepsilon^2}$$

- ซึ่งลู่เข้าสู่หนึ่งเมื่อ $n$ ลู่เข้าสู่อนันต์แล้ว $\lim_{n\to\infty} Pr\left(\left|\overline{X}_n - \mu\right| \leq \varepsilon\right) \geq \lim_{n\to\infty} 1 - \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = 1$
- เนื่องจากความน่าจะเป็นมีค่าไม่เกินหนึ่ง ดังนั้น จึงสามารถสรุปได้ว่า สำหรับทุกๆ ค่าจำนวนจริง $\varepsilon > 0$ แล้ว

$$\lim_{n\to\infty} Pr\left(\left|\overline{X}_n - \mu\right| \leq \varepsilon\right) = 1$$

# การประยุกต์ใช้กฎว่าด้วยจำนวนมาก (Law of Large Numbers)

- กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ระบุว่า " **ค่าคาดหมายจากตัวอย่าง (sample mean)** $\overline{X}_n$ **จะมีค่าใกล้เคียงกับค่าคาดหมาย (mean)** $E[X] = \mu$ **เมื่อกลุ่มตัวอย่างมีขนาดใหญ่**"
- หมายความว่า หากมี**ข้อมูลขนาดใหญ่มากพอ**แล้ว ค่าเฉลี่ยเลขคณิต (arithmetic average) $\bar{x}_n$ ที่คำนวณจากตัวอย่างสุ่ม (random sample) จะ**มีค่าใกล้เคียงกับ** ค่าคาดหมาย (mean) $E[X] = \mu$
- การประมาณค่าคาดหมาย (mean) ของตัวแปรต่างๆ นิยมใช้ค่าเฉลี่ยเลขคณิต (arithmetic average) ของตัวแปรนั้น
- การประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ที่น่าสนใจอันหนึ่ง คือ การสร้างกราฟแสดงความถี่ (histogram)

# นิยาม: กราฟแสดงความถี่ (histogram)

## Definition

กำหนดให้ $x_1, \ldots, x_n$ คือค่าจำนวนจริงที่ $a < x_i < b$ สำหรับทุก $i$ และ $a, b$ คือค่าคงที่ เราสามารถสร้างกราฟแสดงความถี่ (histogram) จากเซ็ตของจำนวนจริงดังกล่าวได้โดยเลือกจำนวนเต็ม $k > 1$ และแบ่งช่วง $[a, b]$ เป็น $k$ ช่วงที่มีความกว้างเท่ากันที่ $\frac{b-a}{k}$ หลังจากนั้น ให้นับดูว่ามีกี่จำนวนจากเซ็ตดังกล่าวที่มีค่าตกอยู่ในช่วงย่อย (subinterval) แต่ละช่วง $j = 1, \ldots, k$ และแทนจำนวนที่นับได้ด้วย $c_j$ ขั้นตอนต่อไปคือการเลือกรูปแบบของกราฟแสดงความถี่ (histogram) ซึ่งทำได้ด้วยการเลือกจำนวนจริง $r$ (โดยทั่วไปใช้ $r = n$ ) เพื่อกำหนดระดับความสูงของแต่ละแท่งตามสมการ

$$\text{ระดับความสูงของแท่ง}_j = \frac{c_j}{n} \tag{3}$$

- ระดับความสูงเท่ากับสัดส่วนของจำนวนที่ตกอยู่ในแต่ละช่วง ซึ่งเป็นกรณีที่ได้รับความนิยมมากที่สุดเพราะมีความสัมพันธ์กับฟังก์ชันความหนาแน่นความน่าจะเป็น (p.d.f.)
- ทฤษฎีบทต่อไปนี้ ประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เพื่อบอกว่า **ในกรณีที่ $r = n$ นั้นลู่เข้าสู่ค่าความน่าจะเป็นที่จะสุ่มได้จำนวนที่มีค่าอยู่ในช่วงนั้น**

# ทฤษฎี: กราฟแสดงความถี่ (histogram)

## Theorem

*กำหนดให้* $X_1, X_2, \ldots$ *เป็นอันดับของตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) และค่าคงที่* $c_1 < c_2$ *พร้อมทั้งกำหนดให้* $Y_i = 1$ *ถ้า* $c_1 \leq X_i < c_2$ *และ* $Y_i = 0$ *ถ้าไม่ใช่ แล้ว* $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ *คือสัดส่วนของ* $X_1, X_2, \ldots$ *ที่มีค่าอยู่ในช่วง* $[c_1, c_2)$ *และ*

$$\overline{Y}_n \xrightarrow{p} Pr\left(c_1 \leq X_i < c_2\right) \tag{4}$$

## Proof.

จากนิยามที่กำหนดให้ จะเห็นได้ว่า $Y_1, Y_2, \ldots$ เป็นตัวแปรสุ่มแบบเบอร์นูลี (Bernoulli) ด้วยค่าพารามิเตอร์ $p = Pr\left(c_1 \leq X_i < c_2\right)$ ที่มีการแจกแจงเหมือนกันและเป็นอิสระต่อกัน (i.i.d.) ดังนั้น สามารถประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) เพื่อสรุปว่า $\overline{Y}_n \xrightarrow{p} p$ เพราะ $p$ คือค่าคาดหมายของ $Y_i$ ในขณะเดียวกัน $p = Pr\left(c_1 \leq X_i < c_2\right)$ □

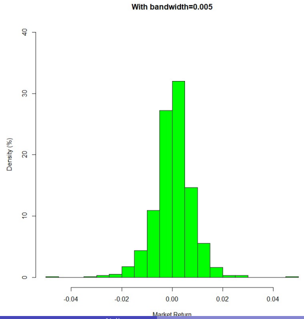
# ตัวอย่าง: การสร้างกราฟแสดงความถี่ (histogram)

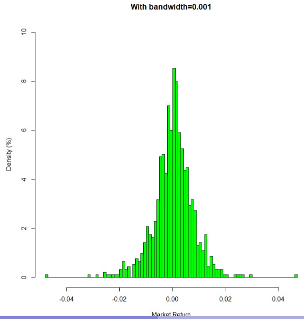
- ตัวอย่างต่อไปนี้แสดงการสร้างกราฟแสดงความถี่ (histogram) ของอัตราผลตอบแทน (returns) ของตลาดหลักทรัพย์แห่งประเทศไทย
- ประเด็นสำคัญของตัวอย่างนี้คือ บทบาทของความกว้างแถบความถี่ (bandwidth) $[c_1, c_2)$ ต่อความราบเรียบ (smoothness) ของกราฟแสดงความถี่ (histogram) โดยจะเห็นได้ว่า กราฟแสดงความถี่จะมีความราบเรียบมากขึ้นเมื่อความกว้างแถบความถี่ (bandwidth) มีค่าสูงขึ้น

# ตัวอย่าง: การสร้างกราฟแสดงความถี่ (histogram) ของอัตราผลตอบแทนต่อปี

## Example

พิจารณาข้อมูลรายวันของอัตราผลตอบแทนต่อปี (annualized returns) ของตลาดหลักทรัพย์แห่งประเทศไทยระหว่าง 1 มกราคม 2015 ถึง 30 กันยายน 2018 โดยในกราฟแรกกำหนดให้ความกว้างแถบความถี่ (bandwidth) แต่ละแถบมีความกว้างเท่าๆ กันที่ 0.005 หรือร้อยละ 0.5 และเพื่อความสะดวก กำหนดให้ค่าต่ำสุดเท่ากับ -0.05 และสูงสุดเท่ากับ 0.05 ส่วนในกราฟที่สองกำหนดให้กว้าง 0.001 หรือร้อยละ 0.1



# ตัวอย่าง: สรุปการสร้างกราฟแสดงความถี่ (histogram) ของอัตราผลตอบแทนต่อปี

## Example

- บทเรียนที่สำคัญจากตัวอย่างนี้คือ **กราฟแรกมีความราบเรียบมากกว่ากราฟที่สอง** ซึ่งสามารถอธิบายได้โดยพิจารณา
  - ตัวแปรสุ่ม $X$ มีการแจกแจงทวินาม (binomial distribution) ด้วยค่าพารามิเตอร์ $n$ และ $p$
  - ตัวแปรสุ่ม $Y$ ซึ่งมีการแจกแจงทวินาม (binomial distribution) ด้วยค่าพารามิเตอร์ $n$ และ $\frac{p}{m}$ โดยที่ $m > 1$
  - เราสามารถสร้างตัวแปรใหม่ $Z = mY$ ที่มีค่าคาดหมาย (mean) เท่ากับ $\mu_Z = np$ ซึ่งมีค่าเท่ากับค่าคาดหมายของ $X$ แต่มีค่าความแปรปรวน (variance) เท่ากับ $\sigma_Z^2 = m^2 n \frac{p}{m} \left(1 - \frac{p}{m}\right) = mnp \left(1 - \frac{p}{m}\right)$ ในขณะเดียวกัน ค่าความแปรปรวน (variance) ของ $X$ มีค่าเท่ากับ $\sigma_X^2 = np\left(1 - p\right)$
- หากพิจารณาในกรณีที่ $p$ มี**ค่าน้อยมาก** (กรณีที่กำหนดให้ความกว้างแถบความถี่ (bandwidth) $[c_1, c_2)$ มีค่าน้อยมาก) จนทำให้สามารถกำหนดให้ค่า $p^2 \approx 0$ ซึ่งมีผลทำให้สามารถแสดงได้ว่า $\sigma_X^2 \approx np$ และ $\sigma_Z^2 \approx mnp$ นั่นคือ ค่าความแปรปรวน (variance) ของ $Z$ มีค่าประมาณ $k$ เท่าของค่าความแปรปรวน (variance) ของ $X$
- ผลการคำนวณในตัวอย่างนี้แสดงให้เห็นว่า ตัวแปร $X$ และ $Z$ มีค่าคาดหมายที่**เท่ากัน** แต่มีช่วงความกว้างแถบความถี่ (bandwidth) ที่**แตกต่างกัน** โดยตัวแปร $X$ นำไปสู่กราฟแสดงความถี่ (histogram) ที่มีช่วงความกว้างแถบความถี่ (bandwidth) ที่กว้างกว่าเพราะ $m > 1$ นอกจากนี้ การที่ค่าความแปรปรวน (variance) ของ $Z$ มีค่ามากกว่าค่าความแปรปรวน (variance) ของ $X$ สะท้อนให้เห็นว่า กราฟแสดงความถี่จะมีความราบเรียบมากขึ้นเมื่อความกว้างแถบความถี่ (bandwidth) มีค่าสูงขึ้น

# การลู่เข้าเชิงความน่าจะเป็น (converge in probability)

- ทฤษฎีบทต่อไปนี้เป็นการขยายผลการลู่เข้าเชิงความน่าจะเป็น (converge in probability) สำหรับฟังก์ชันเชิงเส้นและฟังก์ชันต่อเนื่อง (continuous) ที่ลู่เข้าเชิงน่าจะเป็น ซึ่งทำให้สามารถประยุกต์ใช้กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) กับค่าสถิติที่เป็นฟังก์ชันของค่าคาดหมายจากตัวอย่าง (sample mean)

# ทฤษฎี: การลู่เข้าเชิงความน่าจะเป็น (converge in probability) สำหรับตัวอย่างขนาดใหญ่ (Law of Large Numbers)

## Theorem

*สมมุติว่า* $X_n \xrightarrow{p} X$ *และ* $Y_n \xrightarrow{p} Y$ *แล้ว* $X_n + Y_n \xrightarrow{p} X + Y$

## Proof.

- กำหนดให้ $\epsilon > 0$ ที่สอดคล้องกับอสมการต่อไปนี้

$$|\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon \tag{5}$$

- ดังนั้น อสมการสามเหลี่ยม (triangular inequality) ช่วยให้สามารถบอกได้ว่า

$$|X_n - X| + |Y_n - Y| \geq |\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon \tag{6}$$

- เราสามารถประยุกต์ใช้กับค่าความน่าจะเป็นได้ว่า

$$Pr\left(|\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon\right) \leq Pr\left(|X_n - X| + |Y_n - Y| \geq \epsilon\right) \tag{7}$$

$\square$

# การพิสูจน์: การลู่เข้าเชิงความน่าจะเป็นสำหรับตัวอย่างขนาดใหญ่

## Proof.

- เพื่อให้สามารถเขียนพจน์ด้านขวาในรูปของ $|X_n - X|$ และ $|Y_n - Y|$ จำเป็นต้องพิสูจน์ว่าเหตุการณ์

$$A = \{|X_n - X| + |Y_n - Y| \geq \epsilon\} \tag{8}$$

- กำหนดให้ $\epsilon > 0$ ที่สอดคล้องกับอสมการ $|\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon$ ดังนั้น **อสมการสามเหลี่ยม (triangular inequality)** ช่วยให้สามารถบอกได้ว่า

$$|X_n - X| + |Y_n - Y| \geq |\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon$$

- เราสามารถประยุกต์ใช้กับค่าความน่าจะเป็นได้ว่า

$$Pr\,(|\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon) \leq Pr\,(|X_n - X| + |Y_n - Y| \geq \epsilon)$$

- เพื่อให้สามารถเขียนพจน์ด้านขวาในรูปของ $|X_n - X|$ และ $|Y_n - Y|$ จำเป็นต้องพิสูจน์ว่าเหตุการณ์

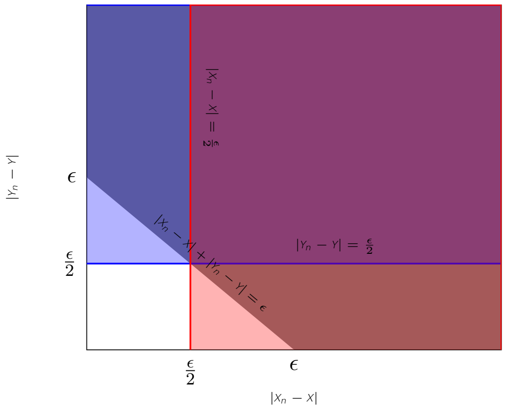
$$A = \{|X_n - X| + |Y_n - Y| \geq \epsilon\}$$

  เป็นสับเซ็ตของเหตุการณ์ $B = \left\{|X_n - X| \geq \frac{\epsilon}{2}\right\} \cup \left\{|Y_n - Y| \geq \frac{\epsilon}{2}\right\}$

- เลือก $\frac{\epsilon}{k}$ สำหรับ $X$ และ $\frac{(k-1)\epsilon}{k}$ สำหรับ $Y$ โดยที่ค่าจำนวนจริง $k > 1$ ซึ่งเห็นได้อย่างชัดเจนจากรูปถัดไป

# การพิสูจน์: รูปเพื่อใช้ในการพิสูจน์ การลู่เข้าเชิงความน่าจะเป็นสำหรับตัวอย่างขนาดใหญ่

Figure: พื้นที่ที่อยู่เหนือเส้น $|X_n - X| + |Y_n - Y| = \epsilon$ แทนเหตุการณ์ $A$ ซึ่งเห็นได้อย่างชัดเจนว่าเป็นส่วนหนึ่งของเหตุการณ์ $B$ ส่วนพื้นที่สามเหลี่ยมสองอันด้านล่างเส้น $|X_n - X| + |Y_n - Y| = \epsilon$ แทนส่วนที่อยู่ใน $B$ แต่ไม่อยู่ใน $A$ ดังนั้น จึงสามารถสรุปได้ว่า $A \subset B$



# การพิสูจน์: สรุปผลการพิสูจน์การลู่เข้าเชิงความน่าจะเป็นสำหรับตัวอย่างขนาดใหญ่

## Proof.

- ผลที่ตามมาก็คือ

$$\begin{aligned} Pr\left(|X_n - X| + |Y_n - Y| \geq \epsilon\right) &\leq Pr\left(|X_n - X| \geq \frac{\epsilon}{2}\right) + Pr\left(|Y_n - Y| \geq \frac{\epsilon}{2}\right) \\ &+ Pr\left(|X_n - X| \geq \frac{\epsilon}{2} \text{ และ } |Y_n - Y| \geq \frac{\epsilon}{2}\right) \\ &\leq Pr\left(|X_n - X| \geq \frac{\epsilon}{2}\right) + Pr\left(|Y_n - Y| \geq \frac{\epsilon}{2}\right) \end{aligned}$$

- ดังนั้น

$$Pr\left(|\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon\right) \leq Pr\left(|X_n - X| \geq \frac{\epsilon}{2}\right) + Pr\left(|Y_n - Y| \geq \frac{\epsilon}{2}\right)$$

- ในขณะเดียวกัน คุณสมบัติการลู่เข้าเชิงความน่าจะเป็นของ $X_n$ และ $Y_n$ ทำให้สรุปได้ว่า

$$\lim_{n\to\infty} Pr\left(|\,(X_n + Y_n) - (X + Y)\,| \geq \epsilon\right) \leq \lim_{n\to\infty} Pr\left(|X_n - X| \geq \frac{\epsilon}{2}\right) + \lim_{n\to\infty} Pr\left(|Y_n - Y| \geq \frac{\epsilon}{2}\right) = 0$$

- หมายความว่า $X_n + Y_n \xrightarrow{p} X + Y$

# ทฤษฎี: การลู่เข้าเชิงความน่าจะเป็น (converge in probability) เมื่อมีค่าคงที่คูณกับตัวแปรสุ่ม

## Theorem

*สมมุติว่า* $X_n \xrightarrow{p} X$ *และ* $a$ *เป็นค่าคงที่ แล้ว* $aX_n \xrightarrow{p} aX$

## Proof.

ถ้า $a = 0$ ค่าลิมิตต้องเท่ากับศูนย์อย่างแน่นอน ดังนั้น จำเป็นต้องพิสูจน์เพิ่มเติมเฉพาะในกรณีที่ $a \neq 0$ โดยเริ่มจากการกำหนดให้ $\epsilon > 0$ ดังนั้น

$$\lim_{n\to\infty} Pr\left(|aX_n - aX| \geq \epsilon\right) = \lim_{n\to\infty} Pr\left(|a|X_n - X| \geq \epsilon\right) = \lim_{n\to\infty} Pr\left(|X_n - X| \geq \frac{\epsilon}{|a|}\right) = 0$$

□

# ทฤษฎี: การลู่เข้าเชิงความน่าจะเป็น (converge in probability) สำหรับฟังก์ชันของตัวแปรสุ่ม

## Theorem

*ถ้า* $X_n \xrightarrow{p} c$ *และ* $g(x)$ *เป็นฟังก์ชันที่ต่อเนื่องที่จุด* $x = c$ *แล้ว* $g(X_n) \xrightarrow{p} g(c)$

## Proof.

- กำหนดให้ $\epsilon > 0$ เนื่องจาก $g(x)$ เป็นฟังก์ชันที่ต่อเนื่องที่จุด $c$ ดังนั้น จะต้องมีค่าคงที่ $\delta > 0$ ที่ทำให้ $|g(x) - g(a)| < \epsilon$ สำหรับทุกๆ ค่า $x$ ที่สอดคล้องกับเงื่อนไข $|x - c| < \delta$ ดังนั้น

$$|g(x) - g(c)| \geq \epsilon \Rightarrow |x - c| \geq \delta \tag{9}$$

- เราสามารถเขียนในรูปของเหตุการณ์ได้ว่า

$$\{|g(x) - g(c)| \geq \epsilon\} \subset \{|x - c| \geq \delta\} \quad \text{ดังนั้น} \quad Pr(|g(x) - g(c)| \geq \epsilon) \leq Pr(|x - c| \geq \delta)$$

  โดยนิยามของการลู่เข้าเชิงความน่าจะเป็น $X_n \xrightarrow{p} c$ ค่าลิมิตของพจน์ด้านขวามือมีค่าเท่ากับ

$$\lim_{n\to\infty} Pr(|x - c| \geq \delta) = 0$$

- สรุปได้ว่า $\lim_{n\to\infty} Pr(|g(x) - g(c)| \geq \epsilon) = 0 \Rightarrow g(X_n) \xrightarrow{p} g(c)$

□

## ทฤษฎี: การลู่เข้าเชิงความน่าจะเป็น (converge in probability) สำหรับการคูณกันของตัวแปรสุ่ม

- ทฤษฎีบทนี้สามารถขยายผลสู่กรณีของเวคเตอร์ของตัวแปรสุ่ม $\mathbf{Z}_n$ นั่นคือ ถ้า $\mathbf{Z}_n \xrightarrow{p} \mathbf{z}$ และ $g(\mathbf{x})$ เป็นฟังก์ชันที่ต่อเนื่องที่จุด $\mathbf{x} = \mathbf{z}$ แล้ว $g(\mathbf{Z}_n) \xrightarrow{p} g(\mathbf{z})$ ดังแสดงตัวอย่างในรูปของฟังก์ชันผลคูณในทฤษฎีบทต่อไปนี้

### Theorem

*สมมุติว่า* $X_n \xrightarrow{p} X$ *และ* $Y_n \xrightarrow{p} Y$ *แล้ว* $X_nY_n \xrightarrow{p} XY$

### Proof.

พิจารณาผลคูณ

$$X_nY_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n + Y_n)^2 \xrightarrow{p} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X - Y)^2 = XY$$

$\square$

- ตัวอย่างต่อไปนี้แสดงการประยุกต์กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) และคุณสมบัติการลู่เข้าเชิงความน่าจะเป็นเพื่อหาค่าลิมิตของค่าความแปรปรวนจากตัวอย่าง (sample variance)

# ตัวอย่าง: การประยุกต์กฎว่าด้วยตัวอย่างขนาดใหญ่

## Example

สมมุติว่า $X_1, \ldots, X_n$ คือตัวอย่างสุ่มที่สุ่มเลือกมาจากการแจกแจงที่ค่าคาดหมายเท่ากับ $\mu$ และค่าความแปรปรวนเท่ากับ $\sigma^2$ พิจารณาค่าความแปรปรวนจากตัวอย่าง (sample variance)

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}_n^2\right]$$

- กฎว่าด้วยตัวอย่างขนาดใหญ่ (Law of Large Numbers) ช่วยให้สรุปได้ว่า

$$\bar{X}_n \xrightarrow{p} \mu$$

$$\frac{\sum_{i=1}^{n} X_i^2}{n} \xrightarrow{p} E\left[X_i^2\right] = \sigma^2 + \mu^2$$

- ในขณะเดียวกัน ทฤษฎีบทก่อนหน้านี้ช่วยให้สามารถสรุปได้ว่า

$$\bar{X}_n^2 \xrightarrow{p} \mu^2$$

- ดังนั้น $$S_n^2 = \frac{n}{n-1}\left[\frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}_n^2\right] \xrightarrow{p} \left(\lim_{n\to\infty}\frac{n}{n-1}\right)\left[\sigma^2 + \mu^2 - \mu^2\right] = \sigma^2$$

# ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem)

- ทฤษฎีบทลิมิตของค่ากลาง (Central Limit Theorem หรือที่เรียกสั้นๆ ว่า CLT) เป็นทฤษฎีทางสถิติที่มีความสำคัญมากอันหนึ่ง ถึงแม้ว่าทฤษฎีบทนี้จะ**มีข้อจำกัดในกรณีที่กลุ่มตัวอย่างมีขนาดเล็ก** (finite sample)
- ทฤษฎีบทนี้ตั้งอยู่บน**พื้นฐานของการหาค่าลิมิต** เมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากพอ (large sample)
- ในปัจจุบัน ความสามารถของคอมพิวเตอร์ที่พัฒนาอย่างก้าวกระโดดได้ช่วยให้เราสามารถคำนวณหาการแจกแจงของค่าสถิติในกรณีที่ตัวอย่างมีขนาดจำกัด (finite sample) ได้สะดวกมากยิ่งขึ้น (ยกตัวอย่างเช่น วิธี bootstrapping)
- ทฤษฎีบทลิมิตของค่ากลาง (CLT) ก็ยังไปสิ่งที่นักสถิติหรือนักวิเคราะห์ข้อมูล **ต้องทำความเข้าใจ เพราะเป็นรากฐานที่สำคัญ** ที่นำไปสู่วิธีการทางสถิติและสูตรที่ใช้ในการคำนวณการแจกแจงของค่าสถิติในโปรแกรมทางสถิติที่ใช้อยู่ในปัจจุบัน

# ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem)

- ทฤษฎีบทนี้ถูกเรียกว่า ทฤษฎีบทลิมิตของค่ากลาง (CLT) ก็เพราะมันเกี่ยวข้องกับค่ากลาง ซึ่งในที่นี้หมายถึงค่าคาดหมายจากตัวอย่าง (sample mean) $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$
- ทฤษฎีบทนี้กล่าวว่า สำหรับตัวอย่างสุ่ม (random sample) ขนาด $n$ ตัวอย่าง ที่ถูกสุ่มมาจากการแจกแจงที่มีค่าคาดหมาย $\mu$ และค่าความแปรปรวน $\sigma^2$ เราสามารถประมาณค่าการแจกแจงของค่าคาดหมายจากตัวอย่าง (sample mean) ได้โดยใช้การแจกแจงปกติ (normal distribution)
    - ค่าคาดหมายเท่ากับ $\mu$ และค่าความแปรปรวน $\frac{\sigma^2}{n}$
    - สุ่มตัวอย่าง $n$ ตัวอย่างมาจากการแจกแจงปกติ (normal distribution)
    - การสุ่มจากการแจกแจงใดๆ ขอเพียงแค่ว่า เป็นการสุ่มอย่างอิสระจากการแจกแจงเดียวกัน

# ทฤษฎี: ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem) ของ Lindeberg and Levy

**Theorem (Lindeberg and Levy)**

*ถ้าตัวแปรสุ่ม $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่ม (random sample) ขนาด n ตัวอย่าง ซึ่งสุ่มมาจากการแจกแจงเดียวกัน ที่มีค่าคาดหมาย $\mu$ และค่าความแปรปรวน $\sigma^2 < \infty$ แล้ว สำหรับค่าคงที่ x ใดๆ*

$$\lim_{n \to \infty} Pr\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi\left(x\right) \tag{10}$$

*โดยที่ $\Phi$ แทนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของการแจกแจงปกติมาตรฐาน (standard normal distribution)*

- ยิ่งไปกว่านั้น ในทางทฤษฎี เราสามารถเขียนทฤษฎีบทลิมิตของค่ากลาง (CLT) โดยใช้หลักการลู่เข้าเชิงการแจกแจง (convergence in distribution) ซึ่งมีนิยามดังต่อไปนี้

# นิยาม: การลู่เข้าเชิงการแจกแจง (Convergence in Distribution)

## Definition (Convergence in Distribution หรือ Asymptotic Distribution)

กำหนดให้ $Y_1, Y_2, \ldots$ เป็นอันดับของตัวแปรสุ่ม และ $F_n$ แทนฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของ $Y_n$ สำหรับ $n = 1, 2, \ldots$ และกำหนดให้ $F^*$ เป็นฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) อันหนึ่ง ดังนั้นอันดับของ $Y_1, Y_2, \ldots$ ลู่เข้าเชิงการแจกแจง (converge in distribution) สู่ $F^*$ ถ้า

$$\lim_{n\to\infty} F_n\left(y\right) = F^*\left(y\right) \tag{11}$$

สำหรับทุกๆ ค่า $y$ ที่ $F^*\left(y\right)$ ต่อเนื่อง (continuous) และเพื่อความกระชับ เราอาจจะกล่าวสั้นๆ แทนได้ว่า $Y_n$ ลู่เข้าเชิงความน่าจะเป็นสู่ $F^*$ ซึ่งอาจจะเขียนแทนได้ด้วย

$$Y_n \xrightarrow{d} F^* \tag{12}$$

และเรียก $F^*$ ว่าการแจกแจงที่ลิมิต (asymptotic distribution) ของ $Y_n$

- ทฤษฎีบทลิมิตของค่ากลาง (CLT) สำหรับตัวอย่างสุ่มสามารถเขียนใหม่ในรูปของการลู่เข้าเชิงการแจกแจงได้ดังต่อไปนี้

# ทฤษฎี: ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem) ของ Lindeberg and Levy

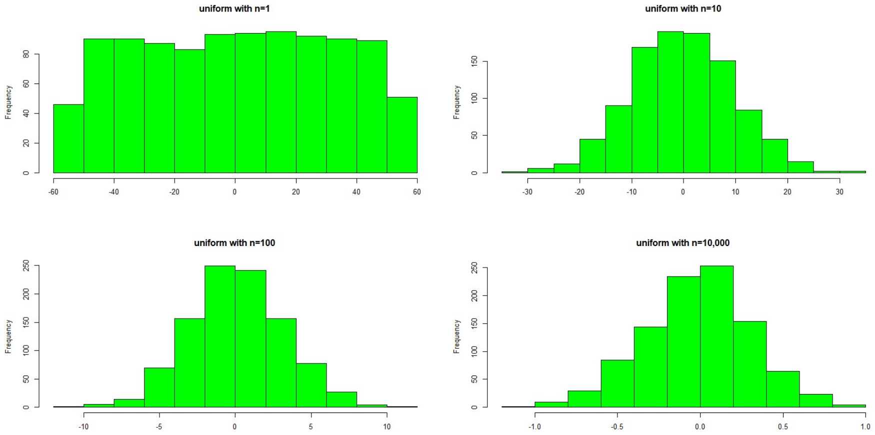
**Theorem (Lindeberg and Levy)**

*ถ้าตัวแปรสุ่ม $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่ม (random sample) ขนาด n ตัวอย่าง ซึ่งสุ่มมาจากการแจกแจงเดียวกัน ที่มีค่าคาดหมาย $\mu$ และค่าความแปรปรวน $\sigma^2 < \infty$ แล้ว สำหรับค่าคงที่ x ใดๆ*

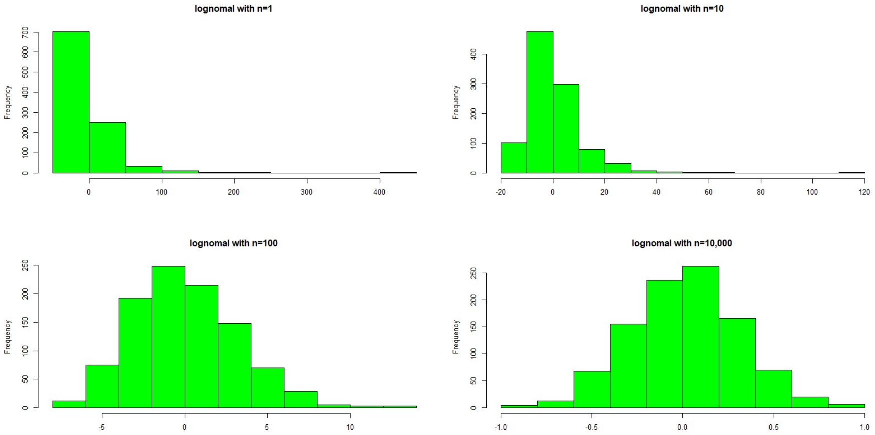
$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} N(0,1) \tag{13}$$

- เพื่อให้เข้าใจทฤษฎีบทลิมิตของค่ากลาง (CLT) ได้ดี ขอนำเสนอผลการจำลอง (simulation) ทั้ง 1000 ครั้ง จากการแจกแจงรูปแบบต่างๆ ดังตัวอย่างต่อไปนี้

ตัวอย่าง: การทำ simulation ของฟังก์ชันการกระจายแบบ uniform: $f(x) = 1$ for $0 \leq x \leq 1$

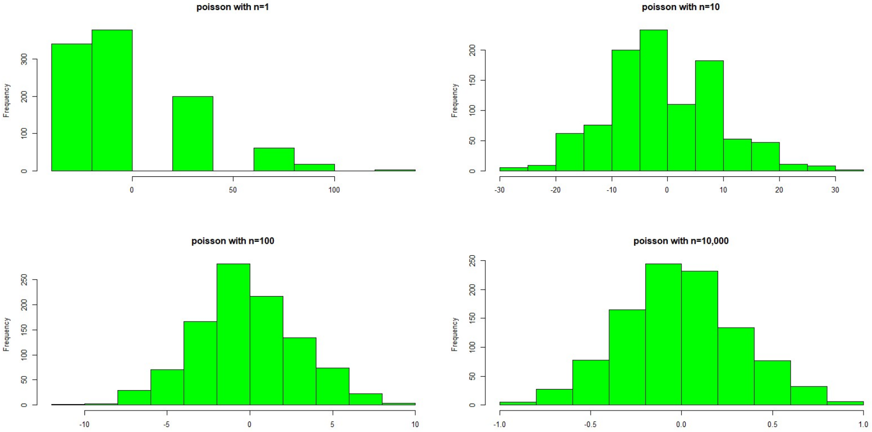


# ตัวอย่าง: การทำ simulation ของฟังก์ชันการกระจายแบบ log-normal : $\log X \sim N(0,1)$ for $x > 0$

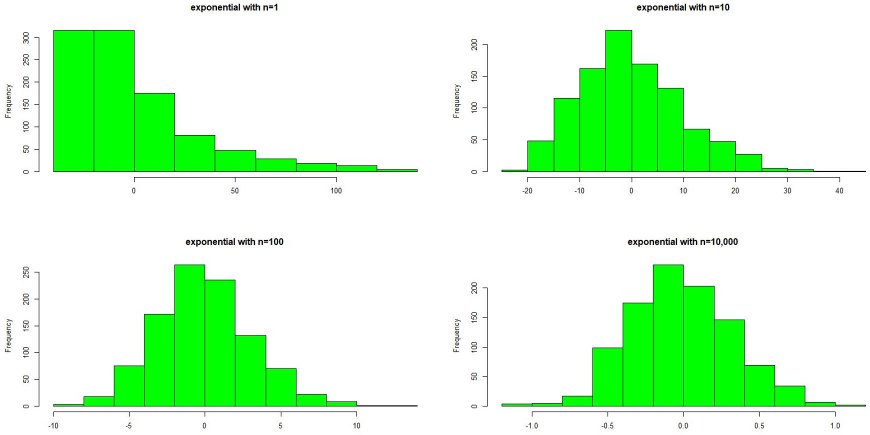


ตัวอย่าง: การทำ simulation ของฟังก์ชันการกระจายแบบ poisson:

$$f(x) = \frac{e^{-1}}{x!} \text{ for } x = 0, 1, 2, \ldots$$



## ตัวอย่าง: การทำ simulation ของฟังก์ชันการกระจายแบบ exponential: $f(x) = e^{-x}$ for $x \geq 0$




ทฤษฎีบทต่อไปนี้ช่วยให้สามารถประยุกต์ทฤษฎีบทลิมิตของค่ากลาง (CLT) ได้กว้างขวางมากยิ่งขึ้น

# ทฤษฎี: ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem)

**Theorem**

ถ้า $X_n \xrightarrow{p} X$ แล้ว $X_n \xrightarrow{d} X$

**Theorem**

ถ้า $X_n \xrightarrow{d} c$ โดยที่ $c$ เป็นค่าคงที่ แล้ว $X_n \xrightarrow{p} c$

**Theorem**

ถ้า $X_n \xrightarrow{d} X$ และ $Y_n \xrightarrow{p} 0$ แล้ว $X_n + Y_n \xrightarrow{d} X$

**Theorem (Slutsky's Theorem)**

กำหนดให้ $X, X_n A_n, B_n$ เป็นตัวแปรสุ่ม (random variable) ส่วน $a$ และ $b$ เป็นค่าคงที่ ดังนั้น ถ้า $X_n \xrightarrow{d} X$, $A_n \xrightarrow{p} a$ และ $B_n \xrightarrow{p} b$ แล้ว

$$A_n + B_n X_n \xrightarrow{d} a + bX \tag{14}$$

# ทฤษฎีบทลิมิตของค่ากลาง (The Central Limit Theorem)

**Theorem**

*สมมุติว่า* $X_n \xrightarrow{d} X$ *และ* $g$ *เป็นฟังก์ชันที่ต่อเนื่องตลอดช่วงค้ำจุน (support) ของ* $X$ *แล้ว* $g(X_n) \xrightarrow{d} g(X)$

- ในทางคณิตศาสตร์ ทฤษฎีบทนี้กล่าวว่า การลู่เข้าเชิงการแจกแจง (convergence in distribution) สามารถส่งผ่านฟังก์ชันต่อเนื่องได้ โดยมีทฤษฎีบทของสลูสกี้ (Slutsky's Theorem) เป็นกรณีพิเศษสำหรับฟังก์ชันเชิงเส้นซึ่งต่อเนื่องตลอดช่วงค้ำจุน
- ถึงแม้ว่าทฤษฎีบทนี้จะช่วยให้สามารถหาการแจกแจงที่ลิมิตของฟังก์ชันต่อเนื่องของตัวแปรสุ่มที่ลู่เข้าเชิงการแจกแจงได้ แต่ในทางปฏิบัติก็ยังมีความยุ่งยากพอสมควร เพราะต้องแปลงการแจกแจงของ $X$ ให้เป็นการแจกแจงของ $g(X)$ ซึ่งก็ขึ้นอยู่กับรูปแบบของฟังก์ชัน $g$
- ในทางปฏิบัติ จึงนิยมใช้วิธีการเดลต้า (Delta method) ซึ่งสามารถประยุกต์ใช้กับการแจกแจงปกติได้อย่างสะดวก ดังแสดงในทฤษฎีบทต่อไปนี้

# วิธีการเดลต้า (Delta method)

**Theorem**

*สมมุติว่า $g$ เป็นฟังก์ชันที่หาค่าอนุพันธ์อันดับหนึ่ง (first-order derivative) ได้ที่จุด $\theta$ และค่าอนุพันธ์อันดับหนึ่ง $g'(\theta) \neq 0$ และ*

$$\sqrt{n}\left(X_n - \theta\right) \xrightarrow{d} N\left(0, \sigma^2\right) \tag{15}$$

*แล้ว*

$$\sqrt{n}\left(g\left(X_n\right) - g\left(\theta\right)\right) \xrightarrow{d} N\left(0, \left[g'\left(\theta\right)\right]^2 \sigma^2\right) \tag{16}$$

- สังเกตว่าวิธีการเดลต้า (Delta method) นี้ใช้ได้กับการแจกแจงปกติ (normal distribution) และฟังก์ชันที่ค่าอนุพันธ์อันดับหนึ่ง (first-order derivative) ไม่เท่ากับศูนย์ ณ จุดที่สนใจเท่านั้น

# ตัวอย่าง: การประยุกต์ใช้วิธีการเดลต้า (Delta method)

## Example

กำหนดให้ $X_1, X_2, \cdots, X_n$ เป็นตัวแปรสุ่ม ที่มี ค่าคาดหวัง คือ $\mu$ และความแปรปรวน คือ $\sigma^2$ จงหาว่า ฟังก์ชันการแจกแจง ของ $\frac{1}{\bar{X}_n}$

- วิธีการเดลต้า (Delta method) เรากำหนดให้

$$g(\bar{X}_n) = \frac{1}{\bar{X}_n} \quad \text{และ} \quad g'(\bar{X}_n) = -\left(\frac{1}{\bar{X}_n}\right)^2$$

- ดังนั้น

$$\sqrt{n}\left(g\left(X_n\right) - g\left(\theta\right)\right) \xrightarrow{d} N\left(0, \left[g'\left(\theta\right)\right]^2 \sigma^2\right)$$

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \left[-\left(\frac{1}{\mu^2}\right)\right]^2 \sigma^2\right)$$

- สรุปได้ว่า $\frac{1}{\bar{X}_n}$ มีการแจกแจงปกติที่มีค่าคาดหวังคือ $\frac{1}{\mu}$ และความแปรปรวน คือ $\frac{\sigma^2}{n\mu^4}$