Random Variable (IV)

Petchara Pattarakijwanich

SCPY208, 5 April 2021

ทบทวน Independent Random Variables

x, y เป็น Random Variable

$$E(ax+by+c) = aE(x) + bE(y) + c$$

$$Var(ax+by+c) = a^2 Var(x) + b^2 Var(y) + 2ab\, Cov(x,y)$$

$$\Downarrow \quad x, y \text{ independent} \Rightarrow cov = 0$$

$$E(ax+by+c) = aE(x) + bE(y) + c$$

$$Var(ax+by+c) = a^2 Var(x) + b^2 Var(y)$$

$$E(x+y) = E(x) + E(y) \quad , \quad Var(x+y) = Var(x) + Var(y)$$

Sample vs Population

<u>ตัวอย่าง</u>   Random variable X = ความสูงคนไทย

Population : คนไทยทั้งประเทศ

ตัวแปร x มี PDF บางอย่าง มี μ, σ^2 บางค่า

[PDF, μ, σ^2 ของ x มีอยู่จริง แต่ไม่รู้ว่าคืออะไร]

Sample : คนไทยจำนวน n คน ($n \approx 10{,}000$)

วัด $x_1, x_2, \ldots, x_n$

ค่าประมาณ (estimator)

จากข้อมูลที่วัดได้ (ความสูง n คน) $\Rightarrow$ หา PDF, μ, σ^2 ยังไง!

Sample vs Population

Probability :

รู้ PDF → คำนวณ prob
ที่จะวัด x ได้เท่าไรเท่าไร

Population

Underlying truth
PDF, μ, σ^2
ของตัวแปร x

อยากรู้ ไม่มีทางรู้เป๊ะๆ
ต้องประมาณหา

Sample

Observable Data
ค่า x วัดมา n ครั้ง
$x_1, \dots, x_n$

วัดได้

Statistics

รู้ข้อมูลที่วัด → reconstruct
underlying truth

Sample vs Population

ปัญหาของวันนี้

รู้ $x_1, x_2, \dots, x_n$ $\rightarrow$ หา estimator ของ μ, σ^2 ยังไง?

ลองหา estimator

$$\bar{x} = \frac{1}{n}\left(x_1 + x_2 + \dots + x_n\right) \Rightarrow \text{estimator ของ } \mu$$

$$s^2 = \frac{1}{n-1}\left[(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2\right] \Rightarrow \text{estimator ของ } \sigma^2$$

คำนวณจาก $x_1, \dots, x_n$
ขึ้นกับข้อมูลจริงๆ

μ, σ^2 = population mean/variance

$\bar{x}, s^2$ = sample mean/variance

<u>ตัวอย่าง</u> จงหาค่า expected value ของตัวแปรสุ่ม $\bar{x}$ ที่นิยามดังนี้

$x_1, x_2, \ldots$ เป็น RV
มีรูปแบบ PDF เดียวกัน
Independent

$$\bar{x} = \frac{1}{n} \sum_i x_i \quad = \frac{1}{n}\left(x_1 + x_2 + \ldots + x_n\right)$$

เมื่อ x_i เป็นตัวแปรสุ่มที่มี expected value และ variance เป็น μ และ σ

$$E(ax+by) = a\,E(x) + b\,E(y)$$

$$E(\bar{x}) = E\left[\frac{1}{n}\left(x_1 + \ldots + x_n\right)\right]$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \ldots + E(x_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \ldots + \mu\right] \qquad \left[E(x_i) = \mu \ (\text{ยังไม่รู้ } \mu!)\right]$$

$E(\bar{x}) = \mu \Rightarrow \bar{x}$ เป็น <u>unbiased estimator</u> ของ μ

<u>ตัวอย่าง</u> จงหาค่า variance ของตัวแปรสุ่ม $\bar{x}$ ที่นิยามดังนี้

$$\bar{x} = \frac{1}{n}\sum_i x_i \quad = \frac{1}{n}\left(x_1 + x_2 + \dots + x_n\right)$$

เมื่อ x_i เป็นตัวแปรสุ่มที่มี expected value และ variance เป็น μ และ σ

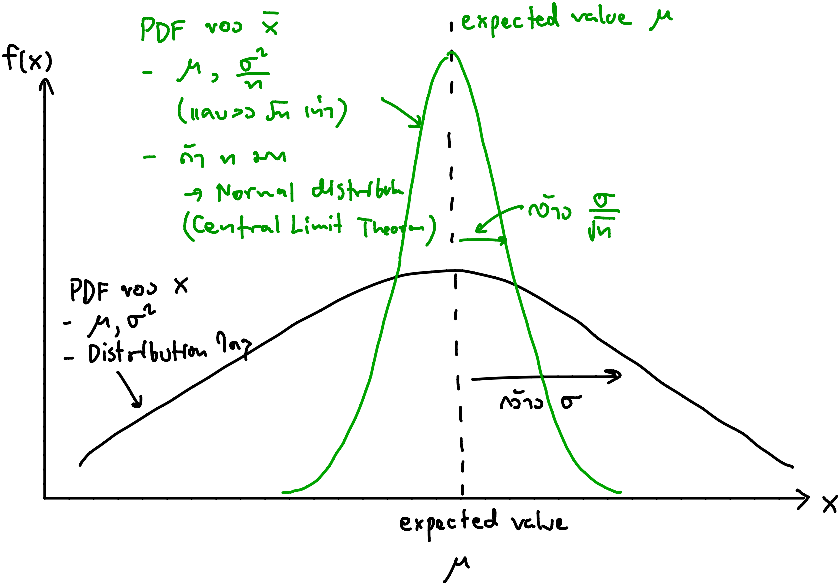
$$\mathrm{Var}(\bar{x}) = \mathrm{Var}\left[\frac{1}{n}\left(x_1 + x_2 + \dots + x_n\right)\right] \qquad \mathrm{Var}(ax+by) = a^2\,\mathrm{Var}(x) + b^2\,\mathrm{Var}(y)$$

$$= \frac{1}{n^2}\left[\mathrm{Var}(x_1) + \mathrm{Var}(x_2) + \dots + \mathrm{Var}(x_n)\right]$$

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \dots + \sigma^2\right]$$

$$= \frac{1}{n^2}\left[n\sigma^2\right]$$

$$\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n} \quad \Rightarrow \quad \text{SD ของ } \bar{x} = \frac{\sigma}{\sqrt{n}}$$


PDF ของ $\bar{x}$
- μ, $\frac{\sigma^2}{n}$
(แคบลง $\sqrt{n}$ เท่า)
- ถ้า n มาก
→ Normal distribution
(Central Limit Theorem)
expected value μ
กว้าง $\frac{\sigma}{\sqrt{n}}$
f(x)
PDF ของ X
- μ, σ^2
- Distribution ใดๆ
กว้าง σ
x
expected value
μ

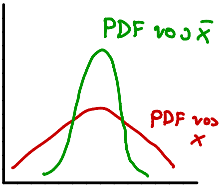
- สุ่มคน 100 คน วัดส่วนสูง

$\boxed{X_1, \dots, X_{100}} \Rightarrow \bar{X}_1$

PDF ของ X

- สุ่มคน 100 คน วัด

$\Rightarrow \bar{X}_2$



- ทำซ้ำหลายครั้ง

$\Rightarrow \boxed{\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_n} \Rightarrow$ PDF ของ $\bar{X}$

ตัวอย่าง จงหาค่า expected value ของตัวแปรสุ่ม s^2 ที่นิยามดังนี้

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

เมื่อ x_i เป็นตัวแปรสุ่มที่มี expected value และ variance เป็น μ และ σ

ตอบ $E(s^2) = \sigma^2 \Rightarrow s^2$ เป็น unbiased estimator ของ variance σ^2

ค่าของ $E(x_i^2)$, $E(x_i x_j)$

$E(x_i) = \mu$, $Var(x_i) = \sigma^2$, x_i independent

$$E(x_i x_j) = E(x_i)E(x_j)$$ [สัปดาห์ก่อน $E(xy) = \mu_x \mu_y$]

$$= \mu \cdot \mu = \mu^2$$

$$\boxed{\begin{aligned} E(x_i x_j) &= \mu^2 \\ E(x_i^2) &= \mu^2 + \sigma^2 \end{aligned}}$$

$$E(x_i^2) = E\left[\left((x_i - \mu) + \mu\right)^2\right]$$

$$= E\left[(x_i - \mu)^2 + 2\mu(x_i - \mu) + \mu^2\right]$$

$$= \underbrace{E\left[(x_i - \mu)^2\right]}_{= \sigma^2} + 2\mu \underbrace{E(x_i - \mu)}_{= 0} + \mu^2$$

$$E(x_i^2) = \sigma^2 + \mu^2$$

$$s^2 = \frac{1}{n-1}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\dots+(x_n-\bar{x})^2\right]$$

$$E(s^2) = \frac{1}{n-1}\left[E\left[(X_1-\bar{X})^2\right]+\dots+E\left[(x_n-\bar{x})^2\right]\right]$$

$$= \frac{n}{n-1}E\left[(X_1-\bar{X})^2\right]$$

$$= \frac{n}{n-1}E\left[X_1^2-2X_1\bar{X}+\bar{X}^2\right]$$

$$= \frac{n}{n-1}E\Bigg[\underbrace{X_1^2-2X_1\left(\frac{X_1+\dots+X_n}{n}\right)+\left(\frac{X_1+\dots+X_n}{n}\right)^2}_{\text{กระจายออกมาให้หมด!}}\Bigg]$$

$$E(S^2) = \frac{n}{n-1} E\left[X_1^2 - 2X_1\left(\frac{X_1 + \cdots + X_n}{n}\right) + \left(\frac{X_1 + \cdots + X_n}{n}\right)^2 \right]$$

$$= \frac{n}{n-1}\Bigg[E(X_1^2) - \frac{2}{n} E\big(X_1^2 + \overbrace{X_1X_2 + \cdots + X_1X_n}^{X_1X_j \text{ } n-1 \text{ พจน์}}\big)$$

$$+ \frac{1}{n^2} E\left(X_1^2 + X_2^2 + \cdots + X_n^2\right) \qquad X_i^2 \text{ มี } n \text{ พจน์}$$

$$+ \frac{1}{n^2} E\left(X_1X_2 + X_1X_3 + \cdots\right)\Bigg] \qquad X_iX_j \text{ มี } n(n-1) \text{ พจน์}$$

$$E(S^2) = \frac{n}{n-1}\Big[E(X_i^2) - \frac{2}{n}E(X_i^2) - \frac{2}{n}(n-1)E(X_iX_j)$$

$$+ \frac{1}{n^2} n E(X_i^2) + \frac{1}{n^2} n(n-1) E(X_iX_j) \Big]$$

$$E(s^2) = \frac{n}{n-1}\Big[E(x_i^2) - \frac{2}{n}E(x_i^2) - \frac{2}{n}(n-1)E(x_i x_j)$$

$$+ \frac{1}{n^2}\,n\,E(x_i^2) + \frac{1}{n^2}\,n(n-1)\,E(x_i x_j)\Big]$$

$$= \frac{n}{n-1}\Big[\Big(1 - \frac{2}{n} + \frac{1}{n}\Big)E(x_i^2)$$

$$+ \Big(\frac{n-1}{n} - 2\,\frac{n-1}{n}\Big)E(x_i x_j)\Big]$$

$$= \frac{n}{n-1}\Big[\Big(1 - \frac{1}{n}\Big)E(x_i^2) - \frac{(n-1)}{n}E(x_i x_j)\Big]$$

$$= E(x_i^2) - E(x_i x_j)$$

$$= (\mu^2 + \sigma^2) - \mu^2$$

$E(s^2) = \sigma^2 \Rightarrow$ s^2 เป็น unbiased estimator ของ σ^2

สรุป sample mean, sample variance, variance of the mean

Population mean μ Population variance σ^2

Sample mean $\bar{x} = \frac{1}{n}\sum x_i$ Sample variance $s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$

$\bar{x}$, s^2 เป็น unbiased estimator ของ μ, σ^2

Variance of the mean

$$\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n}$$ [$\bar{x}$ กระจายแบบ Gaussian]

$\Rightarrow$ Error ของ $\bar{x}$ จะน้อยกว่า Error ของ x อยู่ $\sqrt{n}$ เท่า

$$\left[\sigma_m = \frac{\sigma}{\sqrt{n}} \right]$$

ตัวอย่าง วัดค่าความต้านทานของลวดสิบครั้งด้วยมัลติมิเตอร์ ได้ค่าดังนี้

$$7.2, 7.1, 6.7, 7.0, 6.8, 7.0, 6.9, 7.4, 7.0, 6.9$$

ควรจะรายงานค่าและความคลาดเคลื่อนอย่างไร