Continuous Random Variables

อ.ปรัชญ์ ปิยะวงศ์วิศาล

Topics

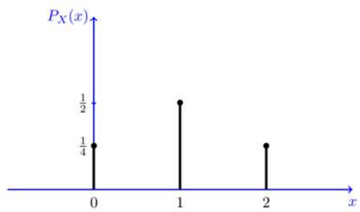
- Review: CDF
- Continuous RV
- PDF
- E[X] and Var(X)
- Uniform Distribution
- Gaussian (Normal) Distribution
- Central Limit Theorem
- Standard Normal Distribution
  - Standard Normal CDF Table

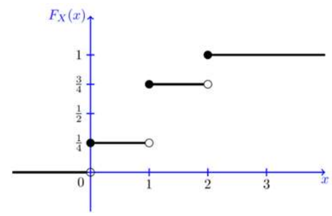
Review: CDF

- นิยาม: Cumulative Distribution Function (CDF) ของตัวแปรสุ่ม X คือ

$$F_X(x) = P(X \leq x) \text{ for all } x \in \mathbb{R}$$

- เช่น กำหนดให้ X เป็นตัวแปรสุ่มที่เป็นจำนวนของ H ที่ออกจากการโยนเหรียญ 2 ครั้ง จงหา CDF ของ X





- คุณสมบัติที่สำคัญของ CDF: $P(a < X \leq b) = F_X(b) - F_X(a)$

CDF Exercise

- กำหนดให้ PMF ของ X เป็นดังตาราง

x	-2	-1	0	1	2
$P_X(x)$	0.2	0.4	0.1	0.2	0.1

  - จงวาดกราฟ CDF ของ X
  - จงหา $P(X \leq 2)$, $P(-1 < X < 1)$

Continuous RV

- **ตัวแปรสุ่มต่อเนื่อง (Continuous RV)** เป็นตัวแปรสุ่มที่
  - มีค่าในช่วงที่กวาดไปบนแกนจำนวน และสามารถมีค่าเป็นทศนิยมที่ละเอียดได้อย่างไม่จำกัด
    - เช่น จำนวนจริงใดๆ ในช่วง $[0,1]$, $(20, 40)$, $(-\infty, \infty)$
  - ค่าที่เป็นไปได้ (ใน S_X) จึงมีมากมาย นับไม่ถ้วน (uncountably infinite)

- ตัวอย่างปริมาณที่เป็น continuous RV

ปัญหาเกี่ยวกับค่า P ในโลก Continuous

- เช่น หากน้ำหนักทารก W เป็นตัวแปรสุ่มต่อเนื่องที่มีค่าอยู่ในช่วง $2.3 - 4.5\ kg$
  - จงหา $P(W = 3.5512187) = ??$
  - ตอบ: ตามหลักการ $P(W = 3.5512187)$ ถือว่ามีค่าเป็น

- เนื่องจาก $P(X = x) =$ สำหรับทุก x ในโลก continuous
  PMF จึงไม่มีประโยชน์สำหรับ continuous RV

- ในการอธิบาย continuous RV เราจึงต้องเปลี่ยนมาใช้
  Probability Density Function (PDF) แทน

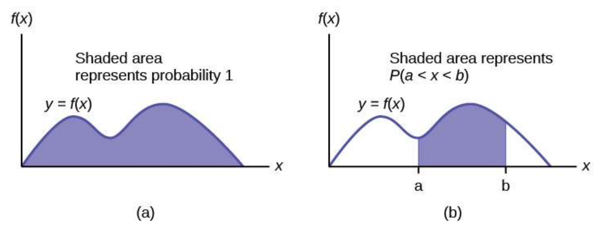
PDF

- หาก X เป็นตัวแปรสุ่มต่อเนื่อง

  $f_X(x)$ คือ ฟังก์ชันความหนาแน่นความน่าจะเป็นของ X
  (Probability Density Function: PDF of X)

สำคัญ

- โดยที่ **พื้นที่ใต้กราฟ** ของ PDF ในช่วง (a, b) ใดๆ คือ **ความน่าจะเป็น**ที่ X จะมีค่าอยู่ในช่วงนั้น



https://math.libretexts.org/Courses/Mt._San_Jacinto_College/Ideas_of_Mathematics/06%3A_Inferential_Statistics/6.02%3A_Continuous_Random_Variables

Area under PDF = probability

- **พื้นที่ใต้กราฟ**ของ PDF ในช่วง ในช่วง (a, b) ใดๆ คือ **ความน่าจะเป็น**ที่ X จะมีค่าอยู่ในช่วงนั้น

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

สำคัญ

- เช่น ถ้า $S = [1, 2)$ จะได้ว่า $P(1 \leq X < 2) = \int_1^2 f_X(x)dx$

ข้อควรระวังเกี่ยวกับ PDF

- ค่าของ $f_X(x)$ **ไม่ใช่**ความน่าจะเป็น
  - แต่ พื้นที่ใต้กราฟ $f_X(x)$ คือความน่าจะเป็น
  - $f_X(x)$ คือความหนาแน่น (density) ของความน่าจะเป็น
- $f_X(x)$ อาจมีค่า >1 ได้
- แม้ความสูงของ $f_X(x)$ จะไม่ใช่ความน่าจะเป็นตรงๆ แต่ความสูงของ $f_X(x)$ พอที่สามารถบ่งบอกถึง relative probability ที่ X จะมีค่าอยู่ในบริเวณแถวๆ นั้น **เมื่อเทียบ**กับบริเวณอื่นได้

Properties of PDF

- $f_X(x) \geq 0; \quad \forall x$
- $P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$ เสมอ (axiom)
- $P(X = a) = \int_a^a f_X(x)dx = 0$
- $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$ (ตามนิยามของ CDF)
- $f_X(x) = \frac{dF_X(x)}{dx}$

Exercise

- กำหนดให้ตัวแปรสุ่มต่อเนื่อง X มี PDF ดังนี้

$$f_X(x) = \begin{cases} C(x - x^2) & ; 0 < x < 1 \\ 0 & ; otherwise \end{cases}$$

- จงหาค่า C ที่เหมาะสม
- จงหา $P(X > 0.5)$

ความสัมพันธ์ระหว่าง PDF กับ CDF

- ทบทวนคุณสมบัติของ CDF: $P(a < X < b) = F_X(b) - F_X(a)$
- นิยามทางการ: หาก X เป็นตัวแปรสุ่มต่อเนื่อง จะได้ว่า PDF ของ X คือ

$$f_X(x) = \lim_{\delta \to 0^+} \frac{P(x < X \leq x + \delta)}{\delta} = \lim_{\delta \to 0^+} \frac{F_X(x+\delta) - F_X(x)}{\delta} = \frac{dF_X(x)}{dx} = F_X'(x)$$

- นั่นคือ PDF เป็นอนุพันธ์ (derivative) ของ CDF

PMF vs PDF

Distribution	Discrete	Continuous
Definition	$P(x) = P\{X = x\}$ (pmf)	
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	
Total probability	$\sum_{x} P(x) = 1$	

PMF vs PDF

Discrete	Continuous
$\mathbf{E}(X) = \sum_x xP(x)$ $\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$ $\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x \sum_y (xy)P(x,y) - \mu_x \mu_y$	

Well-Known Probability Distributions

- **Discrete**
  - Bernoulli
  - Geometric
  - Binomial
  - Poisson
  - Discrete Uniform
  - Negative Binomial
  - Zipf
- **Continuous**
  - Uniform
  - Exponential
  - Gaussian (Normal)
  - Beta
  - Gamma
  - Student’s t
  - Chi-Squared

More distributions: https://en.wikipedia.org/wiki/List_of_probability_distributions

1) Continuous Uniform Distribution

$$\boxed{X \sim Uniform(a, b)}$$

- การทดลองสุ่ม:
- ค่าของ X คือ:
- parameter ของการแจกแจง:

1) Continuous Uniform Distribution

- PDF ของการแจกแจงแบบ Uniform

- $E[X] =$

- $Var(X) =$

Uniform RV Exercise

- กำหนดให้ $X \sim Uniform(-2, 4)$
  - จงหา $P(X < 2)$
  - จงหา $E[X]$

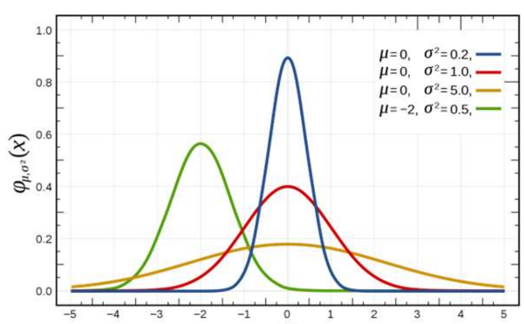
2) Gaussian (Normal) Distribution

$$X \sim Normal(\mu, \sigma^2)$$

- การทดลองสุ่ม:
- ค่าของ X คือ:
- parameter ของการแจกแจง:

2) Gaussian (Normal) Distribution

- PDF
- $E[X] =$
- $Var(X) =$
- เป็นการแจกแจงที่พบได้บ่อยในธรรมชาติ เนื่องด้วย Central Limit Theorem



https://en.wikipedia.org/wiki/Normal_distribution

Central Limit Theorem (CLT)

- Under some conditions, the average of many [i.i.d.] samples (observations) of a random variable with finite mean and variance is itself a random variable—whose distribution **converges to a normal distribution** as the number of samples increases.
- Therefore, physical quantities that are expected to be the sum of many independent processes, such as **measurement errors**, often have distributions that are nearly normal.

https://en.wikipedia.org/wiki/Normal_distribution

Central Limit Theorem (CLT)

- หากเรามีตัวแปรสุ่ม X ใดๆ
  เราสามารถนำ X ไปชักตัวอย่างได้ เช่น:
- จากกลุ่มตัวอย่างนี้ เราสามารถหา sample mean $\bar{X}$
- เนื่องจากถ้าเรา resample X ใหม่ เราก็อาจจะได้ค่า sample mean $\bar{X}$ ที่เปลี่ยนไปจากเดิมได้
  - ดังนั้น $\bar{X}$ เองจึงต้องเป็นตัวแปรสุ่มด้วย

Central Limit Theorem (CLT)
การแจกแจงของ sample mean $\bar{X}$ จะลู่เข้าหา Gaussian หาก $n \to \infty$

CLT ช่วยอธิบายว่าทำไมในธรรมชาติจึงพบเจอ Gaussian ได้บ่อย


สำคัญ

https://en.wikipedia.org/wiki/Normal_distribution

การหาค่า P สำหรับ Gaussian

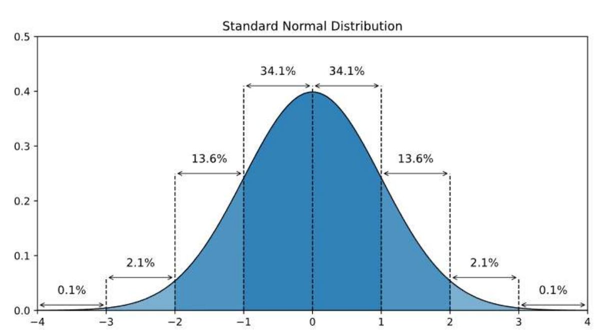
- เช่น กำหนดให้ $X \sim Normal(30, 4)$
- จงหา $\mathrm{P}(29 < X < 32) = \int_{29}^{32} f_X(t)dt$
  - หาตรงๆ ไม่ได้ เนื่องจาก integral แกะไม่ออก

- จำเป็นต้องเปิดตาราง Standard Normal CDF Table
  - ปัญหา: X ในโจทย์ยังไม่ได้มีการแจกแจงแบบ Standard Normal
  - ต้องแปลง (standardize) X ให้เป็น Standard Normal ก่อน

Standard Normal Distribution

- Standard Normal Distribution คือ Normal Distribution ที่มีค่า
  - $\mu = 0$
  - $\sigma^2 = 1$
- นั่นคือ $Z \sim Normal(0, 1)$
  - จะได้ว่า Z มีการแจกแจงดังภาพ
  - CDF ของ Z คือ $\Phi(z)$ คำนวณได้จากตาราง
- เราสามารถแปลง $\mathrm{X} \sim Normal(\mu, \sigma^2)$ ใดๆ เป็น Z ได้โดยให้
  - $\boxed{Z = \frac{X - \mu}{\sigma}}$

สำคัญ



Standard Normal CDF Table

Table A.3. The Cumulative Distribution Function for the Standard Normal Distribution: Values of $\Phi(z)$ for nonnegative z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Exercise:

- เช่น กำหนดให้ $X \sim Normal(30, 4)$
- จงหา $\mathrm{P}(29 < X < 32)$