Random Variables and Distributions

รศ.ดร. วีระชาติ กิเลนทอง
มหาวิทยาลัยหอการค้าไทย

Main issue

- ตัวแปรสุ่ม (Random Variables) คืออะไร?
- เมื่อเราพิจารณาตัวแปรสุ่ม เราควรต้องถามว่า ตัวแปรสุ่มนั้นมีการแจกแจงอย่างไร? เพราะ
  - การแจกแจง (distribution) คือผลของการรวบรวมคุณสมบัติเชิงความน่าจะเป็น (probabilistic properties) ทั้งหมดของตัวแปรสุ่ม
- การแจกแจง (distribution) สามารถเขียนได้ในรูปของ p.f. หรือ p.d.f. และ C.D.F.

ตัวแปรสุ่ม (Random Variables)

Definition (ตัวแปรสุ่ม (random variable))

ตัวแปรสุ่ม (random variable) X คือฟังก์ชันค่าจริง (real-valued function) ซึ่งกำหนดค่าจำนวนจริงสำหรับแต่ละผลลัพธ์ (outcomes) หรือเหตุการณ์ (events) ในปริภูมิตัวอย่าง (sample space) ซึ่งสามารถเขียนในรูปแบบทางคณิตศาสตร์ได้เป็น $X : S \to \mathbb{R}$

- เรามักจะแทนค่าที่เกิดขึ้นจริง (realization) ของตัวแปรสุ่ม X สำหรับผลลัพธ์ $s \in S$ ในรูป $X(s) = x$ ซึ่งเป็นค่าจำนวนจริง
- Key Benefit: เราสามารถวิเคราะห์บนค่าจำนวนจริงโดยไม่ต้องสนใจว่าเหตุการณ์ที่อยู่เบื้องหลังคืออะไร ซึ่งช่วยให้การวิเคราะห์มีความสะดวกอย่างมาก

ตัวอย่าง: ตัวแปรสุ่มแบบไม่ต่อเนื่อง

Example

พิจารณาการทดลองโยนเหรียญ 5 ครั้ง ดังนั้น ปริภูมิตัวอย่าง (sample space) ในกรณีนี้คือ

$$S = \left\{ s : s \text{ ผลของการโยนเหรียญ 5 ครั้ง} \right\}$$

- ถ้าสิ่งที่เราสนใจคือ จำนวนครั้งที่ผลการโยนเป็นก้อย ซึ่งสามารถแทนได้ด้วยตัวแปรสุ่ม X ซึ่งเป็น "**ฟังก์ชันที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นก้อย**"
  - ผลลัพธ์ $s = HTTHT$ (H แทนหัว และ T แทนก้อย) คือ $X(s) = 3$
- ถ้าสิ่งที่เราสนใจคือ จำนวนครั้งของการโยนที่ได้ผลออกมาเป็นหัว เราสามารถแทนด้วยตัวแปรสุ่ม Y ซึ่งเป็น "**ฟังก์ชันที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นหัว**"
  - ผลลัพธ์ $s = HTTHT$ (H แทนหัว และ T แทนก้อย) คือ $Y(s) = 2$

เราสามารถสร้างตัวแปรสุ่มสำหรับการทดลองอันหนึ่งได้ไม่จำกัด

- ตัวอย่างที่ผ่านมาแสดงให้เห็นว่า "เราสามารถสร้างตัวแปรสุ่มสำหรับการทดลองอันหนึ่งได้**ไม่จำกัด**"
- เราจะเลือกใช้ตัวแปรสุ่มที่ช่วยให้เราวิเคราะห์ปัญหาของเราได้สะดวกที่สุด (ควรเลือกนิยามตัวแปรสุ่มให้เหมาะสม)
- เราสามารถใช้การดำเนินการทางพีชคณิต (algebraic operation) กับตัวแปรสุ่มได้ เช่น $Y(s) = 5 - X(s)$
  - ในขณะที่เราอาจจะไม่สามารถทำได้กับผลลัพธ์จากการทดลองโดยตรง

ตัวอย่าง: ตัวแปรสุ่มแบบไม่ต่อเนื่องที่มาจากผลลัพธ์ที่เป็นแบบต่อเนื่อง

Example

พิจารณาการทดลองที่ผลลัพธ์คือความต้องการใช้ไฟฟ้าและน้ำประปา โดยในที่นี้สมมุติให้มีค่าตั้งแต่ 0 ถึง 1 ทั้งสองอย่าง กำหนดให้ผลลัพธ์ของการทดลองนี้อยู่ในรูป (x, y) โดยที่ $x \in [0, 1]$ คือความต้องการใช้ไฟฟ้า และ $y \in [0, 1]$ คือความต้องการใช้น้ำประปา สมมุติว่าเราสนใจว่าความต้องการใช้ไฟฟ้าและน้ำประปาสูงเกินไปหรือไม่ โดยกำหนดว่า ความต้องการที่สูงเกินไปคือระดับความต้องการที่มากกว่า 0.5 สำหรับการใช้ไฟฟ้า และ 0.75 สำหรับการใช้น้ำประปา เราสามารถสร้างตัวแปรสุ่มเพื่อแทนสิ่งที่เราต้องการได้เป็น

$$Z(x, y) = \begin{cases} 1, & \text{ถ้า } x > 0.5 \text{ และ } y > 0.75 \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

โดยที่ $s = (x, y)$ คือผลลัพธ์ของการทดลองใดๆ

- บทเรียน: เราสามารถนิยามตัวแปรสุ่มแบบไม่ต่อเนื่อง (discrete random variable) จากปริภูมิตัวอย่างที่ค่าของผลลัพธ์เป็น**แบบต่อเนื่อง**ได้

การกำหนดตัวแปรสุ่มและความน่าจะเป็นของตัวแปรสุ่ม

- จากบทเรียนก่อนหน้านี้ ได้เรียนรู้แล้วว่า เราสามารถ **นิยามความน่าจะเป็น** ที่ถูกต้องตามหลักการ (well-defined probability) สำหรับปริภูมิตัวอย่าง S ของการทดลองอันหนึ่งได้มากมาย
- เพื่อให้การอภิปรายมีความชัดเจน เราจำเป็นต้องเริ่มด้วย "**การกำหนด**" ก่อนว่าเรากำลังใช้ความน่าจะเป็นอันใด ดังที่นำเสนอใน เรื่องความน่าจะเป็น
  - เราแทนความน่าจะเป็นของเหตุการณ์ A ซึ่งเป็นสับเซ็ตของปริภูมิตัวอย่าง S ด้วย $Pr(A)$
  - กำหนดให้ C คือสับเซ็ตของจำนวนจริงที่ $A = \{s : X(s) \in C\}$ หรือเขียนแบบย่อได้เป็น $A = \{X \in C\}$ เป็นเหตุการณ์ ซึ่งเป็นสับเซ็ตของปริภูมิตัวอย่าง S
- ตัวอย่างเช่น กรณีของการทดลองโยนเหรียญห้าครั้ง สมมุติว่า $C = \{1\}$ และ "X คือ **ตัวแปรสุ่ม** " ที่บอกถึงจำนวนครั้งของการโยนที่ได้ผลออกมาเป็นก้อย
  - เหตุการณ์ที่ได้จาก C คือ $A = \{THHHH, HTHHH, \ldots, HHHHT\}$ โดยที่ H แทนหัว และ T แทนก้อย นั่นคือ $X(s) = 1$ สำหรับทุกๆ ผลลัพธ์ $s \in A$
  - กำหนดความน่าจะเป็นจากตัวแปรสุ่ม X ได้เป็น $Pr(X \in C) = Pr(X \in A) = \frac{5}{32}$ สำหรับ $C = \{1\}$ ผลของการรวบรวมความน่าจะเป็นในรูปแบบนี้ทั้งหมดคือ "**การแจกแจงของตัวแปรสุ่ม** X"

นิยาม: การแจกแจงของตัวแปรสุ่มไม่ต่อเนื่อง

Definition (การแจกแจงของตัวแปรสุ่ม (distribution of random variable))

สำหรับตัวแปรสุ่ม X และความน่าจะเป็น $Pr(\cdot)$ ที่นิยามบนปริภูมิตัวอย่าง S การแจกแจงของตัวแปรสุ่ม (distribution of random variable) X คือเซ็ตของความน่าจะเป็นทุกอันที่อยู่ในรูปแบบ $Pr(X \in C)$ สำหรับเซ็ตของจำนวนจริง C ทุกเซ็ต โดยที่ $\{X \in C\}$ เป็นเหตุการณ์

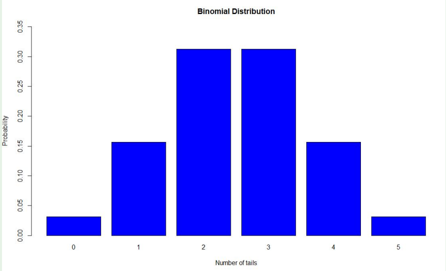
- ข้อสังเกตอันหนึ่งคือ การแจกแจงของ X ก็คือ **มาตรวัดความน่าจะเป็น (probability measure)** ที่นิยามบน **เซ็ตของจำนวนจริง** นั้นเอง ซึ่งสะท้อนข้อเท็จจริงที่ว่า **ตัวแปรสุ่มคือฟังก์ชันจำนวนจริง** ที่แปลงผลลัพธ์ของการทดลองเป็นค่าจำนวนจริง

ตัวอย่าง: การกำหนดตัวแปรสุ่ม และความน่าจะเป็นของตัวแปรสุ่ม

Example

พิจารณาการทดลองโยนเหรียญที่เที่ยงตรงจำนวน 5 ครั้ง เราสนใจในตัวแปรสุ่ม X ที่บอกถึง "**จำนวนครั้งของการโยนที่ได้ผลออกมาเป็นก้อย**"

- ค่าที่เป็นไปได้ทั้งหมดของ X คือ $0, 1, 2, 3, 4, 5$
- ในการหาความน่าจะเป็นเราสามารถใช้ค่าสัมประสิทธิ์ทวินาม (binomial coefficient)
  - จำนวนผลลัพธ์ที่จะมีก้อย 0 ครั้งจากการโยนทั้งหมด 5 ครั้งเท่ากับ $\binom{5}{0} = \frac{5!}{0!5!} = 1$
  - จำนวนผลลัพธ์ที่จะมีก้อย 3 ครั้งจากการโยนทั้งหมด 5 ครั้งเท่ากับ $\binom{5}{3} = \frac{5!}{3!2!} = 10$
- โดยสรุป การแจกแจงของตัวแปรสุ่ม X สามารถรวบรวมได้ดังนี้
  - $Pr(X \in \{0\}) = \frac{1}{32}$
  - $Pr(X \in \{1\}) = \frac{5}{32}$
  - $Pr(X \in \{2\}) = \frac{10}{32}$
  - $Pr(X \in \{3\}) = \frac{10}{32}$
  - $Pr(X \in \{4\}) = \frac{5}{32}$
  - $Pr(X \in \{5\}) = \frac{1}{32}$



นิยาม: ตัวแปรสุ่มไม่ต่อเนื่องและฟังก์ชันความน่าจะเป็น

Definition

ตัวแปรสุ่ม X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) หรือเป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable) ถ้าค่าจำนวนจริงที่เป็นไปได้ของ X มีจำนวนจำกัด $x_1, \ldots, x_n$ หรืออย่างมากต้องเป็นอนุกรมอนันต์ที่นับได้ $x_1, x_2, \ldots$

Definition (ฟังก์ชันความน่าจะเป็น (probability function))

ถ้าตัวแปรสุ่ม X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) ฟังก์ชันความน่าจะเป็น (probability function หรือ p.f.) ของ X คือฟังก์ชัน f โดยที่สำหรับทุกค่าจำนวนจริง x

$$f(x) = Pr(X = x) \tag{1}$$

นอกจากนี้ เราเรียกส่วนปิดคลุม (closure) ของเซ็ต $\{x : f(x) > 0\}$ ว่า ส่วนค้ำจุน (support) ของ X

ตัวอย่าง: ตัวแปรสุ่มไม่ต่อเนื่องและฟังก์ชันความน่าจะเป็น

Example

ฟังก์ชันความน่าจะเป็น (p.f.) ของตัวแปรสุ่ม X ในตัวอย่างที่แล้ว สามารถเขียนได้เป็น

$$f(0) = \frac{1}{32}, f(1) = \frac{5}{32}, f(2) = \frac{10}{32}, f(3) = \frac{10}{32}, f(4) = \frac{5}{32}, f(5) = \frac{1}{32}$$

ในขณะที่ส่วนค้ำจุน (support) ของ X คือ $\{0, 1, 2, 3, 4, 5\}$

- ผลรวมของฟังก์ชันความน่าจะเป็น (p.f.) จากทุกค่าในส่วนค้ำจุนจะต้องเท่ากับหนึ่ง
  - ค่าแต่ละค่าในส่วนค้ำจุนเป็นผลมาจากเหตุการณ์ที่ไม่มีส่วนร่วมต่อกัน (disjoint) แต่ ทุกค่าในส่วนค้ำจุนรวมกันเป็นปริภูมิตัวอย่าง
  - ผลรวมของความน่าจะเป็นของทุกค่าจากส่วนค้ำจุนจึงเท่ากับความน่าจะเป็นของปริภูมิตัวอย่างซึ่งมีค่าเท่ากับหนึ่ง

คุณสมบัติของฟังก์ชันความน่าจะเป็น (p.f.)

Theorem

กำหนดให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable) ที่มี f แทนฟังก์ชันความน่าจะเป็น (probability function) ถ้า x ไม่ใช่ค่าที่เป็นไปได้ $f(x) = 0$ และ ถ้าอนุกรม $x_1, x_2, \ldots$ ประกอบด้วยค่าที่เป็นไปได้ทั้งหมด แล้ว

$$\sum_{i=1}^{\infty} f(x_i) = 1 \tag{2}$$

Theorem

ถ้า X มีการแจกแจงไม่ต่อเนื่อง (discrete distribution) แล้ว เราสามารถคำนวณหาค่าความน่าจะเป็นของสับเซ็ตของจำนวนจริง $C \subset R$ ได้จาก

$$Pr(X \in C) = \sum_{x_i \in C} f(x_i) \tag{3}$$

ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable)

Definition

ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีค่าพารามิเตอร์ p คือ ตัวแปรสุ่ม X ที่มีค่าได้แค่สองค่าคือ 0 และ 1 โดยที่ $Pr(X = 1) = p$ และมีการแจกแจงแบบเบอร์นูลี (Bernoulli distribution) ที่มีค่าพารามิเตอร์ p ซึ่งสามารถเขียนแทนได้ด้วย

$$f(x) = p^x (1-p)^{1-x}, \text{ สำหรับ } x = 0, 1 \tag{4}$$

- สังเกตได้ว่า การระบุเพียงชื่อของตัวแปรสุ่มหรือการแจกแจงไม่เพียงพอที่จะบอกถึงคุณสมบัติของตัวแปรสุ่ม
  - ค่าพารามิเตอร์ที่ต่างกันย่อมเป็นตัวแปรสุ่มคนละตัวกัน
- สิ่งที่ต้องการทราบเพื่อบอกถึงคุณสมบัติของตัวแปรสุ่มใดๆ คือ “**การแจกแจง**” ซึ่งในที่นี้สามารถทราบได้จาก
  - ชื่อของตัวแปรสุ่ม
  - ค่าพารามิเตอร์ที่เกี่ยวข้อง

ตัวแปรสุ่มทวินาม (binomial random variable)

- ตัวแปรสุ่มไม่ต่อเนื่องที่มีชื่อเสียงอีกอันหนึ่งคือ "**ตัวแปรสุ่มทวินาม (binomial random variable)**"
  - ที่มีพารามิเตอร์ n และ p
  - เกิดจากการสุ่มเลือกตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีพารามิเตอร์ p ต่อเนื่องกันจำนวน n ครั้ง
  - ดังนั้น การแจกแจงแบบตัวแปรสุ่มทวินาม (binomial random variable) มีพารามิเตอร์ p ต่อเนื่องกันจำนวน n ครั้ง คือ

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ สำหรับ } x = 0, 1, ..., n \tag{5}$$

ตัวอย่าง: ตัวแปรสุ่มทวินาม (binomial random variable)

Example

พิจารณาการตรวจสอบคุณภาพของบะหมี่กึ่งสำเร็จรูป โดยที่ความน่าจะเป็นที่จะเกิดผลิตภัณฑ์ด้อยคุณภาพเท่ากับ p ดังนั้นความน่าจะเป็นที่จะได้ผลิตภัณฑ์ที่ได้คุณภาพเท่ากับ $1-p$ สมมุติว่าเจ้าหน้าที่ทำการตรวจสอบผลิตภัณฑ์ทั้งหมด n ครั้ง และเหตุการณ์ที่บอกถึงผลการตรวจสอบเป็นอิสระต่อกันหรือการตรวจสอบแต่ละครั้งเป็นอิสระต่อกัน

- ผลการทดสอบจะพบว่า
  - มีผลิตภัณฑ์ด้อยคุณภาพทั้งหมด x ครั้ง จาก n ครั้ง และความน่าจะเป็นที่จะเกิดผลิตภัณฑ์ด้อยคุณภาพแต่ละชิ้นเท่ากับ p
  - อีก $n-x$ ชิ้นจะพบผลิตภัณฑ์ที่ได้คุณภาพ และความน่าจะเป็นที่จะเกิดผลิตภัณฑ์ที่ได้คุณภาพแต่ละชิ้นเท่ากับ $1-p$
  - ความน่าจะเป็นของแต่ละรูปแบบเท่ากับ $p^x\left(1-p\right)^{n-x}$

ความสัมพันธ์ระหว่างตัวแปรสุ่มแบบเบอร์นูลีและตัวแปรสุ่มแบบทวินาม

- การตรวจสอบแต่ละครั้งคือ ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีพารามิเตอร์ p หนึ่งตัว
- การตรวจสอบทั้งหมด n ครั้งจึงเสมือนกับ ตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) ที่มีพารามิเตอร์ p และเป็นอิสระต่อกันทั้งหมด n ตัว ดังนั้น ความน่าจะเป็นของผลการตรวจสอบ n ครั้งมีค่าเท่ากับ

$$\begin{aligned}
Pr\left(Y_1 = y_1, \ldots, Y_n = y_n\right) &= Pr\left(Y_1 = y_1\right) \cdots Pr\left(Y_n = y_n\right) \\
&= p^{y_1}\left(1-p\right)^{1-y_1} \cdots p^{y_n}\left(1-p\right)^{1-y_n} \\
&= p^{\sum_{i=1}^{n} y_i}\left(n-p\right)^{1-\sum_{i=1}^{n} y_i}
\end{aligned}$$

- กรณีที่มีผลิตภัณฑ์ด้อยคุณภาพทั้งหมด x ชิ้น จะได้ว่า $\sum_{i=1}^{n} y_i = x$ และ ความน่าจะเป็นของรูปแบบนี้มีค่าเท่ากับ $p^x\left(1-p\right)^{n-x}$
- ตัวแปรสุ่มแบบทวินามที่มีพารามิเตอร์ n และ p เกิดจากตัวแปรสุ่มเบอร์นูลีที่มีพารามิเตอร์ p และเป็นอิสระต่อกันทั้งหมด n ตัว

นิยาม: ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variables)

Definition

ตัวแปรสุ่ม X เป็นตัวแปรสุ่มต่อเนื่อง (continuous random variable) และมีการแจกแจงต่อเนื่อง (continuous distribution) ถ้ามีฟังก์ชันที่ไม่เป็นลบ (non-negative function) f นิยามบนเส้นจำนวนจริง โดยที่ ความน่าจะเป็นที่ X จะมีค่าที่เกิดขึ้นจริงในช่วงจำนวนจริง (interval of real numbres) ใดๆ จะมีค่าเท่ากับผลการอินทิเกรตของ f บนช่วงของจำนวนจริงนั้น นั่นคือ ความน่าจะเป็นที่ X จะมีค่าอยู่ในช่วง $[a, b]$ เท่ากับ

$$Pr\left(a \leq X \leq b\right) = \int_a^b f\left(x\right) dx \tag{6}$$

ส่วนกรณีที่เป็นช่วงที่ไม่มีขอบเขตบน (unbounded above) จะได้ว่า

$$Pr\left(X \geq a\right) = \int_a^\infty f\left(x\right) dx \tag{7}$$

และกรณีที่เป็นช่วงที่ไม่มีขอบเขตล่าง (unbounded below) จะได้ว่า

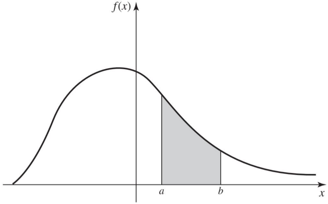
$$Pr\left(X \leq b\right) = \int_{-\infty}^b f\left(x\right) dx \tag{8}$$

นิยาม: ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function)

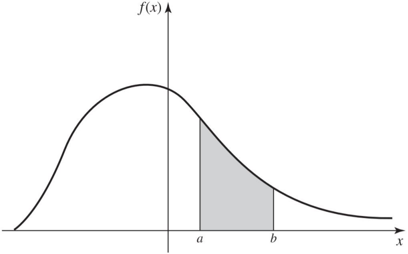
Definition

ถ้าตัวแปรสุ่ม X มีการแจกแจงต่อเนื่อง (continuous distribution) เราจะเรียกฟังก์ชัน f ในนิยามก่อนหน้าว่า ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) หรือเขียนย่อสั้นๆ เป็น p.d.f. นอกจากนี้ เราเรียกส่วนปิดคลุม (closure) ของเซ็ต $\{x : f(x) > 0\}$ ว่า ส่วนค้ำจุน (support) ของ X

- ความสัมพันธ์ระหว่างความน่าจะเป็นและพื้นที่ใต้เส้นฟังก์ชันความหนาแน่นของความน่าจะเป็น
- ความน่าจะเป็นที่ X จะมีค่าอยู่ในช่วง $[a, b]$ มีค่าเท่ากับพื้นที่ที่แรเงาในรูป คือ ผลการอินทิเกรตฟังก์ชันความหนาแน่นของความน่าจะเป็นระหว่างช่วง $[a, b]$



ตัวแปรสุ่มอันหนึ่งสามารถมีฟังก์ชันความหนาแน่นของความน่าจะเป็นได้มากมาย



- ตัวแปรสุ่มอันหนึ่งสามารถมีฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ได้มากมาย โดยทั่วไปเรามักเรียกคุณสมบัตินี้ว่า "**การไม่มีความเป็นหนึ่งเดียวของฟังก์ชันความหนาแน่นของความน่าจะเป็น (non-uniqueness of p.d.f.)**" อย่างไรก็ตาม ถึงแม้ว่าจะมีฟังก์ชันจำนวนมากมายที่แทนตัวแปรสุ่มตัวหนึ่งได้ แต่เรามักจะเลือกใช้อันที่เป็นฟังก์ชันต่อเนื่อง (continuous function) เพื่อให้สะดวกต่อการวิเคราะห์

ตัวอย่าง: ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variables)

Example

พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นดังต่อไปนี้

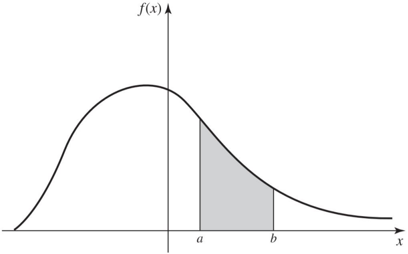
$$f(x) = \begin{cases} \frac{x}{2}, & \text{ถ้า } x \in [0, 2] \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

ความน่าจะเป็นที่ค่าที่เกิดขึ้นจริงจะอยู่ในช่วง $[1, 3]$ เท่ากับ

$$Pr(1 \leq X \leq 3) = \int_1^2 \frac{x}{2} dx + \int_2^3 0 dx = \left. \frac{x^2}{4} \right|_1^2 = \frac{3}{4}$$

- ในทางปฏิบัติ เรามักจะอ้างถึงคุณสมบัติที่สำคัญสองประการของฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) ดังต่อไปนี้

ความน่าจะเป็นเท่ากับศูนย์ไม่ได้หมายความว่าเป็นไปไม่ได้



- ความน่าจะเป็นของค่าจำนวนจริงค่าใดค่าหนึ่งมีค่าเท่ากับศูนย์หากฟังก์ชันเป็นแบบต่อเนื่อง เพราะพื้นที่ใต้กราฟของฟังก์ชันต่อเนื่อง ณ จุดใดจุดหนึ่งมีค่าเท่ากับศูนย์เสมอ นั่นคือ $Pr\left(a \leq X \leq a\right) = Pr\left(X = a\right) = \int_a^a f\left(x\right) dx = 0$
- ไม่ได้หมายความว่า $X = a$ เป็นไปไม่ได้ ดังที่ได้อภิปรายมาแล้ว อย่างไรก็ตาม ความน่าจะเป็นระหว่างช่วง $[a - \epsilon, a + \epsilon]$ สำหรับ $\epsilon > 0$ ใดๆ ย่อมมีค่ามากกว่าศูนย์เสมอ นั่นคือ $Pr\left(a - \epsilon \leq X \leq a + \epsilon\right) = \int_{a-\epsilon}^{a+\epsilon} f\left(x\right) dx \approx 2\epsilon f\left(a\right) > 0$

คุณสมบัติพื้นฐานของตัวแปรสุ่มต่อเนื่อง

Theorem

ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) f ใดๆ จะต้องสอดคล้องกับเงื่อนไขสองข้อต่อไปนี้

$$f(x) \geq 0, \ \text{สำหรับทุกค่าจำนวนจริง } x, \tag{9}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{10}$$

- ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ไม่มีขอบเขตบน ซึ่ง “**แตกต่าง**” จากความน่าจะเป็นที่มีทั้งขอบเขตล่างที่ต้องมีค่ามากกว่าศูนย์และขอบเขตบนที่ต้องมีค่าไม่เกินหนึ่ง
- ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) “**ไม่ใช่ความน่าจะเป็น**” ทำให้ไม่จำเป็นต้องมีค่าน้อยกว่าหนึ่ง
- ประโยชน์อย่างหนึ่งของทฤษฎีบทนี้ คือ หาค่าคงที่มาตรฐาน (normalizing constant) สำหรับฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ดังตัวอย่างต่อไปนี้

ตัวอย่างการหาค่าคงที่มาตรฐาน (normalizing constant) ของการแจกแจงเอกรูป(uniform distribution)

Example (การแจกแจงเอกรูป (uniform distribution))

พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นดังต่อไปนี้

$$f(x) = \begin{cases} c, & \text{ถ้า } x \in [a, b] \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

คำถามก็คือ ค่าคงที่มาตรฐาน (normalizing constant) c ต้องมีค่าเท่าใดฟังก์ชัน f จึงจะเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.)?

- จากสมการที่ 9 ส่งผลให้ค่าคงที่ c จะต้องไม่น้อยกว่าศูนย์ ดังนั้น $c \geq 0$
- ส่วนสมการที่ 10 มีผลทำให้

$$\int_a^b c dx = 1 \Rightarrow c(b-a) = 1 \Rightarrow c = \frac{1}{b-a}$$

- ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของตัวแปรสุ่มที่มีการแจกแจงเอกรูป (uniform distribution) คือ

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{ถ้า } x \in [a, b] \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

ตัวอย่างการหาค่าคงที่มาตรฐาน (normalizing constant) ของการแจกแจงแบบปกติ (normal distribution)

Example (การแจกแจงแบบปกติ (normal distribution))

พิจารณาตัวแปรสุ่ม X ที่มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) เป็นแบบปกติ (normal) ดังต่อไปนี้

$$f(x) = ce^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

โดยที่ μ คือค่าจำนวนจริงที่แทนค่าความคาดหวัง (expectation) และ σ คือค่าจำนวนจริงที่เป็นบวกที่แทนค่าความเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งจะอภิปรายอย่างละเอียดในบท Expectation and Moments โดยหลักการ เราสามารถหาค่าคงที่มาตรฐาน (normalizing constant) สำหรับการแจกแจงแบบปกติ (normal distribution) ได้โดยใช้สมการที่ 10 ดังต่อไปนี้

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} ce^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = c\sigma \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2} d\left(\frac{x-\mu}{\sigma}\right) \\ &= c\sigma\sqrt{2\pi} \Rightarrow c = \frac{1}{\sqrt{2\pi\sigma^2}} \end{aligned}$$

ตัวอย่างการหาค่าคงที่มาตรฐาน (normalizing constant) ของการแจกแจงแบบปกติ (normal distribution): การใช้เทคนิคพิกัดเชิงขั้ว

Example (การแจกแจงแบบปกติ (normal distribution))

โดยในที่นี้ เราได้ประยุกต์ใช้เทคนิคการใช้พิกัดเชิงขั้ว (polar coordinates) เพื่อหาแสดงว่า $\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$ ซึ่งมีขั้นตอนดังต่อไปนี้ เริ่มด้วยการกำหนดให้

$$I = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

ดังนั้น

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{y^2+z^2}{2}} dz dy$$

ขั้นตอนต่อไปคือการแปลง (y, z) ให้อยู่ในรูปของพิกัดเชิงขั้ว (polar coordinates) โดยใช้

$$y = r\cos\theta, z = r\sin\theta$$

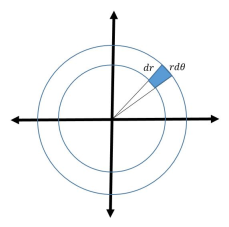
ตัวอย่างการหาค่าคงที่มาตรฐาน (normalizing constant) ของการแจกแจงแบบปกติ (normal distribution): สรุปผล

Example (การแจกแจงแบบปกติ (normal distribution))

$$y^2 + z^2 = r^2, dzdy = rdrd\theta$$

ดังนั้น

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{y^2+z^2}{2}} dzdy = \int_{0}^{2\theta}\int_{0}^{\infty} e^{-\frac{r^2}{2}} rdrd\theta \\
&= \int_{0}^{2\theta}\int_{0}^{\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) d\theta = \int_{0}^{2\theta} -e^{-\frac{r^2}{2}}\Big|_{0}^{\infty} d\theta \\
&= \int_{0}^{2\theta} d\theta = 2\pi \\
I &= \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}
\end{aligned}
$$



ฟังก์ชันความน่าจะเป็นสะสม (The Cumulative Distribution Function)

- ที่ผ่านมา เราอธิบายการแจกแจงของตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่องด้วยฟังก์ชันที่แตกต่างกัน ซึ่งในทางคณิตศาสตร์แล้วไม่ค่อยสะดวกนัก
- ฟังก์ชันที่สามารถอธิบายการแจกแจงของตัวแปรสุ่มทั้งสองประเภทให้ได้ โดยเรียกฟังก์ชันอันใหม่นี้ว่า ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function)
  - มีคุณสมบัติที่ตัวแปรสุ่มตัวใดตัวหนึ่งจะมีฟังก์ชันความน่าจะเป็นสะสมเพียงอันเดียว (uniqueness)
  - ฟังก์ชันความน่าจะเป็นสะสมเป็นจุดเชื่อมระหว่างทฤษฎีความน่าจะเป็นพื้นฐานและทฤษฎีมาตรวัดความน่าจะเป็น (probability measure theory) ซึ่งใช้ฟังก์ชันความน่าจะเป็นสะสมเป็นมาตรวัด (measure) สำหรับการอินทิเกรต

นิยาม: ฟังก์ชันความน่าจะเป็นสะสม (The Cumulative Distribution Function)

Definition (ฟังก์ชันความน่าจะเป็นสะสม (Cumulative Distribution Function: C.D.F.))

ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) หรือฟังก์ชันความน่าจะเป็น (distribution function) ของตัวแปรสุ่ม X ซึ่งเรียกสั้นๆ ว่า C.D.F. ของ X คือฟังก์ชัน

$$F(x) = Pr(X \leq x), \text{ สำหรับ } -\infty < x < \infty \tag{11}$$

- ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) ที่นิยามนี้ใช้ได้กับทั้งตัวแปรสุ่มต่อเนื่องและไม่ต่อเนื่อง

ตัวอย่าง: ฟังก์ชันความน่าจะเป็นสะสม (The Cumulative Distribution Function)

Example

ตัวอย่างนี้แสดงให้เห็นถึงวิธีการแปลงฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ให้เป็นฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) โดยใช้ตัวแปรสุ่ม X ที่มีการแจกแจงดังต่อไปนี้

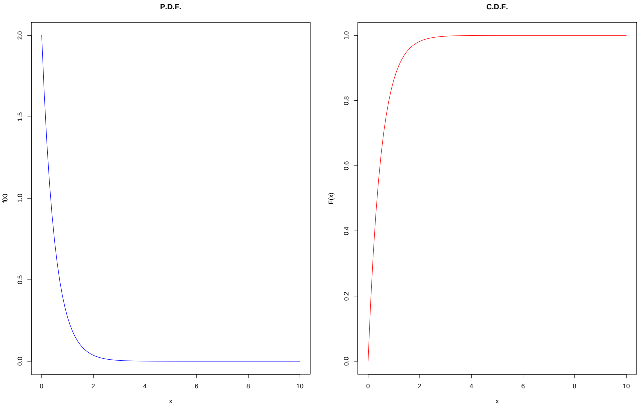
$$f(x) = \begin{cases} 2e^{-2x}, & \text{ถ้า } x > 0 \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

- ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) หรือ C.D.F. ของตัวแปรสุ่ม X คือฟังก์ชัน

$$F(x) = Pr(X \leq x) = \begin{cases} \int_0^x 2e^{-2y}dy = 1 - e^{-2x}, & \text{สำหรับ } x > 0 \\ 0, & \text{ถ้าเป็นอย่างอื่น} \end{cases}$$

- สังเกตว่า เพื่อความสะดวกจึงเปลี่ยนตัวแปรสำหรับการอินทิเกรตจาก x เป็น y เพราะฟังก์ชันความน่าจะเป็นสะสม หรือ C.D.F. นิยามโดยใช้ x ไปแล้ว

รูปของฟังก์ชันความน่าจะเป็นสะสม (The Cumulative Distribution Function)



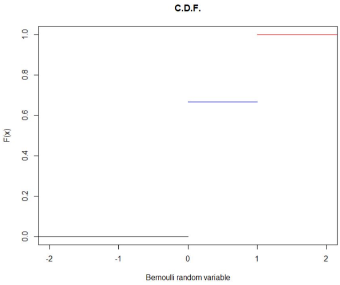
ตัวอย่าง: ฟังก์ชันความน่าจะเป็นสะสมของตัวแปรสุ่มเบอร์นูลี

Example

พิจารณาตัวแปรสุ่มเบอร์นูลี (Bernoulli random variable) X ที่มีพารามิเตอร์ p ดังนั้น $Pr(X = 0) = 1 - p$ และ $Pr(X = 1) = p$ ในกรณีนี้ เราสามารถเขียนฟังก์ชันความน่าจะเป็นสะสม หรือ C.D.F. ได้เป็น

$$F(x) = \begin{cases} 0, & \text{สำหรับ } x < 0 \\ 1 - p, & \text{สำหรับ } 0 \leq x < 1 \\ 1, & \text{สำหรับ } x \geq 1 \end{cases}$$

ตัวอย่างเช่น เมื่อกำหนด เมื่อ $p = \frac{1}{3}$



คุณสมบัติพื้นฐานของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.)

Theorem

ฟังก์ชัน $F(x)$ *เป็นฟังก์ชันไม่ลดลง (decreasing function) เมื่อค่า* x *เพิ่มขึ้น ซึ่งหมายความว่า ถ้า* $x_1 < x_2$ *แล้ว* $F(x_1) \leq F(x_2)$

Proof.

ถ้า $x_1 < x_2$ แล้ว เหตุการณ์ $\{X \leq x_1\}$ จะเป็นสับเซ็ตของ $\{X \leq x_2\}$ ดังนั้น
$Pr(X \leq x_1) \leq Pr(X \leq x_2)$ ซึ่งหมายความว่า $F(x_1) \leq F(x_2)$ $\square$

Theorem

$\lim_{x\to-\infty} F(x) = 0$ *และ* $\lim_{x\to\infty} F(x) = 1$

คุณสมบัติพื้นฐานของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) เมื่อ ตัวแปรสุ่มมีค่ามากกว่าค่าที่สนใจ

Theorem

สำหรับค่าจำนวนจริง x ใดๆ

$$Pr\left(X > x\right) = 1 - F\left(x\right) \tag{12}$$

Proof.

เริ่มจากการที่เหตุการณ์ $\{X > x\}$ และ $\{X \leq x\}$ ไม่มีส่วนร่วมกัน (disjoint) และยูเนียนของทั้งสองเหตุการณ์ครอบคลุมปริภูมิตัวอย่าง (sample space) นั่นคือทั้งสองเหตุการณ์นี้เป็นการแบ่งส่วน (partition) ของปริภูมิตัวอย่าง ดังนั้น

$$Pr\left(X \leq x\right) + Pr\left(X > x\right) = 1 \Rightarrow Pr\left(X > x\right) = 1 - F\left(x\right)$$

□

คุณสมบัติพื้นฐานของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) เมื่อ ตัวแปรสุ่มอยู่ระหว่างค่าที่สนใจ

Theorem

สำหรับค่าจำนวนจริง x_1 *และ* x_2 *ที่* $x_1 < x_2$

$$Pr\left(x_1 < X \leq x_2\right) = F\left(x_2\right) - F\left(x_1\right) \tag{13}$$

Proof.

กำหนดให้ $A = \{x_1 < X \leq x_2\}$, $B = \{x_1 \leq X\}$ และ $C = \{X \leq x_2\}$ เมื่อพิจารณาอย่างรอบคอบจะเห็นได้ว่า เหตุการณ์ A และ B เป็นการแบ่งส่วน (partition) ของ C นั่นคือ $A \cap B = \emptyset$ และ $A \cup B = C$ ดังนั้น

$$Pr\left(x_1 < X \leq x_2\right) + Pr\left(X \leq x_1\right) = Pr\left(X \leq x_2\right) \Rightarrow Pr\left(x_1 < X \leq x_2\right) = F\left(x_2\right) - F\left(x_1\right)$$

ซึ่งเป็นผลมาจากการแทนค่า $Pr\left(X \leq x_n\right) = F\left(x_n\right)$ $\square$

ทฤษฎีความต่อเนื่องของฟังก์ชันความน่าจะเป็นสะสม (C.D.F.)

Theorem

กำหนดให้ตัวแปรสุ่ม X มีการกระจายต่อเนื่อง โดยมี $f(x)$ และ $F(x)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function) และฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function) ดังนั้น

1. *ฟังก์ชันความน่าจะเป็นสะสม (cumulative distribution function)*

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \tag{14}$$

   และมีความต่อเนื่อง (continuous) ทุกๆ ค่าจำนวนจริง x

2. *ฟังก์ชันความหนาแน่นของความน่าจะเป็น (probability density function)*

$$f(x) = \frac{dF(x)}{dx} \tag{15}$$

   สำหรับทุกๆ ค่าจำนวนจริง x ที่ฟังก์ชัน $f(x)$ มีความต่อเนื่อง (continuous)

ตัวอย่าง: การหา P.D.F. จาก C.D.F.

Example

พิจารณาตัวแปรสุ่ม X ที่มีการแจกแจงแบบเลขชี้กำลัง (exponential distribution) ด้วยค่าพารามิเตอร์ $\lambda > 0$ ดังนี้

$$F(x) = \begin{cases} 0, \text{ สำหรับ } x < 0 \\ 1 - e^{-\lambda x}, \text{ สำหรับ } x \geq 0 \end{cases}$$

ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) ของตัวแปรสุ่ม X ที่มีการแจกแจงแบบเลขชี้กำลัง คือ

$$f(x) = \begin{cases} 0, \text{ สำหรับ } x < 0 \\ \frac{d\left[1-e^{-\lambda x}\right]}{dx} = \lambda e^{-\lambda x}, \text{ สำหรับ } x \geq 0 \end{cases}$$

บทเรียน: ฟังก์ชันความน่าจะเป็นสะสม (C.D.F.) ของตัวแปรสุ่มต่อเนื่องนั้นมีความต่อเนื่องที่ทุกค่าจำนวนจริง ในขณะที่ ฟังก์ชันความหนาแน่นของความน่าจะเป็น (p.d.f.) อาจจะไม่มีความต่อเนื่องที่บางค่าได้