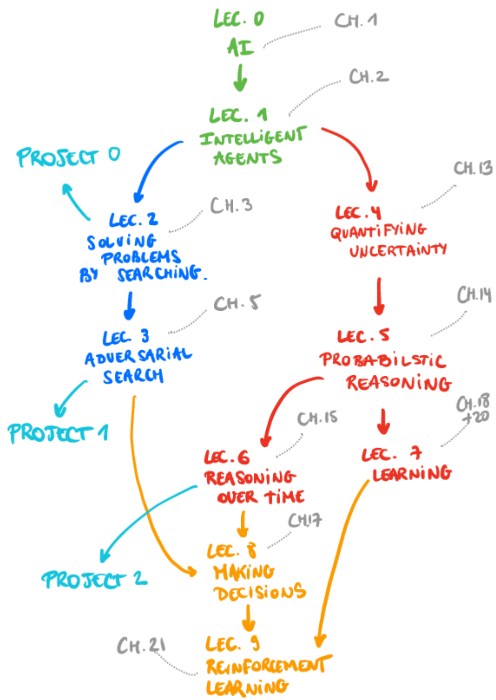
ปัญญาประดิษฐ์เบื้องต้น

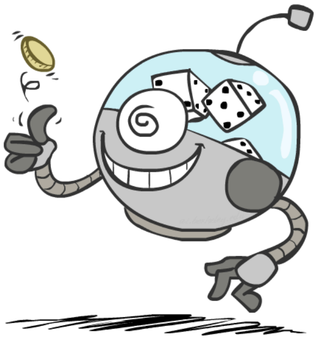
บรรยาย 4: การวัดปริมาณความไม่แน่นอน (Quantifying uncertainty)

ผศ. ดร. อิทธิพล ฟองแก้ว
[ittipon@g.sut.ac.th]


LEC. 0
AI
CH. 1
LEC. 1
INTELLIGENT AGENTS
CH. 2
PROJECT 0
LEC. 2
SOLVING PROBLEMS BY SEARCHING.
CH. 3
LEC. 3
ADVERSARIAL SEARCH
CH. 5
PROJECT 1
LEC. 4
QUANTIFYING UNCERTAINTY
CH. 13
LEC. 5
PROBABILSTIC REASONING
CH. 14
LEC. 6
REASONING OVER TIME
CH. 15
LEC. 7
LEARNING
CH. 18 +20
PROJECT 2
LEC. 8
MAKING DECISIONS
CH. 17
LEC. 9
REINFORCEMENT LEARNING
CH. 21

วันนี้เราจะเรียนเรื่อง

- Random variables (ตัวแปรสุ่ม)
- Probability distributions (การแจกแจงความน่าจะเป็น)
- Inference (การอนุมาน)
- Independence (ความเป็นอิสระ)
- The Bayes' rule (กฎของเบย์)



อย่ามองข้ามการบรรยายนี้!

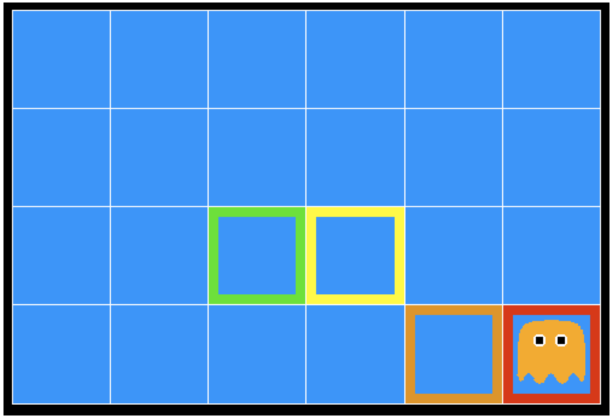
ที่มา: CS188, UC Berkeley.

การวัดปริมาณความไม่แน่นอน (Quantifying uncertainty)

มีผี ซ่อน อยู่ที่ไหนสักแห่งในตาราง

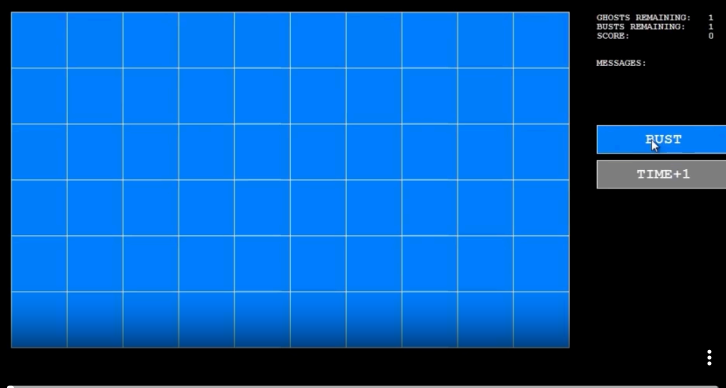
ค่าที่อ่านได้จากเซ็นเซอร์บอกว่าช่องสี่เหลี่ยมอยู่ใกล้กับผีแค่ไหน:

- ตรงตำแหน่งผี: สีแดง
- ห่าง 1 หรือ 2 ช่อง: สีส้ม
- ห่าง 3 ช่อง: สีเหลือง
- ห่าง 4+ ช่อง: สีเขียว

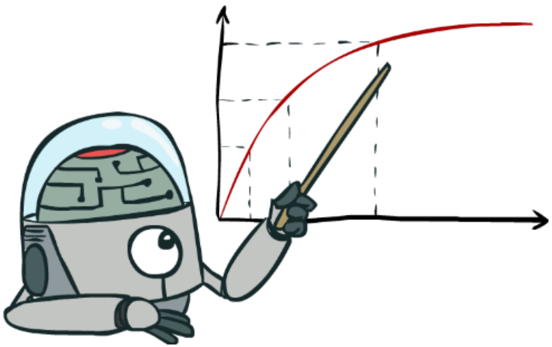


เซ็นเซอร์มี สัญญาณรบกวน (noisy) แต่เรารู้ค่าความน่าจะเป็น $P(\text{color}|\text{distance})$ สำหรับทุกสีและทุกระยะทาง

—
ที่มา: CS188, UC Berkeley.



ที่มา: CS188, UC Berkeley.



หลักการของอรรถประโยชน์คาดหวังสูงสุด (Principle of maximum expected utility)

เอเจนต์จะถูกเรียกว่ามีเหตุผล (rational) ถ้าหากเลือกกระทำที่ให้ อรรถประโยชน์คาดหวัง (expected utility) สูงสุด โดยเฉลี่ยจากผลลัพธ์ที่เป็นไปได้ทั้งหมดของการกระทำนั้น

คำว่า "คาดหวัง" (expected) หมายถึงอะไรกันแน่?

ความไม่แน่นอน (Uncertainty)

สถานการณ์ทั่วไป:

- ตัวแปรที่สังเกตได้หรือหลักฐาน (Observed variables or evidence): เอเจนต์รู้ข้อมูลบางอย่างเกี่ยวกับสถานะของโลก (เช่น ค่าที่อ่านได้จากเซ็นเซอร์)
- ตัวแปรที่สังเกตไม่ได้ (Unobserved variables): เอเจนต์ต้องให้เหตุผลเกี่ยวกับแง่มุมอื่นๆ ที่ไม่แน่นอน (เช่น ผีอยู่ที่ไหน)
- แบบจำลอง (เชิงความน่าจะเป็น) (Probabilistic model): เอเจนต์รู้หรือเชื่อบางอย่างเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้กับตัวแปรที่สังเกตไม่ได้

การให้เหตุผลเชิงความน่าจะเป็น (Probabilistic reasoning) เป็นกรอบการทำงานสำหรับจัดการความรู้และความเชื่อของเรา

การยืนยันเชิงความน่าจะเป็น (Probabilistic assertions)

การยืนยันเชิงความน่าจะเป็นแสดงถึงการที่เอเจนต์ไม่สามารถตัดสินใจได้อย่างชัดเจนเกี่ยวกับความจริงของข้อเสนอหนึ่งๆ

- ค่าความน่าจะเป็น สรุป ผลกระทบของ
  - ความไม่รู้ (ignorance) (ทางทฤษฎี, ทางปฏิบัติ)
  - ความขี้เกียจ (laziness) (ขาดเวลา, ทรัพยากร)
- ความน่าจะเป็นเชื่อมโยงข้อเสนอกับสถานะความรู้ (หรือการขาดความรู้) ของตนเอง
  - เช่น $\mathrm{P}(\text{ผีในช่อง } [3,2]) = 0.02$

Frequentism vs. Bayesianism

ค่าความน่าจะเป็นแสดงถึงอะไร?

- มุมมองของ frequentist ที่เป็นรูปธรรมคือ ความน่าจะเป็นเป็นแง่มุมที่แท้จริงของจักรวาล
    - กล่าวคือ แนวโน้มของวัตถุที่จะมีพฤติกรรมในลักษณะบางอย่าง
    - เช่น การที่เหรียญเที่ยงตรงจะออกหัวด้วยความน่าจะเป็น 0.5 เป็นแนวโน้มของเหรียญเอง
- มุมมองของ Bayesian ที่เป็นอัตวิสัยคือ ความน่าจะเป็นเป็นวิธีบ่งบอกความเชื่อหรือความไม่แน่นอนของเอเจนต์
    - กล่าวคือ ความน่าจะเป็นไม่มีนัยสำคัญทางกายภาพภายนอก
    - นี่คือการตีความความน่าจะเป็นที่เราจะใช้!

เราจะกำหนดค่าตัวเลขให้กับความเชื่อได้อย่างไร?

สัจพจน์ของคอลโมโกรอฟ (Kolmogorov's axioms)

เริ่มต้นด้วยเซต Ω ซึ่งคือ ปริภูมิตัวอย่าง (sample space)

$\omega \in \Omega$ คือ จุดตัวอย่าง (sample point) หรือโลกที่เป็นไปได้

ปริภูมิความน่าจะเป็น (probability space) คือปริภูมิตัวอย่างที่มีฟังก์ชันความน่าจะเป็นติดตั้งอยู่ด้วย กล่าวคือ การกำหนดค่า $\mathrm{P} : \mathcal{P}(\Omega) \to \mathbb{R}$ โดยที่:

- สัจพจน์ข้อที่ 1: $\mathrm{P}(\omega) \in \mathbb{R}, 0 \leq \mathrm{P}(\omega)$ สำหรับทุก $\omega \in \Omega$
- สัจพจน์ข้อที่ 2: $\mathrm{P}(\Omega) = 1$
- สัจพจน์ข้อที่ 3: $\mathrm{P}(\{\omega_1, \ldots, \omega_n\}) = \sum_{i=1}^{n} \mathrm{P}(\omega_i)$ สำหรับเซตของตัวอย่างใดๆ

โดยที่ $\mathcal{P}(\Omega)$ คือเพาเวอร์เซตของ Ω

ตัวอย่าง

- Ω = ผลลัพธ์ที่เป็นไปได้ 6 อย่างจากการทอยลูกเต๋า
- ω_i (สำหรับ $i = 1, \ldots, 6$) คือจุดตัวอย่าง ซึ่งแต่ละจุดสอดคล้องกับผลลัพธ์ของลูกเต๋า
- การกำหนดค่า P สำหรับลูกเต๋าที่เที่ยงตรง:

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

ตัวแปรสุ่ม (Random variables)

- ตัวแปรสุ่ม (random variable) คือฟังก์ชัน $X : \Omega \to D_X$ จากปริภูมิตัวอย่างไปยังโดเมนบางอย่างที่กำหนดผลลัพธ์ของมัน
  - เช่น $\mathrm{Odd} : \Omega \to \{\mathrm{true}, \mathrm{false}\}$ โดยที่ $\mathrm{Odd}(\omega) = (\omega \bmod 2 = 1)$
- P ก่อให้เกิด การแจกแจงความน่าจะเป็น (probability distribution) สำหรับตัวแปรสุ่ม X ใดๆ
  - $P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$
  - เช่น $P(\mathrm{Odd} = \mathrm{true}) = P(1) + P(3) + P(5) = \frac{1}{2}$
- เหตุการณ์ (event) E คือเซตของผลลัพธ์ $\{(x_1, \ldots, x_n), \ldots\}$ ของตัวแปร $X_1, \ldots, X_n$ โดยที่

$$P(E) = \sum_{(x_1,\ldots,x_n) \in E} P(X_1 = x_1, \ldots, X_n = x_n)$$

สัญกรณ์ (Notations)

- ตัวแปรสุ่มจะเขียนด้วยอักษรโรมันตัวใหญ่: X, Y , ฯลฯ
- การปรากฏ (realization) ของตัวแปรสุ่มจะเขียนด้วยอักษรตัวเล็กที่สอดคล้องกัน เช่น x_1, x_2, ..., x_n อาจเป็นผลลัพธ์ของตัวแปรสุ่ม X
- ค่าความน่าจะเป็นของการปรากฏ x จะเขียนเป็น $P(X = x)$
- เมื่อชัดเจนจากบริบท จะย่อเป็น $P(x)$
- การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม (แบบไม่ต่อเนื่อง) X จะเขียนแทนด้วย $\mathbf{P}(X)$ ซึ่งสอดคล้องกับเวกเตอร์ของตัวเลข โดยแต่ละตัวสำหรับค่าความน่าจะเป็น $P(X = x_i)$ (และไม่ใช่ค่าสเกลาร์ค่าเดียว!)

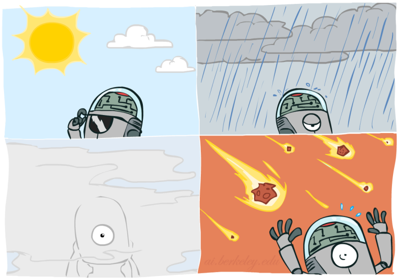
การแจกแจงความน่าจะเป็น (Probability distributions)

สำหรับตัวแปรแบบไม่ต่อเนื่อง การแจกแจงความน่าจะเป็น สามารถเข้ารหัสได้ด้วยรายการของความน่าจะเป็นของผลลัพธ์ต่างๆ หรือที่เรียกว่า ฟังก์ชันมวลของความน่าจะเป็น (probability mass function)

เราสามารถคิดว่าการแจกแจงความน่าจะเป็นเป็น ตาราง ที่เชื่อมโยงค่าความน่าจะเป็นกับแต่ละ ผลลัพธ์ ของตัวแปร

$\mathbf{P}(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0



ที่มา: CS188, UC Berkeley.

การแจกแจงร่วม (Joint distributions)

การแจกแจงความน่าจะเป็นร่วม (joint probability distribution) ของชุดตัวแปรสุ่ม $X_1, ..., X_n$ จะระบุความน่าจะเป็นของแต่ละผลลัพธ์ (ที่รวมกัน):

$$P(X_1 = x_1, ..., X_n = x_n) = \sum_{\{\omega : X_1(\omega) = x_1, ..., X_n(\omega) = x_n\}} P(\omega)$$

$$\mathbf{P}(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

การแจกแจงตามขอบ (Marginal distributions)

การแจกแจงตามขอบ (marginal distribution) ของเซตย่อยของกลุ่มตัวแปรสุ่ม คือการแจกแจงความน่าจะเป็นร่วมของตัวแปรที่อยู่ในเซตย่อยนั้น

$\mathbf{P}(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$\mathbf{P}(T)$

T	P
hot	0.5
cold	0.5

$$P(t) = \sum_w P(t, w)$$

$\mathbf{P}(W)$

W	P
sun	0.6
rain	0.4

$$P(w) = \sum_t P(t, w)$$

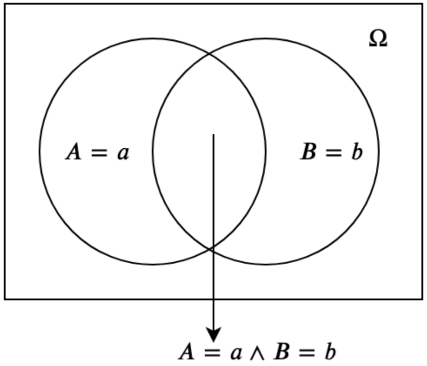
ตามสัญชาตญาณ การแจกแจงตามขอบคือตารางย่อยที่กำจัดตัวแปรบางตัวออกไป

การแจกแจงมีเงื่อนไข (Conditional distributions)

ความน่าจะเป็นมีเงื่อนไข (conditional probability) ของการปรากฏ a เมื่อกำหนดให้มีการปรากฏ b ถูกนิยามว่าเป็นอัตราส่วนของความน่าจะเป็นของการปรากฏร่วมกันของ a และ b ต่อความน่าจะเป็นของ b:

$$\mathrm{P}\left(\mathrm{a}|\mathrm{b}\right) = \frac{\mathrm{P}\left(\mathrm{a},\mathrm{b}\right)}{\mathrm{P}\left(\mathrm{b}\right)}.$$

แท้จริงแล้ว การสังเกต $\mathrm{B} = \mathrm{b}$ จะตัดโลกที่เป็นไปได้ทั้งหมดที่ $\mathrm{B} \neq \mathrm{b}$ ออกไป เหลือเพียงเซตที่มีความน่าจะเป็นรวมเป็น $\mathrm{P}\left(\mathrm{b}\right)$ ภายในเซตนั้น โลกที่ $\mathrm{A} = \mathrm{a}$ จะสอดคล้องกับ $\mathrm{A} = \mathrm{a} \wedge \mathrm{B} = \mathrm{b}$ และคิดเป็นสัดส่วน $\mathrm{P}\left(\mathrm{a},\mathrm{b}\right)/\mathrm{P}\left(\mathrm{b}\right)$



การแจกแจงมีเงื่อนไขคือการแจกแจงความน่าจะเป็นของตัวแปรบางตัว โดยกำหนดให้ค่าของตัวแปรอื่น คงที่

$$\mathbf{P}(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$\mathbf{P}(W \mid T = \text{hot})$$

W	P
sun	0.8
rain	0.2

$$\mathbf{P}(W \mid T = \text{cold})$$

W	P
sun	0.4
rain	0.6

การอนุมานเชิงความน่าจะเป็น (Probabilistic inference)

การอนุมาน เชิงความน่าจะเป็นคือปัญหาของการคำนวณความน่าจะเป็นที่ต้องการจากความน่าจะเป็นอื่นที่ทราบค่าอยู่แล้ว (เช่น การคำนวณความน่าจะเป็นมีเงื่อนไขจากการแจกแจงร่วม)

- โดยทั่วไปเราจะคำนวณความน่าจะเป็นมีเงื่อนไข
  - เช่น $\mathrm{P}\left(\text{ตรงเวลา} | \text{ไม่มีรายงานอุบัติเหตุ}\right) = 0.9$
  - สิ่งเหล่านี้แสดงถึง ความเชื่อ ของเอเจนต์เมื่อพิจารณาจากหลักฐาน
- ความน่าจะเป็นเปลี่ยนแปลงไปตามหลักฐานใหม่:
  - เช่น $\mathrm{P}\left(\text{ตรงเวลา} | \text{ไม่มีรายงานอุบัติเหตุ}, 5\mathrm{AM}\right) = 0.95$
  - เช่น $\mathrm{P}\left(\text{ตรงเวลา} | \text{ไม่มีรายงานอุบัติเหตุ}, \text{ฝนตก}\right) = 0.8$
  - เช่น $\mathrm{P}\left(\text{ผีในช่อง } [3,2] | \text{สีแดงในช่อง } [3,2]\right) = 0.99$
  - การสังเกตหลักฐานใหม่ทำให้ ความเชื่อถูกปรับปรุง

กรณีทั่วไป

- ตัวแปรหลักฐาน (Evidence variables): $\mathrm{E}_1, \ldots, \mathrm{E}_k = \mathrm{e}_1, \ldots, \mathrm{e}_k$
- ตัวแปรที่ต้องการหา (Query variables): Q
- ตัวแปรซ่อนเร้น (Hidden variables): $\mathrm{H}_1, \ldots, \mathrm{H}_r$
- $(\mathrm{Q} \cup \mathrm{E}_1, \ldots, \mathrm{E}_k \cup \mathrm{H}_1, \ldots, \mathrm{H}_r)$ = ตัวแปรทั้งหมด $\mathrm{X}_1, \ldots, \mathrm{X}_n$

การอนุมาน (Inference) คือปัญหาของการคำนวณ $\mathbf{P}(\mathrm{Q} | \mathrm{e}_1, \ldots, \mathrm{e}_k)$

การอนุมานโดยการแจกแจง (Inference by enumeration)

เริ่มต้นจากการแจกแจงร่วม $\mathbf{P}(\mathrm{Q}, \mathrm{E}_1, \ldots, \mathrm{E}_k, \mathrm{H}_1, \ldots, \mathrm{H}_r)$

1. เลือกรายการที่สอดคล้องกับหลักฐาน $\mathrm{E}_1, \ldots, \mathrm{E}_k = \mathrm{e}_1, \ldots, \mathrm{e}_k$
2. กำจัดตัวแปรซ่อนเร้นออกไปโดยการหาผลรวม (Marginalize out) เพื่อให้ได้การแจกแจงร่วมของตัวแปรที่ต้องการหากับตัวแปรหลักฐาน:

$$\mathbf{P}(\mathrm{Q}, \mathrm{e}_1, \ldots, \mathrm{e}_k) = \sum_{\mathrm{h}_1, \ldots, \mathrm{h}_r} \mathbf{P}(\mathrm{Q}, \mathrm{h}_1, \ldots, \mathrm{h}_r, \mathrm{e}_1, \ldots, \mathrm{e}_k)$$

3. ทำให้เป็นมาตรฐาน (Normalize):

$$\mathrm{Z} = \sum_{\mathrm{q}} \mathrm{P}\left(\mathrm{q}, \mathrm{e}_1, \ldots, \mathrm{e}_k\right) \quad \mathbf{P}(\mathrm{Q} \mid \mathrm{e}_1, \ldots, \mathrm{e}_k) = \frac{1}{\mathrm{Z}} \mathbf{P}(\mathrm{Q}, \mathrm{e}_1, \ldots, \mathrm{e}_k)$$

ตัวอย่าง

- $\mathbf{P}(W)$?
- $\mathbf{P}(W \mid \text{winter})$?
- $\mathbf{P}(W \mid \text{winter}, \text{hot})$?

S	T	W	P
summer	hot	sun	0.3
summer	hot	rain	0.05
summer	cold	sun	0.1
summer	cold	rain	0.05
winter	hot	sun	0.1
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.2

ความซับซ้อน (Complexity)

- การอนุมานโดยการแจกแจงสามารถใช้ตอบคำถามเชิงความน่าจะเป็นสำหรับ ตัวแปรแบบไม่ต่อเนื่อง (คือ มีจำนวนค่าจำกัด)
- อย่างไรก็ตาม การแจกแจง ไม่สามารถขยายขนาดได้ (does not scale)!
    - สมมติว่าโดเมนถูกอธิบายด้วยตัวแปร n ตัว ซึ่งแต่ละตัวมีค่าได้ไม่เกิน d ค่า
    - ความซับซ้อนด้านพื้นที่ (Space complexity): $O(d^n)$
    - ความซับซ้อนด้านเวลา (Time complexity): $O(d^n)$

เราสามารถลดขนาดการนำเสนอของการแจกแจงร่วมได้หรือไม่?

กฎผลคูณ (Product rule)

$$P(a, b) = P(b)P(a|b)$$

ตัวอย่าง

$\mathbf{P}(W)$

W	P
sun	0.8
rain	0.2

$\mathbf{P}(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$\mathbf{P}(D, W)$

D	W	P
wet	sun	?
dry	sun	?
wet	rain	?
dry	rain	?

กฎลูกโซ่ (Chain rule)

โดยทั่วไปแล้ว การแจกแจงร่วมใดๆ สามารถเขียนเป็นผลคูณของความน่าจะเป็นมีเงื่อนไขที่เพิ่มขึ้นทีละขั้นได้เสมอ:

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \; P(x_1, ..., x_n) \quad = \prod_{i=1}^{n} P(x_i|x_1, ..., x_{i-1})$$

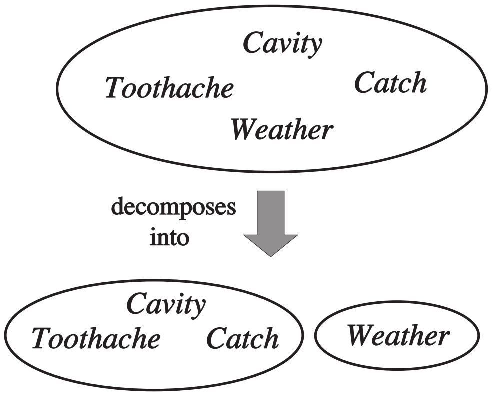
ความเป็นอิสระ (Independence)

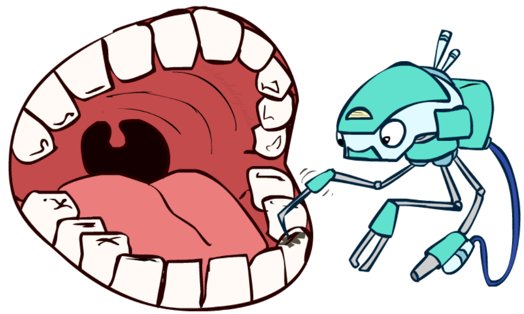
A และ B เป็น อิสระ (independent) ต่อกัน ก็ต่อเมื่อ สำหรับทุก $a \in D_A$ และ $b \in D_B$,

- $P(a|b) = P(a)$, หรือ
- $P(b|a) = P(b)$, หรือ
- $P(a, b) = P(a)P(b)$

ความเป็นอิสระเขียนแทนด้วย $A \perp B$

ตัวอย่างที่ 1





$$\mathrm{P}(\mathrm{toothache}, \mathrm{catch}, \mathrm{cavity}, \mathrm{weather}) = \mathrm{P}(\mathrm{toothache}, \mathrm{catch}, \mathrm{cavity})\mathrm{P}(\mathrm{weather})$$

ตารางเดิมที่มี 32 รายการจะลดลงเหลือตาราง 8 รายการหนึ่งตารางและตาราง 4 รายการอีกหนึ่งตาราง (สมมติว่า $\mathrm{Weather}$ มี 4 ค่า และตัวแปรอื่นเป็นค่าบูลีน)

ที่มา: CS188, UC Berkeley.

ตัวอย่างที่ 2

สำหรับการโยนเหรียญที่เป็นอิสระต่อกัน n ครั้ง การแจกแจงร่วมสามารถ แยกตัวประกอบ (factored) ได้ทั้งหมดและแสดงเป็นผลคูณของตารางขนาด 1 รายการจำนวน n ตาราง

- $2^n \rightarrow n$

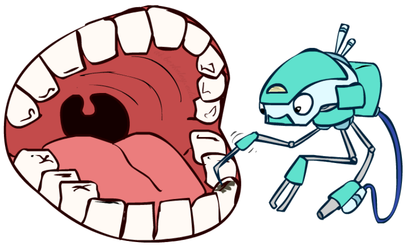
ความเป็นอิสระมีเงื่อนไข (Conditional independence)

A และ B เป็น อิสระมีเงื่อนไข (conditionally independent) ต่อกันเมื่อกำหนด C ก็ต่อเมื่อ สำหรับทุก $a \in D_A$, $b \in D_B$ และ $c \in D_C$,

- $P(a|b,c) = P(a|c)$, หรือ
- $P(b|a,c) = P(b|c)$, หรือ
- $P(a,b|c) = P(a|c)P(b|c)$

ความเป็นอิสระมีเงื่อนไขเขียนแทนด้วย $A \perp B|C$

- การใช้กฎลูกโซ่ ทำให้การแจกแจงร่วมสามารถแยกตัวประกอบเป็นผลคูณของความน่าจะเป็นมีเงื่อนไขได้
- ความน่าจะเป็นมีเงื่อนไขแต่ละตัวอาจสามารถ ทำให้ง่ายขึ้นได้โดยใช้ความเป็นอิสระมีเงื่อนไข
- การยืนยันความเป็นอิสระมีเงื่อนไขช่วยให้แบบจำลองความน่าจะเป็น ขยายขนาดขึ้น (scale up) ได้



ตัวอย่างที่ 1

สมมติตัวแปรสุ่มสามตัว Toothache (ปวดฟัน), Catch (ตรวจพบค้น) และ Cavity (ฟันผุ)

Catch เป็นอิสระมีเงื่อนไขจาก Toothache เมื่อกำหนด Cavity ดังนั้น เราสามารถเขียนได้ว่า:

$$\mathrm{P}(\text{toothache}, \text{catch}, \text{cavity}) \quad = \mathrm{P}(\text{toothache} | \text{catch}, \text{cavity}) \mathrm{P}(\text{catch} | \text{cavity}) \mathrm{P}(\text{cavity}) \;= \mathrm{P}(\text{tootha}$$

ในกรณีนี้ การนำเสนอการแจกแจงร่วมจะลดลงเหลือตัวเลขที่ไม่ขึ้นต่อกัน $2 + 2 + 1$ ตัว (แทนที่จะเป็น $2^n - 1$)

—

ที่มา: CS188, UC Berkeley.

ตัวอย่างที่ 2 (Naive Bayes)

โดยทั่วไป จากกฎผลคูณ เราจะได้

$$\mathrm{P}\left(\mathrm{cause}, \mathrm{effect}_1, \ldots, \mathrm{effect}_n\right) = \mathrm{P}\left(\mathrm{effect}_1, \ldots, \mathrm{effect}_n | \mathrm{cause}\right) \mathrm{P}\left(\mathrm{cause}\right)$$

เมื่อสมมติ ความเป็นอิสระมีเงื่อนไขแบบจับคู่ (pairwise conditional independence) ระหว่างผลกระทบ (effects) เมื่อกำหนดสาเหตุ (cause) จะได้ว่า:

$$\mathrm{P}\left(\mathrm{cause}, \mathrm{effect}_1, \ldots, \mathrm{effect}_n\right) = \mathrm{P}\left(\mathrm{cause}\right) \prod_i \mathrm{P}\left(\mathrm{effect}_i | \mathrm{cause}\right)$$

แบบจำลองความน่าจะเป็นนี้เรียกว่าแบบจำลอง Naive Bayes

- ความซับซ้อนของแบบจำลองนี้คือ $\mathrm{O}(\mathrm{n})$ แทนที่จะเป็น $\mathrm{O}(2^{\mathrm{n}})$ หากไม่มีสมมติฐานความเป็นอิสระมีเงื่อนไข
- Naive Bayes สามารถทำงานได้ดีอย่างน่าประหลาดใจในทางปฏิบัติ แม้ว่าสมมติฐานจะไม่ถูกต้องก็ตาม

ศึกษาสไลด์ถัดไป **สองรอบ**

กฎของเบย์ (The Bayes' rule)

กฎผลคูณนิยามสองวิธีในการแยกตัวประกอบการแจกแจงร่วมของตัวแปรสุ่มสองตัว

$$\mathrm{P}(\mathrm{a}, \mathrm{b}) = \mathrm{P}(\mathrm{a}|\mathrm{b})\mathrm{P}(\mathrm{b}) = \mathrm{P}(\mathrm{b}|\mathrm{a})\mathrm{P}(\mathrm{a})$$

ดังนั้น,

$$\mathrm{P}(\mathrm{a}|\mathrm{b}) = \frac{\mathrm{P}(\mathrm{b}|\mathrm{a})\mathrm{P}(\mathrm{a})}{\mathrm{P}(\mathrm{b})}$$

- $\mathrm{P}(\mathrm{a})$ คือความเชื่อก่อนหน้า (prior belief) เกี่ยวกับ a
- $\mathrm{P}(\mathrm{b})$ คือความน่าจะเป็นของหลักฐาน b
- $\mathrm{P}(\mathrm{a}|\mathrm{b})$ คือความเชื่อภายหลัง (posterior belief) เกี่ยวกับ a เมื่อพิจารณาหลักฐาน b
- $\mathrm{P}(\mathrm{b}|\mathrm{a})$ คือความน่าจะเป็นมีเงื่อนไขของ b เมื่อกำหนด a ขึ้นอยู่กับบริบท คำนี้เรียกว่า likelihood

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

กฎของเบย์เป็น รากฐาน ของระบบ AI จำนวนมาก

ตัวอย่างที่ 1: ความน่าจะเป็นเชิงวินิจฉัยจากความน่าจะเป็นเชิงสาเหตุ

$$\mathrm{P}(\mathrm{cause}|\mathrm{effect}) = \frac{\mathrm{P}(\mathrm{effect}|\mathrm{cause})\mathrm{P}(\mathrm{cause})}{\mathrm{P}(\mathrm{effect})}$$

โดยที่

- $\mathrm{P}(\mathrm{effect}|\mathrm{cause})$ วัดปริมาณความสัมพันธ์ในทิศทาง เชิงสาเหตุ (causal)
- $\mathrm{P}(\mathrm{cause}|\mathrm{effect})$ อธิบายทิศทาง เชิงวินิจฉัย (diagnostic)

ให้ S=คอแข็ง และ M=เยื่อหุ้มสมองอักเสบ กำหนดให้ $\mathrm{P}(\mathrm{s}|\mathrm{m}) = 0.7, \mathrm{P}(\mathrm{m}) = 1/50000, \mathrm{P}(\mathrm{s}) = 0.01,$ จะได้

$$\mathrm{P}(\mathrm{m}|\mathrm{s}) = \frac{\mathrm{P}(\mathrm{s}|\mathrm{m})\mathrm{P}(\mathrm{m})}{\mathrm{P}(\mathrm{s})} = \frac{0.7 \times 1/50000}{0.01} = 0.0014$$

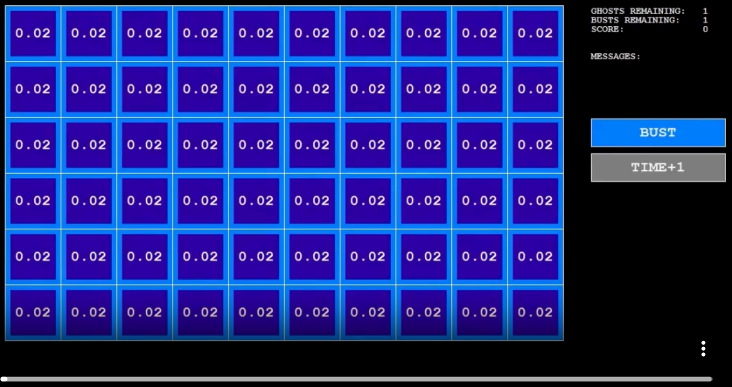
ตัวอย่างที่ 2: Ghostbusters, ทบทวนอีกครั้ง

- สมมติตัวแปรสุ่ม G สำหรับตำแหน่งของผี และชุดของตัวแปรสุ่ม $R_{i,j}$ สำหรับค่าที่อ่านได้แต่ละค่า
- เราเริ่มต้นด้วย การแจกแจงก่อนหน้า (prior distribution) ที่เป็นเอกรูป (uniform) $\mathbf{P}(G)$ สำหรับตำแหน่งของผี
- เราสมมติแบบจำลองการอ่านของเซ็นเซอร์ reading model $\mathbf{P}(R_{i,j} | G)$
  - นั่นคือ เรารู้ว่าเซ็นเซอร์ทำงานอย่างไร
  - $R_{i,j}$ = สีที่อ่านได้ ณ ตำแหน่ง $[i, j]$
    - เช่น $P(R_{1,1} = \text{yellow} | G = [1, 1]) = 0.1$
  - ค่าที่อ่านได้สองค่าเป็นอิสระมีเงื่อนไขต่อกัน เมื่อกำหนดตำแหน่งของผี

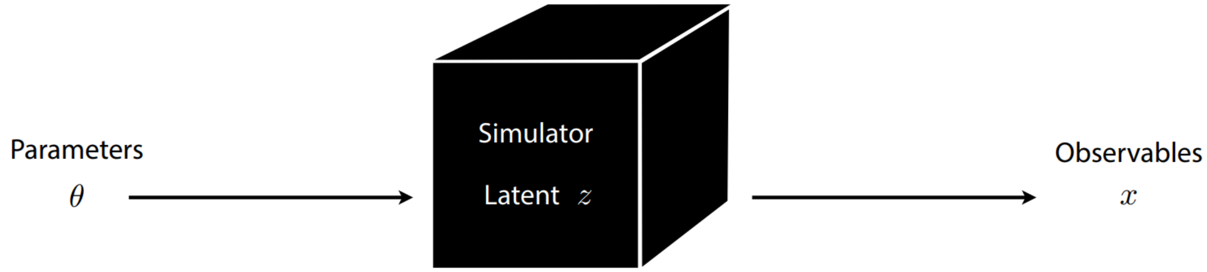
- เราสามารถคำนวณ การแจกแจงภายหลัง (posterior distribution) $\mathbf{P}(\mathrm{G}|\mathrm{R}_{\mathrm{i,j}})$ โดยใช้กฎของเบย์:

$$\mathbf{P}(\mathrm{G}|\mathrm{R}_{\mathrm{i,j}}) = \frac{\mathbf{P}(\mathrm{R}_{\mathrm{i,j}}|\mathrm{G})\mathbf{P}(\mathrm{G})}{\mathbf{P}(\mathrm{R}_{\mathrm{i,j}})}$$

- สำหรับการอ่านค่าครั้งต่อไป $\mathrm{R}_{\mathrm{i',j'}}$ การแจกแจงภายหลังนี้จะกลายเป็นการแจกแจงก่อนหน้าสำหรับตำแหน่งของผี ซึ่งเราจะปรับปรุงในทำนองเดียวกัน


0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02
GHOSTS REMAINING: 1
BUSTS REMAINING: 1
SCORE: 0
MESSAGES:
BUST
TIME+1

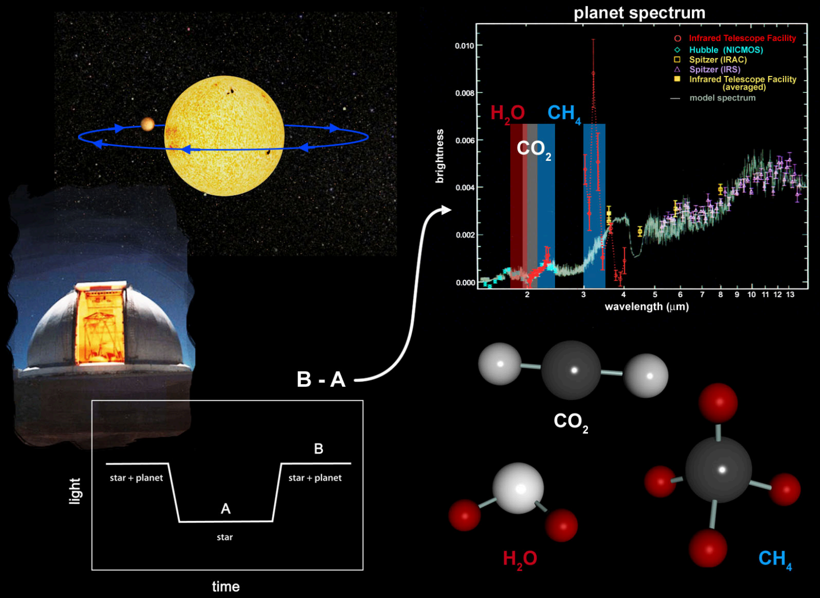
ตัวอย่างที่ 3: AI สำหรับวิทยาศาสตร์ (AI for Science)



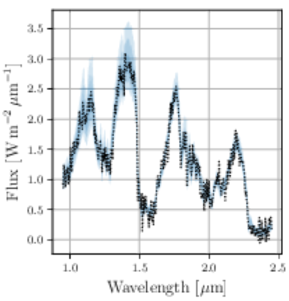
เมื่อมีการสังเกต x และความเชื่อก่อนหน้า $\mathrm{p}(\theta)$ วิทยาศาสตร์คือการปรับปรุงความรู้ของตนเอง ซึ่งสามารถมองในกรอบของการคำนวณ

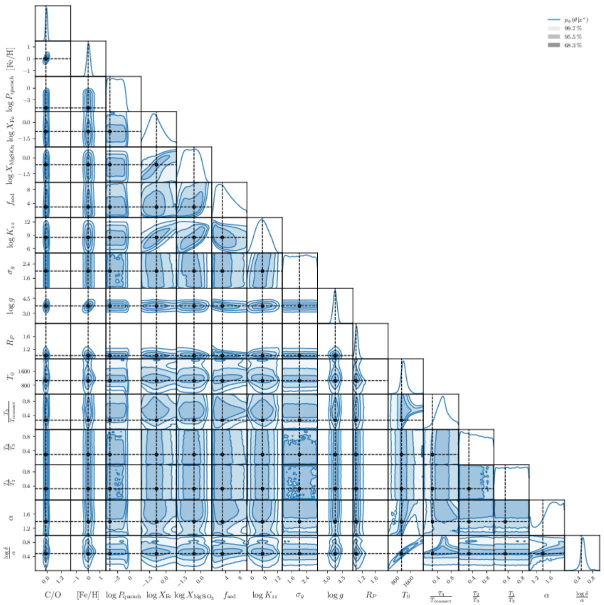
$$\mathrm{p}(\theta|\mathrm{x}) = \frac{\mathrm{p}(\mathrm{x}|\theta)\mathrm{p}(\theta)}{\mathrm{p}(\mathrm{x})}$$

การจำแนกลักษณะชั้นบรรยากาศของดาวเคราะห์นอกระบบสุริยะ (Exoplanet atmosphere characterization)



ที่มา: NSA/JPL-Caltech, 2010.


Flux [W m^-2 μm^-1]
Wavelength [μm]



สรุป (Summary)

- ความไม่แน่นอนเกิดขึ้นจากความขี้เกียจและความไม่รู้ มัน หลีกเลี่ยงไม่ได้ ในสภาพแวดล้อมที่ซับซ้อน ไม่เป็นเชิงกำหนด หรือสังเกตได้เพียงบางส่วน
- การให้เหตุผลเชิงความน่าจะเป็นเป็นกรอบการทำงานสำหรับจัดการความรู้และ ความเชื่อ ของเรา โดยมีกฎของเบย์เป็นเครื่องมือหลักสำหรับการอนุมาน

จบ