1179: Probability
Lecture 3 — Continuity of Probability Function and Conditional Probability

Ping-Chun Hsieh (謝秉均)

September 22, 2021

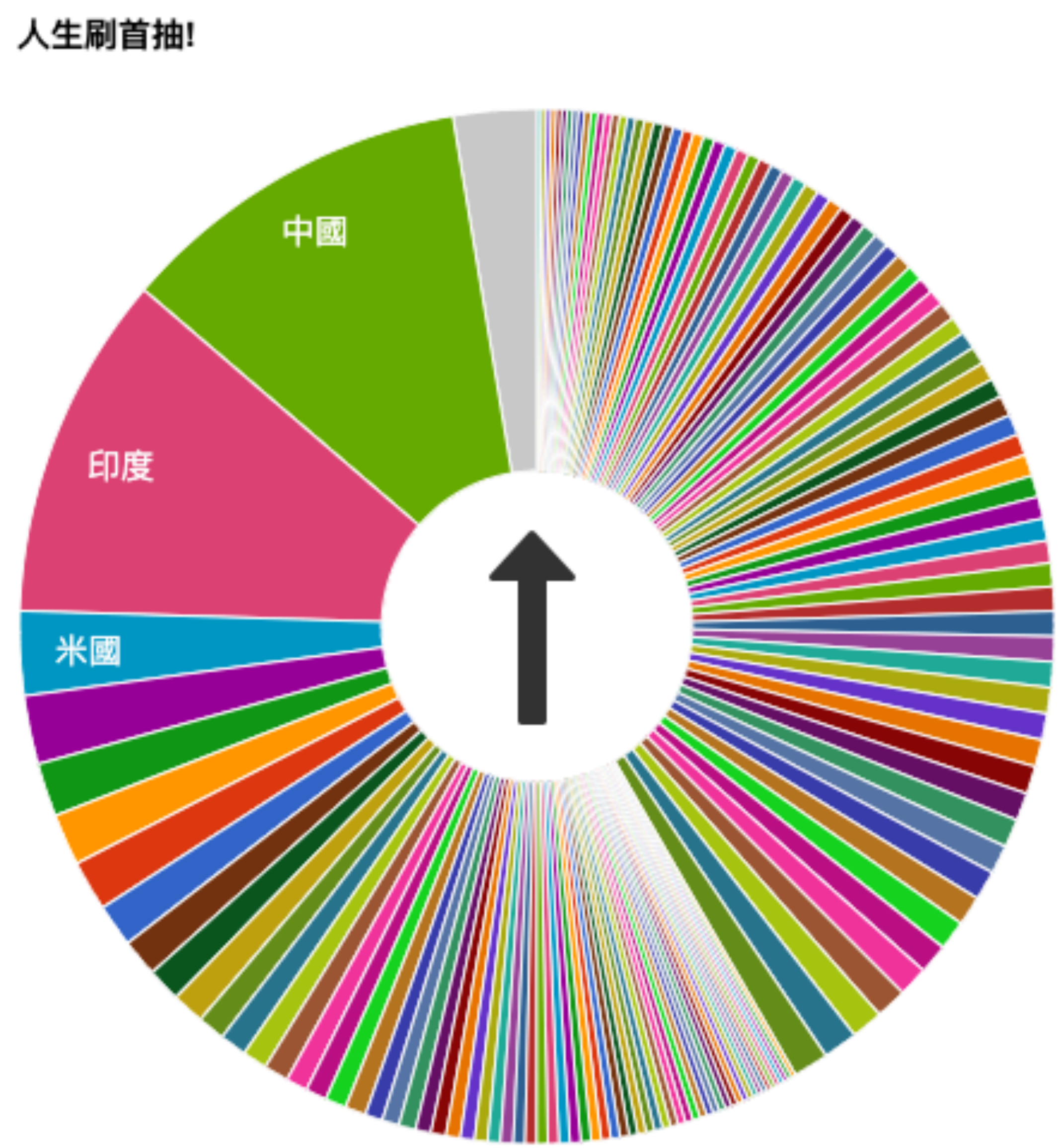
This Lecture

1. Continuity of Probability Functions

2. Conditional Probability and 3 Useful Tools

- Reading material: Chapter 1.5~1.6 & 3.1~3.4

Example: The Lottery of Birth



- Sample space = ?

  {all the countries}

- Probability assignment?

  $$P(\text{country } A) = \frac{\text{population in } A}{\text{population in the world}}$$

- P(born in Taiwan) = ?

  $$\approx \frac{2.3\times 10^7}{7\times 10^9} \approx 0.32\%$$

- [Veil of ignorance]

John Rawls

http://z358z358.github.io/Roulette/#-LOw_CgDNn8VxaIzdtBE

Discrete Uniform Probability Law

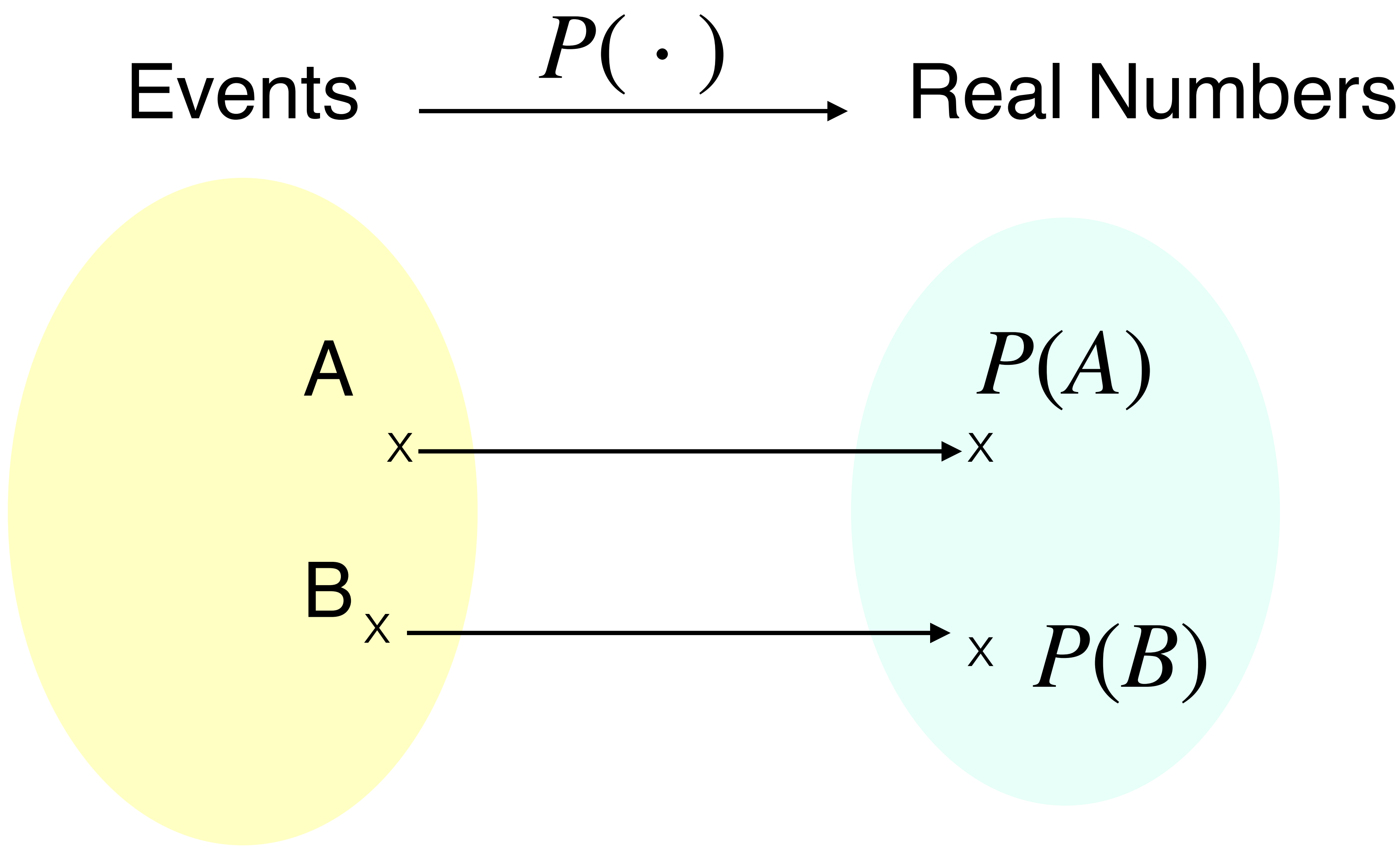
Theorem: Let Ω be the sample space of an experiment. If Ω has N elements that are equally likely to occur, then for any event A of Ω, we have

$$P(A) = \frac{\text{Number of elements in A}}{N}$$

- How to verify this by using the axioms?

1. Continuity of Probability Functions

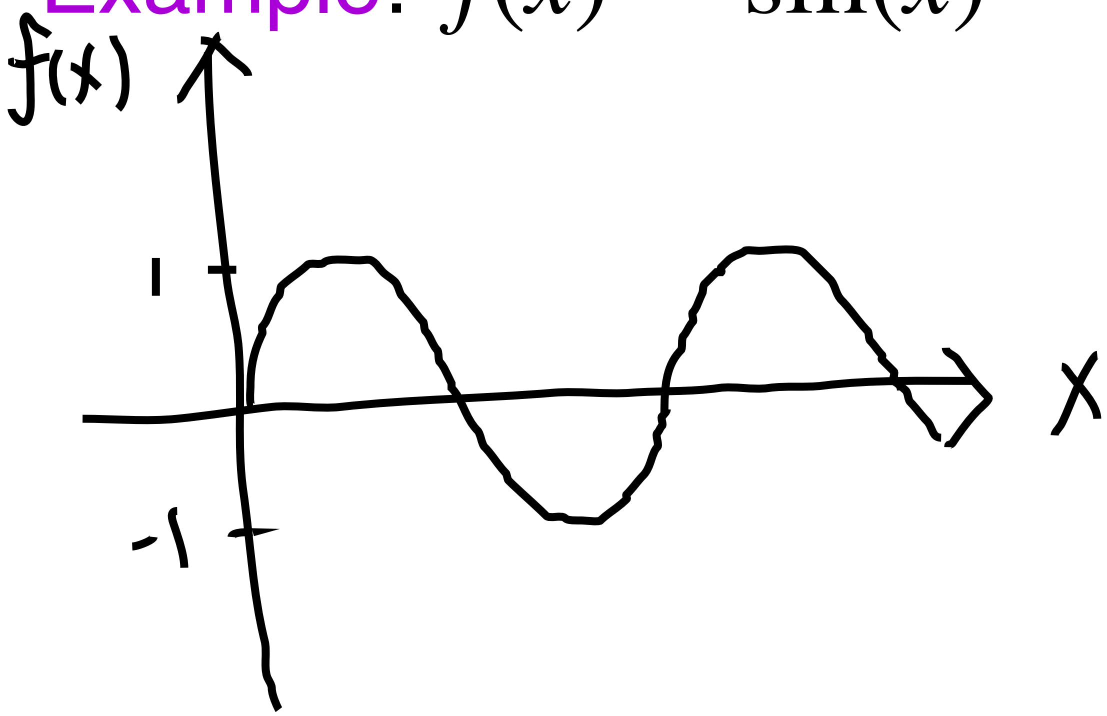
Probability Assignment is a Function of Events



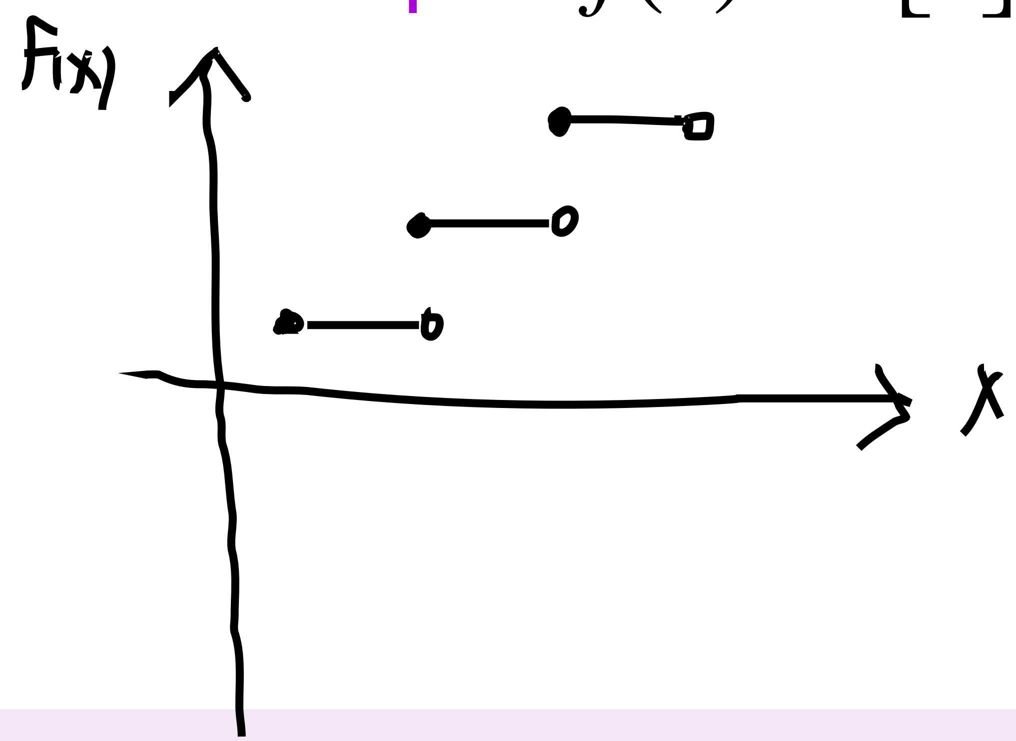
- The function $P(\cdot)$ needs to satisfy the 3 axioms

Review: Continuity of Functions

- What is a continuous function?
- Example: $f(x) = \sin(x)$



- Example: $f(x) = \lfloor x \rfloor$



Definition: A function $f : \mathbb{R} \to \mathbb{R}$ is **continuous** on $\mathbb{R}$ if and only if, for every convergent sequence $\{x_n\}_{n=1}^{\infty}$ with limit $\lim_{n\to\infty} x_n = x$, we have:

ex. $x_n = \frac{1}{n}$

$$\lim_{n\to\infty} f(x_n) = f(x)$$

Continuity of Probability Function

- A sequence of events $E_1, E_2, \cdots$ is **increasing** if

$$E_1 \subseteq E_2 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$$

Theorem: For any increasing sequence of events $E_1, E_2, \cdots$, we have

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

- Is this trivial? Do we need a proof?

- Issue: Interchange of limiting operations

Interchange of Limiting Operations

- Example: $f_n(x) = (\sin nx)/\sqrt{n},\ n = 1,2,3,\cdots$

Do we have $\lim\limits_{n\to\infty} \frac{d}{dx} f_n(x) = \frac{d}{dx}\Big(\lim\limits_{n\to\infty} f_n(x)\Big)$? No !

$\lim\limits_{n\to\infty} \frac{d}{dx} f_n(x)$

$= \lim\limits_{n\to\infty} \frac{n\cos nx}{\sqrt{n}} = \lim\limits_{n\to\infty} \sqrt{n}\cos nx$

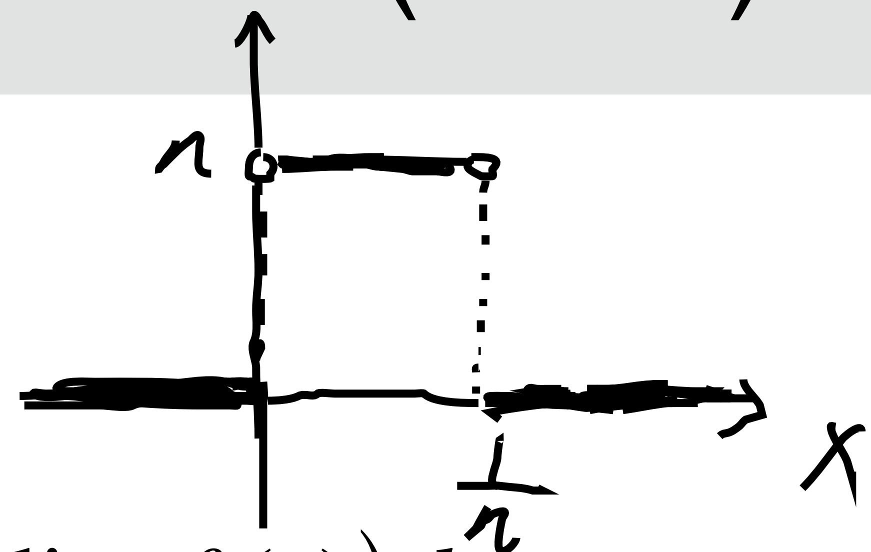
does not exist

$\frac{d}{dx} \lim\limits_{n\to\infty} \frac{\sin nx}{\sqrt{n}} = \frac{d}{dx}(0)$

$= 0$

Interchange of Limiting Operations (Cont.)

- Example: $f_n(x) = \begin{cases} n, & \text{if } x \in (0, \frac{1}{n}) \\ 0, & \text{otherwise} \end{cases}$



Do we have $\lim_{n\to\infty} \int_0^1 f_n(x)dx = \int_0^1 \left(\lim_{n\to\infty} f_n(x) \right) dx$?

No !

$\int_0^1 f_n(x)\, dx = n \cdot \frac{1}{n} = 1$

$\lim_{n\to\infty} = 1$

$\lim_{n\to\infty} f_n(x) = 0$

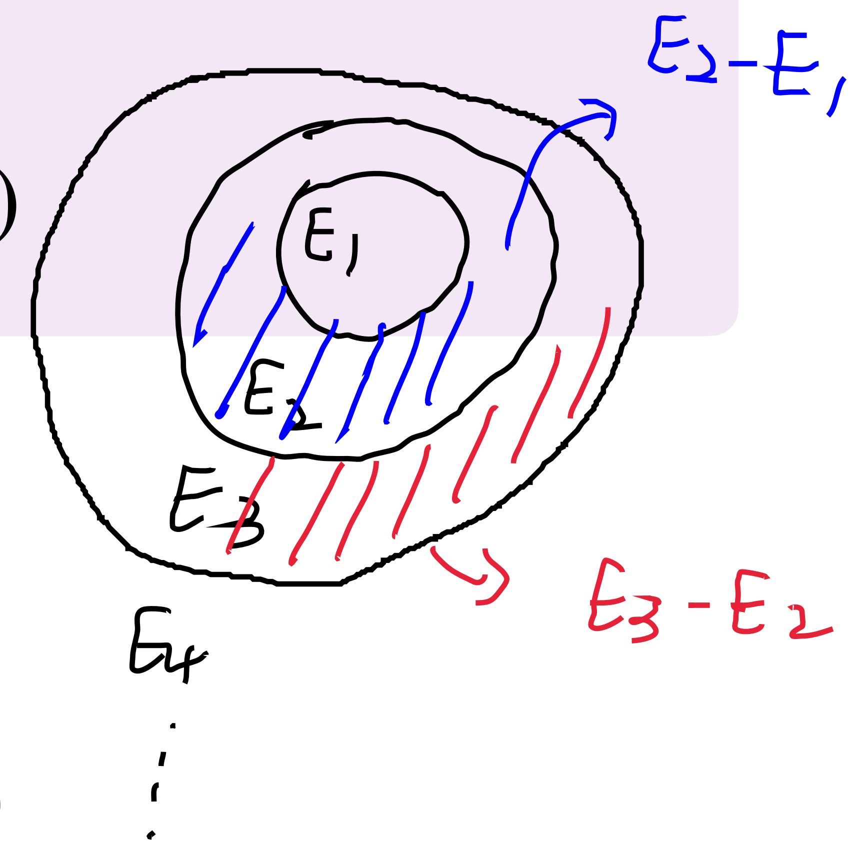
$\int_0^1 (0) = 0$

So it's not a trivial proof for $\lim_{n\to\infty} P(E) = P(\lim_{n\to\infty} E_n)$

Proof: Continuity of Probability Function

Theorem: For any increasing sequence of events $E_1, E_2, \cdots$, we have

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$



- **Proof:** (Proof this by 3 axioms)

(RHS $\longrightarrow$ LHS)

Define $G_1 = E_1$

$G_2 = E_2 - E_1$

$G_3 = E_3 - E_2$

$\vdots$

$G_n = E_n - E_{n-1}$

$G_1, G_2 \cdots G_n$ are mutually exclusive

$$P(\lim_{n\to\infty} E_n) = P(\bigcup_{n=1}^{\infty} G_n)$$

$$= \sum_{n=1}^{\infty} P(G_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(G_k)$$

$$= \lim_{n\to\infty} P(\bigcup_{k=1}^{n} G_k) = \lim_{n\to\infty} P(E_n)$$

#

Continuity of Probability Function (Cont.)

- A sequence of events $E_1, E_2, \cdots$ is **decreasing** if

$$E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq E_{n+1} \supseteq \cdots$$

Theorem: For any decreasing sequence of events $E_1, E_2, \cdots$, we have

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

$E_1^c \subseteq E_2^c \subseteq \cdots \subseteq E_n^c \subseteq E_{n+1}^c \subseteq \cdots$

$\lim_{n\to\infty} P(E_n^c) = P(\lim_{n\to\infty}(E_n)^c)$ by increasing sequence

$LHS = \lim_{n\to\infty} P(U - E_n)$

$= \lim_{n\to\infty} (\underbrace{P(U)}_{1} - P(E_n))$ By Axiom

$= 1 - \lim_{n\to\infty} P(E_n)$ — ①

$RHS = P(\lim_{n\to\infty}(U - E_n))$

$= P(U - \lim_{n\to\infty} E_n)$

$= P(U) - P(\lim_{n\to\infty} E_n)$ By Axiom 3

$= 1 - P(\lim_{n\to\infty} E_n)$ — ②

By ①

Recap

- Can $\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right)$ be viewed as the limit of an increasing / decreasing sequence?

$$\lim_{n\to\infty} P\left(\bigcap_{k=1}^{n}\bigcup_{n=k}^{\infty}A_n\right) = P\left(\lim_{n\to\infty}\bigcap_{k=1}^{n}\bigcup_{n=k}^{\infty}A_n\right)$$

$$\left\{\begin{array}{l} \bigcap_{k=1}^{1}\bigcup_{n=k}^{\infty}A_n \\ \bigcap_{k=1}^{2}\bigcup_{n=k}^{\infty}A_n \\ \bigcap_{k=1}^{3}\bigcup_{n=k}^{\infty}A_n \end{array}\right. \Bigg\downarrow$$

decreasing

- How about $\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}A_n\right)$?

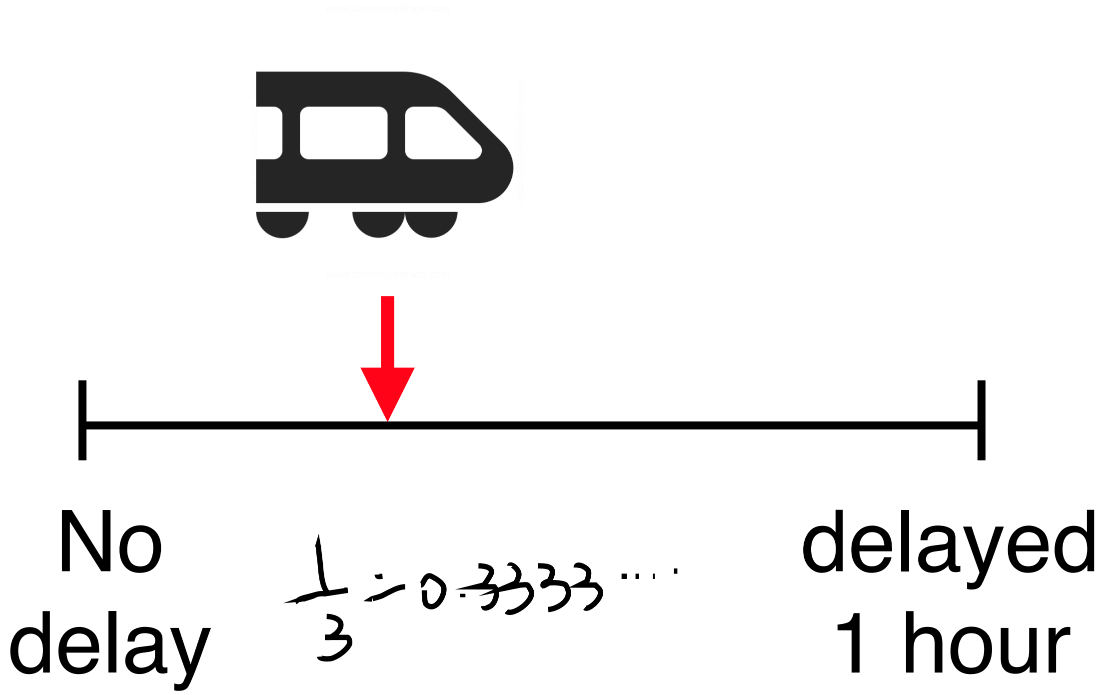
$$\lim_{n\to\infty} P\left(\bigcup_{k=1}^{n}\bigcap_{n=k}^{\infty}A_n\right) = P\left(\lim_{n\to\infty}\bigcup_{k=1}^{n}\bigcap_{n=k}^{\infty}A_n\right)$$

$$\left\{\begin{array}{l} \bigcup_{k=1}^{1}\bigcap_{n=k}^{\infty}A_n \\ \bigcup_{k=1}^{2}\bigcap_{n=k}^{\infty}A_n \\ \bigcup_{k=1}^{3}\bigcap_{n=k}^{\infty}A_n \end{array}\right. \Bigg\downarrow$$

increasing

Example: Train Arrival Time

- Suppose all outcomes are equally likely to happen



$\frac{1}{3} = 0.3333\cdots$

$A_1 = \{\text{1st digit is } 3\}$

$A_2 = \{\text{the first two digits are } 3\}$

$\vdots$

$A_n = \{\text{the first } n \text{ digits are } 3\}$

- Sample space = ? $[0, 1]$

- P(delay is between 0.1 and 0.5 hours) = ? 0.4

- P(delay is exactly 1/3 hours) = ?

0

$$\Rightarrow P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

$$= \lim_{n\to\infty} \frac{1}{10^n} = 0 \quad \#$$

Probability 0 and 1

- If E is an event with $P(E) = 1$, can we say $E = \Omega$?

No!

In train example : $E = \Omega - \{\frac{1}{3}\}$. $P(E) = 1$

- If F is an event with $P(F) = 0$, can we say $F = \emptyset$?

No!

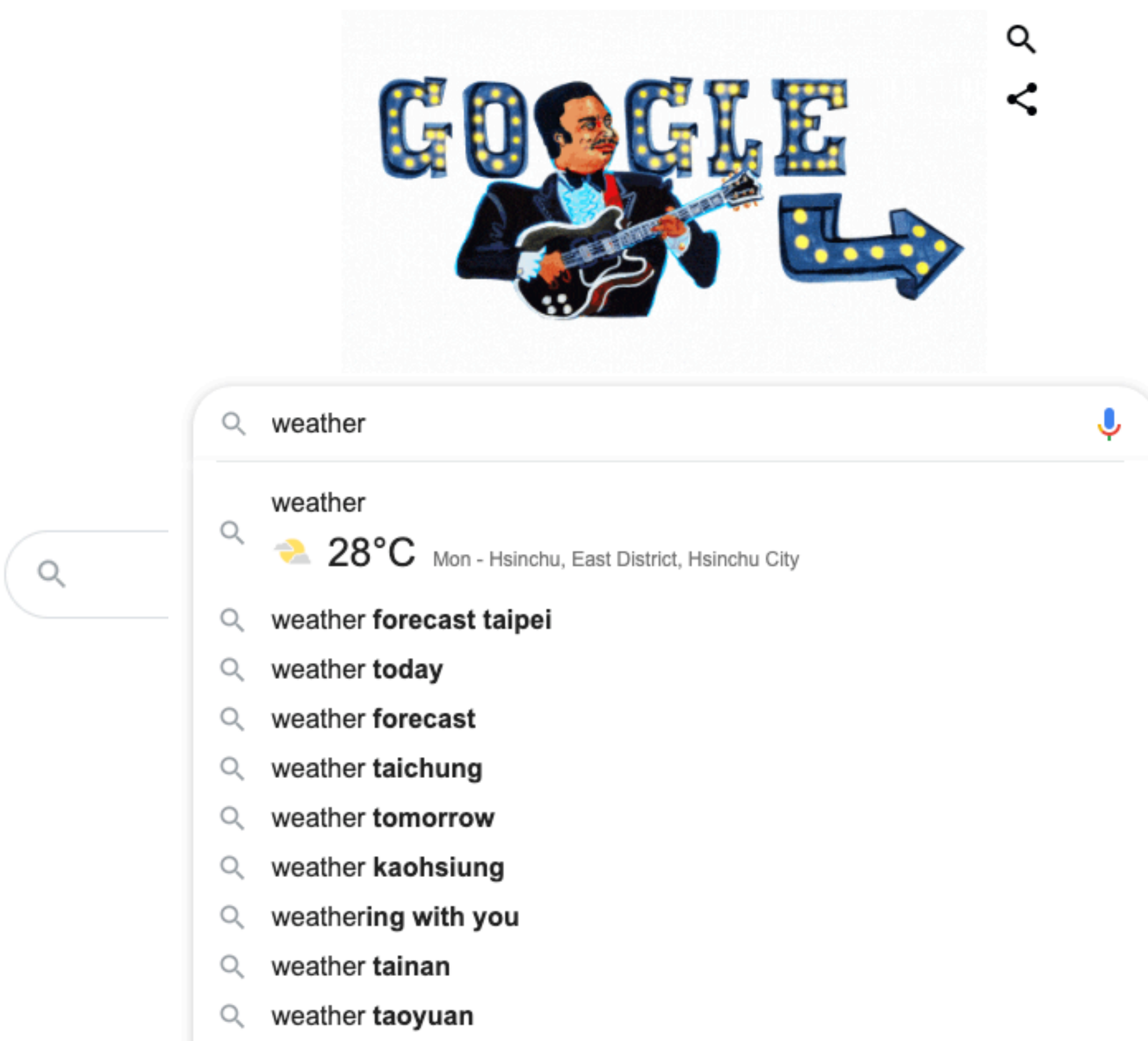
It's improbable

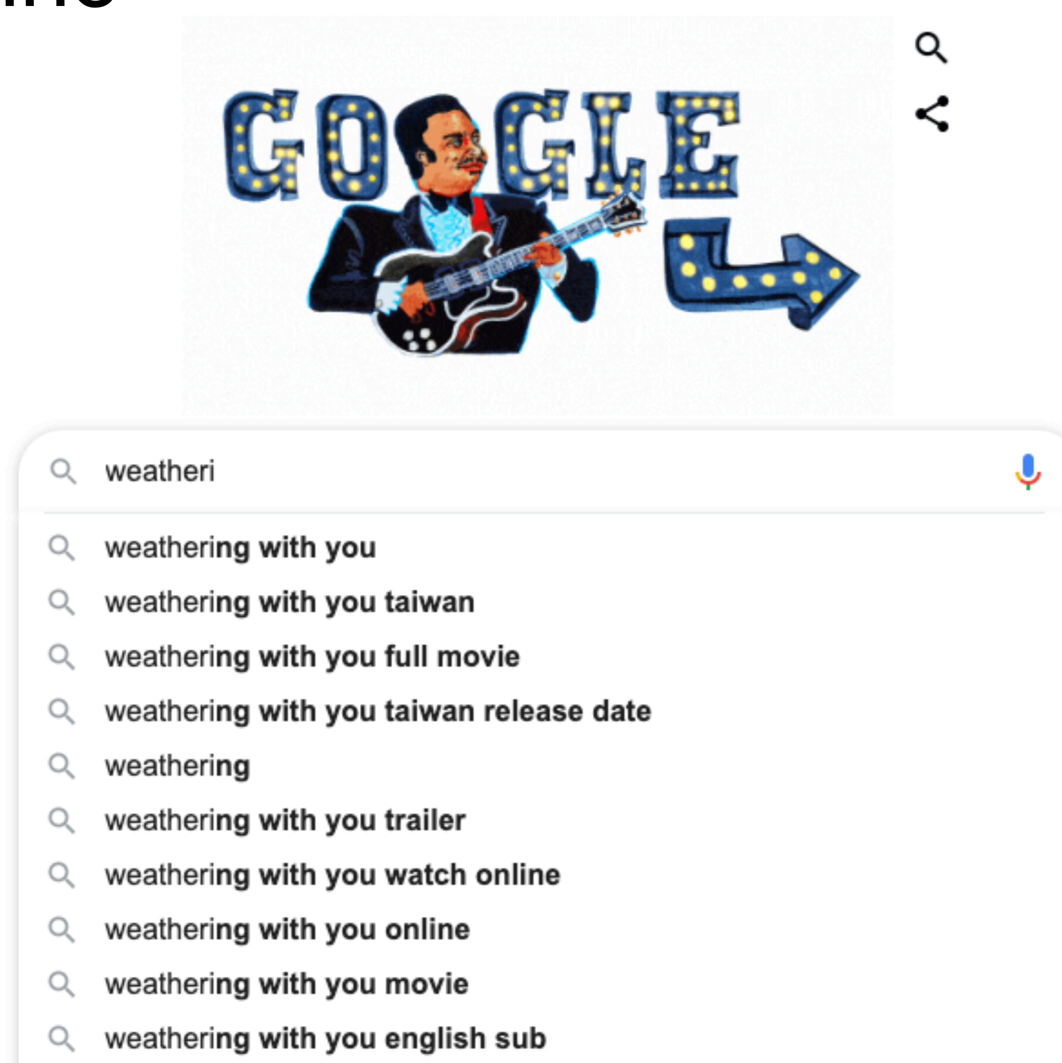
In train example : $E = \{\frac{1}{3}\}$, $P(E) = 0$

2. Conditional Probability

Why Conditional Probability?

- Probability reflects our beliefs on something
- Example: Google search engine





- Partial information can reshape the probability function

Conditional Probability

- Example: Given that today is rainy, what is the probability that no waiting in line when buying a waffle at Shinemood?
- Example: If the oil price goes down, how likely is that the US stocks also drop?
- Suppose we know an event B with $P(B) > 0$ does happen (partial information)
- We want to know the probability of another event A

$P(A \mid B) =$ conditional probability of event A given B

How to Calculate Conditional Probability?

$P(A \mid B) =$ conditional probability of event A given B

Definition:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Is this a natural definition?
- Example: Roll a fair 6-sided die once. Suppose outcome is known to be even. Then, the probability of a 6 = ?

$A = \{6\}$

$B = \{\text{even } (2, 4, 6)\}$

$P(A \mid B) = \frac{1}{3}$

Example

- Example: Roll 2 four-sided dice (outcome is X_1, X_2). Let B be the event that $\max(X_1, X_2) = 3$. What is $P(\min\{X_1, X_2\} = 1 \mid B)$ = ?

$B = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$

$\min\{x_1, x_2\} = 1 = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (3,1), (4,1), (5,1), (6,1)\}$

$P(\min\{x_1, x_2\} = 1 \mid B) = \frac{5}{9}$ #

Conditional Probability Defines a New Probability Assignment

Theorem (Reduction of Sample Space):

Let Ω be the sample space and let B be an event with $P(B) > 0$. Then, we have:

1. $P(A \mid B) \geq 0$, for any event A
2. $P(\Omega \mid B) = 1$
3. $A_1, A_2, \cdots$ is an infinite sequence of mutually exclusive events, then $$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

- Conditional universe!

3. Three Useful Tools

Tool #1: Multiplication Rule

Assuming that all of the conditioning events have positive probability, we have:

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2 \,|\, A_1)\cdots P(A_n \,|\, A_1 \cap A_2 \cap \cdots A_{n-1})$$

- How to intuitively interpret this?
- How to prove this?

Example: Find the Defective Fuses

- Example: Suppose that 7 good and 2 defective fuses are mixed up. To find the defective ones, we test them one by one. $P(\text{we find both defective fuses in exactly 3 tests})$ = ?

Tool #2: Total Probability Theorem

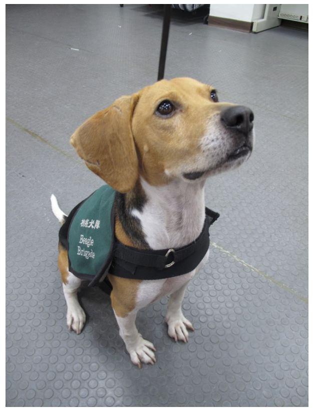
Theorem: Let $A_1, A_2, \cdots, A_n$ be mutually exclusive events that form a partition of Ω, and $P(A_i) > 0$, for all i. Then, for any event B, we have

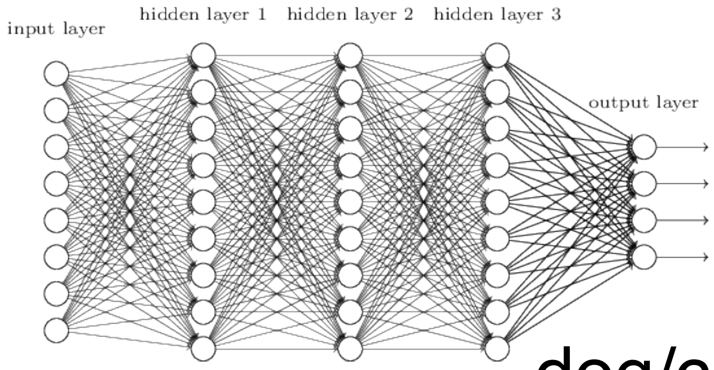
$$
\begin{aligned}
P(B) &= P(A_1 \cap B) + \cdots + P(A_n \cap B) \\
&= P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)
\end{aligned}
$$

- Idea: Divide and conquer

Example: Image Classifier

- Example: Suppose we have a well-trained image classifier
  - Each input image is either a dog/cat/bird with $P(\text{dog}) = P(\text{cat}) = 2 \cdot P(\text{bird})$
  - The probability that a dog is misclassified is 0.1
  - The probability that a cat is misclassified is 0.05
  - The probability that a bird is misclassified is 0.15
  - P(an image is correctly classified) = ?





dog/cat/bird

Example: Gambler's Ruin

- Example: Two gamblers A and B keep tossing a fair coin
  - If ''head'': A pays \$1 to B
  - If ''tail'': B pays \$1 to A
  - Initially, A has a dollars, and B has b dollars
  - The game ends when either A or B has zero dollar
  - What is the probability that A wins the game?

Tool #3: Bayes' Rule

Theorem: Let $A_1, A_2, \cdots, A_n$ be mutually exclusive events that form a partition of Ω, and $P(A_i) > 0$, for all i. Then, for any event B, we have

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)}$$

- Why is Bayes' rule useful? → Inference

Bayesian Inference: Crush and Dates

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)}$$

- Example: Bill has a crush on Amy, and Bill wants to ask Amy out to see whether Amy likes him or not.
  - $A_1 = \{\text{Amy likes Bill}\}$, $A_2 = \{\text{Amy does not like Bill}\}$
  - $B = \{\text{Amy looks happy during the date}\}$
  - $P(B \mid A_1) = 0.9$, and $P(B \mid A_2) = 0.3$

Example: Answer an Exam Question

- Example: Bill answers a question with 4 choices (A, B, C, D)
  - Bill either knows the correct answer or makes a random guess
  - $P(\text{Bill knows the correct answer}) = 2/3$
  - $P(\text{Bill does not make a random guess} \mid \text{answer is correct}) = ?$

1-Minute Summary

1. Continuity of Probability Functions

- Increasing sequence of events
- Interchange of limiting operations
- Probability 0 and 1

2. Conditional Probability

- Multiplication rule / Total probability theorem / Bayes' rule
- Bayesian inference