# Conditional Probability — 條件机率

- **Q**: Should the following probabilities be different? ← new information.
  - ➢ Event = 王建民本季戰績至少獲得6次勝投
    - ▪ $P$=?? in the beginning of the season
    - ▪ $P$=?? in the middle of the season
    - ▪ $P$=?? now → altered.
  - ➢ Event = rain tomorrow
    - ▪ $P$=?? if no information about where you are staying
    - ▪ $P$=?? if you are staying in a desert
    - ▪ $P$=?? if a typhoon will hit the place you stay tomorrow

  altered ↓

- **Q**: What causes the differences?
  - ➢ For an event, *new information* (i.e., some other event has occurred) could change its probability
  - ➢ We call the altered probability a conditional probability
- Mathematical Definition: If $A$ and $B$ are two events in a sample space $\Omega$ and $P(A)>0$, then

  event has occured — $A$; interested. — $B$

$$P(B|A) \equiv \frac{P(A\cap B)}{P(A)}$$

is called the *conditional probability* of $B$ given $A$.



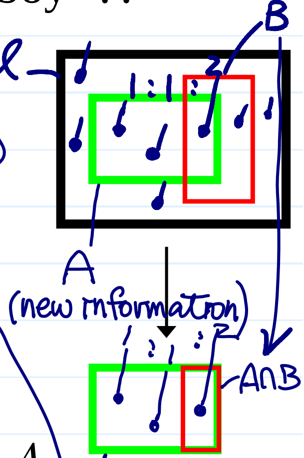
➢ In the classical model,

$$P(A) = \#A/\#\Omega \quad \text{and} \quad P(A\cap B)=\#(A\cap B)/\#\Omega$$

$$\Rightarrow P(B|A) = \frac{\#(A\cap B)/\cancel{\#\Omega}}{\#A/\cancel{\#\Omega}} = \frac{\#(A\cap B)}{\#A} \quad \text{c.f.} \longleftrightarrow P(B) = \frac{\#B}{\#\Omega}$$

- ▪ Example: A family is known to have 2 children, *at least one of whom is a girl*. **Q**: Probability that the other is a boy=??
  - ▫ $\Omega = \{bb, bg, gb, gg\}$ (A: at least one boy.)
  - ▫ $A=\{bg, gb, gg\}$ and $B=\{bb, bg, gb\}$
  - ▫ $P(B|A) = \#(A\cap B)/\#A = 2/3$
- ▪ Note: $\#\Omega$ is reduced to $\#A$.

Ω (original sample space)




A (new sample space)

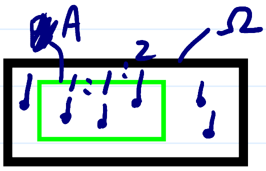
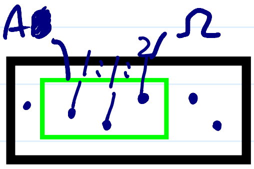
➢ In effect by conditioning,
- ▪ we are restricting the sample space from $\Omega$ to $A$, i.e., $\Omega \to A$,
- ▪ and, for an arbitrary event $B$ in $\Omega$ to occur when $A$ has occurred, we need that both $A$ and $B$ occur together, i.e., $B \to B\cap A$.

➢ The division by $P(A)$ in the definition above rescales all probabilities from the entire sample space $\Omega$ to being relative to the new sample space $A$ ∵ P(sample space)=1

➢ $P(B|A)$ is a probability measure defined on $B$, but not $A$.

can be regarded as a prob. measure defined on $2^\Omega$

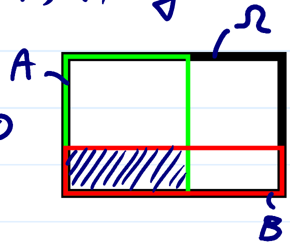
- $P(\cdot|A)$ satisfies the 3 axioms of probability (exercise) for details — check textbook sec 3.5

$P(\cdot)$  $P(\cdot|A)$ 



- Any propositions developed in Chapter 2 for probability measures can be applied on $P(\cdot|A)$.
  (For example, $P(B^c|A)=1-P(B|A)$.) cf. $P(B|A)+P(B|A^c) \neq 1$ in general

• 3 Useful Formulas for Calculating Probabilities (for 2-events case)



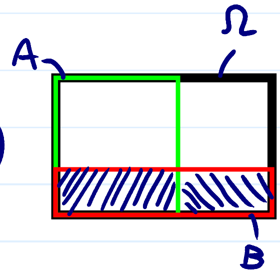


to garantee the cond. prob. well defined

1. If $P(A)>0$, then $P(A\cap B) = P(A)P(B|A)$. —①

$= P(B)\cdot P(A|B)$ if $P(B)>0$

from the definition of cond. prob. $P(B|A)=P(A\cap B)/P(A)$.





2. If $0<P(A)<1$, then

weighted average

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c). \text{ —②}$$

by① $= P(A\cap B)$ by① $= P(A^c\cap B)$

$B = B\cap\Omega$
$= B\cap(A\cup A^c)$
$= (B\cap A)\cup(B\cap A^c)$

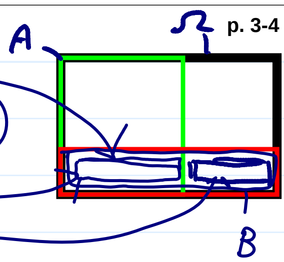
---



3. If $0<P(A)<1$ and $P(B)>0$, then

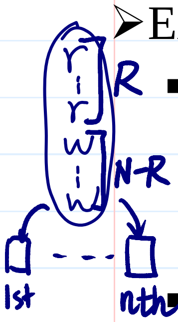
$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)}. \text{ —③}$$

$P(A|B) = P(A\cap B)$, numerator by①, denominator $= P(B)$ by②





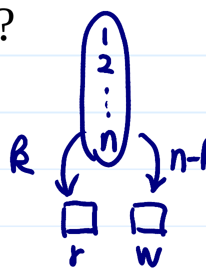
➢ Example (Urn Problem) $P(A) \xrightarrow{\text{update}} P(A|B)$





- The Story. $n$ balls sequentially and randomly chosen, *without replacement*, from an urn containing $R$ red and $N-R$ white balls ($n\le N$). **Q**: Given that $k$ of the $n$ balls are red ($k\le R$), probability that the 1st ball chosen is red = ??

do not put the drawn ball into the urn





- Let $A$={$k$ of the $n$ balls are red}
  $B$={1st ball chosen is red}

- Method 1: $P(B|A) = \binom{n-1}{k-1} \Big/ \binom{n}{k} = k/n$

- Method 2:

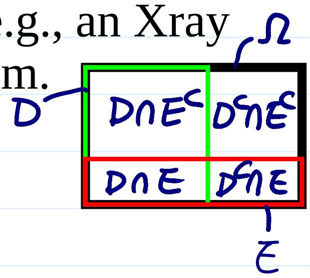
hyper-geometric

▫ $P(A) = \left[\binom{R}{k}\times\binom{N-R}{n-k}\right] \Big/ \binom{N}{n}$

▫ $P(A\cap B) \overset{\text{by①}}{=} P(B)P(A|B) = \dfrac{R}{N} \times \dfrac{\binom{R-1}{k-1}\times\binom{N-R}{n-k}}{\binom{N-1}{n-1}}$

▫ $P(B|A) = P(A\cap B)/P(A) = k/n$

- Example (Diagnostic Tests)
  - The Story. A diagnostic test for a rare disease (e.g., an Xray for lung cancer) is part of a routine physical exam.




  - Let Ω = {the whole population of Taiwan}

    $D$={disease is present}

    $E$={test indicates disease present}
  - Suppose that $P(D)=0.001$, $P(E|D)=0.98$, $P(E|D^c)=0.01$

    large ≈1 small ≈0

    **Q**: Do you think the test is effective?
  - Let us examine it from an alternative viewpoint

$$P(E) \overset{\text{by (2)}}{=} P(D)P(E|D) + P(D^c)P(E|D^c)$$
$$= 0.001 \times 0.98 + 0.999 \times 0.01 = 0.01097.$$
$$P(D|E) \overset{\text{by (3)}}{=} \frac{P(D)P(E|D)}{P(D)P(E|D) + P(D^c)P(E|D^c)} = \frac{0.00098}{0.01097} = 0.0893.$$

C.P. P(D)

$$P(D^c|E) = 1 - P(D|E) = 0.9107$$

"rare disease" P(D) very small

**Q**: Now, do you still think the test is effective?

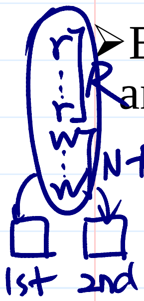
**Q**: But, why the 2 interpretations so different? What causes it?

new information

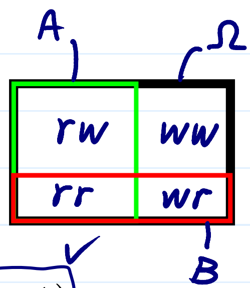
- The probability of $D$ increased by a factor of roughly 90 (0.001 → 0.0893) when $E$ occurs, but 0.0893 is still **small**



- The $P(E|D^c)$ (=0.01) and $P(E^c|D)$ (=1–$P(E|D)$=0.02) are called the *false positive* and *false negative rates*, respectively.




- Example (Sampling Experiments): An urn contains $R$ red balls and $N$–$R$ white balls. Sample 2 balls from the urn.




  - $A$ = {red on the first draw}

    $B$ = {red on the second draw}
  - Sampling *Without Replacement*: — LNp.3-4

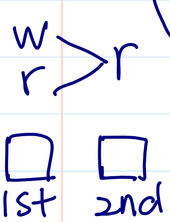
$$P(A) = \frac{R(N-1)}{N(N-1)} = \frac{R}{N}, \qquad P(A\cap B) \overset{\text{by (1)}}{=} \frac{R(R-1)}{N(N-1)}, = P(A)P(B|A)$$

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{(R(R-1))/(N(N-1))}{R/N} = \frac{R-1}{N-1}.$$

update

Similarly, $P(B|A^c) = \dfrac{R}{N-1}$. (exercise)

Note: $P(B|A)+P(B|A^c) = \frac{R-1}{N-1} + \frac{R}{N-1} \neq 1$ in general




$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$$
$$\overset{\text{by (2)}}{=} \frac{R}{N}\cdot\boxed{\frac{R-1}{N-1}} + \frac{N-R}{N}\cdot\boxed{\frac{R}{N-1}}$$
$$= \frac{R^2 - R + NR - R^2}{N(N-1)} = \frac{R(N-1)}{N(N-1)} = \frac{R}{N}. = P(A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{(R(R-1))/(N(N-1))}{R/N} = \frac{R-1}{N-1}.$$

$= P(B|A)$

- Notes:
  - The probabilities are proportional to # of red balls left
  - $\frac{P(A\cap B)}{P(B)} = P(A|B) = P(B|A) = P(A \cap B)/P(A) \Rightarrow$ Symmetry.
- Sampling *With Replacement*: put the drawn ball back into the urn.

$$P(A) = \frac{R}{N} \cdot \frac{N}{N} = \frac{R}{N}, \quad P(B) = \frac{N}{N} \cdot \frac{R}{N} = \frac{R}{N}, \text{ update}$$

$$P(A \cap B) = \frac{R}{N} \cdot \frac{R}{N} = \frac{R^2}{N^2}, \quad P(B|A) = \frac{R^2/N^2}{R/N} = \frac{R}{N} = P(B).$$

independent LNp.3-10

R, W, N−R, 1st, 2nd

- Extensions of the 3 Useful Formulas (for $m$-events case)

1.(Multiplication Law) If $A_1, \ldots, A_m$ are events for which $P(A_1 \cap \cdots \cap A_{m-1}) > 0$, then

to garantee the cond. prob. well defined

$$P(A_1 \cap \cdots \cap A_m) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_m|A_1 \cap \cdots \cap A_{m-1}). \quad (4)$$

by definition $= P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \cdots \cap A_m)}{P(A_1 \cap \cdots \cap A_{m-1})}$



- Example (Birthday Problem, LNp.2-3). Let $A_k$ be the event that the $k^{\text{th}}$ birthday differs from the first $k$–1. Then, $P(A_1)$=1,

1, ⋮, 365; 1st, kth, nth

  - $$P(A_k|A_1 \cap \cdots \cap A_{k-1}) = \frac{365-k+1}{365}$$ Q: What benefit does cond. prob. bring?
  - $$P(A_1 \cap \cdots \cap A_n) \overset{\text{by (4)}}{=} P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

$$= \prod_{k=1}^{n} \frac{365-k+1}{365} = \frac{365!}{365^n \cdot (365-n)!} = \frac{(365)_n}{365^n}.$$

- Example (Matching Problem, LNp.2-11). **Q**: Probability that exactly $k$ of $n$ members have matches = ?? — $A$.

1, ⋮, n; 1st, nth

  - Let Ω be all permutations ω = $(i_1, \ldots, i_n)$ of 1, 2, …, $n$.
  - Let $A_j = \{\omega: i_j = j\}$ and

$j_1$th, $j_2$th, $j_k$th

$$A = \bigcup_{1 \le j_1 < \cdots < j_k \le n} \underbrace{(A_1^c \cap \cdots \cap A_{j_1-1}^c \cap A_{j_1} \cap A_{j_1+1}^c \cap \cdots \cap A_n^c)}_{\text{mutually exclusive}}$$

  - By symmetry, $P(A) = \binom{n}{k} \times P(\underbrace{A_1 \cap \cdots \cap A_k}_{E} \cap \underbrace{A_{k+1}^c \cap \cdots \cap A_n^c}_{G})$
  - Let $E = A_1 \cap \cdots \cap A_k$ and $G = A_{k+1}^c \cap \cdots \cap A_n^c$ (1)
  - Then, $P(E \cap G) = P(E)P(G|E)$, where

Note: in ②,③ $A, A^c$ is a partition of $\Omega$

$$P(E) \overset{\text{by ④}}{=} P(A_1)P(A_2|A_1)\cdots P(A_k|A_1\cap\cdots\cap A_{k-1})$$

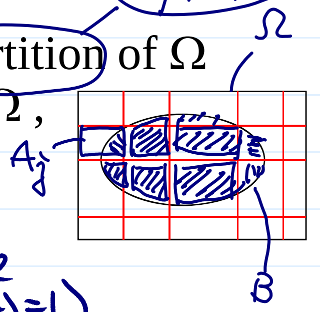
$$= \frac{1}{n}\times\frac{1}{n-1}\times\cdots\times\frac{1}{n-k+1} = \frac{(n-k)!}{n!} = \frac{1}{(n)_k}$$

and, $P(G|E) = \sum_{i=0}^{n-k}(-1)^i\frac{1}{i!} \equiv p_{n-k}$ check LN p.2-12.

□ $P(A) = \binom{n}{k}\frac{1}{(n)_k}p_{n-k} = \frac{p_{n-k}}{k!} \approx \frac{e^{-1}}{k!}$, when $n$ is large

$\bigcup_{i=1} A_i = \Omega$, $A_i \cap A_j = \phi$ LN p.2-12

② (Law of Total Probability) Let $A_1, \ldots, A_m$ be a partition of $\Omega$ and $P(A_i)>0$, $i=1, \ldots, m$, then for any event $B \subset \Omega$,



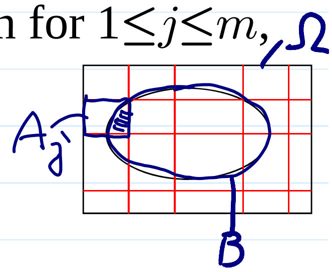

$B = B\cap\Omega$
$= B\cap(\bigcup_{i=1} A_i)$
$= \bigcup_{i=1}(B\cap A_i)$ mutually exclusive

$$P(B) = \sum_{i=1}^m P(A_i)P(B|A_i). \quad ⑤$$

$= \sum P(B\cap A_i)$

weighted average ($\because \sum P(A_i)=1$)

③ (Bayes' Rule) Let $A_1, \ldots, A_m$ be a partition of $\Omega$ and $P(A_i)>0$, $i=1, \ldots, m$. If $B$ is an event such that $P(B)>0$, then for $1\le j\le m$,




$$\frac{P(A_j\cap B)}{P(B)} = P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^m P(A_i)P(B|A_i)} \quad ⑥$$

by ①, by ⑤

$P(A_j) \xrightarrow{\text{update}} P(A_j|B)$



■ From Bayesians' viewpoint,

e.g. drug test.

data.

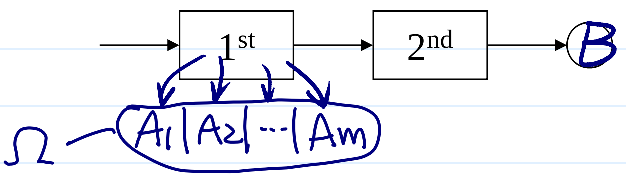
$P(A_j)$=probability of $A_j$ *before* $B$ occurs → prior prob. 事前机率

↓ update

$P(A_j|B)$=probability of $A_j$ *after* $B$ occurs → posterior prob. 事後机率

⇒ The Bayes' rule tells how to update the probabilities of $A_j$ in light of the new information (i.e., $B$ occurs)

■ An Application of Bayes' Rule. Suppose that a random experiment consists of two random stages




□ The probabilities of the 2nd-stage results depend on what happened in the 1st stage $P(B|A_j)$

□ We never see the result of the 1st stage, only the final result (B occur or not)

□ We may be interested in finding the probability for outcomes in the 1st stage given the final result B $P(A_j|B)$

p. 3-11

➢Example (Gold Coins):

- The Story.
  - Box 1 contains 2 silver coins.
  - Box 2 contains 1 gold and 1 silver coin.
  - Box 3 contains 2 gold coins.
  - Experiment: (i) Select a box at random and, (ii) Examine the 2 coins in order (assuming all choices are equally likely at each stage)

1st → 2nd → G's 1st 2nd

$\Omega$ = Box1 | Box2 | Box3

- **Q**: Given that 1st coin is gold, what is the probability that Box $k$ is selected, $k$=1, 2, 3?
  - Let $A_k$= {Box $k$ is selected}, $B$ = {1st coin is gold},
  - $$P(B|A_k) = \begin{cases} 0, & \text{if } k=1, \\ 1/2, & \text{if } k=2, \\ 1, & \text{if } k=3. \end{cases}$$

by ⑤

$$P(B) = \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 1 = \frac{1}{2}.$$

  - $$P(A_1|B) = \frac{(1/3)\cdot 0}{1/2} = 0.$$

by ⑥

Similarly, $P(A_2|B) = 1/3,\ P(A_3|B) = 2/3.$

p. 3-12

- **Q**: Given that 1st coin is gold (B), what is the probability that 2nd coin is gold?
  - Let $C$ = {2nd coin is gold}. $$P(B\cap C|A_k) = \begin{cases} 0, & \text{if } k=1, \\ 0, & \text{if } k=2, \\ 1, & \text{if } k=3. \end{cases}$$
  - $$P(B\cap C) = \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1 = \frac{1}{3}.$$ ✗ wrong calculation

by ⑤

$$P(C|B) = \frac{P(B\cap C)}{P(B)} = \frac{1/3}{1/2} = \frac{2}{3}$$ c.f. $\Longleftrightarrow P(A_3|A_2\cup A_3) = \frac{1/3}{2/3} = \frac{1}{2}$

by ⑥

➢ **Q**: Why are the 3 formulas useful in calculating probabilities? (Note: They all benefit from conditional probabilities.)

Ans: (i) 繁→簡&簡&⋯&簡;

(ii) 簡=conditioning because the sample space is reduced from Ω to a smaller set. (For example, in many cases, $P(B|A_j)$'s are known or are easier to find)

10/6

- Odds and Conditional Odds

➢The odds of an event $A$:

check LNp.2-16

$$o(A) \equiv \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)} \iff P(A) = \frac{o(A)}{1+o(A)}$$

➢The odd of event $B$ given $A$:

$$o(B|A) \equiv \frac{P(B|A)}{P(B^c|A)} = \frac{P(A\cap B)/\cancel{P(A)}}{P(A\cap B^c)/\cancel{P(A)}} = \frac{\boxed{P(B)}\,P(A|B)}{\boxed{P(B^c)}\cdot P(A|B^c)}$$

and

$$o(B|A) = o(B) \times \frac{P(A|B)}{P(A|B^c)}$$

evaluated under the prob. measure $P(\cdot|A)$

❖ **Reading**: textbook, Sec 3.1, 3.2, 3.3, 3.5

## Independence — 獨立.

• Definition (independence for 2-events case): Two events $A$ and $B$ are said to be *independent* if and only if

property defined on events
↕ c.f.
property defined on random variable (~~see~~ future lecture)

for calculation purpose → $P(A\cap B) = P(A)P(B)$.

Otherwise, they are said to be *dependent*.

➢Notes. For independent events $A$ and $B$, if $P(A) > 0$, then

for interpretation purpose → $P(B|A) = \dfrac{P(A\cap B)}{P(A)} = \dfrac{\cancel{P(A)}P(B)}{\cancel{P(A)}} = P(B)$,

similarly, if $P(B) > 0$, $P(A|B) = P(A)$.

**Q**: How to interpret the equality of the conditional and unconditional probabilities?



➢Example (Sampling 2 balls, LNp.3-6) $A$ and $B$ were "independent for" sampling with replacement, but "dependent" for sampling without replacement.

➢Example (Cards): If a card is selected from a standard deck, let

▪ $A$ = {ace} and $B$={spade}. Then,

→ can be changed to any face

can be changed to any other suit.

▪ $P(A) = \dfrac{4}{52} = \dfrac{1}{13}$, $\quad P(B) = \dfrac{13}{52} = \dfrac{1}{4}$,

$P(A\cap B) = \dfrac{1}{52} = P(A)P(B)$

⇒Face and Suit are independent

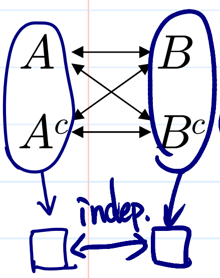
點 花色

$P(A\cap B) > P(A)\cdot P(B)$

exercise.
If $P(B|A) > P(B)$
then $P(B|A^c) < P(B)$

➢Theorem (Independence and Complements, 2-events case).

If $A$ and $B$ are independent, then so are $A^c$ and $B$.




proof: $B = B\cap\Omega = B\cap(A\cup A^c) = \underbrace{(B\cap A)\cup(B\cap A^c)}_{\text{mutually exclusive}}$

$P(B) = P(B\cap A) + P(B\cap A^c)$

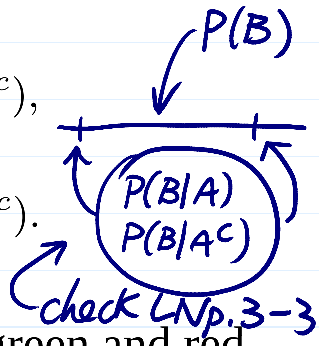
indep. $\overset{(>)}{=} P(A)P(B) + P(B\cap A^c)$

$\Rightarrow P(B\cap A^c) \overset{(<)}{=} P(B) - P(A)P(B) = P(B)[1-P(A)] = P(B)\cdot P(A^c)$

▪ Corollary: If $A$ and $B^c$ are independent, as are $A^c$ and $B^c$

- Corellary: If $A$ and $B$ are independent and $0<P(A)<1$, $0<P(B)<1$, then

$$\boxed{P(B)=P(B|A)}=P(B|A^c),$$
$$P(B^c)=P(B^c|A)=P(B^c|A^c),$$
$$\boxed{P(A)=P(A|B)}=P(A|B^c),$$
$$P(A^c)=P(A^c|B)=P(A^c|B^c).$$

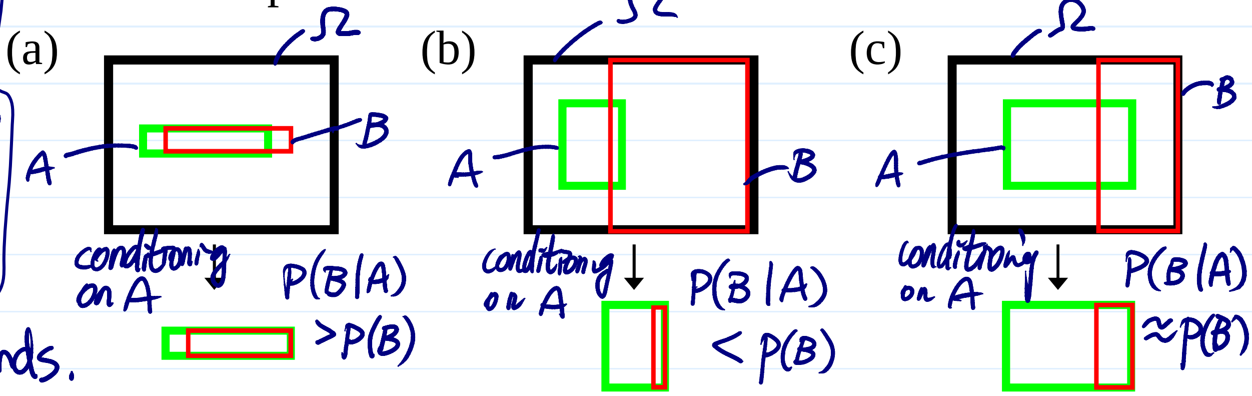



**Q**: What do these equalities say?

**Q**: Which of the following graphs represents "green and red events are independent"? (Assume prob ∝ area) Ans. (c)

Q: What do we prefer? indep.? dep.?

It depends.

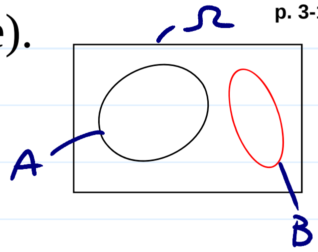



**Q**: Let green event={graduate from Tsing-Hua University}, red event={your future dream will come true}.

Which of the graphs would you prefer? usually (a)



➢ Theorem (Independence and Mutually Exclusive). If $A$ and $B$ are mutually exclusive and $P(A)>0$, $P(B)>0$, then $A$ and $B$ are dependent since




$$0=\frac{P(\phi)=P(A\cap B)}{P(A)}=P(B|A)=0\neq P(B).$$

- Definition (independence for $n$-events case). Events $A_1, \ldots, A_n$ are said to be *pairwise independent* iff weak

$$P(A_i\cap A_j)=P(A_i)P(A_j),$$

for all $1\le i<j\le n$; $A_1, \ldots, A_n$ are said to be *mutually independent* iff strong

equality must hold for any k sets in $\{A_1, \cdots, A_n\}$

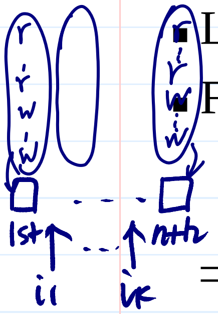
$$P(A_i\cap A_j)=P(A_i)P(A_j),\quad \text{for } 1\le i<j\le n,$$
$$P(A_i\cap A_j\cap A_k)=P(A_i)P(A_j)P(A_k),\quad \text{for } 1\le i<j<k\le n,$$
$$\ldots,$$
$$P(A_{i_1}\cap\cdots\cap A_{i_k})=P(A_{i_1})\cdots P(A_{i_k}),\quad \text{for } 1\le i_1<\cdots<i_k\le n,$$

where $k=2, \ldots, n$. For different $t_1, t_2, \cdots, t_r, t_{r+1}, \cdots, t_k \subset \{1, \cdots, n\}$

➢ Note: $P(A_{t_1}\cap\cdots\cap A_{t_r} \mid A_{t_{r+1}}\cap\cdots\cap A_{t_k}) = P(A_{t_1}\cap\cdots\cap A_{t_r})$, $k=2,\cdots,n$. exercise

- Mutual independence implies pairwise independence; but, the converse statement is usually not true.
- "$n$ events are independent" means "mutually independent"

- Example (Sampling With Replacement)
  - A sample of $n$ balls is drawn with replacement from an urn containing $R$ red and $N$–$R$ white balls
  - Let $A_k$={red on the $k^{\text{th}}$ draw}, then $P(A_k)=R/N$, $k$=1, …, $n$.
  - For all $1\le i_1<\cdots<i_k\le n$, where $k$=2, …, $n$,




$$P(A_{i_1}\cap\cdots\cap A_{i_k})=\frac{R^kN^{n-k}}{N^n}=\left(\frac{R}{N}\right)^k=P(A_{i_1})\cdots P(A_{i_k}),$$

← # of lists if balls are labelled.

$\Rightarrow A_1, \ldots, A_n$ are mutually independent

- Example. Draw one card from a standard deck.
  - Let $A$={Spades or Clubs},
    $B$={Hearts or Clubs},
    $C$={Diamonds or Clubs}.
  - $P(A)=26/52=1/2$, similarly, $P(B)=P(C)=1/2$.
  - $P(A\cap B)=P(\{\text{Clubs}\})=\dfrac{13}{52}=\dfrac{1}{4}=P(A)P(B)$, similarly,

  $P(A\cap C)=1/4=P(A)P(C)$, $P(B\cap C)=1/4=P(B)P(C)$.

  $\Rightarrow$ $A$, $B$, and $C$ are pairwise independent



- However,

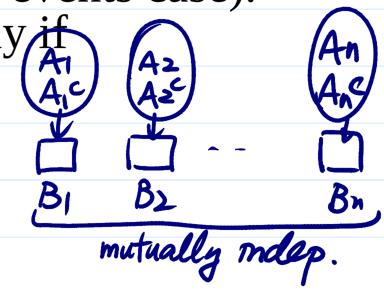
$P(A|B\cap C)=1\ne P(A)$ ⇐ $P(A\cap B\cap C)=P(\{\text{Clubs}\})=\dfrac{1}{4}\ne\dfrac{1}{8}=P(A)P(B)P(C)$,

$\Rightarrow$ A, B, and C are *not* mutually independent

- Theorem (Independence and Complements, $n$-events case). $A_1, \ldots, A_n$ are mutually independent if and only if

$$P(B_1\cap\cdots\cap B_n)=P(B_1)\cdots P(B_n),$$

where $B_i$ is either $A_i$ or $A_i^c$, for $i$=1, …, $n$.




Outline of proof.

(⇒only if) Apply Thm in LNp.3-14 & by induction.

$A_1\cap\cdots\cap A_n \to A_1\cap\cdots\cap A_{i-1}\cap A_i^c\cap A_{i+1}\cap\cdots\cap A_n \to \cdots$

(⇐if) $P(A_1\cap\cdots\cap A_{n-1})=P(A_1\cap\cdots\cap A_n)+P(A_1\cap\cdots\cap A_{n-1}\cap A_n^c)$

$=P(A_1)\cdots P(A_n)+P(A_1)\cdots P(A_{n-1})\cdot P(A_n^c)$

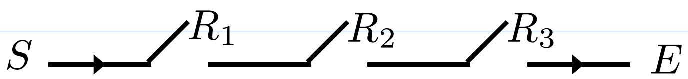
$=P(A_1)\cdots P(A_{n-1})\underbrace{[P(A_n)+P(A_n^c)]}_{=1}$.

Similarly: $P(A_1\cap\cdots\cap A_{n-2}\cap A_{n-1}^c)=P(A_1)\cdots P(A_{n-2})\cdot P(A_{n-1}^c)$

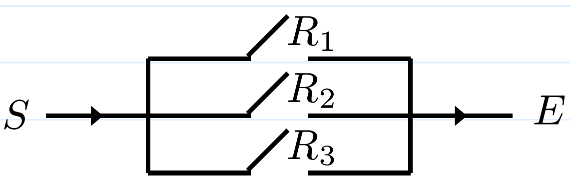
Then: $P(A_1\cap\cdots\cap A_{n-2})=P(A_1)\cdots P(A_{n-2})$, ……

- Example (Series and Parallel Connections of Relays).

Series Connection 串聯




Parallel Connection 並聯



- The Story. For $n$ electrical relays $R_1, \ldots, R_n$, let

$P(A_k) \Longleftrightarrow A_k = \{R_k \text{ works properly}\}$, → bring in a simpler system.

$k=1,\ldots, n$, and suppose that $\boxed{A_1, \ldots, A_n \text{ are independent}}$.

Recall (LNp.3-6)
$P(A_1\cap\cdots\cap A_n)$
$=P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots$

- Series Connection. The probability that current can flow from $S$ to $E$ (which corresponds to the event $A_1\cap\cdots\cap A_n$) is

$$P(A_1\cap\cdots\cap A_n) \overset{\text{c.f.}}{=} P(A_1)\cdots P(A_n).$$

- Parallel Connection. The probability that current can flow from $S$ to $E$ (which corresponds to the event $A_1\cup\cdots\cup A_n$) is

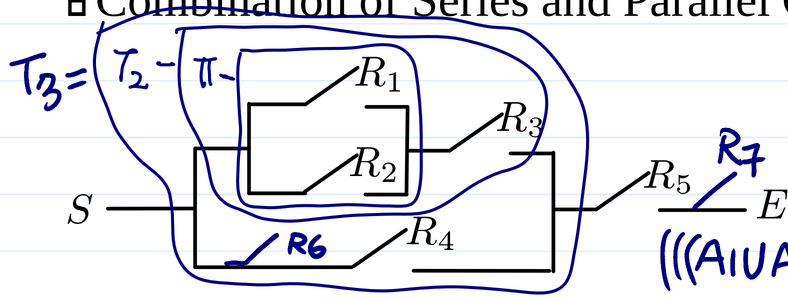
Recall: (LNp.2-9)
$P(A_1\cup\cdots\cup A_n)$
$=\sigma_1-\sigma_2+\sigma_3-\sigma_4 \cdots$

$$P(A_1\cup\cdots\cup A_n) = 1-P(A_1^c\cap\cdots\cap A_n^c)$$
$$= 1-P(A_1^c)\cdots P(A_n^c) = 1-\prod_{k=1}^{n}[1-P(A_k)]$$

$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$

independence often simplifies the calculation of probability. c.f.

- Combination of Series and Parallel Connections




$B_1: \{T_1 \text{ work}\}$
$P(B_1) = 1-(1-P(A_1))(1-P(A_2))$
$B_2: \{T_2 \text{ work}\}$
$P(B_2) = P(B_1)\cdot P(A_3)$

$(((A_1\cup A_2)\cap A_3)\cup(A_4\cap A_6))\cap(A_5\cap A_7)$



➢ Theorem. If $A_1, \ldots, A_n$ are mutually independent and $B_1, \ldots, B_m$, $m\le n$, are formed by taking unions or intersections of mutually exclusive subgroups of $A_1, \ldots, A_n$, then $B_1, \ldots, B_m$ are independent.

$B_1 \xleftarrow{\cup/\cap} \boxed{A_1, A_2}\ \boxed{A_3, A_4, A_5}\ \cdots,\ A_n$
$B_2 \xleftarrow{\cup/\cap}$

Sketch of proof.

For $m=2$, WLOG, suppose that
$B_1 = A_1\cap\cdots\cap A_j$, $B_2 = A_k\cap\cdots\cap A_n$, $1<j<k\le n$
$P(B_1) = P(A_1)P(A_2)\cdots P(A_j)$ $P(B_2)=P(A_k)\cdots\cdots P(A_n)$
$P(B_1\cap B_2) = P(A_1\cap\cdots\cap A_j\cap A_k\cap\cdots\cap A_n)$
$= P(A_1)\cdots\cdots P(A_j)\cdot P(A_k)\cdots\cdots P(A_n) = P(B_1)P(B_2)$

Next, if $B_1 = A_1\cap\cdots\cap A_j$, $1<j<k<n$,
$B_2 = A_k\cup\cdots\cup A_n \Rightarrow B_2^c = A_k^c\cap\cdots\cap A_n^c$
$\Rightarrow B_1$ & $B_2^c$ independent.
$\Rightarrow B_1$ & $B_2$ independent.
the other cases are similar. ※

$(((A_1\cup A_2)\cap A_3)\cup(A_4\cap A_6))\cap A_5$ $A_7$
$C_1$ $C_2$ $C_3$ $C_4$
$D_1$ $D_2$ $D_3$
$E_1$ $E_2$
mutually indep

- Definition (conditional independence): Events $B_1, \ldots, B_n$ are (*pairwise* or *mutually*) *independent* under the probability measure $P(\cdot|A)$ ← LNp.3-3

- e.g., $B_1$ and $B_2$ are conditionally independent given $A$ iff

$$P(B_1 \cap B_2|A) = P(B_1|A)P(B_2|A),$$

or, equivalently, $\frac{P(B_1 \cap B_2|A)}{P(B_2|A)} = \frac{P(B_1 \cap B_2 \cap A)/\cancel{P(A)}}{P(B_2 \cap A)/\cancel{P(A)}} = P(B_1|B_2 \cap A)$

$$P(B_1|B_2 \cap A) = P(B_1|A).$$

- Example (Gold Coins) c.f. Example in LN p.3-11.
  - The Story.
    - Box $i$ contains $i$ gold coins and $k-i$ silver coins, $i=0,1,\ldots,k$.
    - Experiment: (i) Select a box at random and, (ii) draw coins *with replacement* from the box

1st — 2nd — ○

$\Omega$ = $A_0 | A_1 | \cdots | A_k$

G G S … 1st … nth

  - **Q**: Given that first $n$ draws are all gold, what is the probability that $(n+1)^{st}$ draw is gold?
    - Let $A_i$ = {Box $i$ is selected}, $B$ = {first $n$ draws are gold}, $C$ = {$(n+1)^{st}$ draw is gold}
    - By applying law of total probability on $P(\cdot|B)$

$$P(C|B) = \sum_{i=0}^{k} P(A_i|B)P(C|A_i \cap B)$$

$P(\cdot|A_i)$

    - Because $B$ and $C$ are conditionally independent given $A_i$,

$$P(C|A_i \cap B) = P(C|A_i) = i/k$$



    - By Bayes' rule,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=0}^{k} P(A_j)P(B|A_j)} = \frac{\cancel{[1/(k+1)]}(i/k)^n}{\sum_{j=0}^{k}\cancel{[1/(k+1)]}(j/k)^n}$$

(exercise)

    - Hence, $P(C|B) = \sum_{i=0}^{k}(i/k)^{n+1} \Big/ \sum_{i=0}^{k}(i/k)^n \overset{?}{=} P(C)$

$P(\cdot)$

Summary

  - **Q**: Are the two events $B$ and $C$ independent? Ans. No

conditional prob.

Concept: update prob. ∵ new information (some events have occurred)

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

Independence

Def (for calculation)
- 2 events $P(A \cap B) = P(A)P(B)$
- n events

Def: (for interpretation) $P(B|A) = P(B)$

c.f.

3 useful formula
- multiplication law
- Total prob.
- Bayes' rule

↳ complex = simpler & ⋯ & simpler.

$\Omega \to A$
$B \to B \cap A$

c.f.

$P(\cdot) \to P(\cdot|A)$ both defined on subsets of $\Omega$

independence & complement

under independence the calculation of prob. is simpler

c.f.

e.g. $P(A_1 \cap \cdots \cap A_n)$, $P(A_1 \cup \cdots \cup A_n)$, $P((A_1 \cup A_2) \cap A_3 \cdots)$

❖ **Reading**: textbook, Sec 3.4, 3.5