Note 1

GOAL Learn the mathematical foundation used in statistical methods.

① Events and algebra of sets.

- **Q** How to model "random experiments" mathematically?
  A Use set theory.

$\boxed{\text{Terminology}}$

(1) • **sample space** S = set of all possible outcomes.
- an **event** A is a subset of S.
- When the outcome of a random expr lies in A, we say that **event A has occurred**.

(2) Recall:
- $e \in A$: e is an element of A.
- ϕ : empty set
- $A \subseteq B$: A is a subset of B
- $A \cup B$: union of A and B
- $A \cap B$: intersection of A and B.
- A' : complement of A.
- $B \setminus A$: set difference of A and B
  = set of elements in B but not in A.

  (In particular, $A' = S \setminus A$.)

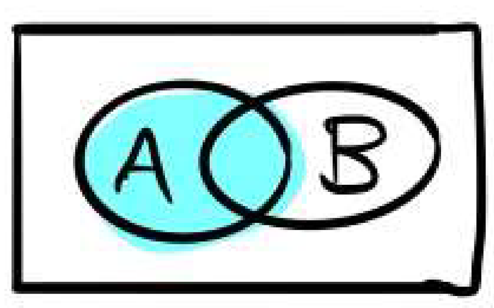
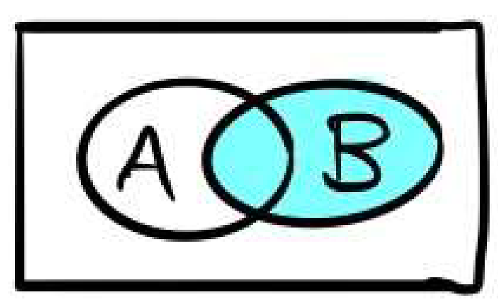
(3) • $A_1, \cdots, A_k$ are **mutually exclusive** if they are disjoint:
$$A_i \cap A_j = \phi \quad \text{whenever } i \neq j.$$

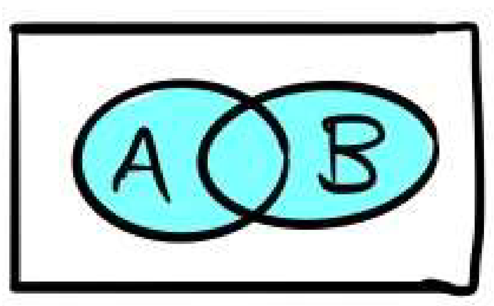
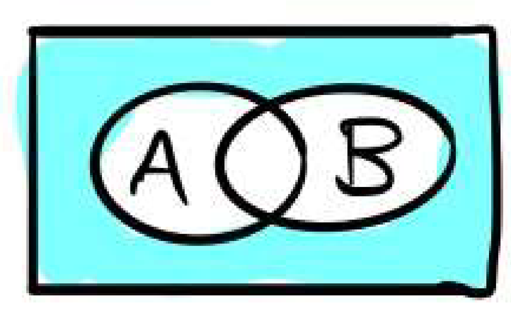
- $A_1, \cdots, A_k$ are **exhaustive** if $S = A_1 \cup \cdots \cup A_k$.

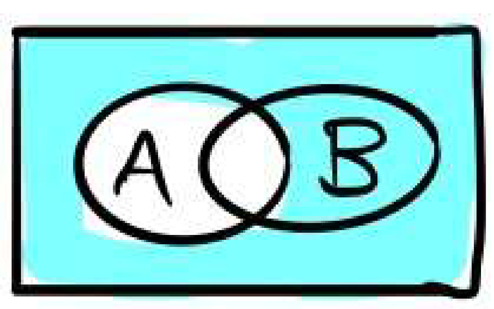
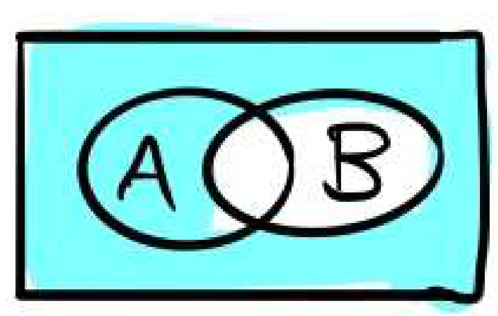
$\boxed{\text{Facts}}$

- (Commutativity) $A \cup B = B \cup A$,
  $A \cap B = B \cap A$.
- (Associativity) $A \cup (B \cup C) = (A \cup B) \cup C$
  $A \cap (B \cap C) = (A \cap B) \cap C$
  (So we can unambiguously write $A \cup B \cup C$ for $A \cup (B \cup C)$ and $A \cap B \cap C$ for $A \cap (B \cap C)$.)
- (Distributivity) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- (De Morgan's Law) $(A \cup B)' = A' \cap B'$
  $(A \cap B)' = A' \cup B'$.

$\boxed{\text{Ex}}$ Using Venn diagrams,

$(A \cup B)' = \Big[$  $\cup$  $\Big]'$

$=$  $'$ $=$ 

$=$  $\cap$ 

$= A' \cap B'$.

② Probability as a set function

$\boxed{\text{Motivation}}$

- Repeat a random expr. n times and write
$$\mathcal{N}(A) = \#\text{ of times } A \text{ occurred.}$$
Then the relative frequency
$$\frac{\mathcal{N}(A)}{n}$$
tends to stabilize to a fixed value p as n increases. We want to associate p to A.
- This p is called the probability of A, denoted by
$$P(A).$$

$\boxed{\text{Ex}}$ Roll a fair dice with 6 faces $1, \cdots, 6$ and let
$$A = [\text{the outcome is an even number.}]$$
Since each face is equally likely and there are 3 faces labeled w/ even number, it makes sense to assign
$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

- We want to "abstractize" the essential properties of relative frequencies:
  - $\mathcal{N}(A)/n$ is defined over all events A
  - $\mathcal{N}(A)/n \geq 0$
  - $\mathcal{N}(S)/n = n/n = 1$
  - If $A_1, A_2, \cdots$ are mutually exclusive,
$$\frac{\mathcal{N}(A_1 \cup A_2 \cup \cdots)}{n} = \frac{\mathcal{N}(A_1)}{n} + \frac{\mathcal{N}(A_2)}{n} + \cdots.$$

So we define

$\boxed{\text{DEF}}$ **Probability** is a real-valued set function

$$P : \left\{ \begin{matrix} \text{collection of} \\ \text{all events} \end{matrix} \right\} \to \mathbb{R}$$

s.t. the following holds:

(a) $P(A) \geq 0$;

(b) $P(S) = 1$;

(c) If $A_1, A_2, \cdots$ are mutually exclusive, then

$$P(A_1 \cup \cdots \cup A_k) = P(A_1) + \cdots + P(A_k)$$

for each positive integer k and

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

$\boxed{\text{Ex}}$ Consider the coin flip. We may set

$$S = \{H, T\}.$$

Then for any $p \in [0,1]$, the set function P specified by

$$P(\phi) = 0, \qquad P(\{H, T\}) = 1,$$
$$P(\{H\}) = p, \qquad P(\{T\}) = 1-p$$

is a probability set function. In particular, there are many different prob. set functions on S.

- Different prob. set functions encode different random experiments.

$\boxed{\text{Ex}}$ (Uniform distribution) If $S = \{e_1, \cdots, e_m\}$, then P defined by

$$P(A) = \frac{N(A)}{N(S)},$$

where $N(A) = \#$ of elements in A, is a probability set function.

This P is uniquely specified by the condition

$$P(\{e_i\}) = \frac{1}{m} \quad \text{for any } i = 1, \cdots, m.$$

In such case, we say outcomes $e_1, \cdots, e_m$ are equally likely.

$\boxed{\text{Ex}^*}$ If $S = [a, b]$, then P defined by

$$P(A) := \frac{[\text{length of } A]}{[\text{length of } S]}$$

is a probability set function. This corresponds to a random expr where an outcome can be "equally likely" to be any value between a and b.

(Rigorously, P may only be defined for "nice" subsets of $[a, b]$. But any sensible sets which we will incounter in this class are indeed nice, so we ignore this technicality.)

③ Properties of P

- We derive various properties of P

$\boxed{\text{THM 1.1-1}}$ $P(A') = 1 - P(A)$.

<u>pf)</u> Noting that $S = A \cup A'$ and $\phi = A \cap A'$,

$$
\begin{aligned}
1 &= P(S) && \text{by DEF. (b)} \\
&= P(A \cup A') \\
&= P(A) + P(A') && \text{by DEF. (c)}
\end{aligned}
$$

□

$\boxed{\text{THM 1.1-2}}$ $P(\phi) = 0$.

<u>1st pf)</u>

$$
\begin{aligned}
P(\phi) &= P(S') \\
&= 1 - P(S) && \text{by THM 1.1-1} \\
&= 1 - 1 && \text{by DEF. (b)} \\
&= 0.
\end{aligned}
$$

□

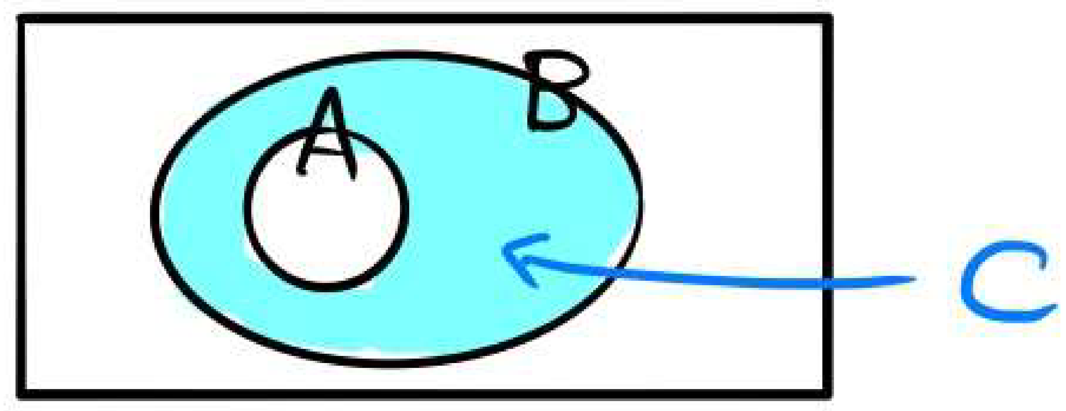
<u>2nd pf)</u> Since $\phi \cup \phi = \phi = \phi \cap \phi$,

$$
\begin{aligned}
P(\phi) &= P(\phi \cup \phi) \\
&= P(\phi) + P(\phi) && \text{by DEF. (c).}
\end{aligned}
$$

Subtracting $P(\phi)$ from both sides proves the thm. □

$\boxed{\text{THM 1.1-3}}$ (Monotonicity) If $A \subseteq B$, then $P(A) \leq P(B)$.

Pf) Write $C = B \setminus A = B \cap A'$.



Then
$$B = A \cup C \quad \text{and} \quad A \cap C = \phi.$$
So
$$\begin{aligned} P(B) &= P(A \cup C) \\ &= P(A) + P(C) && \text{by DEF. (c)} \\ &\geq P(A) + 0 && \text{by DEF. (a)} \\ &= P(A). \end{aligned}$$
$\square$

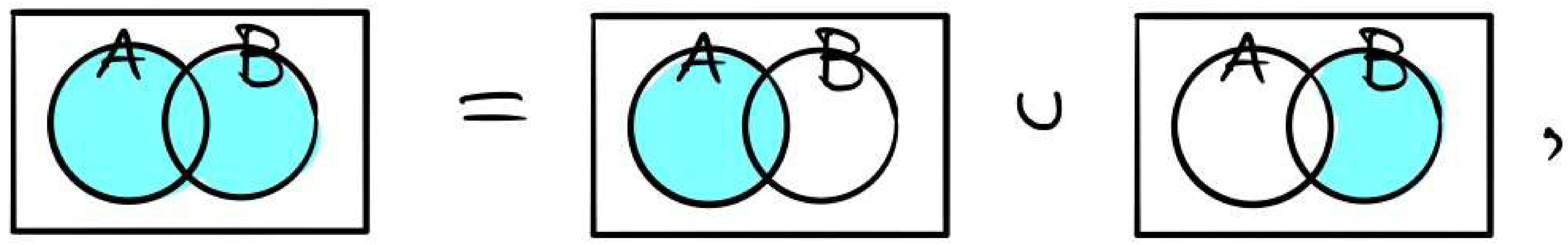
$\boxed{\text{THM 1.1-4}}$ $P(A) \leq 1$.

Pf) Since $A \subseteq S$
$$\begin{aligned} P(A) &\leq P(S) && \text{by THM 1.1-3} \\ &= 1 && \text{by DEF. (b)} \end{aligned}$$
$\square$

Remark. This shows: $0 \leq P(A) \leq 1$ for any event A.

$\boxed{\text{THM 1.1-5}}$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

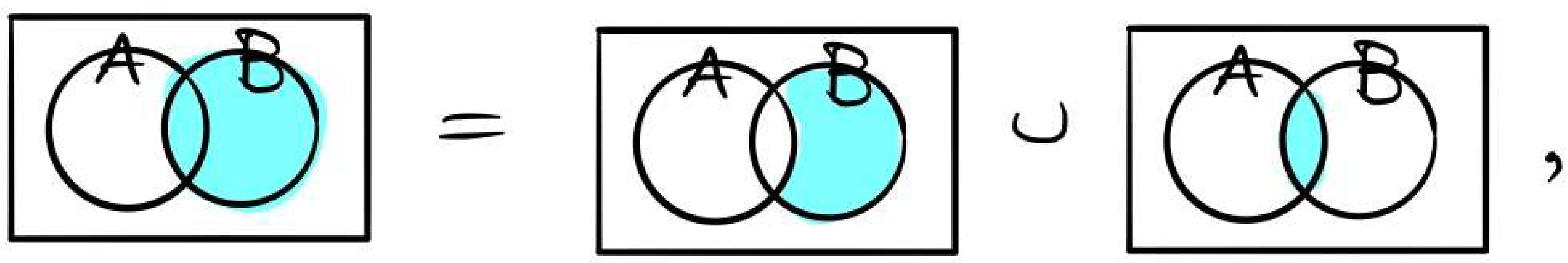
Pf) (1) Since $A \cup B$ is the union of mutually exclusive events A and $B \setminus A$



we get

$$P(A \cup B) = P(A) + P(B \setminus A).$$

(2) Since B is the union of mutually exclusive events $B \setminus A$ and $A \cap B$



we have

$$P(B) = P(B \setminus A) + P(A \cap B).$$

(3) From both equalities,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \setminus A) && \text{by (1)} \\ &= P(A) + (P(B) - P(A \cap B)) && \text{by (2)} \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

$\square$

Rmk. This is a generalization of the identity

$$N(A \cup B) = N(A) + N(B) - N(A \cap B),$$

where the term $N(A \cap B)$ cancels out the "over-counting" of elements in $A \cap B$, which are counted twice in $N(A) + N(B)$.

$\boxed{\text{THM 1.1-6}}$ (Inclusion - Exclusion Principle)

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\
&\quad - P(A \cap B) - P(B \cap C) - P(C \cap A) \\
&\quad + P(A \cap B \cap C).
\end{aligned}
$$

Pf)

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B \cup C) - P(A \cap (B \cup C)) && \text{by THM 1.1-5} \\
&= P(A) + P(B \cup C) - P((A \cap B) \cup (A \cap C)) && \text{De Morgan}
\end{aligned}
$$

$\boxed{\text{HW}}$ Complete the proof by applying THM 1.1-5 again. $\square$