STAT 25100 Lecture 4
Inclusion Exclusion Formula

Yibi Huang
Department of Statistics
University of Chicago

Inclusion Exclusion Formula

Lecture 4 will introduce formula for calculating the probability for the union of several events $E_1, E_2, \dots, E_n$:

$$\mathrm{P}(E_1 \cup E_2 \cup \dots \cup E_n)$$

when $E_1, E_2, \dots, E_n$ are NOT disjoint.

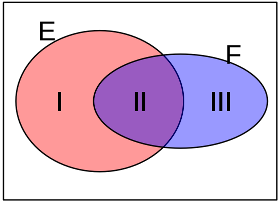
Inclusion Exclusion Formula for Two Events

To find $\mathrm{P}(E \cup F)$, we can split $E \cup F$ into 3 disjoint parts

- I $= E \cap F^c$
- II $= E \cap F$
- III $= E^c \cap F$

By Axiom 3 of Probability,

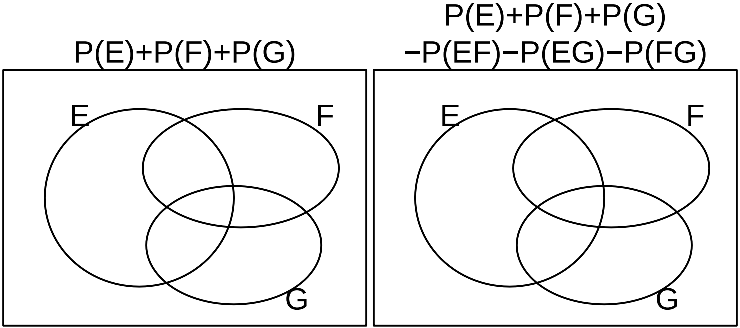
$$\begin{aligned} \mathrm{P}(E \cup F) &= \mathrm{P}(\text{I}) + \mathrm{P}(\text{II}) + \mathrm{P}(\text{III}) \\ \mathrm{P}(E) &= \mathrm{P}(\text{I}) + \mathrm{P}(\text{II}) \\ \mathrm{P}(F) &= \mathrm{P}(\text{II}) + \mathrm{P}(\text{III}) \end{aligned}$$



which shows that

$$\begin{aligned} \mathrm{P}(E \cup F) &= \mathrm{P}(E) + \mathrm{P}(F) - \overbrace{\mathrm{P}(\text{II})}^{=\mathrm{P}(E \cap F)} \\ &= \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \cap F) \end{aligned}$$

Inclusion Exclusion Formula for 3 Events



Thus

$$\begin{aligned} P(E \cup F \cup G) = & \; P(E) + P(F) + P(G) \\ & - P(EF) - P(EG) - P(FG) \\ & + P(EFG) \end{aligned}$$

Inclusion Exclusion Formula for 4 Events

$$
\begin{aligned}
\mathrm{P}(E \cup F \cup G \cup H) = {} & \mathrm{P}(E) + \mathrm{P}(F) + \mathrm{P}(G) + \mathrm{P}(H) \\
& - \mathrm{P}(EF) - \mathrm{P}(EG) - \mathrm{P}(EH) - \mathrm{P}(FG) - \mathrm{P}(FH) - \mathrm{P}(GH) \\
& + \mathrm{P}(EFG) + \mathrm{P}(EFH) + \mathrm{P}(EGH) + \mathrm{P}(FGH) \\
& - \mathrm{P}(EFGH)
\end{aligned}
$$

Inclusion Exclusion Formula for n Events

$$
\begin{aligned}
\mathrm{P}(E_1 \cup E_2 \cup \ldots \cup E_n) = & \sum_{i=1}^{n} \mathrm{P}(E_i) \\
& - \sum_{1 \le i_1 < i_2 \le n} \mathrm{P}(E_{i_1} E_{i_2}) \\
& + \vdots \\
& + (-1)^{r+1} \sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} \mathrm{P}(E_{i_1} E_{i_2} \ldots E_{i_r}) \\
& + \vdots \\
& + (-1)^{n+1} \mathrm{P}(E_1 E_2 \ldots E_n),
\end{aligned}
$$

where the summation

$$
\sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} \mathrm{P}(E_{i_1} E_{i_2} \ldots E_{i_r})
$$

is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \ldots, n\}$.

Example 1 — Sports

Of students in a certain college:

- 36% of the students play tennis,
- 28% play squash,
- 18% play badminton.
- 22% play both tennis and squash
- 12% play both tennis and badminton,
- 9% play both squash and badminton, and
- 4% play all three sports.

What percentage of students play at least one of three sports?

$$
\begin{aligned}
\mathrm{P}(T \cup S \cup B) &= \mathrm{P}(T) + \mathrm{P}(S) + \mathrm{P}(B) - \mathrm{P}(TS) - \mathrm{P}(TB) - \mathrm{P}(SB) + \mathrm{P}(TSB) \\
&= 36\% + 28\% + 18\% - 22\% - 12\% - 9\% + 4\% \\
&= 43\%.
\end{aligned}
$$

Example 2 — "Void" in Bridge

- Recall that a Bridge hand consists of 13 cards.
- A "*void*" in Bridge is a hand with no cards in at least one of the four suits.

TRUE or FALSE and explain:
the probability that a bridge hand is "*void*" in at least one suit

$$\mathrm{P}(\text{void}) = \frac{(\text{4 ways to choose the "void" suit})\binom{39}{13}}{\binom{52}{13}} = \frac{4\binom{39}{13}}{\binom{52}{13}} \approx 0.05116.$$

Example 2 — "Void" in Bridge (2)

Let

- E_1 = the bridge hand includes no clubs (♣)
- E_2 = the bridge hand includes no diamonds (♢)
- E_3 = the bridge hand includes no hearts (♡)
- E_4 = the bridge hand includes no spades (♠)

Note $\{\text{void}\} = \mathrm{E}_1 \cup E_2 \cup E_3 \cup E_4$, and

$$\mathrm{P}(E_i) = \frac{\binom{39}{13}}{\binom{52}{13}}, \quad i = 1, 2, 3, 4.$$

However,

$$\mathrm{P}(\text{void}) \neq \mathrm{P}(E_1) + \mathrm{P}(E_2) + \mathrm{P}(E_3) + \mathrm{P}(E_4)$$

since E_1, E_2, E_3, and E_4 are NOT disjoint.

Example 2 — "Void" in Bridge (3)

Note

$$\mathrm{P}(E_iE_j) = \frac{\binom{26}{13}}{\binom{52}{13}}, \quad \mathrm{P}(E_iE_jE_k) = \frac{\binom{13}{13}}{\binom{52}{13}}, \quad \mathrm{P}(E_1E_2E_3E_4) = 0.$$

By Inclusion Exclusion Formula for 4 Events,

$$\begin{aligned}
\mathrm{P}(\text{void}) &= \mathrm{P}(E_1 \cup E_2 \cup E_3 \cup E_4) \\
&= \mathrm{P}(E_1) + \mathrm{P}(E_2) + \mathrm{P}(E_3) + \mathrm{P}(E_4) \\
&\quad - \mathrm{P}(E_1E_2) - \mathrm{P}(E_1E_3) - \mathrm{P}(E_1E_4) - \mathrm{P}(E_2E_3) - \mathrm{P}(E_2E_4) - \mathrm{P}(E_3E_4) \\
&\quad + \mathrm{P}(E_1E_2E_3) + \mathrm{P}(E_1E_2E_4) + \mathrm{P}(E_1E_3E_4) + \mathrm{P}(E_2E_3E_4) \\
&\quad - \mathrm{P}(E_1E_2E_3E_4) \\
&= 4 \cdot \frac{\binom{39}{13}}{\binom{52}{13}} - \binom{4}{2} \cdot \frac{\binom{26}{13}}{\binom{52}{13}} + \binom{4}{3} \cdot \frac{\binom{13}{13}}{\binom{52}{13}} - 0 \approx 0.05107.
\end{aligned}$$

Example 3 — The Matching Problem

Suppose that each of N men at a party throws his hat into the center of the room.
The hats are first mixed up, and then each man randomly selects a hat.
What is the probability that none of the men selects his own hat?

Example 3 — The Matching Problem — Sol (1)

Let E_i = the ith man gets his own hat.

- $\{\text{None gets his own hat}\}^c = \{\text{at least one gets his own hat}\} = E_1 \cup E_2 \cup \ldots \cup E_N$
- $\mathrm{P}(E_i) =$
- $\mathrm{P}(E_i E_j) =$
- $\mathrm{P}(E_{i_1} E_{i_2} \ldots E_{i_r}) =$

Example 3 — The Matching Problem — Sol (1)

Let E_i = the ith man gets his own hat.

- $\{\text{None gets his own hat}\}^c = \{\text{at least one gets his own hat}\} = E_1 \cup E_2 \cup \ldots \cup E_N$
- $\mathrm{P}(E_i) =$
- $\mathrm{P}(E_i E_j) =$
- $\mathrm{P}(E_{i_1} E_{i_2} \ldots E_{i_r}) =$

Thus,

$$\sum_{1 \le i_1 < i_2 < \ldots < i_r \le N} \mathrm{P}(E_{i_1} E_{i_2} \ldots E_{i_r}) = \binom{N}{r} \frac{(N-r)!}{N!} = \frac{N!}{r!(N-r)!} \frac{(N-r)!}{N!} = \frac{1}{r!}.$$

since there are $\binom{N}{r}$ terms in the sum above.

Example 3 — The Matching Problem — Sol (2)

By the Inclusion-Exclusion Formula,

$$\begin{aligned}
\mathrm{P}(E_1 \cup E_2 \cup \ldots \cup E_N) &= \sum_{i=1}^{N} \mathrm{P}(E_i) - \sum_{1 \le i_1 < i_2 \le n} \mathrm{P}(E_{i_1} E_{i_2}) + \cdots \\
&\quad + (-1)^{r+1} \sum_{1 \le i_1 < i_2 < \ldots < i_r \le N} \mathrm{P}(E_{i_1} E_{i_2} \ldots E_{i_r}) \\
&\quad + \cdots + (-1)^{N+1} \mathrm{P}(E_1 E_2 \ldots E_N) \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{N+1} \frac{1}{N!}
\end{aligned}$$

Finally, P(None gets his own hat) is

$$1 - \mathrm{P}(E_1 \cup E_2 \cup \ldots \cup E_N) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} - \cdots + (-1)^N \frac{1}{N!} = \sum_{k=0}^{N} \frac{(-1)^k}{k!}.$$

When $N \uparrow \infty$, the probability approaches $\sum_{k=0}^{-\infty} \frac{(-1)^k}{k!} = e^{-1} \approx 0.3679$.