

Q10 — Frequent Itemsets (10 points)

Setup

Itemset	Frequency
a	7
b	10
c	17
d	13
(a,b)	7
(b,d)	2
(b,c)	6

Support threshold: $s = 5$. Item e has **no known information**.

Key property: Downward closure (monotonicity) — if any subset of an itemset is infrequent, the itemset is guaranteed infrequent.

Answers

1. (a, b, c, d) — Guaranteed infrequent

$(b,d) \subset (a,b,c,d)$ and $\text{freq}(b,d) = 2 < 5$. By downward closure, (a,b,c,d) must be infrequent.

2. (a, b, d) — Guaranteed infrequent

$(b,d) \subset (a,b,d)$ and $\text{freq}(b,d) = 2 < 5$. Same reasoning as above.

3. (a, b, e) — Could be frequent or infrequent

The subset (a,b) has $\text{freq} = 7 \geq 5$ (frequent), so downward closure does not rule it out. However, we have **no information about item e** — it could co-occur with (a,b) in 5+ baskets (making (a,b,e) frequent) or in fewer than 5 (making it infrequent).

4. (a, d) — Guaranteed infrequent

Critical observation: $\text{freq}(a) = 7$ and $\text{freq}(a,b) = 7$. Since every basket containing a also contains b, we have:

$$\text{freq}(a, d) = \text{freq}(a, b, d) \leq \text{freq}(b, d) = 2 < 5$$

Every basket with a *must* also have b, so any basket with both a and d is also a basket with b and d. Since there are only 2 such baskets, $\text{freq}(a,d) \leq 2$.

5. (c, d) — Could be frequent or infrequent

Both c and d are individually frequent (17 and 13). We lack direct information about their co-occurrence. From the given data:

- Baskets with c but not b: $17 - 6 = 11$
- Baskets with d but not b: $13 - 2 = 11$

These non-b baskets could overlap significantly (pushing $\text{freq}(c,d)$ well above 5) or not at all. We cannot determine $\text{freq}(c,d)$ from the available information.

Summary

Itemset	Classification	Key Reasoning
(a,b,c,d)	Guaranteed infrequent	$(b,d) \subset \text{it}$, $\text{freq}(b,d) = 2 < 5$
(a,b,d)	Guaranteed infrequent	$(b,d) \subset \text{it}$, $\text{freq}(b,d) = 2 < 5$
(a,b,e)	Could be either	(a,b) is frequent but e is unknown
(a,d)	Guaranteed infrequent	$\text{freq}(a) = \text{freq}(a,b) = 7 \Rightarrow a \subset b \text{ always} \Rightarrow \text{freq}(a,d) = \text{freq}(a,b,d) \leq \text{freq}(b,d) = 2$
(c,d)	Could be either	Both frequent individually; co-occurrence unknown