1. Direct Sampling

a. Strengths:

- Simple to implement for known prob.
- Works well for easily sampled distributions

Weakness:

- Inefficient for rare events or constrained datasets

Context:

- Generates travel preferences and stress levels directly but struggles with low-probability events (eg. low-stress but realistic)

Rejection Sampling

Strengths:

- Handles constrained distributions by rejecting unsuitable samples

Weakness:

① Inefficient when many samples are discarded

Context:

① Useful for isolating specific scenarios (eg. skewed troubles) but can waste computational resources.

Gibbs Sampling:

Strengths:

① Efficient for high-dimensional data and complex joint probabilities

② Iterative conditioning generates meaningful samples

Weakness:

① Slow convergence; depends on initial values.

② Requires conditional probabilities

Context:

① Suitable for estimating relationships (eg. travel mode, purpose, stress) but may take longer for precise results.

(b)

Prob. of leisure travel given train preference:

$$P(\text{leisure} \mid \text{train}) = 0.400$$

no. of people preferring train travel:

$$N_{\text{train}} = 30$$

Expected leisure travelers $= P(\text{leisure} \mid \text{train}) \times N_{\text{train}}$

Expected leisure travelers $= 0.40 \times 30$

$= 12$

So, 12 travelers for leisure by train.

(c) Prob. of preferring air travel = $P(\text{Air}) = 0.800$

Prob. of business travel given air travel =

$P(\text{Business} \mid \text{Air}) = 0.200$

By Chain Rule, we have

$P(\text{Air} \cap \text{Business}) = P(\text{Air}) \times P(\text{Business} \mid \text{Air})$

$P(\text{Air} \cap \text{Business}) = 0.800 \times 0.200 = 0.160$

So, Prob. that a randomly selected person prefers air travel and travels for business is 0.160.

(d) Answer:

① Impossible only if sampling process is unbiased.

② Larger samples reduce random variability, leading to estimates closer to the true value.

Precision:

① Increases as larger samples lower the standard error.

② Reduces more consistent and reliable estimates.

Implications On Dataset:

① Stabilizes probabilities like P(Rain=Yes) making more reliable.

② Enhances representation of rare events (eg. low bus travelers)

③ Increases computational cost, requiring a balance between precision and resource.

2. Random Variables:

R: Person read books

J: Person writes academic journals

C: Person participates in book clubs

S1: $P(R \vee J) = 0.910$

S2: $P(J|R) = 0.400$, $P(\neg J|R) = 0.600$

S3: $P(C|R) = 0.320$

S4: $P(J \wedge \neg R) = 0.227$

S5: $P(\neg R \wedge \neg J) = 0.090$

S6: $P(J|\neg R) = 0.716$, $P(\neg J|\neg R) = 0.284$

S7: $P(C \wedge \neg J) = 0.088$

S8: $P(C \vee J) = 0.631$

S9: $P(J|C) = 0.400$

S10: $P(J) = 0.500$

S11: $P(C|\neg R) = 0.0044$

(b.) Axioms of probability satisfied:

① All propositions have $0 \leq P(a) \leq 1$

Non-Negativity

② Normalization: $P(S) = 1$, S: sample space

③ Additivity: $P(A \wedge B) = P(A) + P(B) - P(A \wedge B)$

$\forall A \wedge B$

Verification:

① $P(\neg B \wedge \neg J) = 0.090$ align with sample space's complement

② $P(B \vee J) = 0.910$ is consistent as

$$P(\neg B \wedge \neg J) + P(B \vee J) = 1$$

③ Additivity holds between overlapping events eg. $P(J) = 0.500$ aligns with given condition

Normalization: prob. for all possible outcomes sum to 1

$$\sum P(\text{outcome}) = 1$$

(c.)

R	J	C	Joint Probability	Probabilities
1	1	1	$P(R \wedge J \wedge C)$	0.08736
1	1	0	$P(R \wedge J \wedge \neg C)$	0.273 − 0.08736 = 0.18564
1	0	1	$P(R \wedge \neg J \wedge C)$	0.04384
1	0	0	$P(R \wedge \neg J \wedge \neg C)$	0.137 − 0.04384 = 0.09316
0	1	1	$P(\neg R \wedge J \wedge C)$	0.001
0	1	0	$P(\neg R \wedge \neg J \wedge \neg C)$	0.227 − 0.001 = 0.226
0	0	1	$P(\neg R \wedge \neg J \wedge C)$	0.000396
0	0	0	$P(\neg R \wedge \neg J \wedge \neg C)$	0.090 − 0.000396 = 0.089604

(i.)

① $P(R \vee J) = 0.910$, so $P(\neg R \wedge \neg J) = 0.090$.

② $P(J) = 0.500$ split b/w $P(J \wedge R)$ & $P(J \wedge \neg R)$

③ $P(J \wedge \neg R) = 0.227$, so $P(J \wedge R) = 0.500 - 0.227$

$= 0.273$

④ $P(\neg J \wedge R) = 1 - P(J \wedge R \vee \neg R) - P(\neg R \wedge \neg J)$

$= 0.910 - 0.273 = 0.137$

⑤ $P(C|R) = 0.320$ split b/w J & $\neg J$

⑥ $P(C \mid \neg R) = 0.0044$ given regardless of J

⑦ $P(C \wedge J) = 0.088$

⑧ $P(J \mid C) = 0.400$, so $P(C \wedge \neg J) = 0.400 \times P(C)$

split R and $\neg R$

$P(R) = P(R \wedge J) + P(R \wedge \neg J) = 0.273 + 0.137$
$= 0.410$

$P(\neg R) = 1 - P(R) = 0.590$

split J and $\neg J$

$P(J) = P(J \wedge R) + P(J \wedge \neg R) = 0.273 + 0.227$
$= 0.500$

$P(\neg J) = 1 - P(J) = 0.500$

Compute C Given R

$P(C \wedge R \wedge J) = P(R \wedge J) \times P(C \mid R)$
$= 0.273 \times 0.320 = 0.08736$

$P(C \wedge R \wedge \neg J) = P(R \wedge \neg J) \times P(C \mid R)$
$= 0.137 \times 0.320 = 0.04384$

Compute C Given $\neg R$

$$P(C \wedge \neg R \wedge J) = P(\neg R \wedge J) \times P(C \mid \neg R)$$

$$= 0.227 \times 0.0044 = 0.001$$

$$P(C \wedge \neg R \wedge \neg J) = P(\neg R \wedge \neg J) \times P(C \mid \neg R)$$

$$= 0.090 \times 0.0044 = 0.000396$$

Finally, verifying Total

$$0.08736 + 0.18564 + 0.04384 + 0.09316 + 0.001 + 0.226 + 0.000396 + 0.089604$$

$= 1.000$ satisfying Normalization

(d) Case $R \mid C/J$

$$P(B \wedge C \mid J) = P(B \mid J)\, P(C \mid J)$$

$$P(R \mid J) = \frac{P(R \wedge J)}{P(J)} = \frac{0.273}{0.500} = 0.546$$

$$P(C \mid J) = \frac{P(C \wedge J)}{P(J)} = \frac{0.088}{0.500} = 0.176$$

Joint $P(R \wedge C | J) = \frac{P(R \wedge J \wedge C)}{P(J)}$

$= \frac{0.08736}{0.500} = 0.17472$

verifying $P(R \wedge C | J) = P(R|J) P(C|J)$

$P(R|J) \times P(C|J) = 0.546 \times 0.176$

$= 0.096096$

$P(R \wedge C | J) = 0.17472 \neq 0.096096$

R and C are not conditionally independent given J

Case 2 $J \perp C | B$

$P(J \wedge C | R) = P(J|R) P(C|R)$

S-1 $P(J|R) = \frac{P(R \wedge J)}{P(R)} = \frac{0.273}{0.410} = 0.659$

$P(C|R) = \frac{P(C \wedge R)}{P(R)} = \frac{0.08736 + 0.04384}{0.410}$

$$= \frac{0.1312}{0.410} = 0.32$$

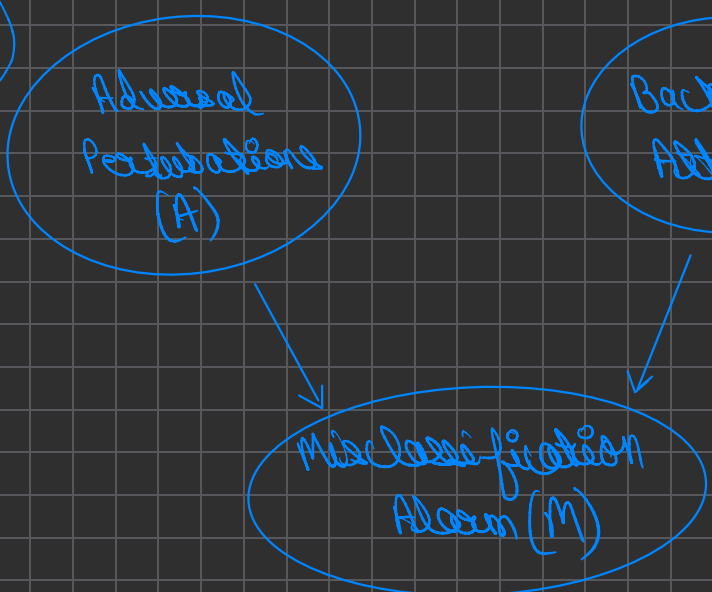
$$P(J \wedge C \mid R) = P(J \mid R)\, P(C \mid R)$$

$$P(J \mid R) \times P(C \mid R) = 0.6659 \times 0.32$$
$$= 0.213$$

$$P(J \wedge C \mid R) = 0.213 = 0.213$$

J and C are conditionally independent given B

8.



Initial Observation, A and B are independent

(a)

$$\text{Posterior Probability} = \frac{P(M|A) \cdot P(A)}{P(M)}$$

$$P(M) = P(M|A,B) \cdot P(A \cap B) + P(M|A,\neg B) \cdot P(A \cap \neg B) + P(M|\neg A,B) \cdot P(\neg A \cap B)$$

(b) Prior Probabilities

$P(A)$, $P(B)$ → independent

$P(A)$: initial/prior belief of adversarial perturbations

$P(B)$: initial/prior belief of backdoor attacks

Likelihood Probabilities are:

$P(M|A)$: Prob. of observing misclassification given that adversarial perturbation has occurred

$P(M|B)$: Prob. of observing misclassification given that backdoor attack has occurred

Posterior Probability

$P(A|M)$: Post. belief OR likelihood of adversarial perturbation given that a misclassification has occurred

$P(B|M)$: Updated belief in backdoor attacks after observing M

(c)

(i) Initial Independence : Before the report Adversarial perturbations (A) and backdoor attacks (B) were independent causes of M. Observing M increased the likelihood of both A and B

(ii) Input of New Information : The report increases $P(B)$, the prior probability of backdoor attacks, making B a more likely explanation of M

(iii) Explaining Away Effect : As B becomes more probable, it explains away M, reducing the posterior prob. of A being the cause :

$$P(A \mid M, \text{high } P(B)) < P(A \mid M, \text{original } P(B))$$

Conclusion: The increased likelihood of burglary attacks shifts the explanation for M towards B, decreasing the belief that A caused the mis-identification. This is a result of the explaining-away effect.