LECTURE 2: Conditioning and Bayes' rule

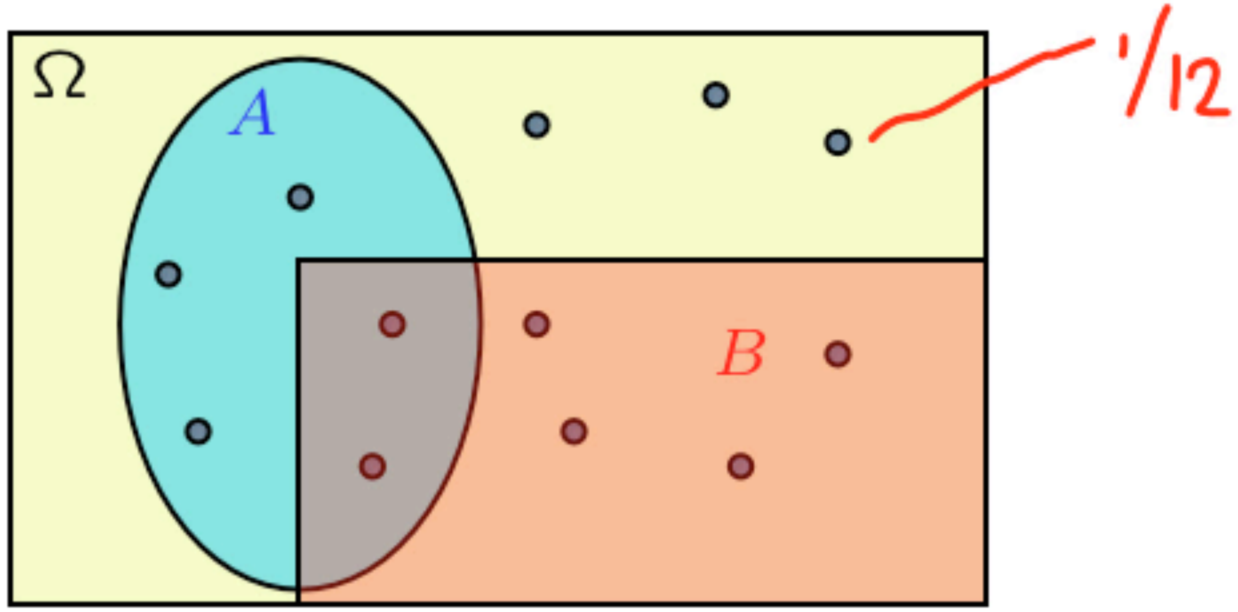
- Conditional probability

- Three **important** tools:
  - Multiplication rule
  - Total probability theorem
  - Bayes' rule ($\longrightarrow$ inference)

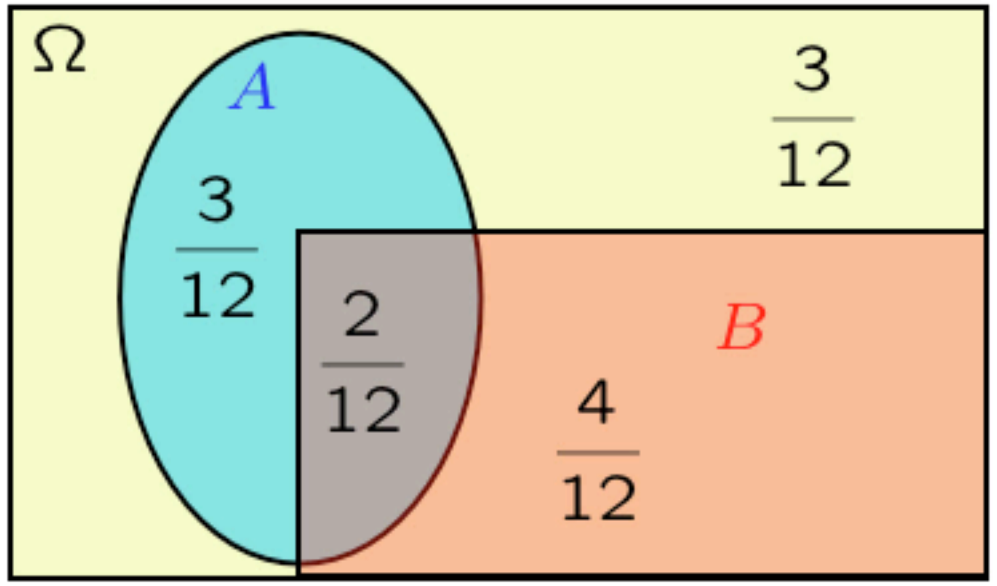
The idea of conditioning

Use new information to revise a model

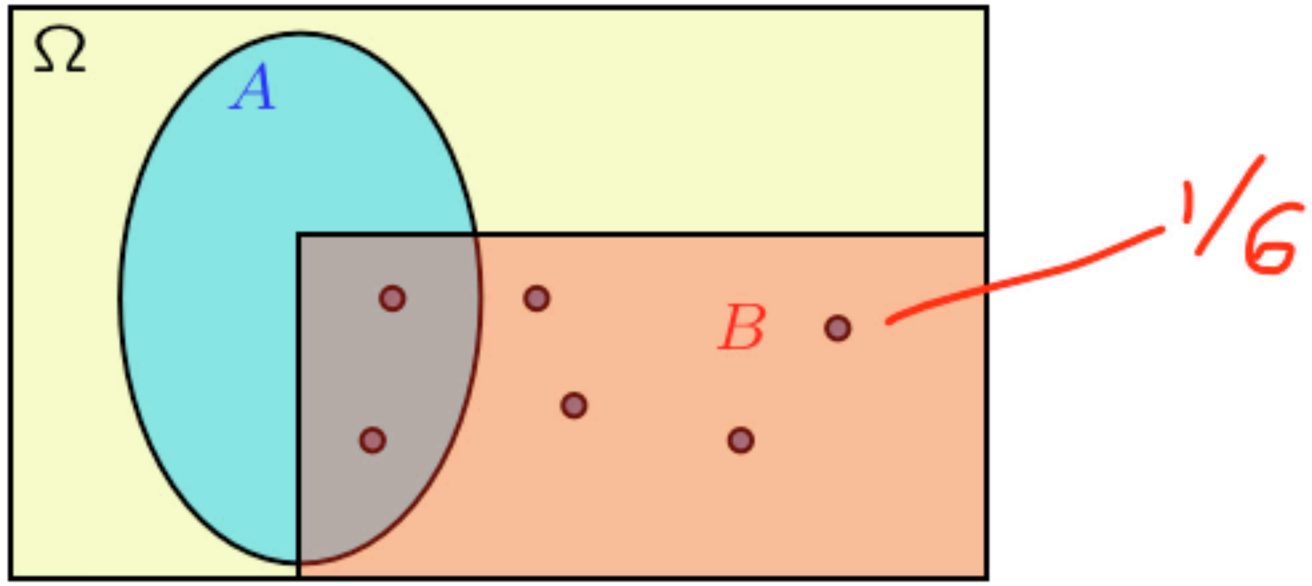
Assume 12 equally likely outcomes



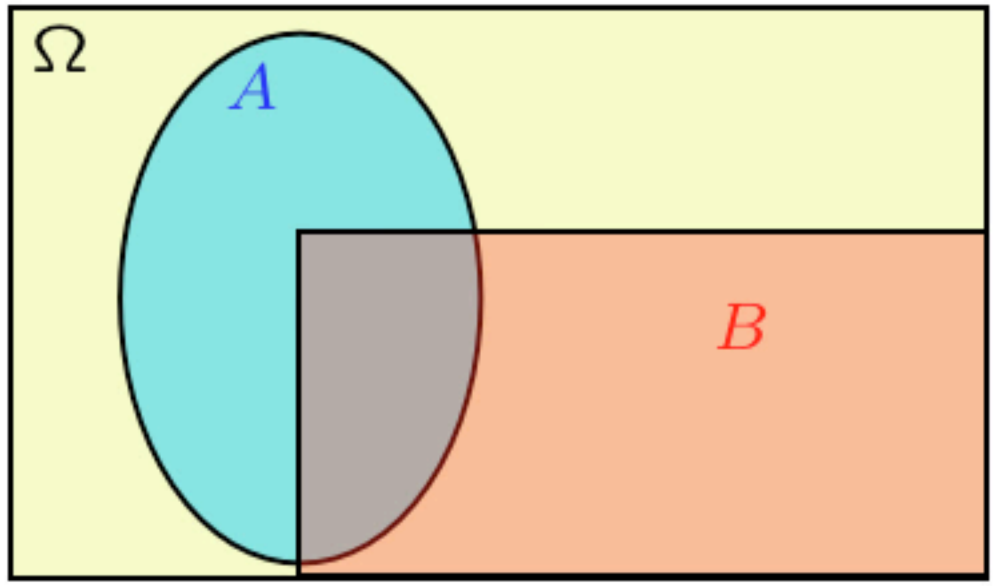
$$\mathbf{P}(A) = \frac{5}{12} \qquad \mathbf{P}(B) = \frac{6}{12}$$



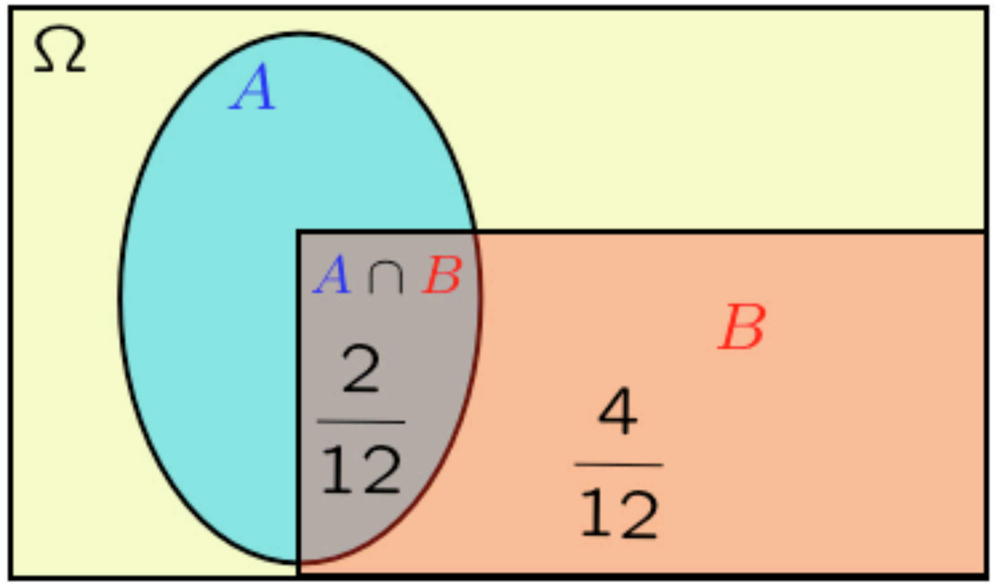
If told B occurred:



$$\mathbf{P}(A \mid B) = \frac{2}{6} = \frac{1}{3} \qquad \mathbf{P}(B \mid B) = 1$$



Definition of conditional probability



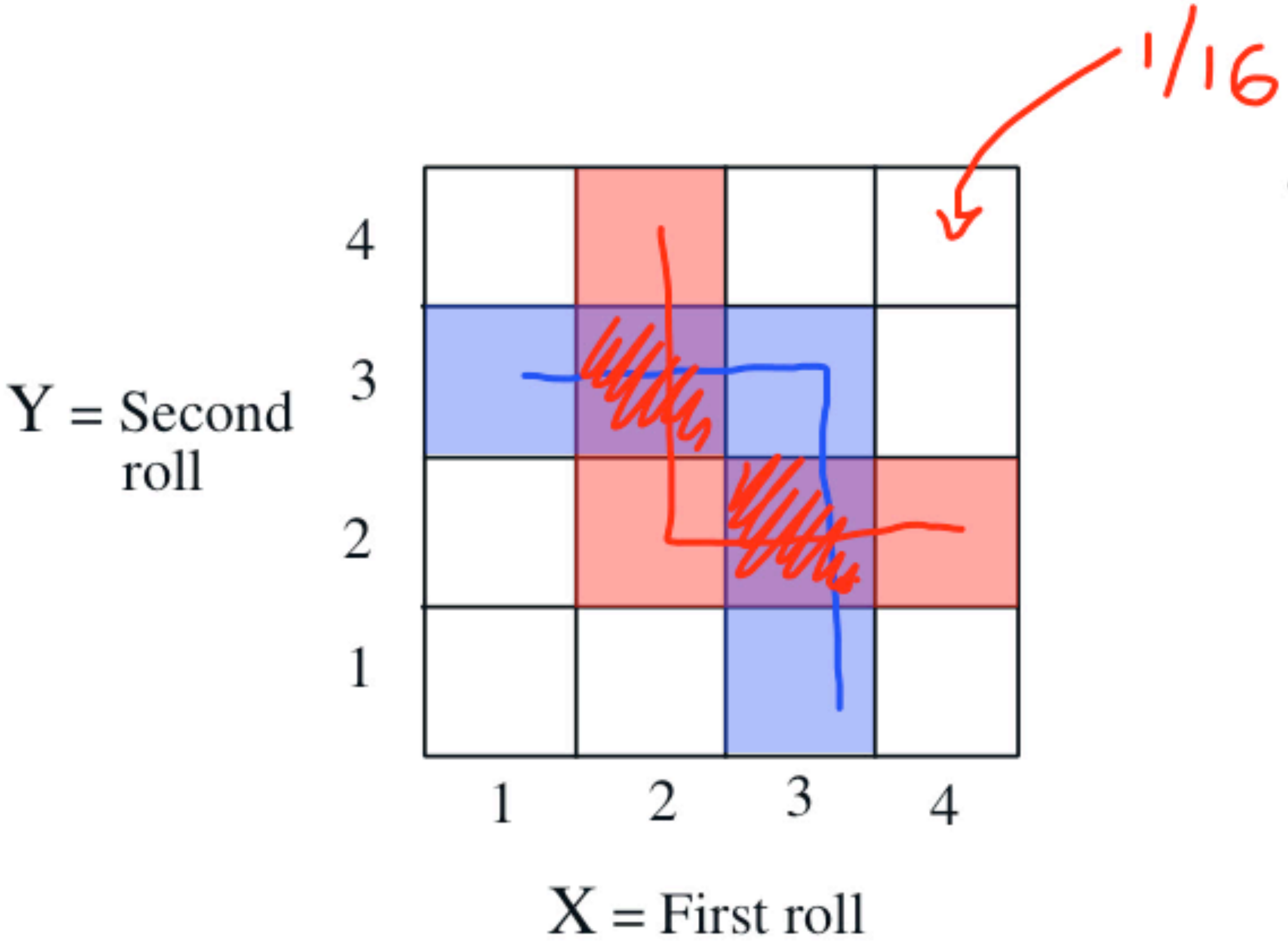
- $\mathbf{P}(A \mid B) =$ "probability of A, given that B occurred"

Def. $\longrightarrow$

$$\mathbf{P}(A \mid B) \stackrel{\Delta}{=} \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \leftarrow \quad = \frac{2/12}{6/12} = \frac{1}{3}$$

defined only when $\mathbf{P}(B) > 0$

Example: two rolls of a 4-sided die



- Let B be the event: $\min(X, Y) = 2$

  Let $M = \max(X, Y)$

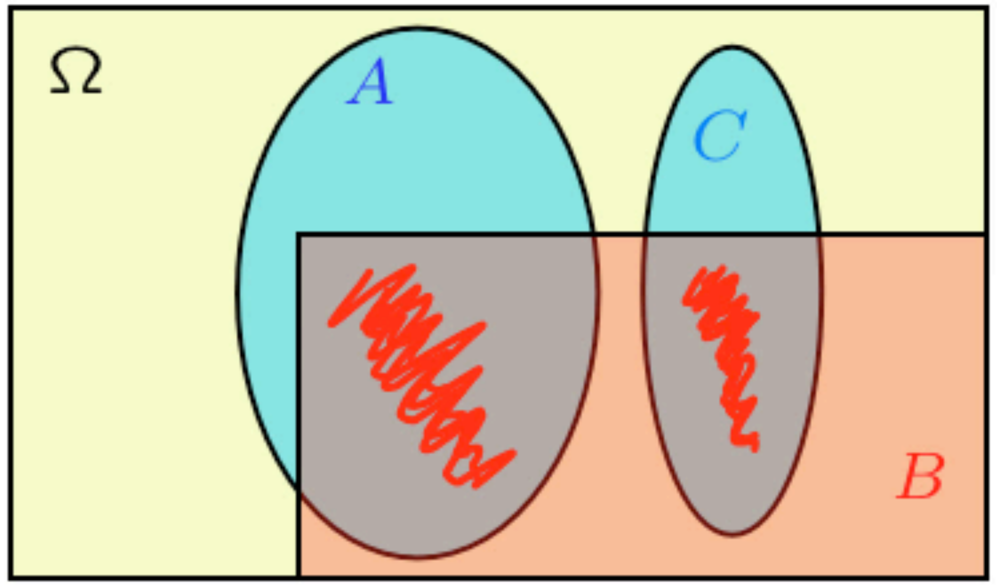
$$\mathbf{P}(M = 1 \mid B) = 0$$

$$\mathbf{P}(\underline{M = 3} \mid B) = \frac{P(M=3 \text{ and } B)}{P(B)}$$

$$= \frac{2/16}{5/16} = \frac{2}{5}$$

Conditional probabilities share properties of ordinary probabilities

$\mathbf{P}(A \mid B) \geq 0$ assuming $\mathbf{P}(B) > 0$



$\mathbf{P}(\Omega \mid B) = \dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

$\mathbf{P}(B \mid B) = \dfrac{P(B \cap B)}{P(B)} = 1$

If $A \cap C = \emptyset$, then $\mathbf{P}(A \cup C \mid B) = \mathbf{P}(A \mid B) + \mathbf{P}(C \mid B)$

$= \dfrac{P((A \cup C) \cap B)}{P(B)} = \dfrac{P((A \cap B) \cup (C \cap B))}{P(B)} = \dfrac{P(A \cap B) + P(C \cap B)}{P(B)} =$

$= P(A \mid B) + P(C \mid B)$

also finite additivity

countable

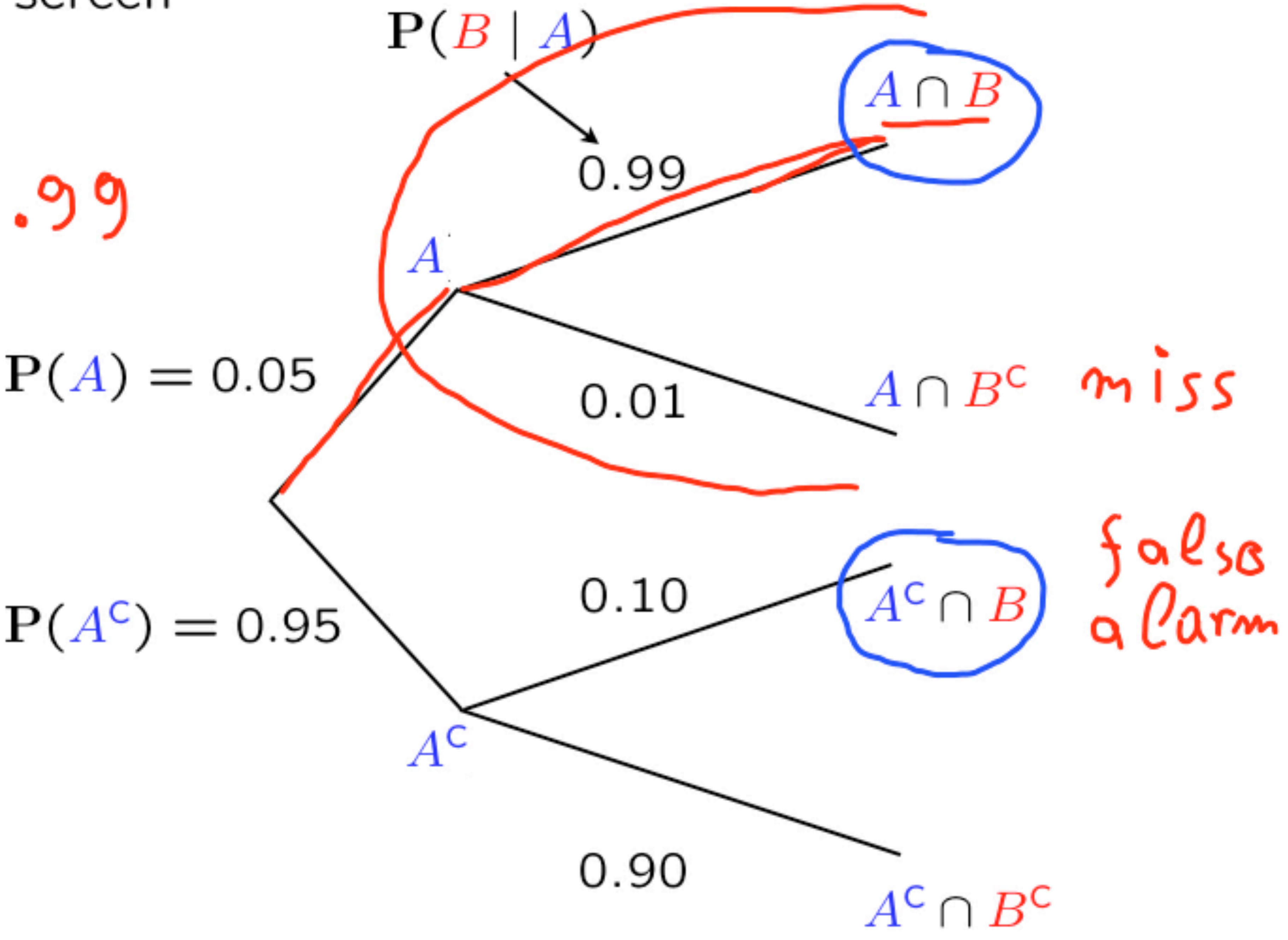
Models based on conditional probabilities

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \qquad \mathbf{P}(B \mid A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$$

Event A: Airplane is flying above

Event B: Something registers on radar screen

- $\mathbf{P}(A \cap B) = P(A) \cdot P(B|A) = 0.05 \cdot 0.99$
- $\mathbf{P}(B) = 0.05 \cdot 0.99 + 0.95 \cdot 0.1 = 0.1445$
- $\mathbf{P}(A \mid B) = \dfrac{0.05 \cdot 0.99}{0.1445} = 0.34$

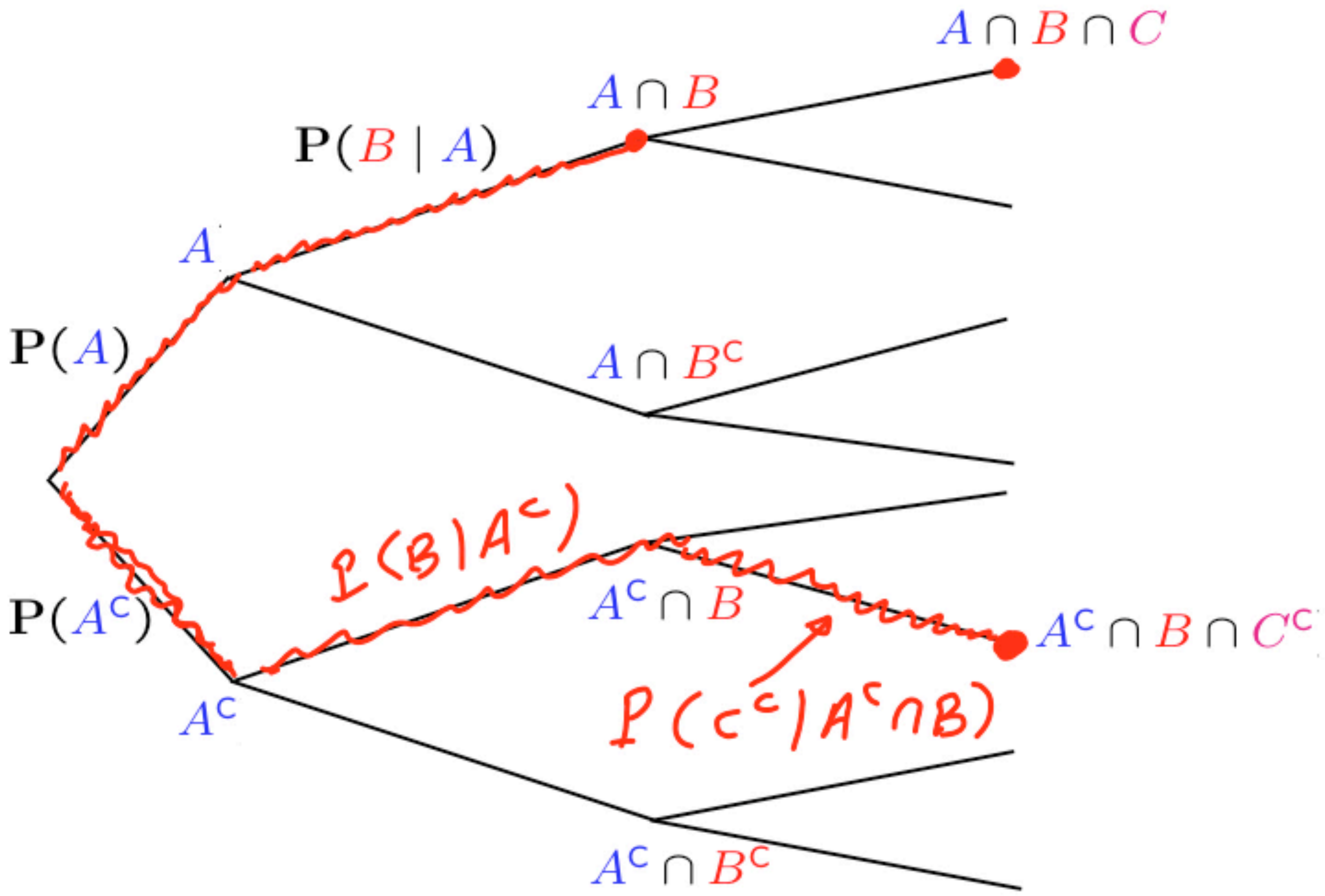


The multiplication rule

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

$$\mathbf{P}(A \cap B) = \mathbf{P}(B)\,\mathbf{P}(A \mid B)$$

$$= \mathbf{P}(A)\,\mathbf{P}(B \mid A)$$



$$\mathbf{P}(\underbrace{(A^c \cap B)}\cap \underbrace{C^c}) =$$

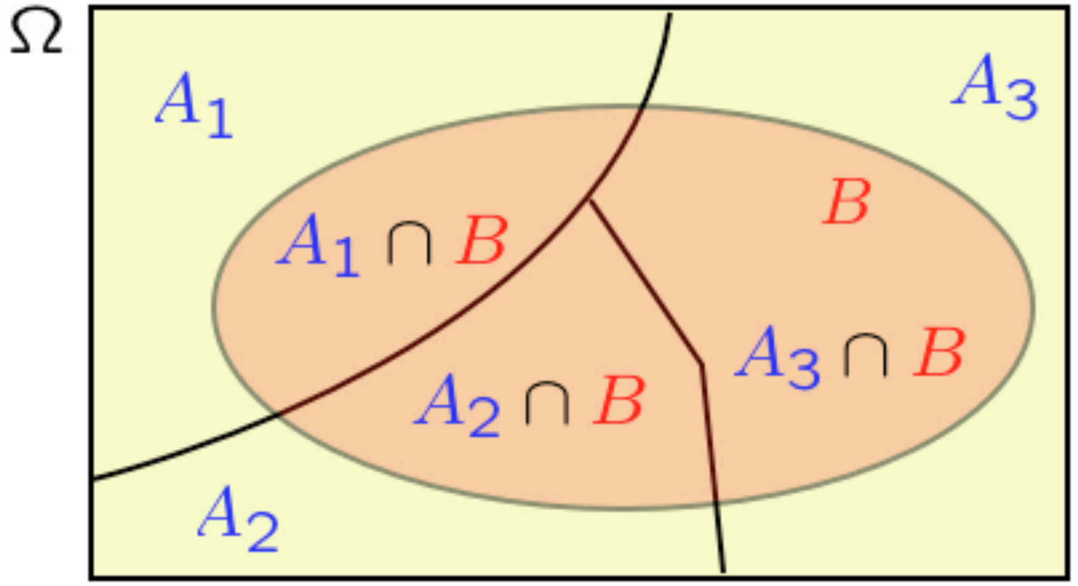
$$= \mathbf{P}(A^c \cap B)\,\mathbf{P}(C^c \mid A^c \cap B)$$

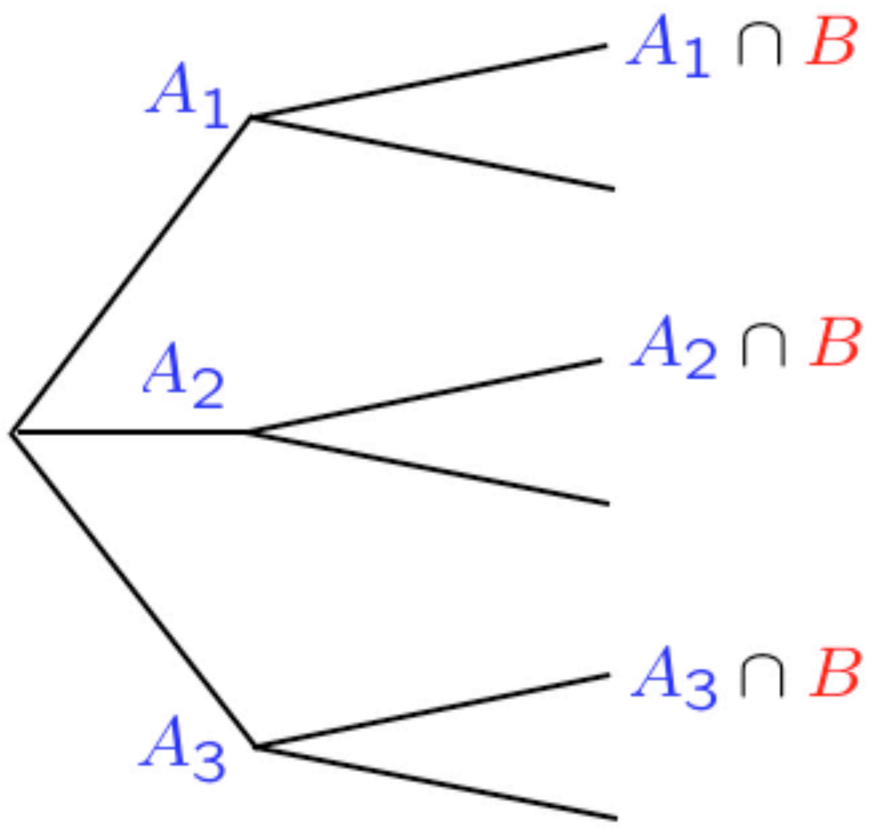
$$= \mathbf{P}(A^c)\cdot\mathbf{P}(B \mid A^c)\,\mathbf{P}(C^c \mid A^c \cap B)$$

$$\mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$

$$= \mathbf{P}(A_1)\prod_{i=2}^{n}\mathbf{P}(A_i \mid A_1 \cap \cdots \cap A_{i-1})$$

Total probability theorem





- Partition of sample space into $A_1, A_2, A_3, \ldots$
- Have $\mathbf{P}(A_i)$, for every i
- Have $\mathbf{P}(B \mid A_i)$, for every i

$$\mathbf{P}(B) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2) + \mathbf{P}(B \cap A_3)$$

$$= \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \cdots$$

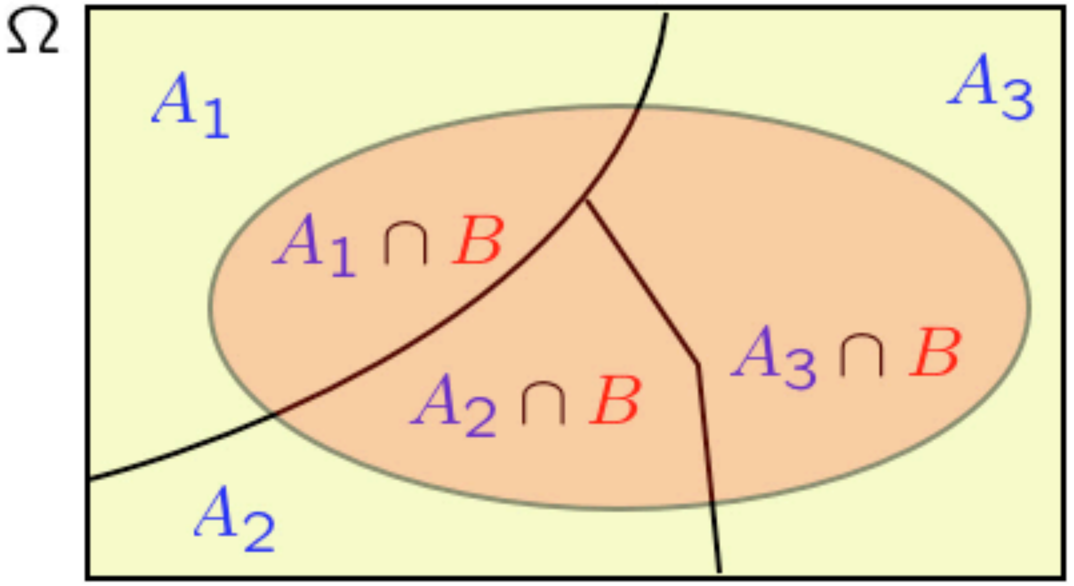
$$\sum_i \mathbf{P}(A_i) = 1$$

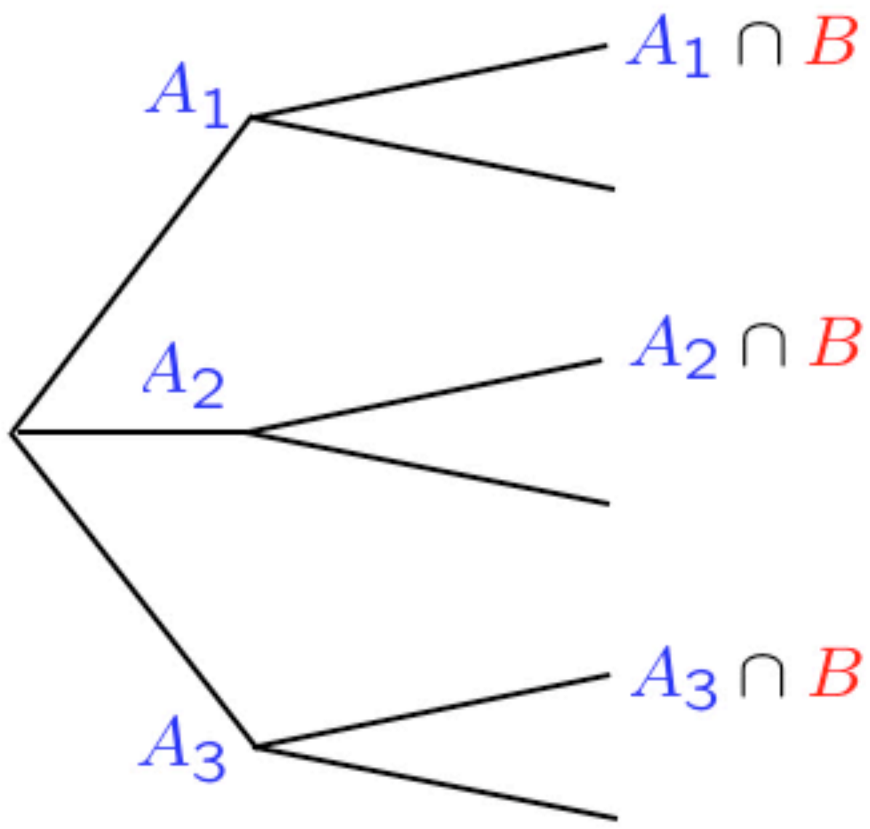
weights

$$\boxed{\mathbf{P}(B) = \sum_i \mathbf{P}(A_i)\,\mathbf{P}(B \mid A_i)}$$

weighted average of $\mathbf{P}(B \mid A_i)$

Bayes' rule





- Partition of sample space into A_1, A_2, A_3
- Have $\mathbf{P}(A_i)$, for every i initial "beliefs"
- Have $\mathbf{P}(B \mid A_i)$, for every i

revised "beliefs," given that B occurred:

$$\mathbf{P}(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

$$\boxed{\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\sum_j \mathbf{P}(A_j)\mathbf{P}(B \mid A_j)}}$$

Bayes' rule and inference

- Thomas Bayes, presbyterian minister (c. 1701-1761)
- "Bayes' theorem," published posthumously
- systematic approach for incorporating new evidence

- Bayesian inference
  - initial beliefs $\mathbf{P}(A_i)$ on possible causes of an observed event B
  - model of the world under each A_i: $\mathbf{P}(B \mid A_i)$

$$A_i \xrightarrow[\mathbf{P}(B \mid A_i)]{\text{model}} B$$

  - draw conclusions about causes

$$B \xrightarrow[\mathbf{P}(A_i \mid B)]{\text{inference}} A_i$$