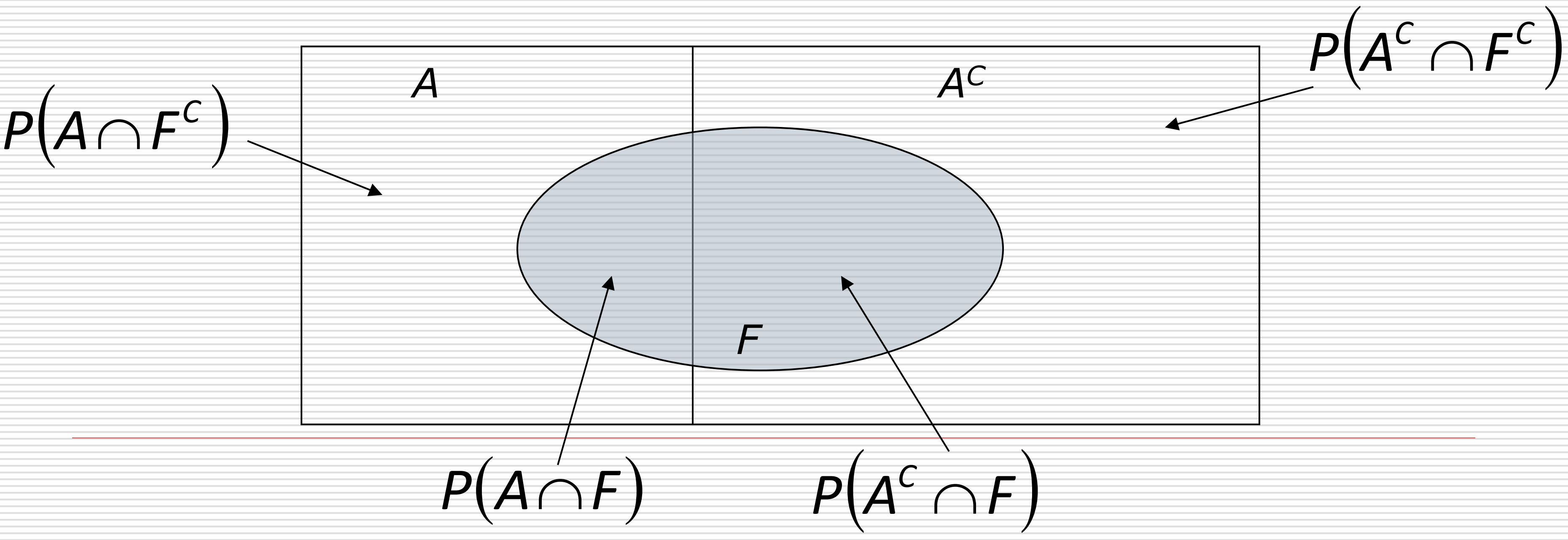
Bayes' Theorem

Bayes' Theorem

- An insurance company divides its clients into two categories: those who are accident prone and those who are not. Statistics show there is a 40% chance an accident prone person will have an accident within 1 year whereas there is a 20% chance non-accident prone people will have an accident within the first year.
- If 30% of the population is accident prone, what is the probability that a new policyholder has an accident within 1 year?

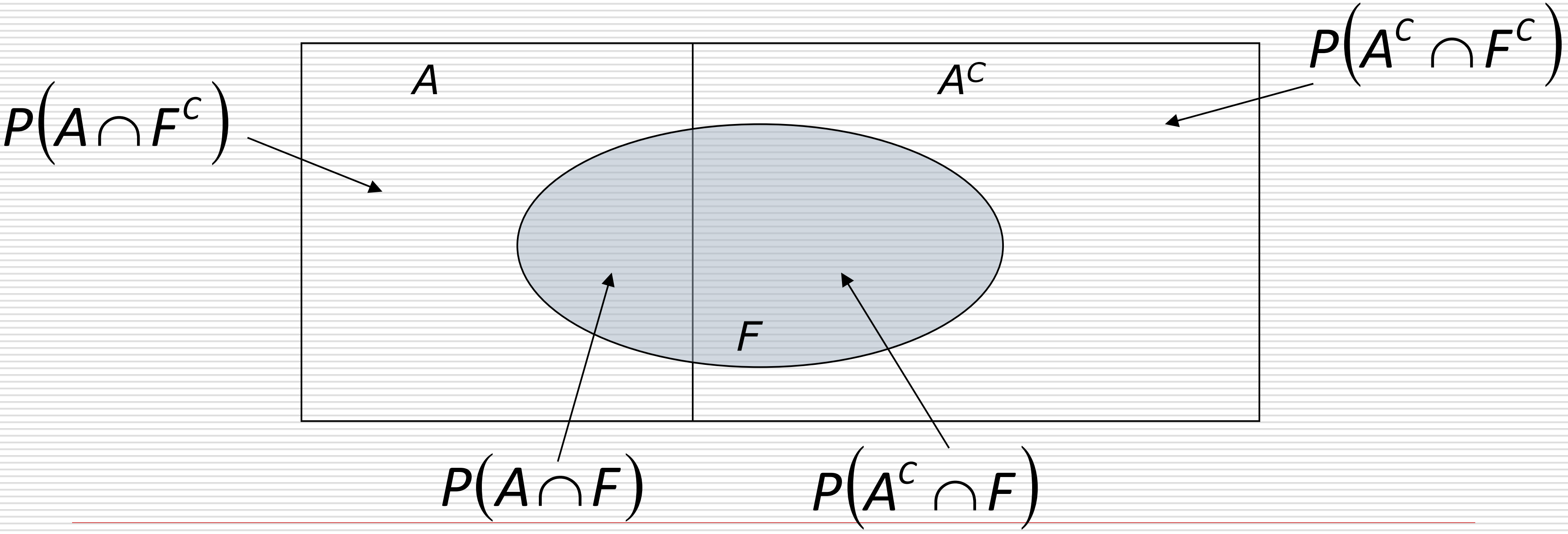
Bayes' Theorem

- Let A be the event a person is accident prone
- Let F be the event a person has an accident within 1 year



Bayes' Theorem

- Notice we've divided up or partitioned the sample space along accident prone and non-accident prone



Bayes' Theorem

- Notice that $(A \cap F)$ and $(A^c \cap F)$ are mutually exclusive events and that
$$(A \cap F) \cup (A^c \cap F) = F$$
- Therefore $P(F) = P(A \cap F) + P(A^c \cap F)$
- We need to find $P(A \cap F)$ and $P(A^c \cap F)$
- How?

Bayes' Theorem

- Recall from conditional probability
  - $$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

$$P(E \mid F) \cdot P(F) = P(E \cap F)$$

  - $$P(F \mid E) = \frac{P(F \cap E)}{P(E)}$$

$$P(F \mid E) \cdot P(E) = P(E \cap F) = P(F \cap E)$$

Bayes' Theorem

- Thus:
  - $P(A \cap F) = \boxed{P(F \mid A)} \cdot \boxed{P(A)}$
  - $P(A^c \cap F) = \underbrace{P(F \mid A^c)} \cdot \underbrace{P(A^c)}$

$P(A)$ = 0.30 since 30% of population is accident prone

$P(F|A)$ = 0.40 since if a person is accident prone, then his chance of having an accident within 1 year is 40%

$P(F|A^C)$ = 0.2 since non-accident prone people have a 20% chance of having an accident within 1 year

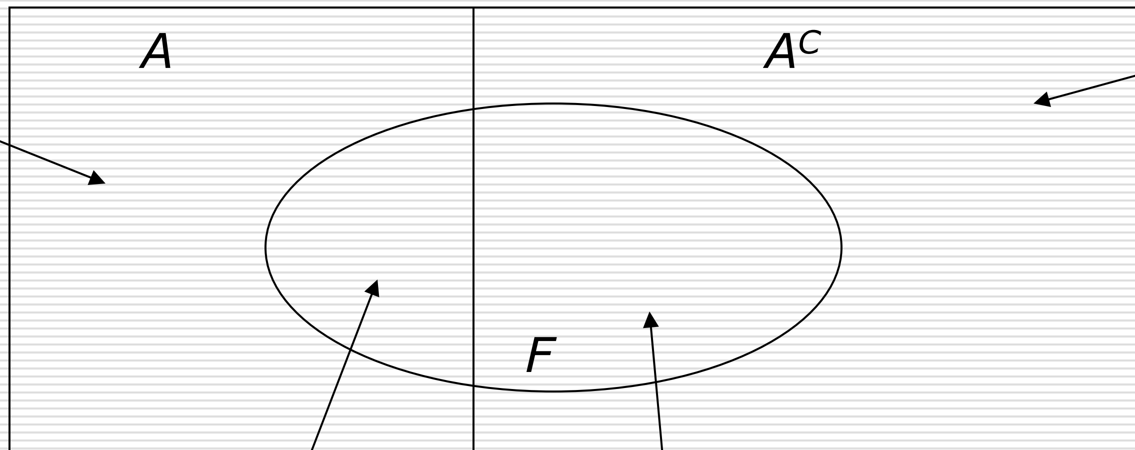
$P(A^C)$ = 1- $P(A)$ = 0.70

Bayes' Theorem

- Updating our Venn Diagram

$$P\left(A^c \cap F^c\right)$$

$$P\left(A \cap F^c\right)$$



$$P(A \cap F) = P(F \mid A)P(A)$$

$$P\left(A^c \cap F\right) = P\left(F \mid A^c\right)P\left(A^c\right)$$

- Notice again that $P(F) = P(A \cap F) + P\left(A^c \cap F\right)$

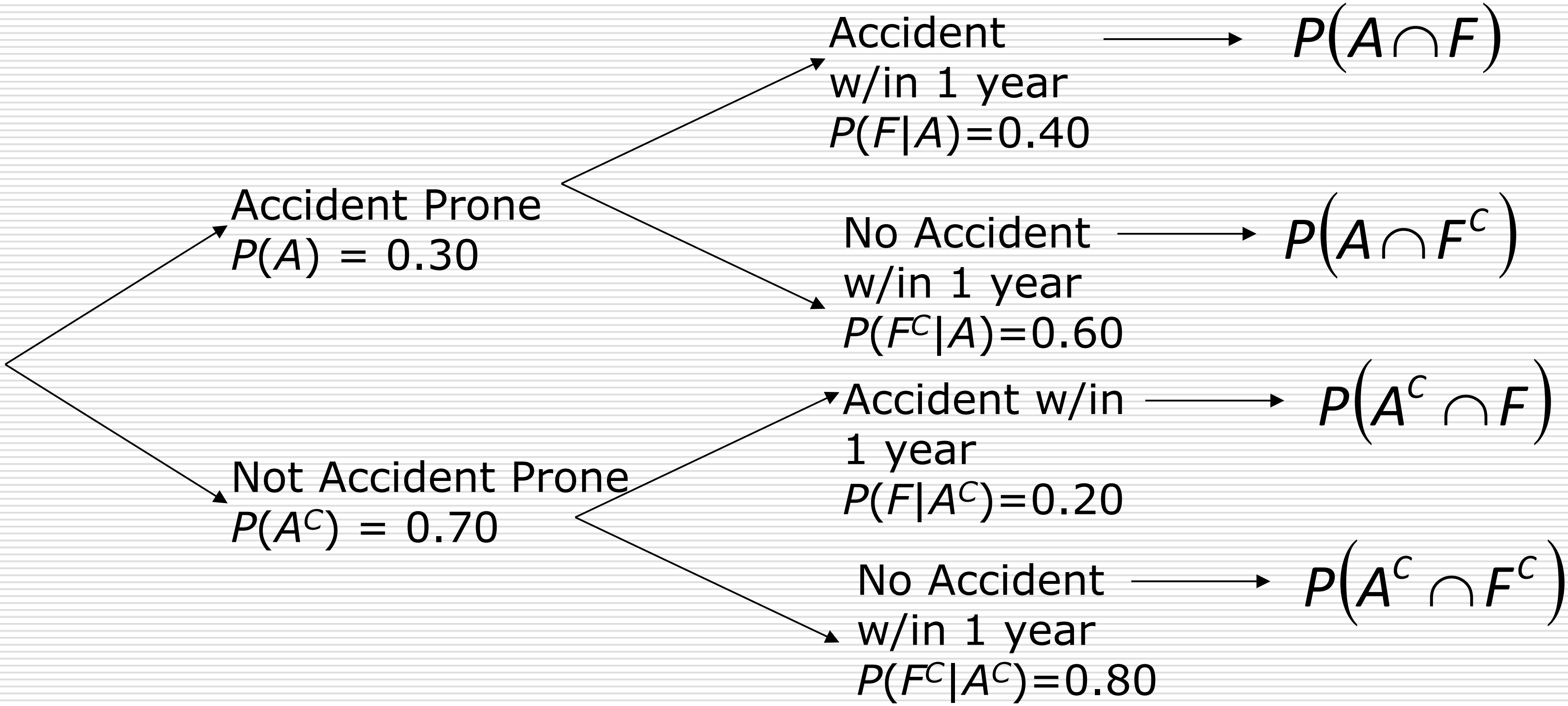
Bayes' Theorem

- So the probability of having an accident within 1 year is:

- $$\begin{aligned} P(F) &= P(A \cap F) + P(A^c \cap F) \\ &= P(F \mid A)P(A) + P(F \mid A^c)P(A^c) \\ &= 0.40(0.30) + 0.20(0.70) = 0.26 \end{aligned}$$

Bayes' Theorem

- Using Tree Diagrams:



Bayes' Theorem

- Notice you can have an accident within 1 year by following branch *A* until *F* is reached
  - The probability that *F* is reached via branch *A* is given by $P(F|A) \cdot P(A)$
  - In other words, the probability of being accident prone and having one within 1 year is

$$P(A \cap F) = P(F|A) \cdot P(A)$$

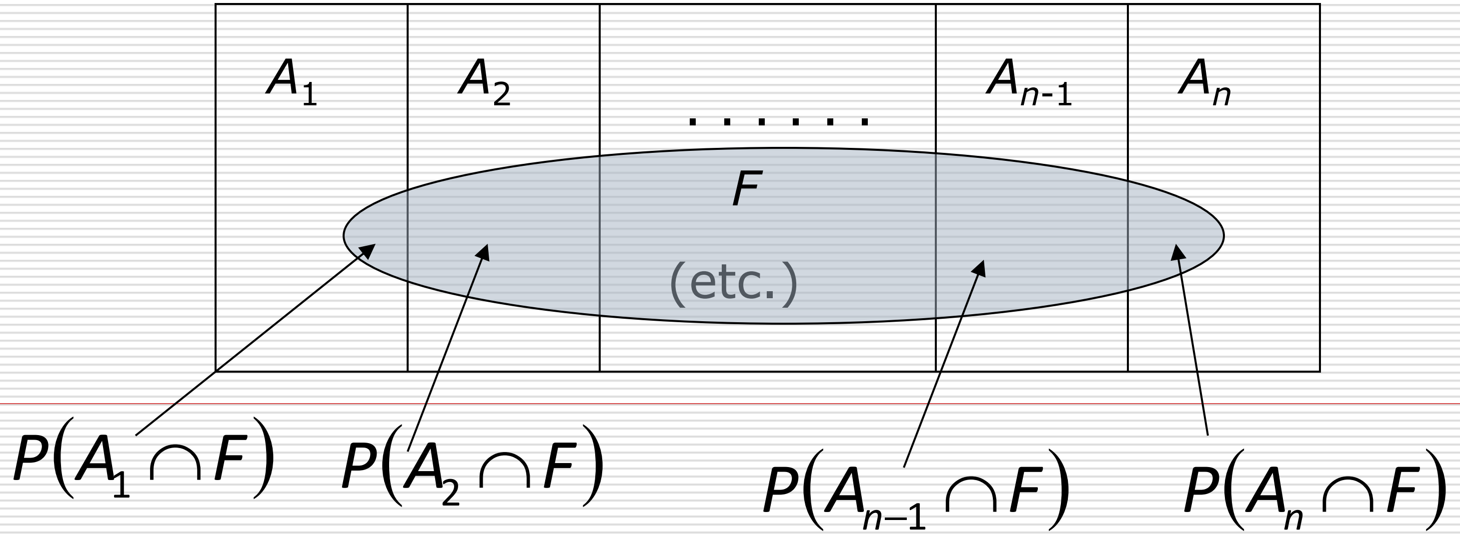
Bayes' Theorem

- You can also have an accident within 1 year by following branch A^C until F is reached
  - The probability that F is reached via branch A^C is given by $P(F \mid A^C) \cdot P(A^C)$
  - In other words, the probability of NOT being accident prone and having one within 1 year is

$$P(A^C \cap F) = P(F \mid A^C) \cdot P(A^C)$$

Bayes' Theorem

- What would happen if we had partitioned our sample space over more events, say $A_1, A_2, \ldots, A_n$, all them mutually exclusive?
- Venn Diagram



Bayes' Theorem

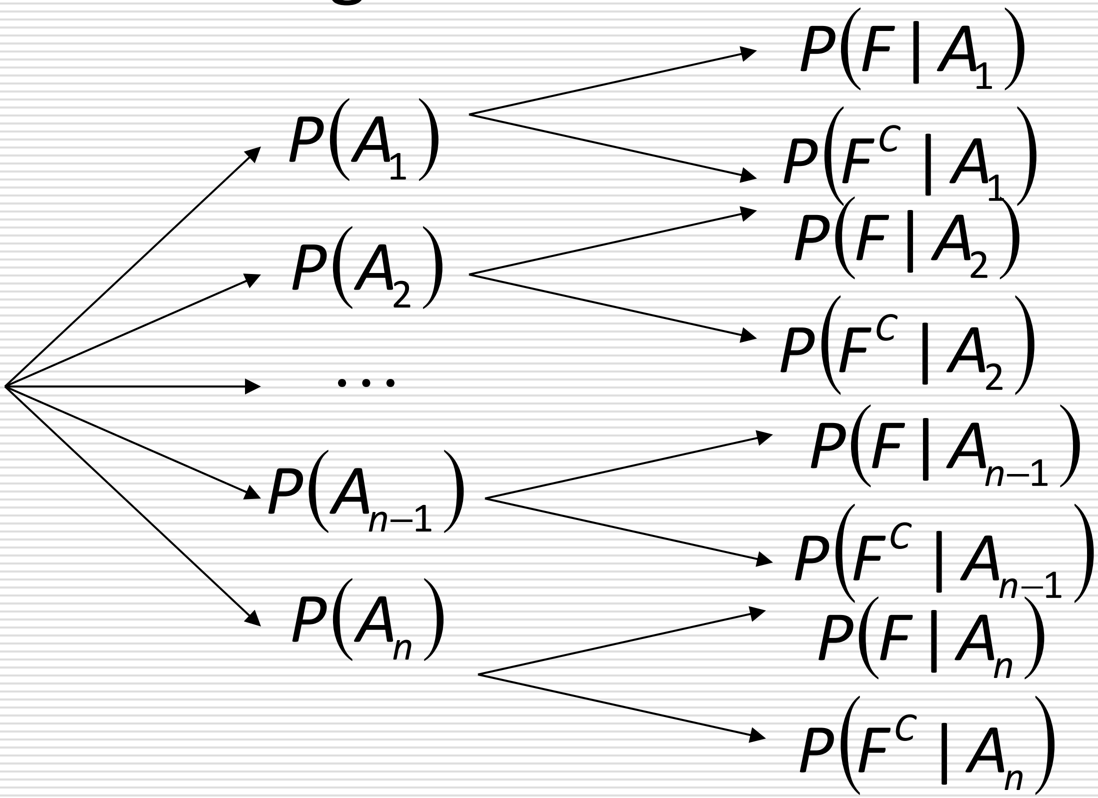
- $P(F) = P(A_1 \cap F) + P(A_2 \cap F) + \cdots + P(A_n \cap F)$

- For each $P(A_i \cap F) = P(F \mid A_i) P(A_i)$

$$
\begin{aligned}
P(F) &= P(A_1 \cap F) + P(A_2 \cap F) + \cdots + P(A_n \cap F) \\
&= P(F \mid A_1) P(A_1) + P(F \mid A_2) P(A_2) + \cdots + P(F \mid A_n) P(A_n) \\
&= \sum_{i=1}^{n} P(F \mid A_i) P(A_i)
\end{aligned}
$$

Bayes' Theorem

- Tree Diagram



Bayes' Theorem

- Notice that F can be reached via $A_1, A_2, \ldots, A_n$ branches
- Multiplying across each branch tells us the probability of the intersection
- Adding up all these products gives:

$$P(F) = \sum_{i=1}^{n} P(F \mid A_i) P(A_i)$$

Bayes' Theorem

- **<u>Ex: 2</u>** (text tractor example) Suppose there are 3 assembly lines: Red, White, and Blue. Chances of a tractor not starting for each line are 6%, 11%, and 8%. We know 48% are red and 31% are blue. The rest are white. What % don't start?

Bayes' Theorem

- **Soln.**

  R: red — $P(R) = 0.48$

  W: white — $P(W) = 0.21$

  B: blue — $P(B) = 0.31$

  N: not starting

  $P(N \mid R) = 0.06$

  $P(N \mid W) = 0.11$

  $P(N \mid B) = 0.08$

Bayes' Theorem

- **Soln.**

$$\begin{aligned} P(N) &= P(N \mid R)\cdot P(R) + P(N \mid W)\cdot P(W) + P(N \mid B)\cdot P(B) \\ &= (0.06)\cdot(0.48) + (0.11)\cdot(0.21) + (0.08)\cdot(0.31) \\ &= 0.0767 \end{aligned}$$

Bayes' Theorem

- Main theorem:

  Suppose we know $P(E \mid F)$. We would like to use this information to find $P(F \mid E)$ if possible.

  Discovered by Reverend Thomas Bayes

Bayes' Theorem

- Main theorem:

  - Ex. Suppose B_1 and B_2 partition a space and A is some event.

  - Use $P(B_1), P(B_2), P(A \mid B_1)$, and $P(A \mid B_2)$ to determine $P(B_1 \mid A)$.

Bayes' Theorem

- Recall the formulas: $P(B_1 \mid A) = \dfrac{P(B_1 \cap A)}{P(A)}$

$$P(B_1 \cap A) = P(A \cap B_1) = P(A \mid B_1) \cdot P(B_1)$$

$$P(A) = P(B_1) \cdot P(A \mid B_1) + P(B_2) \cdot P(A \mid B_2)$$

- So, $P(B_1 \mid A) = \dfrac{P(B_1 \cap A)}{P(A)}$

$$= \frac{P(A \mid B_1) \cdot P(B_1)}{P(B_1) \cdot P(A \mid B_1) + P(B_2) \cdot P(A \mid B_2)}$$

Bayes' Theorem

- Bayes' Theorem:

$$P(B_k \mid A) = \frac{P(A \mid B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)}$$