Data Reliability

Interpreting Sensor Data

Notes and Reminders

- *MPs start next week*. If you have a 4-person group, send the names to me and TA *today*:
  - zaher@Illinois.edu
  - CC: maamin2@illinois.edu

  If you do not have one (or do not send email today), a group will be assigned to you.
- *HW2 will be out tonight*. (HW1 solutions out tonight too.) It is due in a week.

Analogy

- System reliability challenge:
  - Building reliable systems from less reliable components
- Data reliability challenge:
  - Making reliable conclusions from less reliable (sensor) data

Making Conclusions from Probabilistic Data

- Cyber-physical systems obtain data about their environment via sensors
- Sensors (or data sources in general) are often imperfect
- The challenge is: how to correctly compute the "state of the environment" given sensor readings?

Review: Things You Should Know About Probabilities

- Probability of multiple simultaneous events
  - What are the odds that it rains and my basement floods? Say P(rains) = 0.2. P(flood) = 0.1

Review: Things You Should Know About Probabilities

- Probability of multiple simultaneous events
  - What are the odds that it rains and my basement floods?
  - Answer: It is the odds that "it rains", times the odds that "my basement floods given that it rains":

$$P(\text{rain, flood}) = P(\text{rain})\, P(\text{flood}|\text{rain})$$

Note: $P(\text{flood}|\text{rain})$ is larger than $P(\text{flood})$

Review: Things You Should Know About Probabilities

- $P(A,B) = P(A|B).P(B)$

- Corollary: If events A and B are independent, the odds of them happening together is the product of their individual probabilities.
- $P(A,B) = P(A).P(B)$

Note: This is because $P(A|B) = P(A)$

Review: Probability versus Conditional Probability

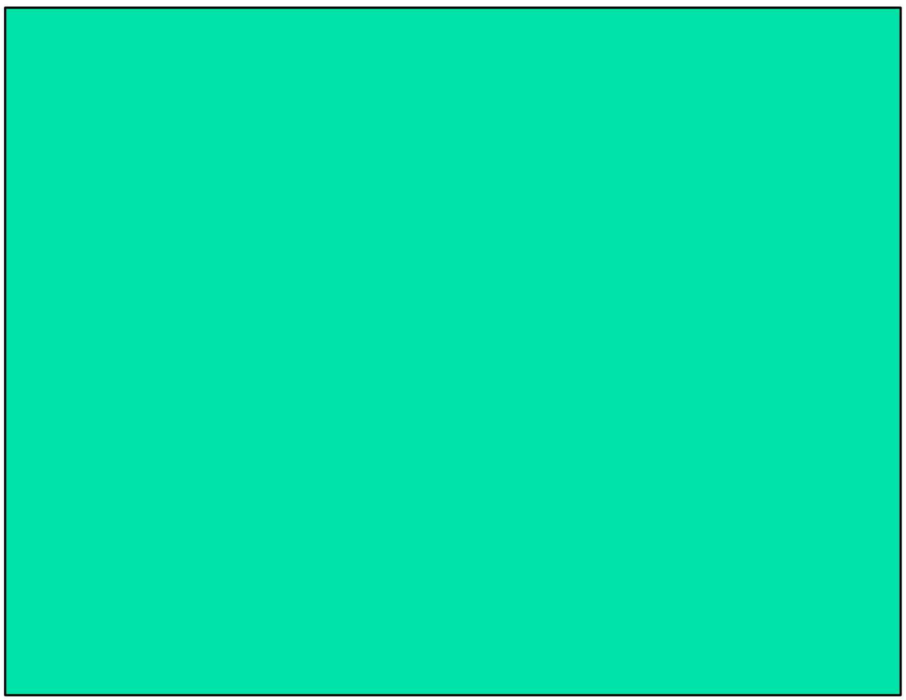
- Probability: the odds that something, say X, happens, P(X).
- Conditional probability: the odds that X happens *given* that a certain condition, C, has occurred, P(X|C).
  - This condition may affect the odds.

Review: Probability versus Conditional Probability

- Probability: the odds that something, say X, happens, P(X).
- Conditional probability: the odds that X happens *given* that a certain condition, C, has occurred, P(X|C).
  - This condition may affect the odds.
- Traffic example:
  - P(Accident) may be low
  - P(Accident|Black ice) is a lot higher!

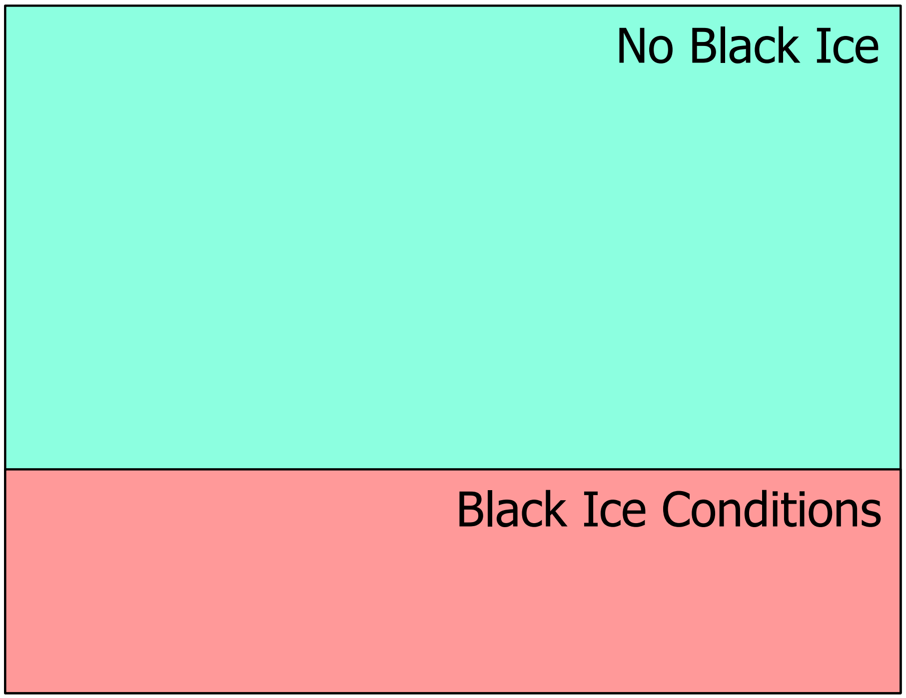
A Visual Interpretation

The universe of all possibilities



A Visual Interpretation

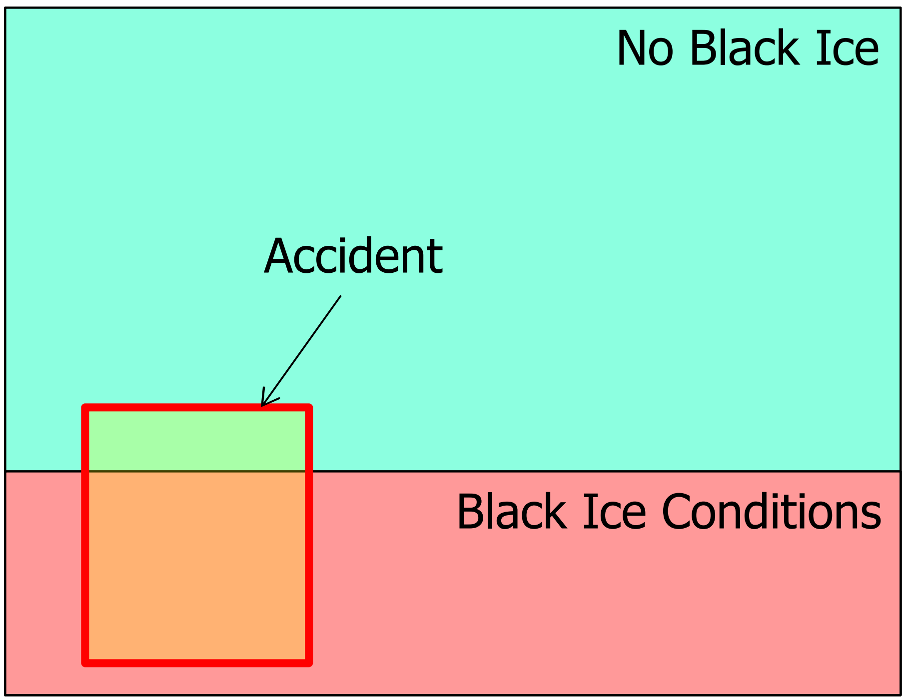
The universe of all possibilities



Consider a graphical visualization of probabilities where area represents probability

A Visual Interpretation

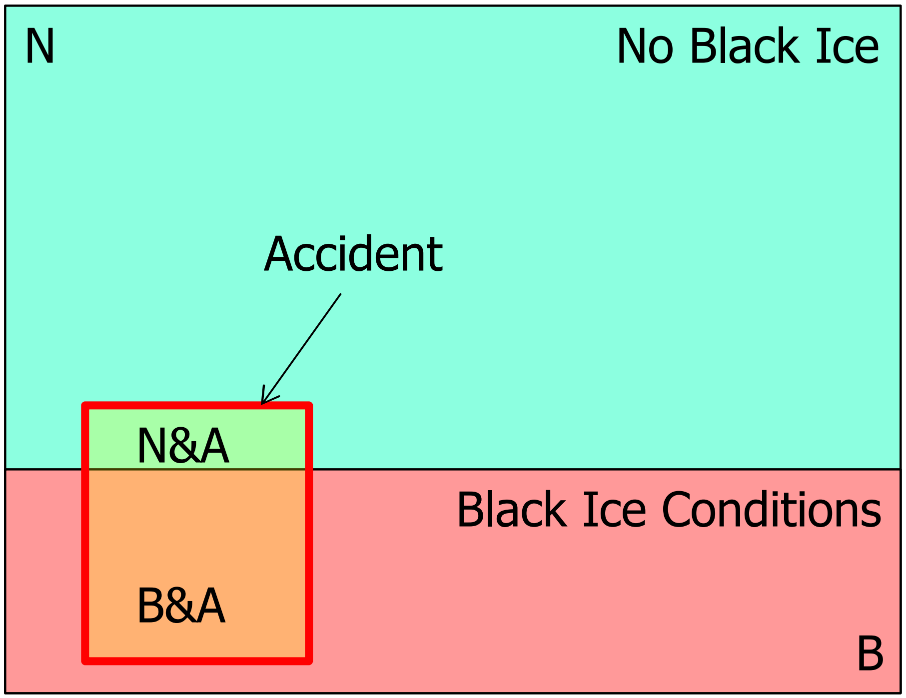
The universe of all possibilities



Consider a graphical visualization of probabilities where area represents probability

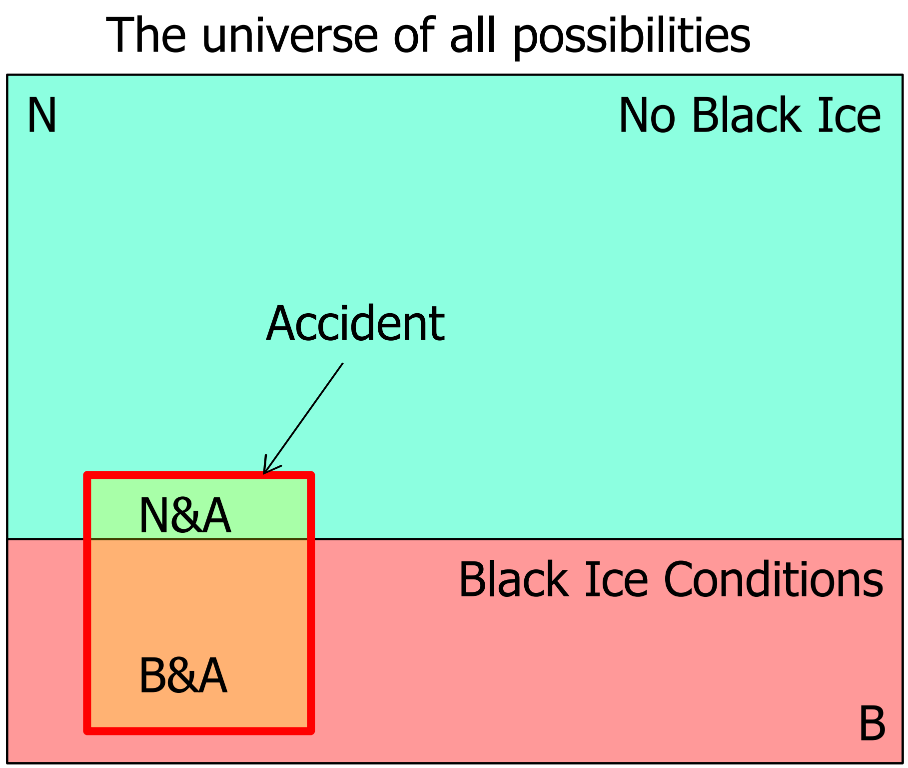
A Visual Interpretation

The universe of all possibilities



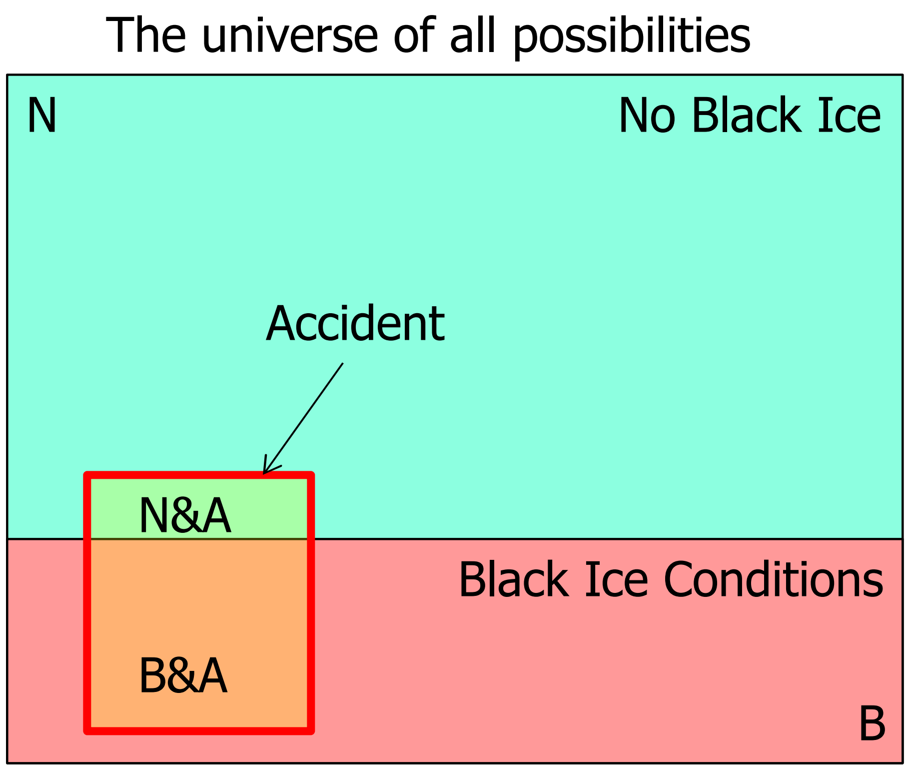
Consider a graphical visualization of probabilities where area represents probability

A Visual Interpretation



- P(Accident) = (N&A+B&A)/(N+B)

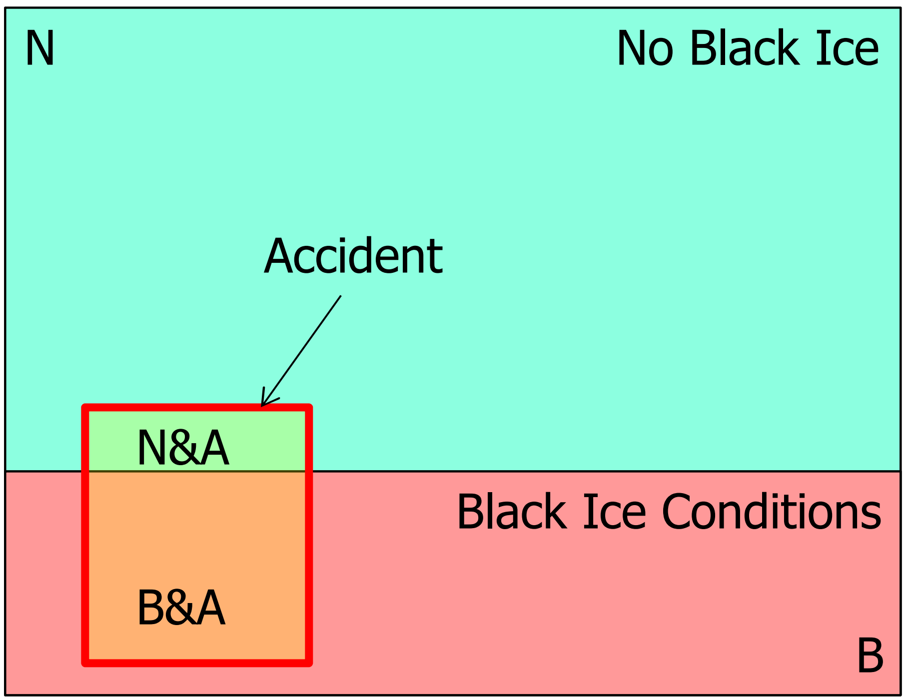
A Visual Interpretation



- P(Accident) = (N&A+B&A)/(N+B)
- P(Accident|Ice) = (B&A)/B
- P(Accident|No Ice) = (N&A)/N

A Visual Interpretation

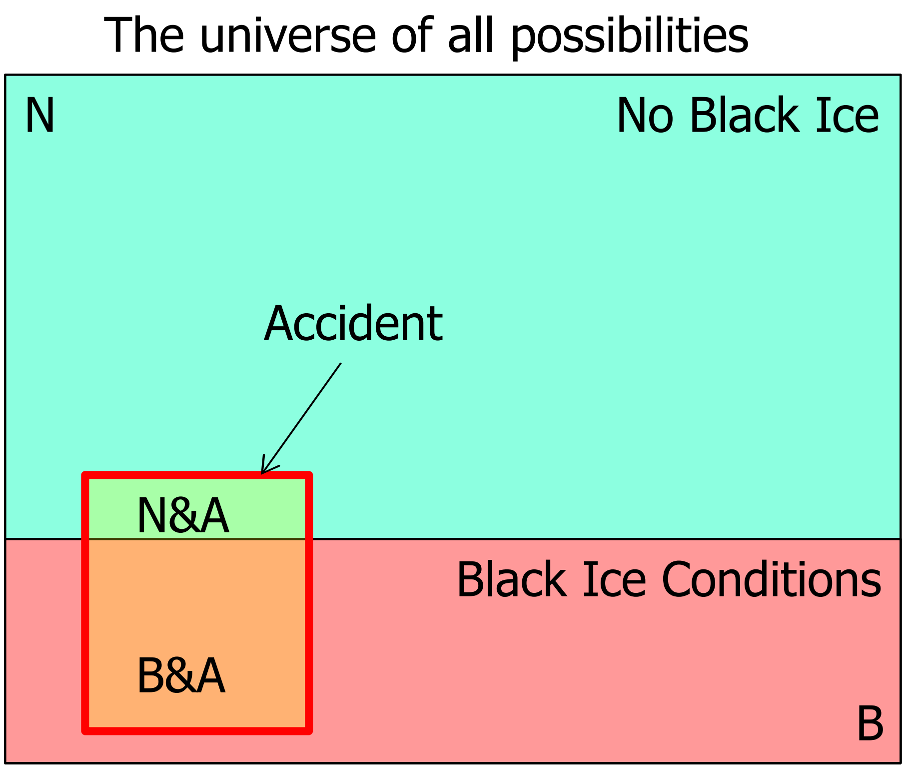
The universe of all possibilities



- P(Accident) = (N&A+B&A)/(N+B)
- P(Accident|Ice) = (B&A)/B
- P(Accident|No Ice) = (N&A)/N

Question: If P(Accident|Ice) = 0.1, P(Accident|No-Ice) = 0.02, P(Ice) = 0.2, what are the odds of accidents in general, P(Accident)?

A Visual Interpretation



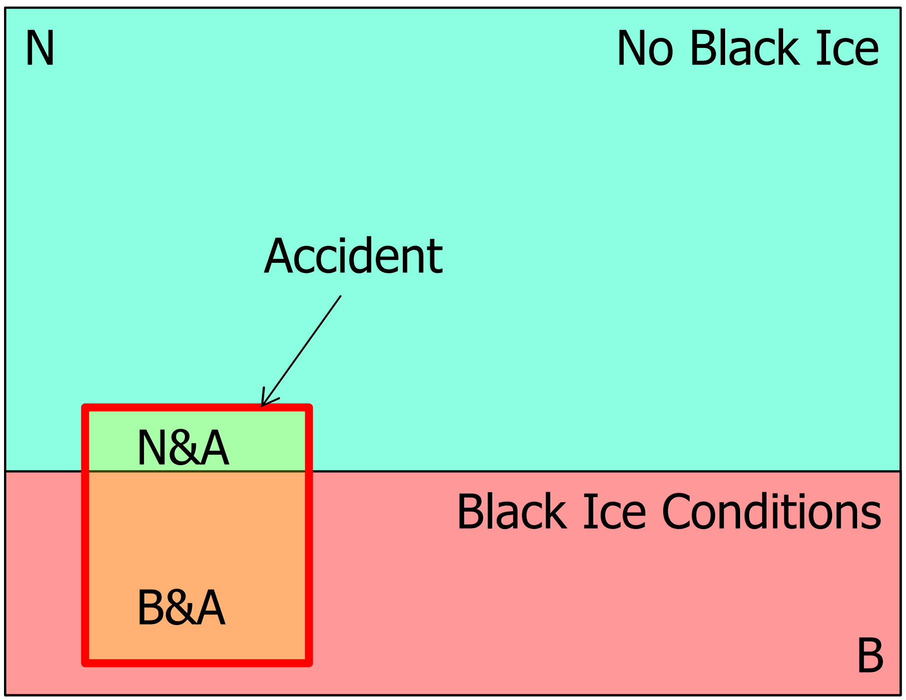
- (B&A)/B = 0.1
- (N&A)/N = 0.02
- B/(B+N) = 0.2
- (N&A+B&A)/(N+B)?

Question: If P(Accident|Ice) = 0.1, P(Accident|No-Ice) = 0.02, P(Ice) = 0.2, what are the odds of accidents in general, P(Accident)?

Review: Total Probability Theorem

$$P(\text{Something}) = P(\text{Something}|X)\ P(X) + P(\text{Something}|\bar{X})\ P(\bar{X})$$

The universe of all possibilities



- P(Accident) =
  0.1 * 0.2
  + 0.02 * 0.8
  = 0.032

Question: If P(Accident|Ice) = 0.1, P(Accident|No-Ice) = 0.02, P(Ice) = 0.2, what are the odds of accidents in general, P(Accident)?

Example: Probability versus Conditional Probability

- A man is accused of murdering his battered wife. The lawyer says that only 2% of men who batter their wives actually end up killing them, so the odds that this is a murder are very low.

- Is this argument statistically valid? If so, explain why (mathematically). If not, why not?

Example: Probability versus Conditional Probability

- A man is accused of murdering his battered wife. The lawyer says that only 2% of men who batter their wives actually end up killing them, so the odds that this is a murder are very low.

- The relevant statistic is: *given that* a battered wife is murdered, what are the odds that the husband did it? (This happens to be 50%)

Example: Probability versus Conditional Probability

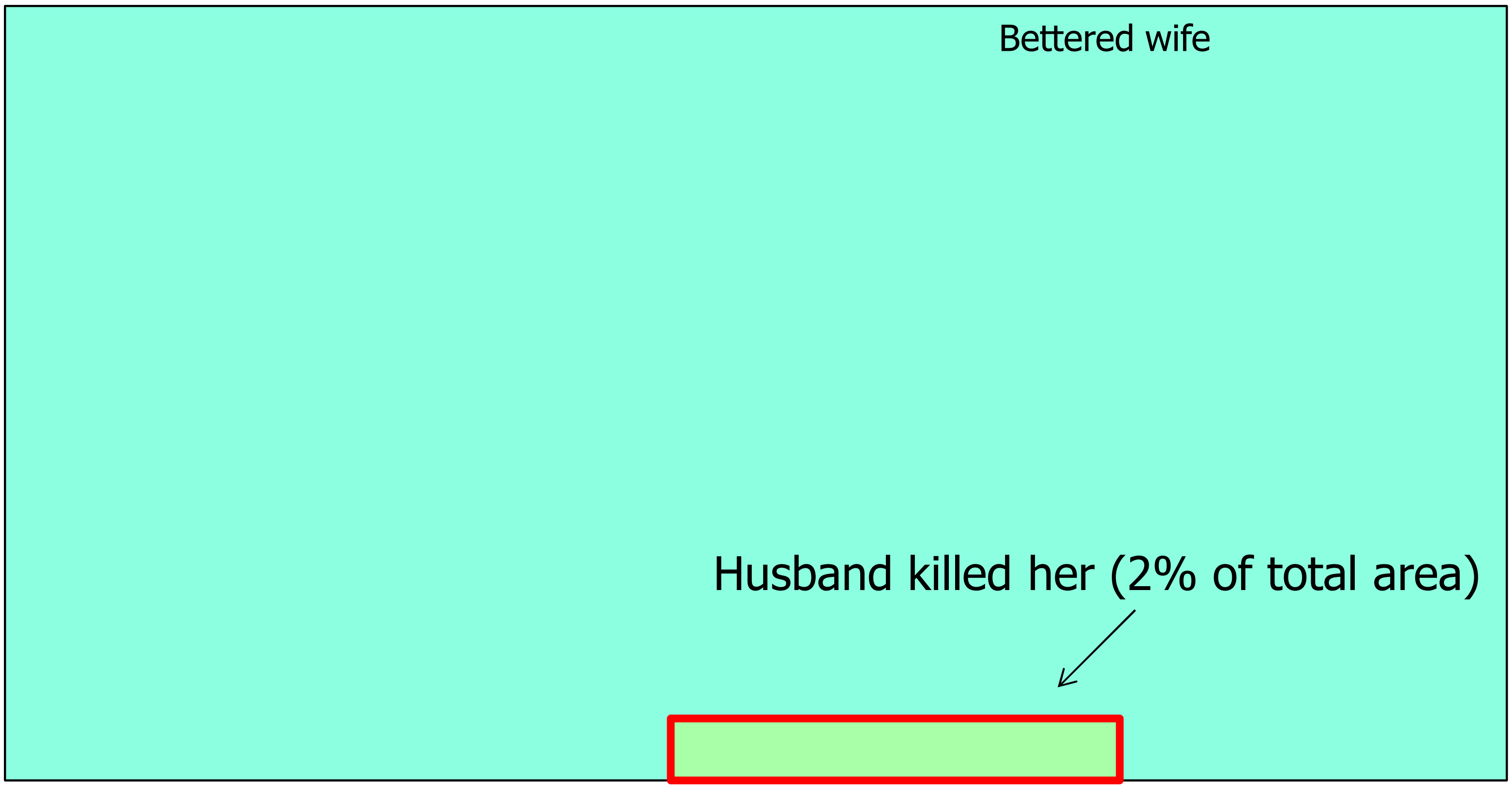
- A man is accused of murdering his battered wife. The lawyer says that only 2% of men who batter their wives actually end up killing them, so the odds that this is a murder are very low.

Conditional probability

- The relevant statistic is: *given that* a battered wife is murdered, what are the odds that the husband did it? (This happens to be 50%)

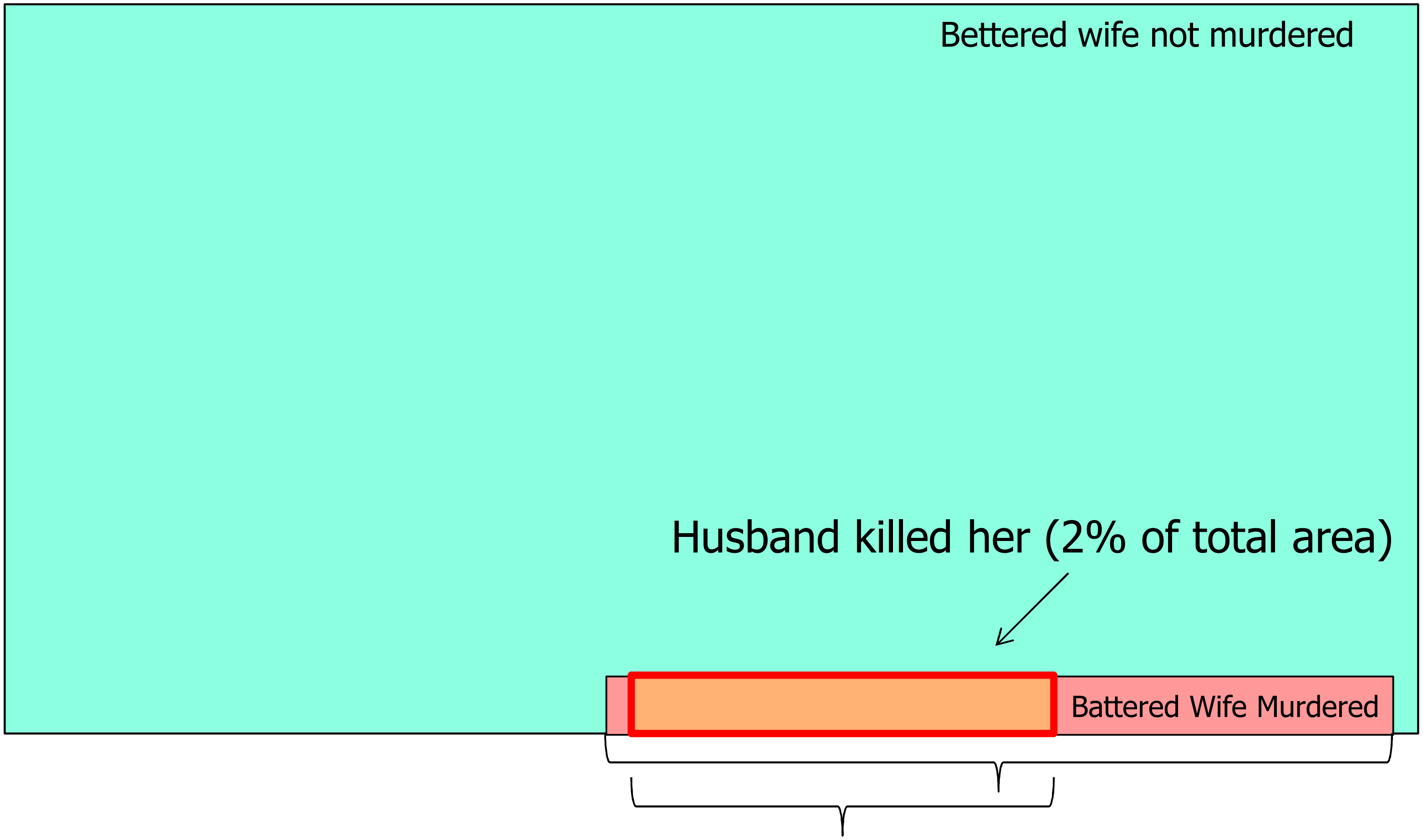
A Visual Interpretation

The universe of all possibilities



A Visual Interpretation

The universe of all possibilities



Husband did it given battered wife murdered is 50%

Example: Intrusion Detection

- A motion alarm is used to detect unauthorized access to a warehouse after hours. The motion sensor is mounted near the only entrance to the warehouse. If a burglar enters the building, there is a 99% chance that the burglar triggers the motion alarm.
- At 9pm, on September 16th, 2013, the alarm was set off. What are the odds that a burglar is in the building?

Example: Asteroid Collision with Earth

- If a major Asteroid collides with Earth in St. Louis, traffic on I-74 will be backed up.

- On Feb 10th, 2014, there was a big back-up on I-74. What are the odds that a major Asteroid collided with Earth?

Example: Asteroid Collision with Earth

- If a major Asteroid collides with Earth in St. Louis (A), traffic on I-74 will be backed up (B).
  - P (B|A) = P (Backup given Asteroid) = 1

- On Feb 10th, 2014, there was a big back-up on I-74 (B). What are the odds that a major Asteroid collided with Earth in St. Louis (A)?
  - P(A|B) = P (Asteroid given Backup) = ?

Factors to Consider

P (Asteroid given Backup) ?

Factors to Consider
P (Asteroid given Backup) ?

- How often do major asteroids hit earth?
  - P(A) = P (Asteroid) = ?
  - The less often it happens, the less likely it is that a traffic jam is attributed to an asteroid.
- How often traffic backs up on I-74
  - P(B) = P (Backup) = ?
  - The more often this happens the less likely it is to be an indicator of asteroid collision

Factors to Consider
P (Asteroid given Backup) ?

- How often do major asteroids hit earth?
  - P(A) = P (Asteroid) = ?
  - The less often it happens, the less likely it is that a traffic jam is attributed to an asteroid.
- How often traffic backs up on I-74
  - P(B) = P (Backup) = ?
  - The more often this happens the less likely it is to be an indicator of asteroid collision
- $P(A|B) = P(B|A).\ P(A)/P(B)$ ← Bayes Theorem

Factors to Consider
P (Asteroid given Backup) ?

- $P(A) = P(\text{Asteroid}) = 0.00001$
- $P(B) = P(\text{Backup}) = 0.01$
- $P(B|A) = 1$

- $P(A|B) = P(B|A).\ P(A)/P(B) = 0.001$

Review of Important Theorems

- Total Probability Theorem:
  $P(A) = P(A|C_1)\ P(C_1) + \ldots + P(A|C_n)\ P(C_n)$
  where $C_1, \ldots, C_n$ cover the space of all possibilities

- Bayes Theorem:
  $P(A|B) = P(B|A).\ P(A)/P(B)$

- Other: $P(A,B) = P(A|B)\ P(B)$

Intrusion Detection, Again

- A motion alarm is used to detect unauthorized access to a warehouse after hours. The motion sensor is mounted near the only entrance to the warehouse. If a burglar enters the building, there is a 99% chance that the burglar triggers the motion alarm.
- At 9pm, on September 16th, 2013, the alarm was set off. What are the odds that a burglar is in the building?

Intrusion Detection, Again

- A motion alarm is used to detect unauthorized access to a warehouse after hours. The motion sensor is mounted near the only entrance to the warehouse. If a burglar enters the building, there is a 99% chance that the burglar triggers the motion alarm.
- At 9pm, on September 16^{th}, 2013, the alarm was set off. What are the odds that a burglar is in the building?
- Assume the alarm goes off about 3 days a year and burglaries happen about once a year

Intrusion Detection, Again

- A motion alarm is used to detect unauthorized access to a warehouse after hours. The motion sensor is mounted near the only entrance to the warehouse. If a burglar enters the building, there is a 99% chance that the burglar triggers the motion alarm.
- At 9pm, on September 16th, 2013, the alarm was set off. What are the odds that a burglar is in the building?
- Assume the alarm goes off about 3 days a year and burglaries happen about once a year

  - $P(A) = P(Alarm) = 3/365$
  - $P(B) = P(Burglar) = 1/365$
  - $P(A|B) = 0.99$

Intrusion Detection, Again

- A motion alarm is used to detect unauthorized access to a warehouse after hours. The motion sensor is mounted near the only entrance to the warehouse. If a burglar enters the building, there is a 99% chance that the burglar triggers the motion alarm.
- At 9pm, on September 16th, 2013, the alarm was set off. What are the odds that a burglar is in the building?
- Assume the alarm goes off about 3 days a year and burglaries happen about once a year

  - $P(A) = P(Alarm) = 3/365$
  - $P(B) = P(Burglar) = 1/365$
  - $P(A|B) = 0.99$
  - $P(B|A) = P(A|B).P(B)/P(A) = 0.33$ (i.e., if alarm sounds, there is a 33% chance it is a burglar)

A Second Sensor

- In the intrusion detection example, assume that there is a vibration sensor on the floor that detects footsteps. If a burglar enters the building, there is a 90% chance that the vibration sensor will fire. If the vibration sensor fires, what are the odds that there is a burglar? Assume that the vibration sensor fires 10 times a year

A Second Sensor

- In the intrusion detection example, assume that there is a vibration sensor on the floor that detects footsteps. If a burglar enters the building, there is a 90% chance that the vibration sensor will fire. If the vibration sensor fires, what are the odds that there is a burglar? Assume that the vibration sensor fires 10 times a year
  - $P(Burg|vib) = P(Vib|Burg).P(Burg)/P(Vib)$

    $= 0.9 * (1/365) / (10/365) = 0.09$

Two Sensor Example

- In the intrusion detection example, what are the odds of burglary if both sensors fire?
- P (Burg|A, Vib) = ?

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Two Sensor Example

- In the intrusion detection example, what are the odds of burglary if both sensors fire?
- $P\ (Burg|A, Vib) = ?$
- $P\ (B|A,V) = P(A,V|B)\ P(B)/P(A,V)$

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Two Sensor Example

- In the intrusion detection example, what are the odds of burglary if both sensors fire?
- $P(Burg|A, Vib) = ?$
- $P(B|A,V) = P(A,V|B)\ P(B)/P(A,V)$

Now what?

Is it OK to say $P(A,V|B) = P(A|B)P(V|B)$?

Is it OK to say $P(A,V) = P(A)P(V)$?

Independence versus Conditional Independence

- John and Sally follow Mike on Twitter.
- When Mike tweets something, John re-tweets it with a 50% probability. Sally re-tweets it with a 30% probability.
- Are John's and Sally's tweets independent?

Independence versus Conditional Independence

- John and Sally follow Mike on Twitter.
- When Mike tweets something, John re-tweets it with a 50% probability. Sally re-tweets it with a 30% probability.
- Are John's and Sally's tweets independent?
  - No. However, given that Mike says something, their decisions to re-tweet it are independent (conditional independence)

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Two Sensor Example

- In the intrusion detection example, what are the odds of burglary if both sensors fire?
- $P(Burg|A, Vib) = ?$
- $P(B|A,V) = P(A,V|B)\,P(B)/P(A,V)$

Now what?

OK to say $P(A,V|B) = P(A|B)P(V|B)$

$\cancel{P(A,V) = P(A)P(V)}$?

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Two Sensor Example

- In the intrusion detection example, what are the odds of burglary if both sensors fire?
- $P(\text{Burg}|A, \text{Vib}) = ?$
- $P(B|A,V) = P(A,V|B)\,P(B)/P(A,V)$ where
  $P(A,V) = P(A,V|B)\,P(B) + P(A,V|\bar{B})\,P(\bar{B})$
  and $P(A,V|B) = P(A|B)P(V|B)$

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Two Sensor Example

- In the intrusion detection example, what are the odds of burglary if both sensors fire?
- $P(\text{Burg}|A, \text{Vib}) = ?$
- $P(B|A,V) = P(A,V|B)\,P(B)/P(A,V)$ where

$P(A,V) = P(A,V|B)\,P(B) + P(A,V|\bar{B})\,P(\bar{B})$

and $P(A,V|B) = P(A|B)P(V|B)$

$P(A,V|\bar{B}) = P(A|\bar{B})P(V|\bar{B})$

Remember: If burglar enters, motion alarm fires 99% of the time and vibration alarm fires 90% of the time. Burglaries occur once a year, motion alarm fires 3 times a year, and vibration alarm fires 10 times a year.

Two Sensor Example

- P (Burg|A, Vib) Solution steps:
  - Find the probability of false alarms from:

    $P(A) = P(A|B)\ P(B) + P(A|\overline{B})\ P(\overline{B})$

    $P(V) = P(V|B)\ P(B) + P(V|\overline{B})\ P(\overline{B})$
  - Find the probability of both sensors firing:

    $P(A,V) = P(A,V|B)\ P(B) + P(A,V|\overline{B})\ P(\overline{B})$

    where $P(A,V|B) = P(A|B)P(V|B)$

    $P(A,V|\overline{B}) = P(A|\overline{B})P(V|\overline{B})$
  - $P\ (B|A,V) = P(A,V|B)\ P(B)/P(A,V) = 94.62\%$