01-0: Notation

- A Random Variable (or just variable) is a variable whose value can be described using probabilities
  - Use Upper Case for variables – \( X, Y, Z, \) etc.
- Random Variables can have discrete or continuous values (for now, we will assume discrete values)
  - use lower case for values of variables – \( x, y, x_1, x_2, \) etc.
- \( P(X = x) \) is the probability that variable \( X \) has the value \( x \)
  - Can also be written as \( P(x) \)

01-1: Notation

- If variable \( X \) can have the values \( x_1, x_2, \ldots, x_n \), then the expression \( P(X) \) stands for a vector which contains \( P(X = x_k) \), for all values \( x_k \) of \( X \)
  \[
P(X) = [P(X = x_1), P(X = x_2), \ldots, P(X = x_n)]
  \]
- If \( D \) is a variable that represents the value of a fair die, then
  \[
P(D) = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6]
  \]

01-2: Notation

- Variable \( W \), represents Weather, which can have values sunny, cloudy, rain, or snow.
  - \( P(W = \text{sunny}) = 0.7 \)
  - \( P(W = \text{cloudy}) = 0.2 \)
  - \( P(W = \text{rain}) = 0.08 \)
  - \( P(W = \text{snow}) = 0.02 \)
  - \( P(W) = [0.7, 0.2, 0.08, 0.02] \)

01-3: Notation – AND

\[
P(x, y) = P(X = x \land Y = y)
\]

- Given two fair dice \( D_1 \) and \( D_2 \):
  \[
P(D_1 = 3, D_2 = 4) = 1/36
  \]
- \( P(X, Y) \) represents the set of \( P(x, y) \) for all values \( x \) of \( X \) and \( y \) of \( Y \). Thus, \( P(D_1, D_2) \) represents 36 different values.

01-4: Notation – Binary Variables

- If \( X \) has two values (false and true), we can represent:
  - \( P(X = \text{false}) \) as \( P(\neg x) \), and
  - \( P(X = \text{true}) \) as \( P(x) \)
01-5: **Conditional Probability**

- \( P(x|y) = \) Probability that \( X = x \) given that all we know is \( Y = y \)
- \( P(\text{cavity}|\text{toothache}) = 0.8 \)
- \( P(\text{Cavity}|\text{Toothache}) \) represents 4 values:

\[
\begin{bmatrix}
P(\neg\text{cavity}|\neg\text{toothache}) & P(\text{cavity}|\neg\text{toothache}) \\
P(\neg\text{cavity}|\text{toothache}) & P(\text{cavity}|\text{toothache})
\end{bmatrix}
\]

01-6: **Conditional Probability**

- We can define conditional probabilities in terms of unconditional probabilities.

\[
P(a|b) = \frac{P(a,b)}{P(b)}
\]

Whenever \( P(b) > 0 \)

- \( P(a,b) = P(a|b)P(b) = P(b|a)P(a) \)
- \( P(A,B) = P(A|B)P(B) \) means \( P(a,b) = P(a|b)P(b) \) for all values \( a,b \)

01-7: **Axioms of Probability**

- \( 0 \leq P(a) \leq 1 \)
- \( P(\text{true}) = 1, P(\text{false}) = 0 \)
- \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)

Everything follows from these three axioms

For instance, prove \( P(x) = 1 - P(\neg x) \)

01-8: **Axioms of Probability**

- \( 0 \leq P(a) \leq 1 \)
- \( P(\text{true}) = 1, P(\text{false}) = 0 \)
- \( P(a \lor b) = P(a) + P(b) - P(a \land b) \)

\[
\begin{align*}
P(x \lor \neg x) &= P(x) + P(\neg x) - P(x \land \neg x) \\
1 &= P(x) + P(\neg x) - 0 \\
1 - P(\neg x) &= P(x) \\
P(x) &= 1 - P(\neg x)
\end{align*}
\]

01-9: **Joint Probability**
• Probability for all possible values of all possible variables
  
  | cavity | ~toothache | 0.04 |
  | cavity | toothache  | 0.06 |
  | ~cavity| toothache | 0.01 |
  | ~cavity| ~toothache| 0.89 |

• From the joint, we can calculate anything
  
  - \( P(\text{cavity}) = 0.04 + 0.06 = 0.10 \)
  - \( P(\text{cavity} \lor \text{toothache}) = 0.04 + 0.06 + 0.01 \)
  - \( = 0.11 \)
  - \( P(\text{cavity}|\text{toothache}) = \frac{P(c, t)}{P(t)} \)
  - \( = \frac{0.04}{(0.04 + 0.01)} = 0.80 \)

01-10: **Joint Probability**

- Joint can tell us everything
- Calculate the joint, read off what you want to know
- This will *not* work!
  - \( x \) different variables, each of which has \( v \) values
  - Size of joint = \( v^x \)
  - 50 variables, each has 7 values, \( 1.8 \times 10^{42} \) table entries

01-11: **Conditional Probability**

- Working with the joint is impractical
- Work with conditional probabilities instead
- Manipulate conditional probabilities based on definition:

\[
P(A|B) = \frac{P(A, B)}{P(B)}
\]

(when \( P(B) \) is always > 0)

01-12: **Bayes Rule**

\[
P(B|A) = \frac{P(A \land B)}{P(A)}
\]

\[
= \frac{P(A|B)P(B)}{P(A)}
\]

Generalize Bayes Rule, with additional evidence \( E \):

\[
P(B|A \land E) = \frac{P(A \land B|E)}{P(A|E)}
\]

\[
= \frac{P(A|B \land E)P(B|E)}{P(A|E)}
\]

01-13: **Using Bayes Rule**
• Rare disease, strikes one in every 10,000
• Test for the disease that is 95% accurate:
  • \( P(t|d) = 0.95 \)
  • \( P(\neg t|\neg d) = 0.95 \)
• Someone tests positive for the disease, what is the probability that they have it?
  • \( P(d|t) = ? \)

01-14: Using Bayes Rule

• \( P(d) = 0.0001 \)
• \( P(t|d) = 0.95 \) (and hence \( P(\neg t|d) = 0.05 \))
• \( P(\neg t|\neg d) = 0.95 \) (and hence \( P(t|\neg d) = 0.05 \))

\[
P(d|t) = \frac{P(t|d)P(d)}{P(t)}
\]
\[= \frac{0.95 \times 0.0001}{(P(t|d)P(d) + P(t|\neg d)P(\neg d))}
\]
\[= \frac{0.95 \times 0.0001}{(0.95 \times 0.0001 + 0.05 \times 0.9999)}
\]
\[= 0.0019
\]

01-15: Conditional Independence

• Variable \( A \) is conditionally independent of variable \( B \), if \( P(A|B) = P(A) \)
• Notation: \( (A \perp B) \)
  • \( D \) – roll of a fair die \( (d_1 \ldots d_6) \)
  • \( C \) – value of a coin flip \( (\text{h or t}) \)
  • \( P(D|C) = P(D) \)
  • \( P(C|D) = P(C) \)
• \( (A \perp B) \Leftrightarrow (B \perp A) \)
• \( P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B) \)

01-16: Conditional Independence

• If \( A \) and \( B \) are independent, then \( P(a, b) = P(a)P(b) \)
  • (Also used as a definition of conditional independence – two definitions are equivalent)
• \( P(a, b) = P(a|b)P(b) = P(a)P(b) \)

01-17: Conditional Independence
At an elementary school, reading scores and shoe sizes are correlated. Why?

- $P(R|S) \neq P(R)$
- $P(R|S,A) = P(R|A)$
- Notation: $(R \perp S|A)$

**01-18: Monte Hall Problem**
From the game show “Let’s make a Deal”

- Pick one of three doors. Fabulous prize behind one door, goats behind other 2 doors
- Monty opens one of the doors you did not pick, shows a goat
- Monty then offers you the chance to switch doors, to the other unopened door
- Should you switch?

**01-19: Monte Hall Problem**
Problem Clarification:

- Prize location selected randomly
- Monty always opens a door, allows contestants to switch
- When Monty has a choice about which door to open, he chooses randomly.

Variables

- Prize: $P = p_A, p_B, p_C$
- Choose: $C = c_A, c_B, c_C$
- Monty: $M = m_A, m_B, m_C$

**01-20: Monte Hall Problem**
Without loss of generality, assume:

- Choose door A
- Monty opens door B

\[ P(p_A|c_A, m_B) = P(m_B|c_A, p_A) \frac{P(p_A|c_A)}{P(m_B|c_A)} \]

**01-21: Monte Hall Problem**

\[ P(p_A|c_A, m_B) = P(m_B|c_A, p_A) \frac{P(p_A|c_A)}{P(m_B|c_A)} \]

- $P(m_B|c_A, p_A) = 1/2$
- $P(p_A|c_A) = 1/3$
01-22: **Monte Hall Problem**

\[ P(p_c|m_b) = P(m_b|p_A)P(p_A) + P(m_b|p_B)P(p_B) + P(m_b|p_C)P(p_C) = \frac{P(p_c|m_A)}{P(m_b|m_A)} = \frac{2}{3} \]

- \( P(m_b|p_A, p_C) = 1 \)
- \( P(p_C|m_A) = 1/3 \)
- \( P(m_b|p_A) = 1/2 \)
- \( P(m_b|p_C) = 0 \) Won’t open prize door
- \( P(m_b|p_C) = 1 \) Monty has no choice

01-23: **Rare Disease Redux**

- Rare disease, strikes one in every 10,000
- Two tests, one 95% accurate, other 90% accurate:
  - \( P(t1|d) = 0.95, P(\neg t1|\neg d) = 0.95 \)
  - \( P(t2|d) = 0.90, P(\neg t2|\neg d) = 0.90 \)
- Tests use independent mechanisms to detect disease, and are conditionally independent, given disease state:
  - \( P(t1|d, t2) = P(t1|d) \)
- Both test are positive, what is the probability of disease?
  - \( P(d|t1, t2) = ? \)

\[
P(d|t1, t2) = P(t1|d, t2) \frac{P(t2)}{P(t1|t2)}
= P(t1|d) \frac{P(t2)}{P(t1|t2)}
= 0.0168
\]

\[
P(t1|t2) = P(t1|t2, d)P(d|t2) + P(t1|t2, \neg d)P(\neg d|t2)
= P(t1|d)P(d|t2) + P(t1|\neg d)P(\neg d|t2)
= 0.05081
\]

\[
P(d|t2) = P(t2|d) \frac{P(d)}{P(t2)}
= 0.0009
\]
\[ P(t2) = P(t2|d)P(d) + P(t2|\neg d)P(\neg d) \]
\[ = 0.1008 \]

**01-25: Probabilistic Reasoning**

- Given:
  - Set of conditional probabilities \( P(t1|d), \text{etc} \)
  - Set of prior probabilities \( P(d) \)
  - Conditional independence information \( P(t1|d, t2) = P(t1|d) \)
- We can calculate any quantity that we like

**01-26: Probabilistic Reasoning**

- Given:
  - Set of conditional probabilities \( P(t1|d), \text{etc} \)
  - Set of prior probabilities \( P(d) \)
  - Conditional independence information \( P(t1|d, t2) = P(t1|d) \)
- We can calculate any quantity that we like
- Problems:
  - Hard to know exactly what data we need
  - Even given sufficient data, calculations can be complex – especially dealing with conditional independence

**01-27: Bayesian Networks**

Bayesian Networks are:

- Clever encoding of conditional independence information
- Mechanical, “turn the crank” method for calculation
  - Can be done by a computer

Nothing “magic” about Bayesian Networks

**01-28: Directed Acyclic Graphs**

- We will encode conditional independence information using Directed Acyclic Graphs (or DAGs)
- While we will use causal language to give intuitive justification, these DAGs are *not necessarily* causal (more on this later)
- Three basic “junctions”
01-29: **Head-to-Tail**

```
A → B → C
```

- “Causal Chain”
- Rain $\rightarrow$ Wet Pavement $\rightarrow$ Slippery Pavement
  - $(A \not\rightarrow C)$
  - $(A \perp C | B)$

01-30: **Tail-to-Tail**

```
A → B → C
```

- “Common Cause”
- Reading Ability $\leftrightarrow$ Age $\rightarrow$ Shoe Size
  - $(A \not\leftrightarrow C)$
  - $(A \perp C | B)$

01-31: **Head-to-Head**

```
A → B → C
```

- “Common Effect”
- Rain $\rightarrow$ Wet Grass $\leftrightarrow$ Sprinkler
  - $(A \perp C)$
  - $(A \not\leftrightarrow C | B)$
01-32: **Head-to-Head**

![Diagram showing relationships between Rain, Sprinkler, Wet Grass, and Slugs]

- Also need to worry about descendants of head-head junctions.
- \((\text{Rain} \perp \!\!\!\!\!\!\!\!\!\perp \text{Sprinkler})\)
- \((\text{Rain} \not\perp \!\!\!\!\!\!\!\!\!\not\perp \text{Sprinkler} | \text{Slugs})\)

01-33: **Markovian Parents**

- \(V\) is an ordered set of variables \(X_1, X_2, \ldots X_n\).
- \(P(V)\) is a joint probability distribution over \(V\)
- Define the set of Markovian Parents of variable \(X_j\), \(PA_j\) as:
  - Minimal set of predecessors of \(X_j\) such that
  - \(P(X_j | X_1, \ldots X_{j-1}) = P(X_j | PA_j)\)
- The Markovian Parents of a variable \(X_j\) are often (*but not always*) the direct causes of \(X_j\)

01-34: **Markovian Parents & Joint**

- For any set of variables \(X_1, \ldots X_n\), we can calculate any row of the joint:
  - \(P(x_1, \ldots x_n) = P(x_1) P(x_2 | x_1) P(x_3 | x_1, x_2) \ldots P(x_n | x_1, x_2, \ldots x_{n-1})\)
  - Using Markovian parents
    - \(P(x_1, \ldots x_n) = P(x_1) P(x_2 | PA_2) P(x_3 | PA_3) \ldots P(x_n | PA_n)\)

01-35: **Markovian Parents & DAGs**

- We can create a DAG which represents conditional independence information using Markovian parents.
  - Each variable is a node in the graph
  - For each variable \(X_j\), add a directed link from all elements in \(PA_j\) to \(X_j\)
- Example: Burglary, Earthquake, Alarm, John Calls, Mary Calls, News Report
01-36: **DAG Example**

![DAG Diagram]

01-37: **DAGs & Cond. Independence**

- Given a DAG of Markovian Parents, we know that every variable $X_i$ is independent of its ancestors, given its parents.
- We also know quite a bit more.

01-38: **d-separation** To determine if a variable $X$ is conditionally independent of $Y$ given a set of variables $Z$:

- Examine all paths between $X$ and $Y$ in the graph.
- Each node along a path can be “open” or “blocked”:
  - A node at a head-to-tail or tail-to-tail junction is open if the node is not in $Z$, and closed otherwise.
  - A node at a head-to-head junction is open if the node or any of its descendants is not in $Z$, and closed otherwise.

Examples (on board)