# 1 *Bayes' Nets*

Unfortunately during spring due to illness and allergies, Billy is unable to distinguish the cause (X) of his symptoms which could be: coughing (C), sneezing (S), and temperature (T). If he is able to determine the cause with a reasonable accuracy, it would be beneficial for him to take either "Robitussin DM" or Benedryll. He has asked you to help him formulate his problem as a Bayesian Network. At any point in time, Billy can either be sick ($X = sick$), allergic ($X = allergic$), or well ($X = well$).

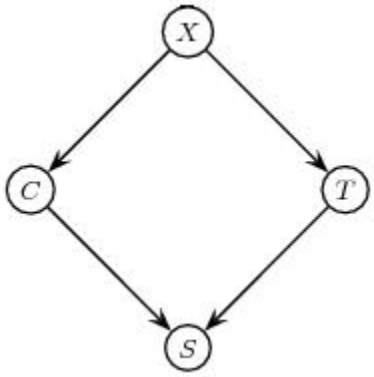


Figure 1: Bayes' Net

**(a) (2 points)** List all independence and conditional independence relationships implied by this Bayes' net.

$$C \perp T | X$$
$$S \perp X | C, T$$

**(b) (2 points)** Write the algebraic expression for $P(C, S)$ in terms of the joint distribution $P(X, C, T, S)$.

$$P(C, S) = \sum_x \sum_t P(X = x, C, T = t, S)$$

**(c) (2 points)** What probability and conditional probability tables must we store for this Bayes' net?

$$P(X), P(C|X), P(T|X), P(S|C, T)$$

**(d) (2 points)** Derive an expression for the posterior distribution over $X$ given whether Billy is coughing ($C$) and sneezing ($S$).

$$P(X|C, S) = \frac{P(X,C,S)}{P(C,S)} = \frac{\sum_t P(X,C,T=t,S)}{\sum_x \sum_t P(X=x,C,T=t,S)}$$

It is also correct to rewrite the above using the fact $P(X, C, T, S) = P(X)P(C|X)P(T|X)P(S|C, T)$.

**(e) (2 points)** Derive an expression for the likelihood of observing that he is coughing ($C$) and sneezing ($S$) given $X$.

$$\begin{aligned} P(C, S|X) &= \frac{P(X,C,S)}{P(X)} \\ &= \frac{\sum_t P(X,C,T=t,S)}{P(X)} \end{aligned}$$

## 2 *HMMs*

You sometimes get colds, which make you sneeze. You also get allergies, which make you sneeze. Sometimes you are well, which doesn't make you sneeze (much). You decide to model the process using the following HMM, with hidden states $X \in \{well, allergy, cold\}$ and observations $E \in \{sneeze, quiet\}$.

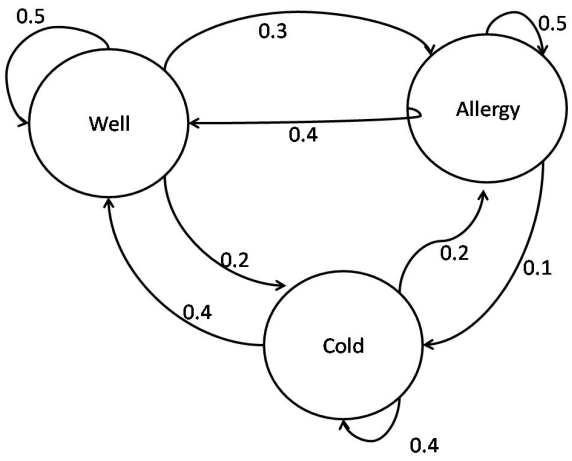


Figure 2: HMM State Transition Diagram

$P(X_1)$

| | |
|---|---|
| *well* | 0.60 |
| *allergy* | 0.25 |
| *cold* | 0.15 |

$P(X_t \mid X_{t-1} = well)$

| | |
|---|---|
| *well* | 0.5 |
| *allergy* | 0.3 |
| *cold* | 0.2 |

$P(X_t \mid X_{t-1} = allergy)$

| | |
|---|---|
| *well* | 0.4 |
| *allergy* | 0.5 |
| *cold* | 0.1 |

$P(X_t \mid X_{t-1} = cold)$

| | |
|---|---|
| *well* | 0.4 |
| *allergy* | 0.2 |
| *cold* | 0.4 |

Transitions

$P(E_t \mid X_t = well)$

| | |
|---|---|
| *quiet* | 0.9 |
| *sneeze* | 0.1 |

$P(E_t \mid X_t = allergy)$

| | |
|---|---|
| *quiet* | 0.1 |
| *sneeze* | 0.9 |

$P(E_t \mid X_t = cold)$

| | |
|---|---|
| *quiet* | 0.3 |
| *sneeze* | 0.7 |

Emissions

Figure 3: HMM Probability Tables

**(a) (1 point)** Fill in the missing entries in the probability tables above.

**(b) (2 points)** Imagine you observe the sequence quiet, sneeze, sneeze. What is the probability that you were well all three days and observed these effects?

$$\begin{aligned} &P(X_1 = w, X_2 = w, X_3 = w | E_1 = q, E_2 = s, E_3 = s) \\ &\qquad = \frac{P(X_1 = w, X_2 = w, X_3 = w, E_1 = q, E_2 = s, E_3 = s)}{P(E_1 = q, E_2 = s, E_3 = s)} \end{aligned}$$

$$P(X_1, X_2, X_3, E_1, E_2, E_3) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(E_1|X_1)P(E_2|X_2)P(E_3|X_3)$$

$$P(X_1 = w, X_2 = w, X_3 = w, E_1 = q, E_2 = s, E_3 = s) = 0.60 * 0.5 * 0.5 * 0.9 * 0.1 * 0.1 = 0.00135$$

$$P(E_1 = q, E_2 = s, E_3 = s) = \sum_{x_1} \sum_{x_2} \sum_{x_3} P(X_1 = x_1, X_2 = x_2, X_3 = x_3, E_1 = q, E_2 = s, E_3 = s)$$

$$P(X_1 = w, X_2 = w, X_3 = w | E_1 = q, E_2 = s, E_3 = s) = 0.00135 / P(E_1 = q, E_2 = s, E_3 = s)$$

**(c) (2 points)** What is the posterior distribution over your state on day 2 ($X_2$) if $E_1$ = quiet, $E_2$ = sneeze? This is the filtering problem.

$$P(X_1|E_1 = q) \propto P(X_1)P(E_1 = q|X_1)$$

Running one step of filtering, we have (in the order of well, allergy, cold):

$$P(X_1|E_1 = q) = [0.885 \quad 0.041 \quad 0.074]$$

$$\begin{aligned} P(X_2|E_1 = q, E_2 = s) \propto & \quad P(X_2, E_1 = q, E_2 = s) \\ = & \quad \textstyle\sum_{x_1} P(X_2, X_1 = x_1, E_1 = q, E_2 = s) \\ = & \quad \textstyle\sum_{x_1} P(E_2 = s|X_2)P(X_2|X_1 = x_1)P(X_1 = x_1)P(E_1 = q|X_1 = x_1) \\ = & \quad P(E_2 = s|X_2) \textstyle\sum_{x_1} P(X_2|X_1 = x_1)P(X_1 = x_1)P(E_1 = q|X_1 = x_1) \\ \propto & \quad P(E_2 = s|X_2) \textstyle\sum_{x_1} P(X_2|X_1 = x_1)P(X_1 = x_1|E_1 = q) \end{aligned}$$

Using the equation above, after two steps of filtering, we have

$$P(X_2|E_1 = q, E_2 = s) = [0.104 \quad 0.580 \quad 0.316]$$

**(d) (2 points)** What is the posterior distribution over your state on day 3 ($X_3$) if $E_1$ = quiet, $E_2$ = sneeze, $E_3$ = sneeze?

We simply need to compute another step of filtering, using our answer from part (c) above.

$$P(X_3|E_1 = q, E_2 = s, E_3 = s) \propto P(E_3 = s|X_3) \sum_{x_2} P(X_3|X_2 = x_2)P(X_2|E_1 = q, E_2 = s)$$

After normalizing, we have

$$P(X_3|E_1 = q, E_2 = s, E_3 = s) = [0.077 \quad 0.652 \quad 0.271]$$

## 3 *MDPs*

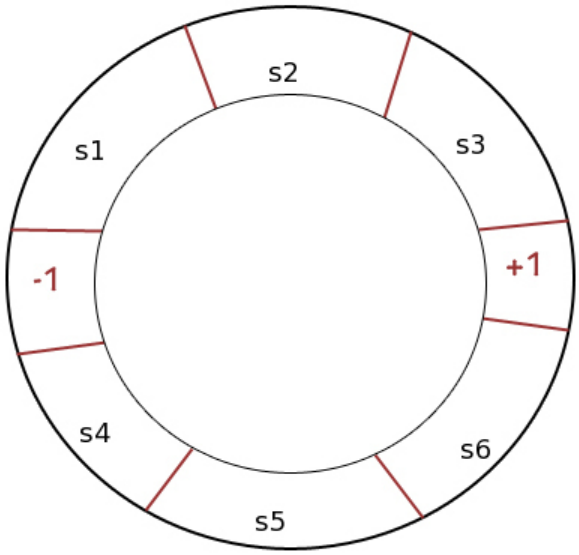


Figure 4: MDP

Consider the above MDP, representing a robot on a circular wheel. The wheel is divided into eight states and the available actions are to move clockwise or counterclockwise. The robot has a poor sense of direction and will move with probability 0.6 in the intended direction and with probability 0.4 in the opposite direction. All states have reward zero, except the terminal states which have rewards $-1$ and $+1$ as shown. The discount factor is $\gamma = 0.9$.

**(5 points)** Compute the numeric values of the state-value function $V(s)$ for states $s1$ through $s6$ (compute $V(s1)$, $V(s2)$, ...). Show your work below.

By symmetry, we know that $V(s1) = V(s4)$, $V(s2) = V(s5)$, and $V(s3) = V(s6)$. To compute $V(s1)$, $V(s2)$, and $V(s3)$, we could use value iteration. However, it is much easier to solve directly using the Bellman equations. It is clear that the optimal policy is to move toward the $+1$ reward.

$$V(s1) = 0.4 * \gamma * (-1) + 0.6 * \gamma * V(s2)$$
$$V(s2) = 0.4 * \gamma * V(s1) + 0.6 * \gamma * V(s3)$$
$$V(s3) = 0.4 * \gamma * V(s2) + 0.6 * \gamma * (+1)$$

Solving this system of equations yields

$$V(s1) = -0.217$$
$$V(s2) = 0.265$$
$$V(s3) = 0.635$$

# 4 *More MDPs and Reinforcement Learning*

**(a) (2 points)** What is an $\epsilon$-greedy policy and why is it important in $Q$-learning?

An $\epsilon$-greedy policy in one that chooses the best action $argmax_a Q(s,a)$ with probability $1-\epsilon$ and chooses a random action with probability $\epsilon$. In $Q$-learning, such a policy provides the required balance between exploration and explotation.

**(b) (2 points)** Why might one want to keep track of an action-value function $Q(s,a)$ rather than a state-value function $V(s)$? Can we compute $V(s)$ from $Q(s,a)$? If so, give an equation. Can we compute $Q(s,a)$ from $V(s)$? If so, give an equation.

We might want to keep track of an action-value function $Q(s,a)$ because we can directly obtain a policy from $Q(s,a)$ (take the action $a$ maximizing $Q(s,a)$ when in state $s$), without needing to know the transition model for our environment.

Yes, we can compute $V(s)$ from $Q(s,a)$. $V(s) = \max_a Q(s,a)$ in the case that our policy $\pi$ is to take the best action. In general, $V(s) = \sum_a \pi(s,a)Q(s,a)$.

If we don't know the transition model, then we can't compute $Q(s,a)$ from $V(s)$. However, if we're given the transition model $T(s,a,s')$, then $Q(s,a) = \sum_{s'} T(s,a,s')V(s')$.

**(c) (2 points)** Consider a two-player adversarial zero-sum game. Players MAX and MIN alternate actions in this MDP. When a transition $(s,a,s')$ occurs, MAX receives reward $R(s,a,s')$ and MIN receives reward $R(s,a,s')$. MAX attempts to maximize his total rewards and MIN attempts to minimize his total rewards. Let $Q_{MAX}(s,a)$ and $Q_{MIN}(s,a)$ denote the q-values for performing action $a$ in state $s$ for players MAX and MIN, respectively. Given discount factor $\gamma$ and transition probabilities $T(s,a,s')$, write two Bellman equations expressing each of $Q_{MAX}$ and $Q_{MIN}$ in terms of adjacent lookahead values of $Q_{MAX}$ and/or $Q_{MIN}$.

Since MAX and MIN alternate turns, MAX must take the minimum over the adjacent lookahead value, as MIN plays next. Similarly, MIN must take the maximum over the adjacent lookahead value.

$$Q_{MAX}(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma \min_{a'} Q_{MIN}(s',a')]$$
$$Q_{MIN}(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma \max_{a'} Q_{MAX}(s',a')]$$