Probability and Bayesian Networks Exercises

Prof. Jim Rehg
CS 3600 Introduction to Artificial Intelligence
School of Interactive Computing
Georgia Institute of Technology

March 25, 2018

These practice exercises are for your benefit in preparing for the exams. They will not be collected or graded. Solutions will be provided. Note that not all questions are representative of questions you will find on the exam, but the material covered by these questions will also be covered by the exams.

Question 1. **Events and Combinatorics**

For each problem below, *state* the sample space Ω, the event space A, and the probability of the event $P(A)$. Assume the standard probability models hold for dice, cards, etc.

(a) (1 pts) You roll two 6-sided dice, what is the probability of getting a 7 or an 11?

SOLUTION:

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad \text{(36 possible outcomes)}$$
$$A_7 = \{(3,4),(4,3),(2,5),(5,2),(1,6),(6,1)\}$$
$$A_{11} = \{(5,6),(6,5)\}$$
$$A = A_7 \cup A_{11}$$
$$P(A) = P(A_7) + P(A_{11}) = (6+2)/36 = 0.22$$

Note that the probabilites add without any correction because the event spaces (sets) A_7 and A_{11} are *disjoint*. The formula for the probability of the union of two events is sometimes referred to as the *addition rule*.

(b) (1 pts) You roll two 6-sided dice, what is the probability that the total is greater than 9 or even? What is the probability that it is greater than 9 *and* even?

SOLUTION:

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad \text{(36 possible outcomes)}$$
$$A_{>9} = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\}$$
$$A_{even} = \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),\ldots,(6,6)\} \qquad \text{(18 possible outcomes)}$$
$$A_{>9} \cap A_{even} = \{(4,6),(6,4),(5,5),(6,6)\} \qquad \text{(Greater than 9 and even)}$$
$$P(A_{>9} \cap A_{even}) = 4/36 = 0.11$$
$$P(A_{>9} \cup A_{even}) = P(A_{>9}) + P(A_{even}) - P(A_{>9} \cap A_{even}) = (6+18-4)/36 = 0.56$$

Note that since the event spaces (sets) $A_{>9}$ and A_{even} are *not* disjoint, we calculate the probability of the union of the events (i.e. the *or* case) by using the addition rule with the correction for the intersection, to avoid double-counting.

(c) (1 pts) You roll *three* 6-sided dice, what is the probability that no two of the numbers are the same (i.e. no combinations such as $(1, 3, 1)$ or $(2, 2, 2)$ occur).

SOLUTION:

$$\Omega = \{(1,1,1),(1,1,2),\ldots,(6,6,6)\} \qquad (6^3 = 216 \text{ possible outcomes})$$
$$A = \{(1,2,3),(1,3,2),\ldots,(6,5,4)\}$$
$$|A| = 6 \times 5 \times 4 = 120$$
$$P(A) = 120/216 = 0.56$$

When events are independent, the cardinality of the event space follows the *product rule*: If there are m_i outcomes for the ith event, then the total number of outcomes is $\prod_i m_i$. We have been using this rule to compute the cardinality of the sample space, e.g. if you roll n 6-sided dice the total number of outcomes is 6^n.

In the event space for this problem, numbers are not allowed to repeat. Thus we can imagine that the first die has 6 possible values, but the second die has only 5 because it can't duplicate the first die, and the third die has only 4 possible values because it can't duplicate the first two. This gives $6 \cdot 5 \cdot 4 = 120$ for the total values. There is a more general way to think about this: For a given valid roll like $(2, 5, 6)$, all permutations of the pattern will also be valid rolls (e.g. $(5, 6, 2)$ and $(6, 2, 5)$). Thus the number of valid outcomes is given by the number of permutations of n things taken k at a time (without replacement):

$$P_k^n = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!},$$

where $P_3^6 = 120$.

(d) (1 pts) What is the probability of being dealt a flush (all cards of the same suit) in 5 card poker?

SOLUTION:

$$\Omega_c = \{2,\ldots,10,J,Q,K,A\} \times \{\clubsuit,\diamondsuit,\heartsuit,\spadesuit\} \qquad \text{(Cartesian product of numbers and suits)}$$
$$\Omega = \Omega_c \times \Omega_c \times \Omega_c \times \Omega_c \times \Omega_c$$
$$A = \text{5 cards of the same suit, either } \clubsuit, \diamondsuit, \heartsuit, \text{ or } \spadesuit$$
$$P(A) = \frac{4C_5^{13}}{C_5^{52}} = 4 \times 1287/2598960 = 0.0019808,$$

where the 4 comes from the fact there are 4 suits, $|\Omega_c| = (13)(4) = 52$, and C_k^n gives the number of *combinations* of n things taken k at a time (without replacement), defined by:

$$C_k^n = \frac{P_k^n}{k!} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k},$$

where the last term is the symbol for binomial coefficients. A more detailed explanation for the formula for $P(A)$: C_5^{13} gives the number of ways you can draw 5 cards of the same suite from all of the cards of that suite. This is multiplied by 4 to give the total number of flush hands. This is then divided by the total number of 5 card hands, given by C_5^{52}.

The difference between combinations and permutations is that for combinations, the order in which the k items are drawn does not matter, whereas for permuations it does. That is why we obtain the number of combinations by dividing the number of permutations by $k!$, which is the number of permutations of k items. How do you know whether to use permutations or combinations? It hinges on whether the order matters. The number of possible 4-digit smartphone pins, for example, is governed by permutations, because $(3, 8, 5, 1)$ is a different pin from $(5, 8, 1, 3)$.

However, for a flush it is only necessary to have all 5 cards with the same suit, and the order in which they are dealt doesn't matter. It turns out that parts (a), (b), and (c) above could also have been solved using combinations, as those questions involved properties of the dice rolls (such as the sum of the faces or the number of pairs) for which the order of the numbers does not matter. For example, in part (c) the number of combinations of three die rolls is $6^3/3! = 216/6 = 36$, reflecting the fact that the dice rolls $\{(1, 3, 5), (1, 5, 3), (3, 1, 5), (3, 5, 1), (5, 3, 1), \text{and } (5, 1, 3)\}$, for example, all represent the same combination. The number of combinations that satisfy the no-pairs constraint is given by the number of ways we can choose 3 unique numbers out of 6, or C_3^6. Then $P(A) = C_3^6/36 = 20/36 = 0.56$.

(e) (1 pts) In the game of blackjack (sometimes called 21), the goal is to beat the dealer by having the highest count of cards without going bust (exceeding 21). The dealer has one card face-down (the hole card) and all other cards are dealt face up. The dealer must hit (take a card) if his total is 16 or less, and stand (end his turn) otherwise. The face cards (jack, queen, and king) are worth 10, assume that the ace is worth 11. Suppose your total is 13, what is the probability of a bust if you hit? Note: Assume that the probabilities for all cards are equal, in other words that the deck is infinite (or equivalently that cards are dealt with replacement).

SOLUTION:

$$\Omega = \{2, \dots, 10, J, Q, K, A\} \qquad \text{(13 possibilities, ignore suits)}$$
$$A = \{9, 10, J, Q, K, A\} \qquad \text{(Cards that cause total to exceed 21)}$$
$$P(A) = 6/13 = 0.46$$

(f) (1 pts) Same setting as in (e), and in addition the dealer is showing a 6. Suppose the dealer ignores the rules and is going to stand no matter what her hole card is. If you hit once and then stand, what is the probability that you will win? Note: In the event of a tie, the dealer wins.

SOLUTION: Let T_d and T_p represent the state for the dealer and player, respectively. The state is their current total, or "bust" if the total exceeds 21. Note that there is no chance that the dealer will bust. We then have:

$$\Omega = \{4, \dots, 21, \text{bust}\} \times \{4, \dots, 21, \text{bust}\} \qquad \text{(Joint state space for player and dealer)}$$
$$A = \{(T_p, T_d) : T_p > T_d \text{ and } T_P \neq \text{bust}\}$$
$$P(T_d) = \begin{cases} 1/13 : 8 \leq T_d \leq 15 & \text{(Hole card is 2-9)} \\ 4/13 : T_d = 16 & \text{(Hole card is worth 10)} \\ 1/13 : T_d = 17 & \text{(Hole card is an ace)} \\ 0 : \text{All other cases} & \end{cases}$$
$$P(T_p) = \begin{cases} 1/13 : 15 \leq T_p \leq 21 & \text{(Next card is 2-8)} \\ 6/13 : T_p = \text{bust} & \text{(From part (e))} \\ 0 : \text{All other cases} & \end{cases}$$

If the player doesn't bust, they will still lose in the following cases:

$$\{(T_p, T_d) : (15, 15), (15, 16), (15, 17), (16, 16), (16, 17), (17, 17)\}.$$

Since all of the events defined by subsets of T_p and T_d are disjoint, by the addition and product rules we can write the probability of losing as:

$$P(A^c) = \frac{6}{13}\frac{13}{13} + \frac{1}{13}\frac{6}{13} + \frac{1}{13}\frac{5}{13} + \frac{1}{13}\frac{1}{13} = 0.53,$$

where $A^c = \Omega \setminus A$ is the complement of A. Then $P(A) = 0.47$.

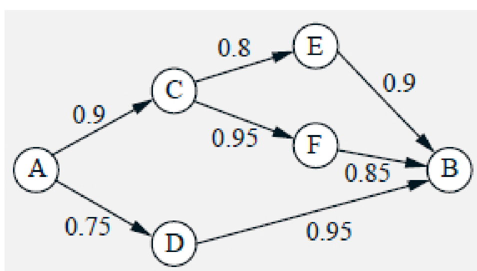
(g) (1 pts) An agent starts at the point $(0, 0)$ in the plane and moves to the goal point $(100, 101)$ by taking a sequence of steps, where each step *either* goes to the right one unit (R) or goes up one unit (U). Clearly, there are many possible valid paths that reach the goal. If the agent selects a valid path at random, what is the probability that it consists of 100 R-steps followed by 101 U-steps?

SOLUTION: Each valid path is a sequence of the form $URUURRURRR\ldots$. The challenge is to identify how many valid paths there are. The desired probability will then be one divided by this number. To reach the goal, there must be 100 R-steps and 101 U-steps, and each step will be either R or U, for a total of 201 steps. Therefore, each path can by uniquely represented by the positions of the 100 R-steps, because the remaining 101 positions will be U's.

$$\Omega = \{\text{All sequences of } R \text{ and } U \text{ of length 201 with 100 } R\text{'s and 101 } U\text{'s}\}$$
$$A = \{RR\ldots RU\ldots U\}$$
$$|\Omega| = C^{201}_{100} \qquad \text{(Each sequence consists of choosing 100 locations for } R\text{'s)}$$
$$P(A) = 1/|\Omega|$$

Note that we use combinations here because the individual R-steps are not distinguished from each other, so all that matters is which of the 201 positions will have an R-step in it.

(h) (1 pts) Consider the following communication network, where each arrow represents a link between two nodes. The number for each link is the probability that it is working. What is the probability that a path exists between A and B such that all of the links along the path are working? Note that the link outcomes are independent events.



Hint: Let p_i denote the probability that the ith link is working. In order to follow a series of links (e.g. $C \to E \to B$), each link in the series must be working (i.e. an *AND* connection). In that case, the probability that a path across m links in series is working is $p_1 p_2 \cdots p_m$. In contrast, when there are multiple links leaving the same node in parallel (e.g. $C \to E$ or $C \to F$), then if any one of the links works, there will be a path out of that node (i.e. an *OR* connection). In that case, the probability of success for m links in parallel is given by: $P(\text{connected}) = 1 - P(\text{not-connected}) = 1 - (1 - p_1)(1 - p_2) \cdots (1 - p_n)$. Note that it's much easier to compute the failure probability than it is to sum all of the successful cases.

SOLUTION: The sample space is a bit vector, with one bit for each link, where the bit is set if the link is working and cleared if the link is not working. The event space is a subset of the bit vectors with the property that B is reachable from A by following links whose bit is set. Let b_{XY} denote the bit for the link $X \rightarrow Y$ and $p_{XY} = P(b_{XY} = T)$.

$\Omega = \{b_{AC}, b_{AD}, b_{CE}, b_{CF}, b_{DB}, b_{EB}, b_{FB}\}$
$A = \{$Bit vectors for: $A \rightarrow C \rightarrow E \rightarrow B,\ A \rightarrow C \rightarrow F \rightarrow B,\ A \rightarrow D \rightarrow B$ active$\}$
$P(C \rightarrow E \rightarrow B) = p_{CE}p_{EB}$ (Series connection, *AND*)
$P(C \rightarrow F \rightarrow B) = p_{CF}p_{FB}$ (Series connection, *AND*)
$P(C \rightarrow B) = 1 - (1 - p_{CE}p_{EB})(1 - p_{CF}p_{FB}) = 0.946$ (Parallel connection, *OR*)
$P(A \rightarrow B) = 1 - (1 - p_{AC}P(C \rightarrow B))(1 - p_{AD}p_{DB}) = 0.957$ (Parallel connection, *OR*)

Question 2. Conditional Probabilities

(a) (1 pts) Suppose two 6-sided dice are rolled. Given that the roll resulted in a sum of 4 or less, what is the probability that the roll was a double?

SOLUTION: Let A be the conditioning event that the sum is 4 or less, and B be the event of a double. Then

$$A = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}$$
$$B = \{(1,1),(2,2)\}$$
$$P(B|A) = 2/6 = 0.33$$

Note that the effect of conditioning is to reduce the outcome space from the original 36 possibilities to 6 possibilities. In this case we can compute the conditional probability directly, without computing $P(A)$ or $P(A \cap B)$.

(b) (1 pts) Suppose two 6-sided dice are rolled. Given that the two dice land on different numbers, what is the probability that at least one die roll is a 6.

SOLUTION: Let A be the conditioning event and B be the event that one die roll is 6. Then:

$$|A| = C_2^6 = 15 \qquad \text{(Different numbers means choosing 2 out of the 6 possibilities)}$$
$$|B| = 5 \qquad \text{(There are 5 combinations in } A \text{ with a six and a number that is not 6)}$$
$$P(B|A) = 5/15 = 0.33$$

(c) (1 pts) Suppose that three 6-sided dice are rolled. What is the probability that no two numbers are the same *and* there is a 1 or a 6 among the numbers? *Hint:* Use the result from problem 1(c).

SOLUTION: Let A be the event that no two numbers are the same and B the event that there is a 1 or a 6 among the numbers. The problem asks for $P(A \cap B)$. We will start by computing $P(B|A)$. Given A, the number of combinations with a 1 is given by $C_2^5 = 10$, as two die face must be chosen from the remaining 5 values. There are similarly 10 combinations which have a single 6. There are 4 combinations which have both a 1 and a 6 with the third die face ranging from 2-5 (i.e. C_1^4). Thus, given A we have $|B| = 2 \times 10 - 4 = 16$. From problem 1, we have that $|A| = C_3^6 = 20$, and therefore:

$$P(B|A) = 16/20 = 0.8$$
$$P(A \cap B) = P(B|A)P(A) = 0.8 \times 0.56 = 0.448,$$

where $P(A)$ comes from problem 1(c).

(d) (1 pts) We are given three coins: one has heads on both faces, one has tails on both faces, and one has a head and a tail and is fair. We choose a coin at random and then toss it, and it comes up heads. What is the probability that the opposite face is a tail (i.e. that it is the coin with a head and a tail)?

SOLUTION: Let the event that each type of coin is selected be denoted by $HH, TT,$ and HT. Let the heads or tails outcome of the coin toss be denoted by O_H and O_T, respectively. Then the sample space is $\Omega = \{HH, TT, HT\} \times \{O_H, O_T\}$. The question asks for $P(HT|O_H)$. We have $P(HH) = P(TT) = P(HT) = 1/3$. We have $P(O_H|HH) = 1$, $P(O_H|TT) = 0$, and $P(O_H|HT) = 1/2$. Then:

$$
\begin{aligned}
P(O_H) &= P(O_H|HH)P(HH) + P(O_H|TT)P(TT) + P(O_H|TH)P(TH) \\
&= 1/3 \times (1 + 0 + 1/2) = 1/2 \\
P(HT|O_H) &= \frac{P(O_H|HT)P(HT)}{P(O_H)} = \frac{1}{2}\,\frac{1}{3}\,2 = 1/3
\end{aligned}
$$

(e) (1 pts) In a game show (Monty Hall), you have to choose one of three doors. Behind one door is a new car, behind the other two are old goats. You choose, but your chosen door is not opened immediately. Instead, the presenter opens a different door, showing you what's behind it. He then gives you the opportunity to choose the third, unopened door *instead* of the one you originally selected. The question is whether you should do so. Answer the question by first assuming that the presenter *always* shows you a goat (in other words, when your door conceals a goat, he always chooses the door with the other goat, and when your door conceals the car, he will choose one of the other two doors at random.) Then answer the question a second time, assuming that the presenter chooses one of the other two doors at random, and opens it for you. If there is a goat behind that door, should you switch?

SOLUTION: The fallacious "intuitive" solution is that since there is a 1/3 chance that the car is behind the door you chose, and also a 1/3 chance of it being behind the other two doors, then switching can't change the outcome. This disregards the fact that by opening a door, the presenter has provided information which changes the marginal probabilities to conditional probabilities. More concretely, 2/3 of the time you will choose a door with a goat initially, and you will win by switching in 100% of those cases, because the presenter reveals the position of the second goat. Only 1/3 of the time will you choose the car initially, and switching will cause you to lose in 100% of those cases. Therefore it is clear that you should always switch. The solution may appear counterintuitive because no matter which door you choose initially, the presenter will always show you a goat and you will always switch doors. The fact that your behavior seems preordained might be troubling, but keep in mind that when you switch you are choosing a *specific* door which is determined by the door the presenter opens, and which has a higher conditional probability of concealing the car than the door you chose initially.

To show this mathematically, we need to define the sample space. Let C_i be the event that door i conceals the car. We assume that all 6 combinations of car and goats are equally likely. Since the problem is symmetric, we can assume that the contestant always selects door 1. Let O_i be the probability that the presenter opens door i. Without loss of generality, we can assume that the presenter opened door 2, revealing a goat. We need to find $P(C_3|O_2)$ and compare it to $P(C_1|O_2)$:

$$P(C_3|O_2) = \frac{P(C_3 \cap O_2)}{P(O_2)} = \frac{P(O_2|C_3)P(C_3)}{P(O_2|C_3)P(C_3) + P(O_2|C_3^c)P(C_3^c)}$$
$$= \frac{(1)(1/3)}{(1)(1/3) + (1/4)(2/3)} = \frac{2}{3}$$

Note that $P(O_2|C_3) = 1$ because if the car is behind door 3, then door 2 must conceal the other goat. $P(O_2|C_3^c) = 1/4$ because if the car is not behind door 3, then with probability 1/2 it will be behind door 2, resulting in zero probability for O_2, and with probability 1/2 it will be behind door 1. In that case, door 2 will be opened half the time, resulting in a total probability of 1/4. It's clear that $P(C_2|O_2) = 0$. Therefore, $P(C_1|O_2) = 1 - P(C_3|O_2) = 1/3$. Therefore you can double your probability of winning by switching. It should be clear that there is a symmetry in the problem and the result will be the same if the presenter opened door 3 instead of 2.

Now consider the case where the presenter chooses a door at random and opens it. Once again we can assume that door 2 is opened without loss of generality. Let G_i denote the event that there is a goat behind door i. The problem is asking for $P(C_3|G_2)$. Then:

$$P(C_3|G_2) = \frac{P(G_2|C_3)P(C_3)}{P(G_2)} = \frac{(1)(1/3)}{2/3} = \frac{1}{2}$$

As before, it's clear that $P(C_2|G_2) = 0$, and therefore $P(C_1|G_2) = 1 - P(C_3|G_2) = 1/2$. It follows that while the probability of choosing the car has gone up (since one of the goat choices was eliminated), the original door and the alternative door have the same probability of concealing the car. BTW, if you search on the web for "Monty Hall Problem" you can see that the correct solution is counterintuitive to many people.

Question 3. **Conditional Independence** You are going to play two games of go against an opponent that you never played before. Your opponent is equally likely to be a beginner, intermediate, or expert player. Depending on which they are, your probability of winning is 90%, 50%, or 30% respectively.

(a) (1 pts) What is the probability that you win your first game?

SOLUTION: Let W_i be the event of winning the ith game. Then,

$$P(W_1) = (0.9 + 0.5 + 0.3)/3 = 17/30 = 0.57$$

(b) (1 pts) Suppose that you win the first game. What is the probability that you will also win the second game? *Note:* you should assume that the outcomes of the games are independent, *given* the skill level of your opponent.

SOLUTION: The question is asking for $P(W_2|W_1)$. We need $P(W_1 \cap W_2)$, which can be obtained by marginalizing out the skill level of the opponent. Let B, I, and E represent the event where skill level of the oponent is beginner, intermediate, and expert, respectively. Then

$$\begin{aligned}
P(W_1 \cap W_2) &= \frac{1}{3}\left[P(W_1 \cap W_2|B) + P(W_1 \cap W_2|I) + P(W_1 \cap W_2|E)\right] \\
&= \frac{1}{3}\left[P(W_1|B)P(W_2|B) + P(W_1|I)P(W_2|I) + P(W_1|E)P(W_2|E)\right] \\
&= (0.9^2 + 0.5^2 + 0.3^2)/3 = 23/60 = 0.383 \\
P(W_2|W_1) &= \frac{P(W_1 \cap W_2)}{P(W_1)} = \frac{23/60}{17/30} = 23/34 = 0.676,
\end{aligned}$$

where we used the fact that W_1 and W_2 are conditionally-independent given the opponent's skill level, and the value of $P(W_1)$ from part (a).

(c) (1 pts) In part (b) you assumed that the outcomes of the two games are conditionally-indpendent given the opponent's skill level. Comment on the validity of this assumption relative to the alternative assumption that the game outcomes are marginally independent.

SOLUTION: The assumption of conditional independence makes sense when you consider that once you know how good your opponent is, you can predict how likely you are to win any game. The fact that you might have won or lost a game to that opponent in the past is unlikely to have much impact (e.g. winning an upset against an expert does not mean that your next game against them is likely to be easier.) The only exception might be if you learned something about effective strategy during one game that you could apply in the next. In contrast, the assumption of marginal indpendence doesn't make sense. If you do not know how strong your opponent is, the first game you play with them will provide evidence for their ability, and this in turn would affect your assessment of the likelihood of winning the next game.

Question 4. **Probability Tables** Consider the following joint probability distribution $P(A, B, C)$ in three discrete random variables A, B, C:

$$p(A=1,B,C) = \begin{array}{c|ccc} & B=1 & B=2 & B=3 \\ \hline C=1 & 0 & 0.05 & 0.05 \\ C=2 & 0.05 & 0.05 & 0.05 \\ C=3 & 0.05 & 0 & 0.05 \\ \hline \end{array}$$

$$p(A=2,B,C) = \begin{array}{c|ccc} & B=1 & B=2 & B=3 \\ \hline C=1 & 0.1 & 0.1 & 0.2 \\ C=2 & 0.1 & 0 & 0 \\ C=3 & 0 & 0.1 & 0.05 \\ \hline \end{array}$$

(a) (2 pts) Compute $P(B|A = 2, C = 1)$.

SOLUTION: As B has three possible values, the result will be a table of three probabilities:

$$P(B|A=2,C=1) = \frac{P(B,A=2,C=1)}{P(A=2,C=1)}$$

$$\begin{aligned} P(A=2,C=1) &= \sum_{B=1}^{3} P(B,A=2,C=1) \\ &= P(B=1,A=2,C=1) + P(B=2,A=2,C=1) + P(B=3,A=2,C=1) \\ &= 0.1 + 0.1 + 0.2 = 0.4 \end{aligned}$$

$$P(B|A=2,C=1) = \frac{1}{0.4}\begin{bmatrix}0.1 & 0.1 & 0.2\end{bmatrix}$$

$$= \begin{array}{ccc} B=1 & B=2 & B=3 \\ \hline 0.25 & 0.25 & 0.5 \\ \hline \end{array}$$

(b) (2 pts) What is the *a priori* probability distribution over C?

SOLUTION: The question asks for $P(C)$ which we obtain via marginalization:

$$P(C) = \sum_{A,B} P(A,B,C) = \sum_{A=1}^{2}\sum_{B=1}^{3} P(A,B,C).$$

There is no shortcut, we just need to do the brute-force computation. First sum out B:

$$P(A=1,C) = \sum_{B=1}^{3} P(A=1,B,C) = \begin{bmatrix}0.1 & 0.15 & 0.1\end{bmatrix}$$

$$P(A=2,C) = \sum_{B=1}^{3} P(A=2,B,C) = \begin{bmatrix}0.4 & 0.1 & 0.15\end{bmatrix}.$$

Each of the probabilities above (corresponding to $C = 1, C = 2, C = 3$ in order) is the result of summing out B by adding up the values in one row from the appropriate table, indexed by A and C. For example,

$$P(A = 2, C = 1) = 0.1 + 0.1 + 0.2 = 0.4.$$

In order to get $P(C)$ we have to marginalize out A, which is equivalent to adding together the two tables for $A = 1$ and $A = 2$ above:

$$P(C) = \sum_{A=1}^{2} P(A, C) = \begin{array}{ccc} C=1 & C=2 & C=3 \\ \hline 0.5 & 0.25 & 0.25 \\ \hline \end{array}$$

Question 5. **Professor Bayes** A professor is heading back into the lab late at night and is trying to guess whether his student is working on their paper. He has noticed that whenever this student is working, their car is in the garage and there is music playing in the lab. Let W, C, M be three binary random variables corresponding to the events "student Working on paper", "Car in parking lot", and "Music playing." Through careful observation, the professor has determined the following conditional probabilities:

(1) $P(C|W,M) = P(C|W) =$

	$W=1$	$W=0$
$C=1$	0.8	0.1
$C=0$	0.2	0.9

(2) $P(M|W) = P(C|W)$

(3) $P(W=1) = 0.8$

Note: Equation (2) says that C and M have the same conditional distribution (e.g. $P(C = 1|W=1) = P(M=1|W=1) = 0.8$.)

(a) (1 pts) Draw a Bayesian network (i.e. a directed graphical model) for this problem.

The graph should have the edges $W \to M$ and $W \to C$.

(b) (2 pts) Compute the marginal distributions $P(C)$ and $P(M)$, before any evidence is availble

SOLUTION: We will use Equation 2 to simplify the computation:

$$P(C,W) = P(C|W)P(W) = P(M|W)P(W) = P(M,W)$$

$P(M,W) = P(M|W)P(W) =$

	$W=1$	$W=0$
$M=1$	0.64	0.02
$M=0$	0.16	0.18

,

where the results in the table are obtained by multiplying the first column in the table from Equation 1 by $P(W=1) = 0.8$, and the second column by $P(W=0) = 0.2$. Marginalizing out W by summing each row gives the result:

$P(M) =$

$M=1$	$M=0$
0.66	0.34

$P(C) =$

$C=1$	$C=0$
0.66	0.34

(c) (2 pts) Upon entering the garage, the professor notices that the student's car is parked there. Calculate $P(W|C=1)$.

SOLUTION: Using Bayes Rule:

$$P(W|C=1) = \frac{P(C=1|W)P(W)}{P(C=1)} = \frac{1}{0.66}[0.64 \quad 0.02] = \begin{array}{cc} W=1 & W=0 \\ \hline 0.97 & 0.03 \\ \hline \end{array}$$

(d) (2 pts) Calculate the updated probability that music is playing in the lab, taking the evidence $C=1$ into account.

SOLUTION:

$$\begin{aligned} P(M|C=1) &= \sum_{W=0}^{1} P(M,W|C=1) = \sum_{W=0}^{1} P(M|W,C=1)P(W|C=1) \\ &= \sum_{W=0}^{1} P(M|W)P(W|C=1), \end{aligned}$$

where the last equality follows from the fact that $M \perp C \mid W$. Using $P(W|C=1)$ from part (c) and multiplying the columns of $P(M|W)$, we obtain:

$$P(M,W|C=1) = \begin{array}{ccc} & W=1 & W=0 \\ \hline M=1 & (0.8)(0.97) & (0.1)(0.03) \\ M=0 & (0.2)(0.97) & (0.9)(0.03) \\ \hline \end{array}$$

$$= \begin{array}{ccc} & W=1 & W=0 \\ \hline M=1 & 0.776 & 0.003 \\ M=0 & 0.194 & 0.027 \\ \hline \end{array}$$

$$= \begin{array}{cc} M=1 & M=0 \\ \hline 0.779 & 0.221 \\ \hline \end{array}$$

Question 6. Safety Monitor

You are worried about the safety of your home, and constantly monitor whether your front door is open (D) using an iphone app. When your door is open there are two possible causes: your spouse is at home (S), and may have left the door open by accident, or your house was robbed (R). D, S, R are binary random variables. R and S are marginally independent (i.e. $R \perp S$) with prior probabilities $P(S = 1) = 0.5$ and $P(R = 1) = 0.1$. The following conditional probability table gives the probability of the door being open:

S	R	$P(D = 1\|S, R)$	$P(D = 0\|S, R)$
0	0	0.01	0.99
0	1	0.25	0.75
1	0	0.05	0.95
1	1	0.75	0.25

(a) (1 pts) Why don't each of the columns in the probability table above sum to 1?

Because they are not a PDF for any random variable.

(b) (1 pts) Draw a Bayes net model (i.e. a directed graphical model) for this problem.

The graph should have the edges $S \to D$ and $R \to D$.

(c) (1 pts) What is the probability that the door is open conditioned on being told that your spouse is at home and your house was not robbed?

From the table, $P(D = 1|S = 1, R = 0) = 0.05$.

(d) (1 pts) Calculate the joint probability $P(D = 1, S = 1, R = 0)$. *Hint:* Use the definition of conditional probability. How does this probability differ from your answer in part (c)? Explain the difference.

$$P(D = 1, S = 1, R = 0) = P(D = 1|S = 1, R = 0)P(S = 1)P(R = 0) = (0.05)(0.5)(0.9) = 0.0225$$

It is much smaller than (c) because it is the joint event without conditioning on the evidence.

(e) (2 pts) Suppose that your only evidence is that your spouse is at home. Compute the conditional probability that the door is open (e.g. $P(D=1|S=1)$).

$$P(D=1,S=1,R=1) = P(D=1|S=1,R=1)P(S=1)P(R=1) = (0.75)(0.5)(0.1) = 0.375$$

$$P(D=1|S=1) = \frac{\sum_{R=0}^{1} P(D=1,S=1,R)}{P(S=1)} = \frac{0.0225+0.375}{0.5} = 0.12$$

(f) (2 pts) Now suppose that your only evidence is that the door is open. Compute the posterior distribution of S and R.

SOLUTION: The problem is asking for $P(S,R|D=1)$ which will be a two-by-two table we can compute using Bayes Rule:

$$P(S,R|D=1) = \frac{P(D=1|S,R)P(S)P(R)}{P(D=1)}$$

$$= \frac{1}{P(D=1)} \times \begin{array}{c|cc} & R=0 & R=1 \\ \hline S=0 & (0.01)(0.5)(0.9) & (0.25)(0.5)(0.1) \\ S=1 & (0.05)(0.5)(0.9) & (0.75)(0.5)(0.1) \end{array}$$

$$= \frac{1}{P(D=1)} \times \begin{array}{c|cc} & R=0 & R=1 \\ \hline S=0 & 0.0045 & 0.0125 \\ S=1 & 0.0225 & 0.0375 \end{array}$$

$P(D=1) = 0.077$ (Sum of the elements in the matrix above)

$$P(S,R|D=1) = \begin{array}{c|cc} & R=0 & R=1 \\ \hline S=0 & 0.0584 & 0.162 \\ S=1 & 0.292 & 0.487 \end{array}$$

Question 7. **Recursive Updating**

Derive the standard recursive updating rule for inference in a Hidden Markov Model that is used to obtain $P(x_t|e_{1:t})$, where x_t is the hidden state at time t and $e_{1:t}$ is the sequence of observations up to time t. The rule should be recursive (or incremental) in the sense that $P(x_t|e_{1:t})$ can be obtained given *only* $P(x_{t-1}|e_{1:t-1})$ and e_t (in other words, there is no need to store the previous measurements $e_{1:t-1}$ or reprocess them.)

SOLUTION: The starting point is the use of Bayes Rule to define the posterior of interest:

$$(1) \quad P(x_t|e_{1:t}) = \frac{P(x_t, e_{1:t})}{P(e_{1:t})}$$

$$(2) \quad = \frac{1}{k} \sum_{x_{t-1}} P(x_{t-1}, x_t, e_{1:t-1}, e_t)$$

$$(3) \quad = \frac{1}{k} \sum_{x_{t-1}} P(x_t, e_t|x_{t-1}, e_{1:t-1}) P(x_{t-1}, e_{1:t-1})$$

$$(4) \quad = \frac{1}{k} \sum_{x_{t-1}} P(e_t|x_t, x_{t-1}, e_{1:t-1}) P(x_t|x_{t-1}, e_{1:t-1}) P(x_{t-1}, e_{1:t-1})$$

$$(5) \quad = \frac{1}{k} \sum_{x_{t-1}} P(e_t|x_t) P(x_t|x_{t-1}) P(x_{t-1}, e_{1:t-1})$$

$$(6) \quad = \frac{1}{k} P(e_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) B(x_{t-1}) = \frac{1}{k} P(e_t|x_t) B'(x_t).$$

$$(7) \quad k = P(e_{1:t}) = \sum_{x_t} P(e_t|x_t) B'(x_t) = \sum_{x_t} B(x_t)$$

$$(8) \quad = \sum_{x_t} P(x_t, e_t, e_{1:t-1}) = \sum_{x_t} P(e_t|x_t, e_{1:t-1}) P(x_t, e_{1:t-1}) = \sum_{x_t} P(e_t|x_t) B'(x_t).$$

In step 2, we introduce the variable x_{t-1} and then marginalize it out with the sum, and in addition separate $e_{1:t}$ into $e_{1:t-1}$ and e_t. This is done in order to be able to create the necessary dependencies for the recursive updating rule. Step 3 follows from the definition of conditional probability, as does step 4. In step 5, the conditional independencies in the model, namely $e_t \perp \{x_{t-1}, e_{1:t-1}\} \mid x_t$ and $x_t \perp e_{1:t-1} \mid x_{t-1}$, are applied to simplify the conditional probabilities. In step 6, we identify the relevant belief functions, $B(x_t) = P(x_t, e_{1:t})$ and $B'(x_t) = P(x_t, e_{1:t-1})$, that define the recursive updating rule. In the step 7, we derive the formula for the probability of the evidence from the consideration that $P(x_t|e_{1:t})$ should sum to one, because k is the normalizing constant in the formula for $P(x_t|e_{1:t})$. In step 8, we show that the same formula can be derived from first principles, by marginalizing x_t out of the joint distribution $P(x_t, e_{1:t})$, which is the normal procedure for obtaining the probability of the evidence when using Bayes Rule.

The belief functions are just a convenient way to reference the relevant intermediate state distributions (or factors) that arise in the recursive updating process. The relationship between the belief functions and the posterior probabilities is summarized in the sequence of steps that make up the recursive updating algorithm:

(1) $B(x_0) = P(x_0)$ (Initialization)

(2) $B'(x_t) = \sum_{x_{t-1}} P(x_t|x_{t-1})B(x_{t-1})$ (Forward Simulation)

(3) $B(x_t) = P(e_t|x_t)B'(x_t)$ (Evidence Updating)

(4) $P(e_{1:t}) = \sum_{x_t} B(x_t)$ (Probability of the Evidence)

(5) $P(x_t|e_{1:t}) = \dfrac{B(x_t)}{P(e_{1:t})}$ (Normalization)

Question 8. Dynamic Dean

The AI faculty decide to build an agent to guess the mood of Dean Zvi Galil from hour to hour (so visits to the Dean's office can be carefully timed). After detailed observation, they determine that when the Dean is in a good mood, there is a 75% chance he will be in a good mood an hour later. But when he is in a bad mood, there is a 50% chance he will be in a bad mood an hour later. Unfortunately the Dean's mood can't be observed directly. However, through careful study the faculty have determined that when the Dean is in a good mode, the likelihood that he will send a reminder email about CIOS is 80%, but there is only a 60% chance of sending a CIOS email when he is in a bad mood.

(a) (1 pts) Draw a dynamic Bayesian network model for this problem. Be sure to show at least two time slices. Label the nodes in your graph. Use the random variable M to indicate the Dean's mood. M can take on the values "good" or "bad." Use the random variable C to indicate whether a CIOS email was sent. C can take on the values "true" or "false." Use subscripts on M and C to indicate time slices.

The graph should have the edges $M_{t-1} \to M_t$, $M_{t-1} \to C_{t-1}$, and $M_t \to C_t$.

(b) (2 pts) Fill in the conditional probability tables for the transition model and sensor model elements of the DBN. Be sure the columns and rows are labeled properly in the spaces provided.

Transition Model:

$P(M_t \mid M_{t-1})$	$M_t =$ bad	$M_t =$ good
$M_{t-1} =$ bad	0.5	0.5
$M_{t-1} =$ good	0.25	0.75

Measurement Model:

$P(C_t \mid M_t)$	$C_t = F$	$C_t = T$
$M_t =$ bad	0.4	0.6
$M_t =$ good	0.2	0.8

(c) (1 pts) The faculty are interested in tracking the Dean's mood carefully during the Final Exam period. The day before Finals start, the faculty have no knowledge of the Dean's mood. In other words, $P(M_0 = \text{good}) = 0.5$. Predict (through forward simulation) the probability of the Dean's mood on the first day of Finals:

$$P(M_1 = \text{good}) = \sum_{M_0} P(M_1 = \text{good}|M_0)P(M_0) = (0.5)(0.5) + (0.75)(0.5) = 0.625$$

Note that this is an example of belief updating with belief function $B(M_0) = P(M_0)$ updated due to passage of time to yield the prediction belief $B'(M_1) = P(M_1)$. The general form of the updated belief for any time t would be $B'(M_t) = P(M_t, C_{1:t-1})$. In this case, evidence has not been introduced yet.

(d) (2 pts) Now suppose that the Dean sends a CIOS reminder email on the first day of Finals, in other words $C_1 = T$. Compute an updated posterior estimate for the Dean's mood M_1.

SOLUTION: The problem is asking for $P(M_1|C_1)$, which we can obtain by updating the prediction belief function $B'(M_1)$ from part (c) with evidence, yielding $B(M_1) = P(M_1, C_1)$, which can then be normalized to obtain $P(M_1|C_1)$. We have:

$$P(M_1|C_1) = \frac{B(M_1)}{P(C_1)} = \frac{P(C_1|M_1)B'(M_1)}{P(C_1)}$$

$$= \frac{1}{P(C_1)} \times \begin{array}{c|c} & P(M_1, C_1) \\ \hline M_1 = \text{bad} & (0.6)(0.375) = 0.225 \\ M_1 = \text{good} & (0.8)(0.625) = 0.5 \end{array}$$

$$= \begin{array}{c|c} & P(M_1|C_1) \\ \hline M_1 = \text{bad} & 0.225/0.725 = 0.31 \\ M_1 = \text{good} & 0.5/0.725 = 0.69 \end{array}$$

Note that we were able to obtain the probability of evidence by summing the (unnormalized) beliefs. Dividing by this number normalizes the beliefs, resulting in a probability density.

Question 9. Particle Filter

This question explores the particle filter algorithm for inference in an HMM using sampling methods. The hidden state is the location of an agent in a 5x5 grid world, which we parameterize as (u, v) for discrete u and v random variables. The transition model for a state (u, v) has 40% probability of staying in the same state and 15% probability of moving in each of the four compass directions N, S, E, and W. The measurement model assigns 60% probability of measuring the discrete state correctly, and 10% probability of generating an erroneous measurement corresponding to a displacement in one of the four compass directions.

(a) (2 pts) Write down the transition model relating (u_{t-1}, v_{t-1}) and (u_t, v_t) and the measurement model relating (u_t, v_t) and the corresponding measurements (x_t, y_t).

Transition Model:

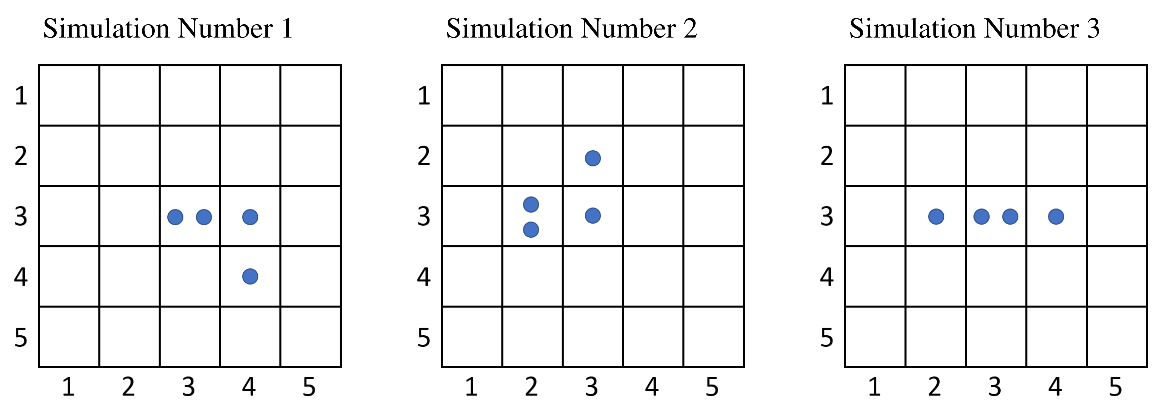
$$P(u_t = i, v_t = j | u_{t-1} = i, v_{t-1} = j) = 0.40$$
$$P(u_t = i + 1, v_t = j | u_{t-1} = i, v_{t-1} = j) = 0.15$$
$$P(u_t = i - 1, v_t = j | u_{t-1} = i, v_{t-1} = j) = 0.15$$
$$P(u_t = i, v_t = j + 1 | u_{t-1} = i, v_{t-1} = j) = 0.15$$
$$P(u_t = i, v_t = j - 1 | u_{t-1} = i, v_{t-1} = j) = 0.15$$

Measurement Model:

$$P(x_t = i, y_t = j | u_t = i, v_t = j) = 0.60$$
$$P(x_t = i + 1, y_t = j | u_t = i, v_t = j) = 0.10$$
$$P(x_t = i - 1, y_t = j | u_t = i, v_t = j) = 0.10$$
$$P(x_t = i, y_t = j + 1 | u_t = i, v_t = j) = 0.10$$
$$P(x_t = i, y_t = j - 1 | u_t = i, v_t = j) = 0.10$$

(b) (2 pts) Suppose that at time $t-1$ we have a set of four particles that are all located at $(3,3)$. Let $s_{t-1}^k = (u_{t-1}^k, v_{t-1}^k)$ denote the kth particle, and let $S_{t-1} = \{s_{t-1}^1, s_{t-1}^2, s_{t-1}^3, s_{t-1}^4\}$ denote the particle set at time $t-1$. Then we have $s_{t-1}^k = (3,3), \quad k = 1,\dots,4$. The three figures below show three *different* possible sets of particles resulting from *one step* of forward simulation (i.e. the result of sampling once from $P(u_t, v_t | u_{t-1}, v_{t-1})$ for each particle). Which of these three particle sets is the *most likely* to result from the transition model in part (a)? Justify your answer numerically. *Hint:* Compute the likelihood of each particle set, which is given by:

$$P(S_t) = \prod_{k=1}^{4} P(u_t^k, v_t^k | u_{t-1}^k, v_{t-1}^k).$$



SOLUTION: By inspection it is clear that simulation number 3 is the most likely. Simulation 1 is impossible, as in one step of forward simulation the particles can only diffuse in the compass directions, and so the particle at $(4,4)$ in simulation 1 has a likelihood of zero. Since the probability of diffusing in each compass direction is the same, the only difference between simulations 2 and 3 is that simulation 3 has two particles at $(3,3)$ and simulation 2 has only one. This will make simulation 3 more likely, since the likelihood of staying at $(3,3)$ is 0.4, which is higher than the likelihood of drifting, 0.15. To show this numerically we compute the likelihood for each case.

Simulation 1:

$$P(S_t) = (0.4)(0.4)(0.15)(0.0) = 0.0$$

Simulation 2:

$$P(S_t) = (0.15)(0.15)(0.15)(0.4) = 0.00135$$

Simulation 3:

$$P(S_t) = (0.15)(0.4)(0.4)(0.15) = 0.0036$$

(c) (2 pts) Now suppose that the result of forward simulation is the distribution of particles in Simulation Number 2 from part (b). We obtain a measurement of the target position at time t, which is $m_t = (2, 3)$. Write the list of particles from Simulation Number 2 with their *likelihood weights*, following the measurement update step.

SOLUTION: The particle list is as follows:

$(2, 3)$ $w = 0.6$
$(2, 3)$ $w = 0.6$
$(3, 2)$ $w = 0.0$
$(3, 3)$ $w = 0.1$

There are two particles located at $(2, 3)$, so that location appears twice. Note that the particle at $(3, 2)$ receives a weight of zero because its probability is zero under the measurement model of part (a).

(d) (2 pts) Suppose that we perform *resampling* using the list of weighted particles given below. Calculate the probability of drawing each of the particles during resampling.

$(1, 2)$ $w = 0.1$
$(2, 1)$ $w = 0.5$
$(2, 2)$ $w = 0.6$
$(2, 2)$ $w = 0.6$
$(2, 4)$ $w = 0.3$
$(3, 2)$ $w = 0.2$
$(3, 3)$ $w = 0.4$
$(3, 3)$ $w = 0.4$
$(3, 3)$ $w = 0.4$
$(3, 5)$ $w = 0.2$
$(4, 3)$ $w = 0.8$
$(4, 4)$ $w = 0.7$

SOLUTION: The probability of drawing a particle during resampling is equal to the total likelihood for that particle divided by the sum of all the likelihoods, which is 5.2. The probabilities are as follows:

$(1, 2)$ $P = 0.0192$
$(2, 1)$ $P = 0.0962$
$(2, 2)$ $P = 0.231$
$(2, 4)$ $P = 0.0577$
$(3, 2)$ $P = 0.0385$
$(3, 3)$ $P = 0.231$
$(3, 5)$ $P = 0.0385$
$(4, 3)$ $P = 0.154$
$(4, 4)$ $P = 0.135$