CS 188 Introduction to Artificial Intelligence **Probability Review 2**

1 Independence

1. Here are two joint probability tables. In which are the variables independent?

A	B	$P(A,B)$
+a	+b	0.42
+a	–b	0.18
–a	+b	0.28
–a	–b	0.12

C	D	$P(C,D)$
+c	+d	0.4
+c	–d	0.15
–c	+d	0.4
–c	–d	0.05

The left table is independent, which we can see if we calculate the conditional probability tables.

2. (a) Given that $A \perp\!\!\!\perp B$, simplify $\sum_a P(a|B)P(C|a)$:
First, we utilize our independence assumption to get $\sum_a P(a|B)P(C|a) = \sum_a P(a)P(C|a)$. Now, we can invoke the chain rule to see $\sum_a P(a)P(C|a) = \sum_a P(a,C)$. Marginalizing / summing over a, we are left with $P(C)$.

(b) Given that $B \perp\!\!\!\perp C|A$, simplify $\frac{P(A)P(B|A)P(C|A)}{P(B|C)P(C)}$:
We can use our independence assumption in the numerator to see that $P(B|A)P(C|A) = P(B,C|A)$. We can then apply the chain rule in both the numerator and denominator to get that $\frac{P(A)P(B|A)P(C|A)}{P(B|C)P(C)} = \frac{P(A,B,C)}{P(B,C)} = P(A|B,C)$.

(c) Given that $A \perp\!\!\!\perp B|C$, simplify $\frac{P(C,A|B)P(B)}{P(C)}$:
Applying the formula for conditional probability to the numerator and then to the entire fraction gets $\frac{P(C,A|B)P(B)}{P(C)} = \frac{P(A,B,C)}{P(C)} = P(A,B|C)$. We can now utilize the independence assumption to simplify to $P(A|C)P(B|C)$.

2 Expression Manipulation

Simplify the following expressions, as much as possible:

1. $\sum_x \sum_y f(x)h(x,y) = \sum_x f(x) \sum_y h(x,y)$

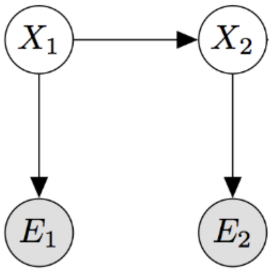
2. $\sum_a \sum_b \sum_c f(c)f(c,b)f(c,b,a) = \sum_c f(c) \sum_b f(c,b) \sum_a f(c,b,a)$

3. $\sum_a \sum_b \sum_c f(b)f(c)f(a,b,c)f(a,j)f(a,m) = \sum_a f(a,j)f(a,m) \sum_b f(b) \sum_c f(c)f(a,b,c)$

4. $\sum_d \sum_k \sum_c \sum_x f(c,k)f(x)f(c,k,x)f(k)f(c,d,k,x)f(k,x) = \sum_x f(x) \sum_k f(k)f(k,x) \sum_c f(c,k)f(c,k,x) \sum_d f(c,d,k,x)$

3 Bayesian Inference

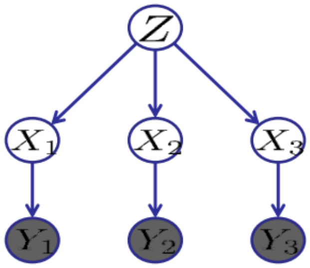
For each of the following Bayes's Nets, identify the (conditional) independence assumptions it makes, and then derive expressions for the specified probabilities:



The key thing to note is that conditioned on X_1, E_1 is independent of the rest of the graph, and conditioned on X_2, E_2 is independent of the rest of the graph. Thus a partial list of independence assumptions includes:

- $E_1 \perp\!\!\!\perp X_2 | X_1$
- $X_1 \perp\!\!\!\perp E_2 | X_2$
- $E_1 \perp\!\!\!\perp E_2 | X_1$
- $E_1 \perp\!\!\!\perp E_2 | X_2$

1. $P(X_1|E_1)$:
   $P(X_1|E_1) = \frac{P(X_1,E_1)}{P(E_1)} = \frac{P(X_1,E_1)}{\sum_{x_1} P(x_1,E_1)} = \frac{P(X_1)P(E_1|X_1)}{\sum_{x_1} P(x_1)P(E_1|x_1)}$
2. $P(X_2|E_1)$:
   $P(X_2|E_1) = \sum_{x_1} P(x_1,X_2|E_1) = \sum_{x_1} P(x_1|E_1)P(X_2|x_1)$
3. $P(X_2|E_1,E_2)$:
   $P(X_2|E_1,E_2) = \frac{P(X_2,E_2|E_1)}{P(E_2|E_1)} = \frac{P(E_2|X_2,E_1)P(X_2|E_1)}{\sum_{x_2} P(x_2,E_2|E_1)} = \frac{P(E_2|X_2)P(X_2|E_1)}{\sum_{x_2} P(x_2|E_2)P(E_2|E_1)}$



Some of the independence assumptions (of which there are many) include: $X_i \perp\!\!\!\perp X_j | Z$, $Y_i \perp\!\!\!\perp Z | X_i$, $Y_i \perp\!\!\!\perp Y_j | X_i$, and $X_i \perp\!\!\!\perp Y_j | X_j$.

Derive $P(X_3|Y_1,Y_2,Y_3)$:

We first calculate some helper expressions:

$$\begin{aligned} P(Y_i|Z) &= \sum_{x_i} P(x_i, Y_i|Z) \\ &= \sum_{x_i} P(x_i|Z) P(Y_i|X_i, Z) \\ &= \sum_{x_i} P(x_i|Z) P(Y_i|X_i) \end{aligned}$$

$$\begin{aligned} P(Z|Y_1,Y_2,Y_3) &= \frac{P(Z,Y_1,Y_2,Y_3)}{P(Y_1,Y_2,Y_3)} \\ &= \frac{P(Z,Y_1,Y_2,Y_3)}{\sum_z P(z,Y_1,Y_2,Y_3)} \\ &= \frac{P(Z)P(Y_1|Z)P(Y_2|Z,Y_1)P(Y_3|Z,Y_1,Y_2)}{\sum_z P(z)P(Y_1|z)P(Y_2|z,Y_1)P(Y_3|z,Y_1,Y_2)} \\ &= \frac{P(Z)P(Y_1|Z)P(Y_2|Z)P(Y_3|Z)}{\sum_z P(z)P(Y_1|z)P(Y_2|z)P(Y_3|z)} \end{aligned}$$

$$\begin{aligned} P(X_3|Y_1,Y_2,Y_3) &= \sum_z P(X_3, z|Y_1,Y_2,Y_3) \\ &= \sum_z P(X_3|z,Y_1,Y_2,Y_3) P(z|Y_1,Y_2,Y_3) \\ &= \sum_z P(X_3|z,Y_3) P(z|Y_1,Y_2,Y_3) \\ &= \sum_z \frac{P(X_3,Y_3|z)}{\sum_{x_3} P(x_3,Y_3|z)} P(z|Y_1,Y_2,Y_3) \\ &= \sum_z \frac{P(Y_3|X_3,z)P(X_3|z)}{\sum_{x_3} P(Y_3|x_3,z)P(x_3|z)} P(z|Y_1,Y_2,Y_3) \\ &= \sum_z \frac{P(Y_3|X_3)P(X_3|z)}{\sum_{x_3} P(Y_3|x_3)P(x_3|z)} P(z|Y_1,Y_2,Y_3) \end{aligned}$$

4 Expectation

Let's say we have a fair, 6-sided dice.

1. Let X_i be the result of the ith dice roll. What is $E[X_1]$? $\sum_i P(X=i)i = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

2. What is $E[X_1^2]$?
   $\sum_i P(X=i)i^2 = \frac{91}{6}$

3. What is $E[X_1+X_2]$?
   By linearity of expectation, $E[X_1+X_2] = E[X_1]+E[X_2] = 7$.

4. Let K be an arbitrary random variable, and let Y_K be the sum of the first K dice rolls. What is $E[Y|K]$?

   $$E[Y|K] = \sum_{i=1}^{K} E[X_1] = 3.5K$$

5. We play a game where we first flip a fair coin. If the coin lands heads, then we roll the dice twice and score points equal to the sum. If the coin lands tails, then we roll three dice instead. What is our expected score?
   Let C be the result of the coin flip, and X as our expected score. We want to calculate:

   $$E[X] = E_C[E[X|C]] = P(C=H)*E[X|C=H]+P(C=T)*E[X|C=T] = \frac{1}{2}*(2*3.5)+\frac{1}{2}*(3*3.5) = 8.75$$

   (Notice that we could have also calculated: (the expected number of dice rolls) * (the expected result per dice roll) to get the same result; this is not a coincidence, but something that is generally true, as per Wald's equation.)