Matthew Ragoza
mtr22@pitt.edu

Assignment 9

ISSP 2160
November 10, 2021

Problem 1. Joint probability distributions

a. The probability distributions $P(A)$ and $P(B)$ are shown below.

	$P(A)$
$A = 1$	0.6
$A = 0$	0.4
Total	1.0

	$P(B)$
$B = 1$	0.3
$B = 0$	0.7
Total	1.0

b. The joint probability distribution $P(A, B)$, assuming $A \perp\!\!\!\perp B$, is shown below.

$P(A,B)$	$B = 1$	$B = 0$	Total
$A = 1$	0.18	0.42	0.60
$A = 0$	0.12	0.28	0.40
Total	0.30	0.70	1.00

c. Using the provided joint distribution $P(A, B)$, we have the following:

$P(A,B)$	$B = 1$	$B = 0$	Total
$A = 1$	0.2	0.4	0.6
$A = 0$	0.1	0.3	0.4
Total	0.3	0.7	1.0

$$
\begin{aligned}
&P(B = 1)\\
&= \sum_{a \in \{1,0\}} P(A = a, B = 1)\\
&= P(A = 1, B = 1) + P(A = 0, B = 1)\\
&= 0.2 + 0.1\\
&= 0.3
\end{aligned}
$$

$$
\begin{aligned}
&P(A = 1|B = 0)\\
&= \frac{P(A = 1, B = 0)}{P(B = 0)}\\
&= \frac{P(A = 1, B = 0)}{\sum_{a \in \{1,0\}} P(A = a, B = 0)}\\
&= \frac{P(A = 1, B = 0)}{P(A = 1, B = 0) + P(A = 0, B = 0)}\\
&= \frac{0.4}{0.4 + 0.3}\\
&= \frac{0.4}{0.7}\\
&\approx 0.57
\end{aligned}
$$

$$
\begin{aligned}
&P(B = 1|A = 0)\\
&= \frac{P(A = 0, B = 1)}{P(A = 0)}\\
&= \frac{P(A = 0, B = 1)}{\sum_{b \in \{1,0\}} P(A = 0, B = b)}\\
&= \frac{P(A = 0, B = 1)}{P(A = 0, B = 1) + P(A = 0, B = 0)}\\
&= \frac{0.1}{0.1 + 0.3}\\
&= \frac{0.1}{0.4}\\
&= 0.25
\end{aligned}
$$

Problem 2. Bayes' theorem

Let D represent the presence of disease A and let T represent a positive test result. We know that the true positive rate and true negative rate are 99%, while the false positive rate and false negative rate are 1%. We also know that the prevalence of the disease in the population is 0.01% and the chance of not having the disease is 99.99%, by the law of total probability. This knowledge can be represented as the following given probabilities:

$$\begin{aligned}
P(T=1|D=1) = P(T=0|D=0) &= 0.99 \\
P(T=0|D=1) = P(T=1|D=0) &= 0.01 \\
P(D=1) &= 0.0001 \\
P(D=0) &= 0.9999
\end{aligned}$$

We are interested in finding the probability that a person has disease A given that they test positive for the disease, which is the probability $P(D=1|T=1)$. We can compute this using Bayes' theorem as follows:

$$\begin{aligned}
P(D=1|T=1) &= \frac{P(T=1|D=1)P(D=1)}{P(T=1)} \\
&= \frac{P(T=1|D=1)P(D=1)}{\sum_{d\in\{1,0\}} P(T=1, D=d)} \\
&= \frac{P(T=1|D=1)P(D=1)}{\sum_{d\in\{1,0\}} P(T=1|D=d)P(D=d)} \\
&= \frac{P(T=1|D=1)P(D=1)}{P(T=1|D=1)P(D=1) + P(T=1|D=0)P(D=0)} \\
&= \frac{(0.99)(0.0001)}{(0.99)(0.0001) + (0.01)(0.9999)} \\
&= \frac{0.000099}{0.000099 + 0.009999} \\
&= \frac{0.000099}{0.010098} \\
&\approx 0.0098
\end{aligned}$$

This indicates that when someone tests positive for the disease, they only have a 0.98% probability of actually having it. Even though 99% test accuracy seems high, the disease is so rare that the probability of having it given a positive test result remains low. I would only recommend this test if it were used as a preliminary screen before using a test with a much lower false positive rate to definitively determine the presence of the disease. Otherwise, relying on this test could result in over-diagnosis and wasted medical resources.

Problem 3. Conditional independence

Proof.

1. $P(A, B|C) = P(A|C)P(B|C)$ (Given)
2. $P(A, B|C) = \frac{P(A,B,C)}{P(C)}$ (Definition of conditional probability)
3. $\frac{P(A,B,C)}{P(C)} = P(A|C)P(B|C)$ (Substitute step 1 with step 2)
4. $P(A, B, C) = P(A|C)P(B|C)P(C)$ (Multiply step 3 by $P(C)$)
5. $P(A, B, C) = P(A|B, C)P(B|C)P(C)$ (Chain rule of probability)
6. $P(A|B, C)P(B|C)P(C) = P(A|C)P(B|C)P(C)$ (Substitute step 4 with step 5)
7. $P(A|B, C) = P(A|C)$ (Divide step 6 by $P(B|C)P(C)$)

$\square$

Problem 4. Bayesian belief network

Let B represent "Battery," R represent "Radio," L represent "Lights," I represent "Ignition," G represent "Gas," E represent "Engine starts," and C represent "Car moves." We will assume that the domain of each random variable is $\mathbb{B} = \{1, 0\}$. Let $n = 7$ represent the number of random variables and let $d = 2$ represent the size of their domain.

a. The total number of parameters in the full joint distribution is $d^n = 2^7 = 128$. The number of free parameters in the full joint distribution is $128 - 1 = 127$, since one parameter is constrained by the law of total probability.

b. The total number of parameters in the Bayesian belief network is $2^1 + 2^2 + 2^2 + 2^2 + 2^1 + 2^3 + 2^2 = 28$. Since one parameter is constrained by the law of total probability in each local conditional distribution, the number of free parameters in the Bayesian belief network is $2^0 + 2^1 + 2^1 + 2^1 + 2^0 + 2^2 + 2^1 = 14$.

c. The expression for the full joint probability defined by the Bayesian belief network is:

$$\begin{aligned}
P(B,R,L,I,G,E,C) &= P(C|B,R,L,I,G,E)P(B,R,L,I,G,E) \\
&= P(C|E)P(B,R,L,I,G,E) \\
&= P(C|E)P(E|B,R,L,I,G)P(B,R,L,I,G) \\
&= P(C|E)P(E|I,G)P(B,R,L,I,G) \\
&= P(C|E)P(E|I,G)P(G|B,R,L,I)P(B,R,L,I) \\
&= P(C|E)P(E|I,G)P(G)P(B,R,L,I) \\
&= P(C|E)P(E|I,G)P(G)P(I|B,R,L)P(B,R,L) \\
&= P(C|E)P(E|I,G)P(G)P(I|B)P(B,R,L) \\
&= P(C|E)P(E|I,G)P(G)P(I|B)P(L|B,R)P(B,R) \\
&= P(C|E)P(E|I,G)P(G)P(I|B)P(L|B)P(B,R) \\
&= P(C|E)P(E|I,G)P(G)P(I|B)P(L|B)P(R|B)P(B)
\end{aligned}$$

For the provided variable outcomes, the joint probability can be expressed as:

$$P(C=0|E=0)P(E=0|I=1,G=1)P(G=1)P(I=1|B=1)P(L=1|B=1)P(R=0|B=1)P(B=1)$$

d. The simplified expression for the marginal probability of the car not moving is:

$$\begin{aligned}
P(C=0) &= \sum_{B\in\mathbb{B}}\sum_{R\in\mathbb{B}}\sum_{L\in\mathbb{B}}\sum_{I\in\mathbb{B}}\sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C=0|E)P(E|I,G)P(G)P(I|B)P(L|B)P(R|B)P(B) \\
&= \sum_{R\in\mathbb{B}}\sum_{L\in\mathbb{B}}\sum_{I\in\mathbb{B}}\sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C=0|E)P(E|I,G)P(G)\left[\sum_{B\in\mathbb{B}} P(I|B)P(L|B)P(R|B)P(B)\right] \\
&= \sum_{L\in\mathbb{B}}\sum_{I\in\mathbb{B}}\sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C=0|E)P(E|I,G)P(G)\left[\sum_{B\in\mathbb{B}} P(I|B)P(L|B)\left[\sum_{R\in\mathbb{B}} P(R|B)\right]P(B)\right] \\
&= \sum_{I\in\mathbb{B}}\sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C=0|E)P(E|I,G)P(G)\left[\sum_{B\in\mathbb{B}} P(I|B)\left[\sum_{L\in\mathbb{B}} P(L|B)\right]\left[\sum_{R\in\mathbb{B}} P(R|B)\right]P(B)\right] \\
&= \sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C=0|E)\left[\sum_{I\in\mathbb{B}} P(E|I,G)P(G)\left[\sum_{B\in\mathbb{B}} P(I|B)\left[\sum_{L\in\mathbb{B}} P(L|B)\right]\left[\sum_{R\in\mathbb{B}} P(R|B)\right]P(B)\right]\right] \\
&= \sum_{E\in\mathbb{B}} P(C=0|E)\left[\sum_{I\in\mathbb{B}}\left[\sum_{G\in\mathbb{B}} P(E|I,G)P(G)\right]\left[\sum_{B\in\mathbb{B}} P(I|B)\left[\sum_{L\in\mathbb{B}} P(L|B)\right]\left[\sum_{R\in\mathbb{B}} P(R|B)\right]P(B)\right]\right] \\
&= \sum_{E\in\mathbb{B}} P(C=0|E)\left[\sum_{I\in\mathbb{B}}\left[\sum_{G\in\mathbb{B}} P(E|I,G)P(G)\right]\left[\sum_{B\in\mathbb{B}} P(I|B)\left[\sum_{L\in\mathbb{B}} P(L|B)\right]P(B)\right]\right] \\
&= \sum_{E\in\mathbb{B}} P(C=0|E)\left[\sum_{I\in\mathbb{B}}\left[\sum_{G\in\mathbb{B}} P(E|I,G)P(G)\right]\left[\sum_{B\in\mathbb{B}} P(I|B)P(B)\right]\right]
\end{aligned}$$

e. The conditional probability of the battery working given that the lights do not work is:

$$
\begin{aligned}
&P(B=1|L=0) \\
&= \frac{P(B=1,L=0)}{P(L=0)} \\
&= \frac{\sum_{C\in\mathbb{B}}\sum_{R\in\mathbb{B}}\sum_{I\in\mathbb{B}}\sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C|E)P(E|I,G)P(G)P(I|B=1)P(L=0|B=1)P(R|B=1)P(B=1)}{\sum_{C\in\mathbb{B}}\sum_{B\in\mathbb{B}}\sum_{R\in\mathbb{B}}\sum_{I\in\mathbb{B}}\sum_{G\in\mathbb{B}}\sum_{E\in\mathbb{B}} P(C|E)P(E|I,G)P(G)P(I|B)P(L=0|B)P(R|B)P(B)} \\
&= \frac{\sum_{R\in\mathbb{B}}\sum_{I\in\mathbb{B}} P(I|B=1)P(L=0|B=1)P(R|B=1)P(B=1)}{\sum_{B\in\mathbb{B}}\sum_{R\in\mathbb{B}}\sum_{I\in\mathbb{B}} P(I|B)P(L=0|B)P(R|B)P(B)} \\
&= \frac{\sum_{I\in\mathbb{B}} P(I|B=1)P(L=0|B=1)\left[\sum_{R\in\mathbb{B}} P(R|B=1)\right]P(B=1)}{\sum_{B\in\mathbb{B}}\sum_{I\in\mathbb{B}} P(I|B)P(L=0|B)\left[\sum_{R\in\mathbb{B}} P(R|B)\right]P(B)} \\
&= \frac{\left[\sum_{I\in\mathbb{B}} P(I|B=1)\right]P(L=0|B=1)\left[\sum_{R\in\mathbb{B}} P(R|B=1)\right]P(B=1)}{\sum_{B\in\mathbb{B}}\left[\sum_{I\in\mathbb{B}} P(I|B)\right]P(L=0|B)\left[\sum_{R\in\mathbb{B}} P(R|B)\right]P(B)} \\
&= \frac{P(L=0|B=1)P(B=1)}{\sum_{B\in\mathbb{B}} P(L=0|B)P(B)}
\end{aligned}
$$