Homework 1 (due: 9/27/2023)

1. (a) True. By the product rule we know $p(b, c)p(a|b, c) = p(a, c)p(b|a, c)$, which by assumption reduces to $p(b, c) = p(a, c)$. Dividing through by $p(c)$ gives the result.

   (b) False. The statement $p(a|b, c) = p(a)$ merely states that a is independent of b and c, it makes no claim regarding the dependence of b and c. A counter-example: a and b record the results of two independent coin flips, and $c = b$.

   (c) False. While the statement $p(a|b) = p(a)$ implies that a is independent of b, it does not imply that a is conditionally independent of b given c. A counter-example: a and b record the results of two independent coin flips, and c equals the xor of a and b.

2. (a) $P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

   (b) $P(Catch) = \langle 0.34, 0.66 \rangle$

   (c) $P(Cavity|catch) = \langle (.108 + .072)/0.34, (0.016 + 0.144)/0.34 \rangle = \langle 0.5294, 0.4706 \rangle$

   (d) $P(Cavity|toothache \vee catch) = \langle (0.108+0.012+0.072)/0.416, (0.016+0.064+0.144)/0.416 \rangle = \langle 0.4615, 0.5384 \rangle$

3. The correct message is received if either zero or one of the $n + 1$ bits are corrupted. Since corruption occurs independently with probability ϵ, the probability that zero bits are corrupted is $(1 - \epsilon)^{n+1}$. There are $n + 1$ mutually exclusive ways that exactly one bit can be corrupted, one for each bit in the message. Each has probability $\epsilon(1 - \epsilon)^n$, so the overall probability that exactly one bit is corrupted is $(n + 1)\epsilon(1 - \epsilon)^n$. Thus, the probability that the correct message is received is

$$(1 - \epsilon)^{n+1} + (n + 1)\epsilon(1 - \epsilon)^n.$$

   The maximum feasible value of n, therefore, is the largest n satisfying the inequality

$$(1 - \epsilon)^{n+1} + (n + 1)\epsilon(1 - \epsilon)^n \geq 1 - \delta$$

   Numerically solving this for $\epsilon = 0.002, \delta = 0.01$, we find $n = 73$.

4. (a) $P(z|y) = \frac{P(y,z)}{P(y)} = \frac{\sum_x P(x,y,z)}{\sum_{x,z'} P(x,y,z')}$

   (b) $P(z|y) = \frac{\sum_x \theta(x)\theta(y|x)\theta(z|y)}{\sum_{x,z'} \theta(x)\theta(y|x)\theta(z'|y)}$

   (c) $P(z|y) = \frac{\theta(z|y)\sum_x \theta(x)\theta(y|x)}{\sum_x \theta(x)\theta(y|x)\sum_{z'}\theta(z'|y)} = \frac{\theta(z|y)\sum_x \theta(x)\theta(y|x)}{\sum_x \theta(x)\theta(y|x)} = \theta(z|y)$

   (d) For any set of variables V, we have the following generalized sum-to-1 rule:

$$\sum_V \prod_i \theta(v_i|pa(V_i)) = 1.$$

Divide variables into Z, Y(parents of Z), U(descendants of Z), and X(all other variables), Then

$$\begin{aligned}
P(z|\mathbf{y}) &= \frac{\sum_{\mathbf{x},\mathbf{u}} P(\mathbf{x},\mathbf{y},z,\mathbf{u})}{\sum_{\mathbf{x},z',\mathbf{u}} P(\mathbf{x},\mathbf{y},z',\mathbf{u})} \\
&= \frac{\sum_{\mathbf{x},\mathbf{u}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j))\theta(z|\mathbf{y}) \prod_k \theta(u_k|pa(U_k))}{\sum_{\mathbf{x},z',\mathbf{u}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j))\theta(z'|\mathbf{y}) \prod_k \theta(u_k|pa(U_k))} \\
&= \frac{\sum_{\mathbf{x}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j))\theta(z|\mathbf{y}) \sum_{\mathbf{u}} \prod_k \theta(u_k|pa(U_k))}{\sum_{\mathbf{x}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j)) \sum_{z'} \theta(z'|\mathbf{y}) \sum_{\mathbf{u}} \sum_k \theta(u_k|pa(U_k))} \\
&(moving\ the\ sums\ in\ as\ far\ as\ possible) \\
&= \frac{\sum_{\mathbf{x}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j))\theta(z|\mathbf{y})}{\sum_{\mathbf{x}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j)) \sum_{z'} \theta(z'|\mathbf{y})} \\
&(using\ the\ generalized\ sum-to-1\ rule\ for\ \mathbf{u}) \\
&= \frac{\sum_{\mathbf{x}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j))\theta(z|\mathbf{y})}{\sum_{\mathbf{x}} \prod_i \theta(x_i|pa(X_i)) \prod_j \theta(y_j|pa(Y_j))} \\
&(using\ the\ sum-to-1\ rule\ for\ z') \\
&= \theta(z|\mathbf{y}).
\end{aligned} \tag{1}$$

5. (a) Yes. Numerically one can compute that $P(B, E) = P(B)P(E)$. Topologically B and E are d-separated by A.

   (b) We check whether $P(B, E|a) = P(B|a)P(E|a)$. First computing $P(B, E|a)$

$$\begin{aligned}
P(B,E|a) &= \alpha P(a|B,E)P(B,E) \\
&= \alpha \begin{cases} .95 \times 0.001 \times 0.002 & \text{if} \quad B = b \text{ and } E = e \\ .94 \times 0.001 \times 0.998 & \text{if} \quad B = b \text{ and } E = \neg e \\ .29 \times 0.999 \times 0.002 & \text{if} \quad B = \neg b \text{ and } E = e \\ .001 \times 0.999 \times 0.998 & \text{if} \quad B = \neg b \text{ and } E = \neg e \end{cases} \\
&= \alpha \begin{cases} 0.0008 & \text{if} \quad B = b \text{ and } E = e \\ 0.3728 & \text{if} \quad B = b \text{ and } E = \neg e \\ 0.2303 & \text{if} \quad B = \neg b \text{ and } E = e \\ 0.3962 & \text{if} \quad B = \neg b \text{ and } E = \neg e \end{cases}
\end{aligned} \tag{2}$$

   where α is a normalization constant. Checking whether $P(b, e|a) = P(b|a)P(e|a)$ we find

$$P(b, e|a) = 0.0008 \neq 0.0863 = 0.3736 \times 0.2311 = P(b|a)P(e|a)$$

   showing that B and E are not conditionally independent given A.

6. (a) *IcyWeather* is not caused by any of the car-related variables, so needs no parents. It directly affects the battery and the starter motor. *StarterMoter* is an additional precondition for *Starts*. The new network is shown in Figure 1.

   (b) Reasonable probabilities may vary a lot depending on the kind of car and perhaps the personal experience of the assessor. The following values indicate the general order of magnitude and relative values that make sense:

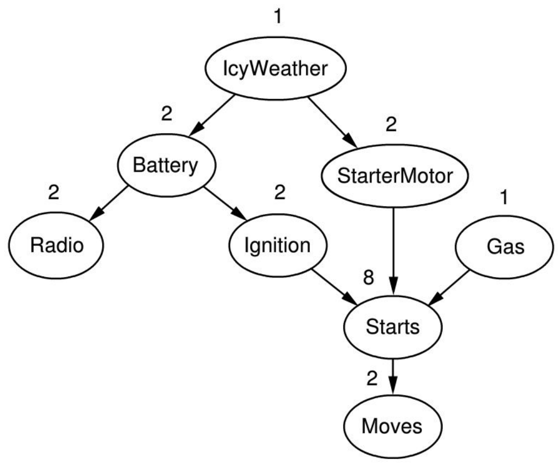

Figure 1: Car network amended to include *IcyWeather* and *StarterMoter*

(c) with 8 boolean variables, the joint has $2^8 - 1 = 255$ independent entries.

(d) Given the topology shown in Figure 1, the total number of independent CPT entries is $1 + 2 + 2 + 2 + 2 + 1 + 8 + 2 = 20$.

(e) The CPT for *Starts* describes a set of nearly necessary conditions that are together almost sufficient. That is, all the entries are nearly zero except for the entry where all the conditions are true. That entry will be not quite 1 (because there is always some other possible fault that we didn't think of), but as we add more conditions it gets closer to 1. If we add a *Leak* node as an extra parent, then the probability is exactly 1 when all parents are true. We can relate noisy-AND to noisy-OR using de Morgan's rule: $A \wedge B \equiv \neg(\neg A \vee \neg B)$. That is, noisy-AND is the same as noisy-OR except that the polarities of the parent and child variables are reversed. In the noisy-OR case, We have

$$P(Y = true|x_1, \dots, x_k) = 1 - \prod_{\{i:x_i=true\}} q_i$$

where q_i is the probability that the *presence* of the ith parent *fails* to cause the child to be *true*. In the noisy-AND case, we can write

$$P(Y = true|x_1, \dots, x - k) = \prod_{\{i:x_i=false\}} r_i$$

where r_i is the probability that the *absence* of the ith parent *fails* to cause the child to be *false* (e.g., it is magically bypassed by some other mechanism).

7. Exact Inference (MATLAB)

(b)

$P(b\|j,m)$	0.3742	$P(e\|j,m)$	0.0573
$P(b\|j,\neg m)$	0.0056	$P(e\|j,\neg m)$	0.0084
$P(b\|\neg j,m)$	0.0881	$P(e\|\neg j,m)$	0.0194
$P(b\|\neg j,\neg m)$	0.0013	$P(e\|\neg j,\neg m)$	0.0078
$P(\neg b\|j,m)$	0.6258	$P(\neg e\|j,m)$	0.9427
$P(\neg b\|j,\neg m)$	0.9944	$P(\neg e\|j,\neg m)$	0.9916
$P(\neg b\|\neg j,m)$	0.9119	$P(\neg e\|\neg j,m)$	0.9806
$P(\neg b\|\neg j,\neg m)$	0.9987	$P(\neg e\|\neg j,\neg m)$	0.9922

8. Approximate Inference

(a) First of all, S and R are not independent. We could prove this from the numerical semantics.

$P(s) = P(s|c)P(c) + P(s|\neg c)P(\neg c) = 0.05 \times 0.7 + 0.65 \times 0.3 = 0.23$
$P(r) = P(r|c)P(c) + P(r|\neg c)P(\neg c) = 0.9 \times 0.7 + 0.3 \times 0.3 = 0.72$
$P(s,r) = P(s|c)P(r|c)P(c) + P(s|\neg c)P(r|\neg c)P(\neg c) = 0.05 \times 0.9 \times 0.7 + 0.65 \times 0.3 \times 0.3 = 0.09$
$\therefore P(s)P(r) \neq P(s,r)$

But, according to the Bayesian networks structure, S and R are conditionally independent given C.

(b) The following table shows the exact $P(S,R|W)$.

$P(s,r\|w)$	0.1227	$P(s,r\|\neg w)$	0.0033
$P(s,\neg r\|w)$	0.1639	$P(s,\neg r\|\neg w)$	0.0767
$P(\neg s,r\|w)$	0.6941	$P(\neg s,r\|\neg w)$	0.4600
$P(\neg s,\neg r\|w)$	0.0193	$P(\neg s,\neg r\|\neg w)$	0.4600

When the grass is wet, the probability of R is higher than that of S. And of course, S and R not independent.

(c) (MATLAB)