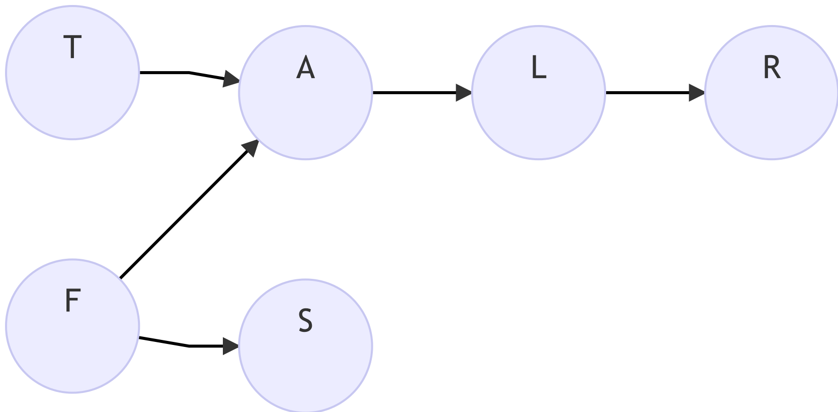
3.1 Variable Elimination Algorithm



(a)

$$
\begin{aligned}
P(T=0|S=1,R=1) &= \frac{P(T=0,S=1,R=1)}{P(S=1,R=1)} \\
&= \frac{P(T=0,S=1,R=1)}{P(T=0,S=1,R=1)+P(T=1,S=1,R=1)}
\end{aligned}
$$

Therefore, all we need to calculate is$P(T,S=1,R=1)$

$$
\begin{aligned}
&P(T,S=1,R=1) \\
&= \sum_{F=0}^{1}\sum_{A=0}^{1}\sum_{L=0}^{1} P(T,A,F,L,S=1,R=1) \\
&= \sum_{F=0}^{1}\sum_{A=0}^{1}\sum_{L=0}^{1} P(T)P(A|T,F)P(F)P(L|A)P(S=1|F)P(R=1|L) \\
&= P(T)\sum_{F=0}^{1} P(S=1|F)P(F)\sum_{A=0}^{1} P(A|T,F)\sum_{L=0}^{1} P(L|A)P(R=1|L) \\
&= f_0(T)\sum_{F=0}^{1} f_3(F,1)P(F)\sum_{A=0}^{1} f_2(T,F,A)\sum_{L=0}^{1} f_4(A,L)f_5(L,1) \\
&= f_0(T)\sum_{F=0}^{1} f_3'(F)f_1(F)\sum_{A=0}^{1} f_2(T,F,A)\sum_{L=0}^{1} f_4(A,L)f_5'(L)
\end{aligned}
$$

Assume that $f_6(A,L)=f_4(A,L)*f_5'(L)$, then

A	L	$f_6(A,L)$
0	0	0.00999

A	L	$f_6(A,L)$
0	1	0.00075
1	0	0.0012
1	1	0.66

and

$$P(T,S=1,R=1) = f_0(T)\sum_{F=0}^{1} f_3'(F)\sum_{A=0}^{1} f_2(T,F,A)\sum_{L=0}^{1} f_6(A,L)$$

Assume that $f_7(A) = \sum_{L=0}^{1} f_6(A,L)$, then

A	$f_7(A)$
0	0.01074
1	0.6612

and

$$P(T,S=1,R=1) = f_0(T)\sum_{F=0}^{1} f_3'(F)\sum_{A=0}^{1} f_2(T,F,A)f_7(A)$$

Assume that $f_8(T,F,A) = f_2(T,F,A) * f_7(A)$, then

T	F	A	$f_8(T,F,A)$
0	0	0	0.10738926
0	0	1	0.00006612
0	1	0	0.0001074
0	1	1	0.654588
1	0	0	0.01611
1	0	1	0.56202
1	1	0	0.0537

T	F	A	$f_8(T,F,A)$
1	1	1	0.3306

and

$$P(T,S=1,R=1)=f_0(T)\sum_{F=0}^{1} f_3'(F)\sum_{A=0}^{1} f_8(T,F,A)$$

Assume that $f_9(T,F)=\sum_{A=0}^{1} f_8(T,F,A)$, then

T	F	$f_9(T,F)$
0	0	0.10745538
0	1	0.6546954
1	0	0.57813
1	1	0.3843

and

$$P(T,S=1,R=1)=f_0(T)\sum_{F=0}^{1} f_3'(F)f_9(T,F)$$

Assume that $f_{10}(T,F)=f_3'(F)*f_9(T,F)$, then

T	F	$f_{10}(T,F)$
0	0	0.0010745538
0	1	0.58922586
1	0	0.0057813
1	1	0.34587

and

$$P(T,S=1,R=1)=f_0(T)\sum_{F=0}^{1} f_{10}(T,F)$$

Assume that $f_{11}(T)=\sum_{F=0}^{1} f_{10}(T,F)$, then

T	$f_{11}(T)$
0	0.5903004138
1	0.3516513

and

$$P(T, S=1, R=1) = f_0(T) f_{11}(T)$$

Assume that $f_{12}(T) = f_0(T) * f_{11}(T)$, then

T	$f_{12}(T)$
0	0.578494405524
1	0.007033026

Therefore

$$P(T, S=1, R=1) = f_{12}(T)$$

In conclusion

$$\begin{aligned} P(T=0|S=1,R=1) &= \frac{P(T=0,S=1,R=1)}{P(T=0,S=1,R=1)+P(T=1,S=1,R=1)} \\ &= \frac{f_{12}(0)}{f_{12}(0)+f_{12}(1)} \\ &= \frac{0.578494405524}{0.578494405524+0.007033026} \\ &\approx 0.9880 \end{aligned}$$

and

$$\begin{aligned} P(T=1|S=1,R=1) &= \frac{P(T=1,S=1,R=1)}{P(T=0,S=1,R=1)+P(T=1,S=1,R=1)} \\ &= \frac{f_{12}(1)}{f_{12}(0)+f_{12}(1)} \\ &= \frac{0.007033026}{0.578494405524+0.007033026} \\ &\approx 0.0120 \end{aligned}$$

(b)

Phase of algorithm	# multiplications	# additions	# divisions

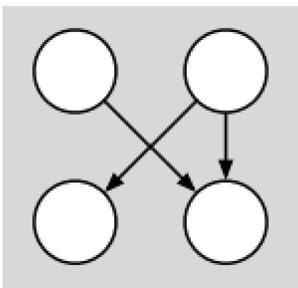
Phase of algorithm	# multiplications	# additions	# divisions
Eliminate S(evidence)	0	0	0
Eliminate R(evidence)	0	0	0
Eliminate L	4	2	0
Eliminate A	8	4	0
Eliminate F	4	2	0
Combine T factors	2	0	0
Normalize distribution over T	0	1	2
Total	18	9	2

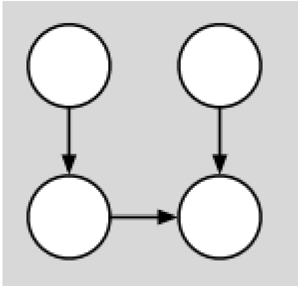
(c)

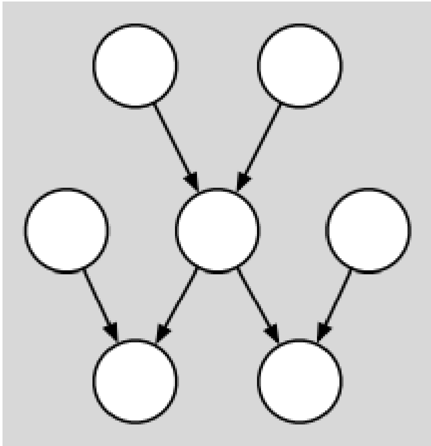
Phase of algorithm	# multiplications	# additions	# divisions
Compute $P(T=0, S=1, R=1)$	40	7	0
Compute $P(T=1, S=1, R=1)$	40	7	0
Normalize distribution over T	0	1	2
Total	80	15	2

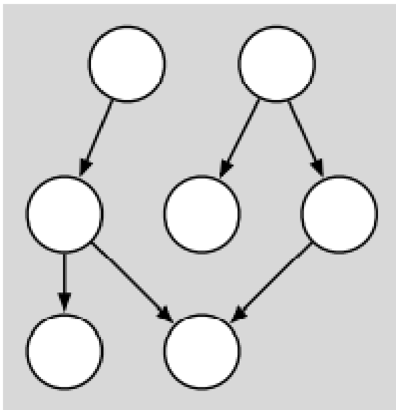
3.2 To be, or not to be, a polytree: that is the question

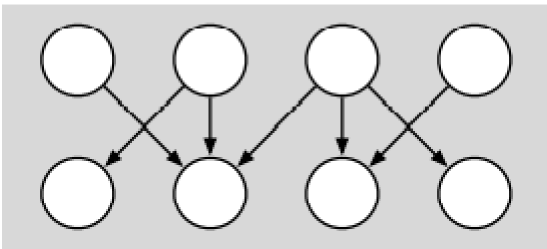
Polytrees:

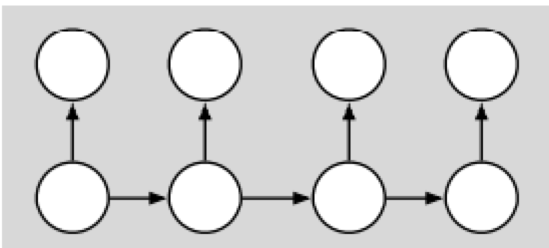






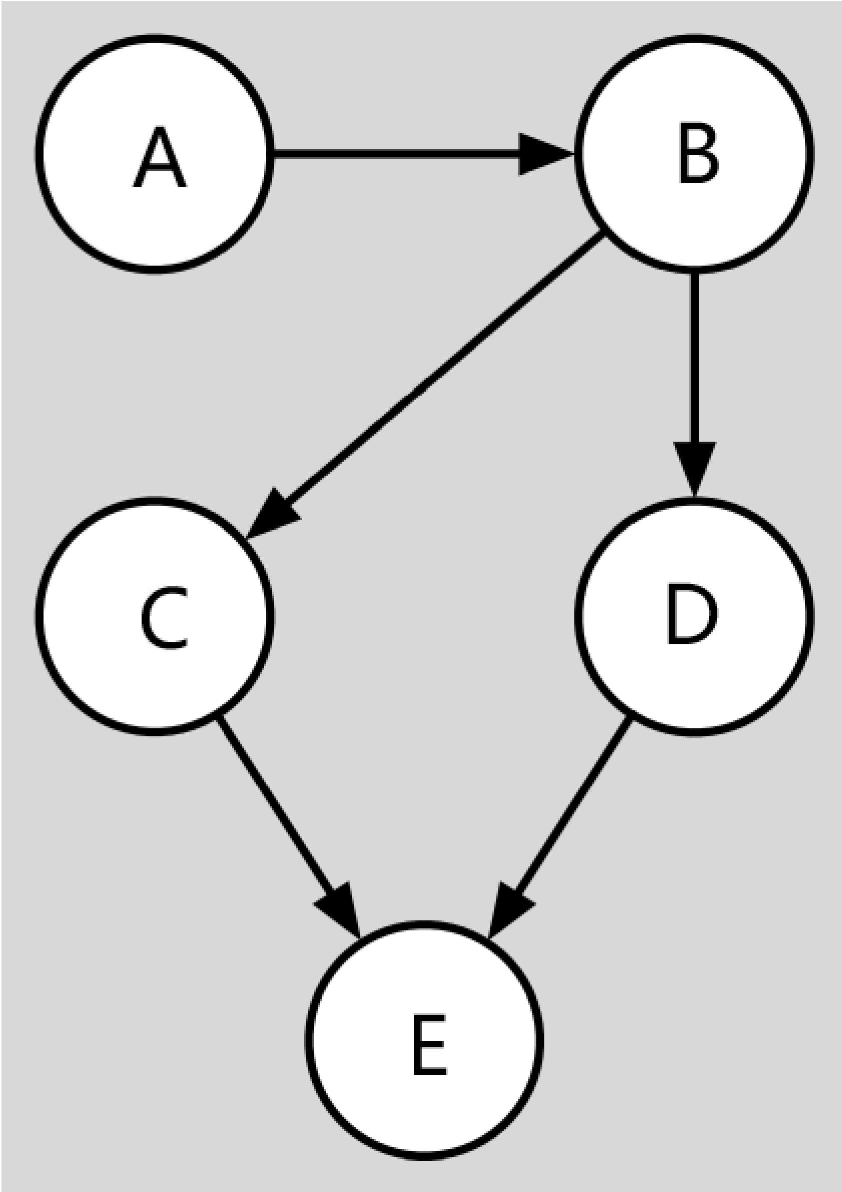






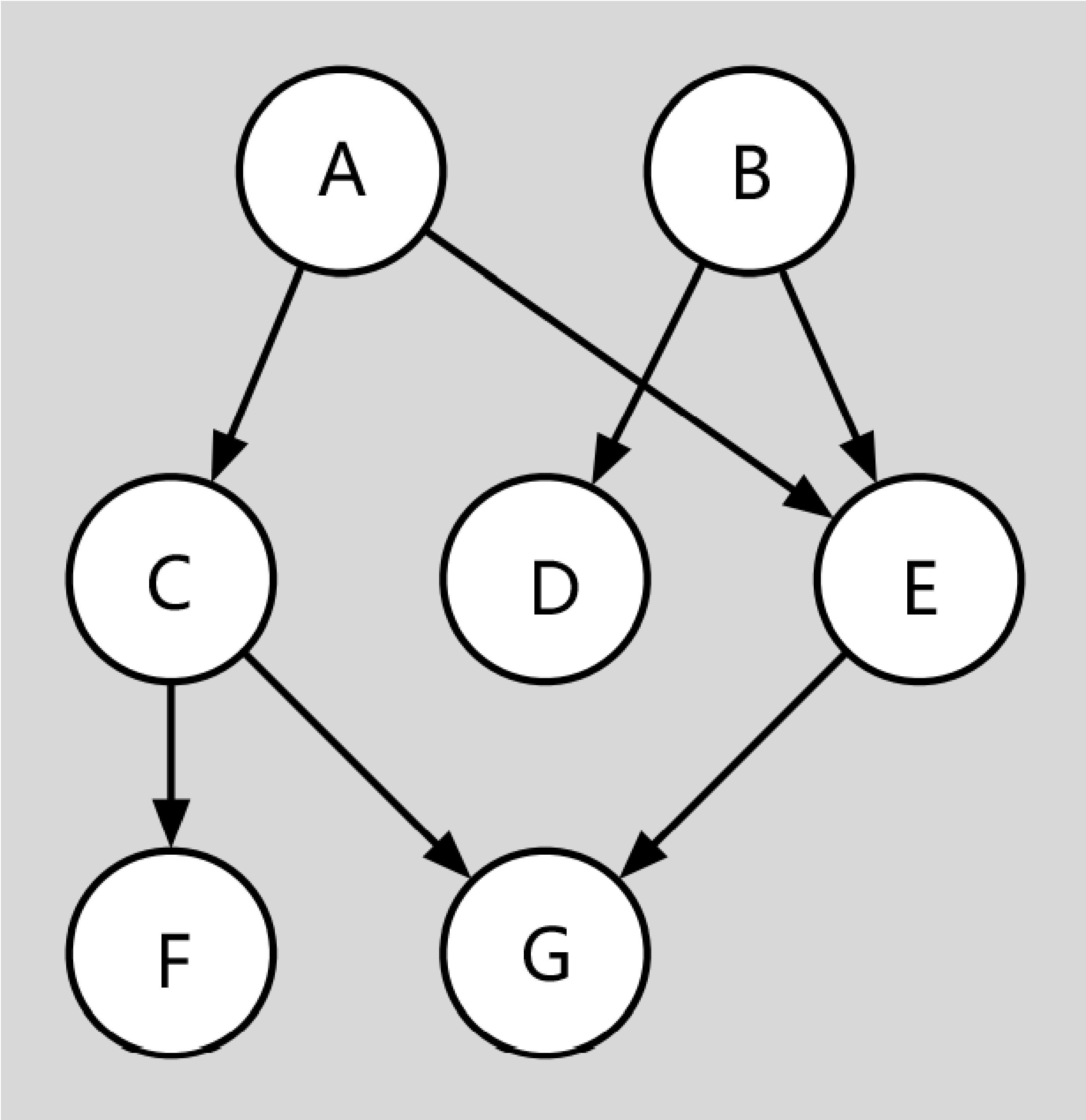
Cluster nodes:

1.



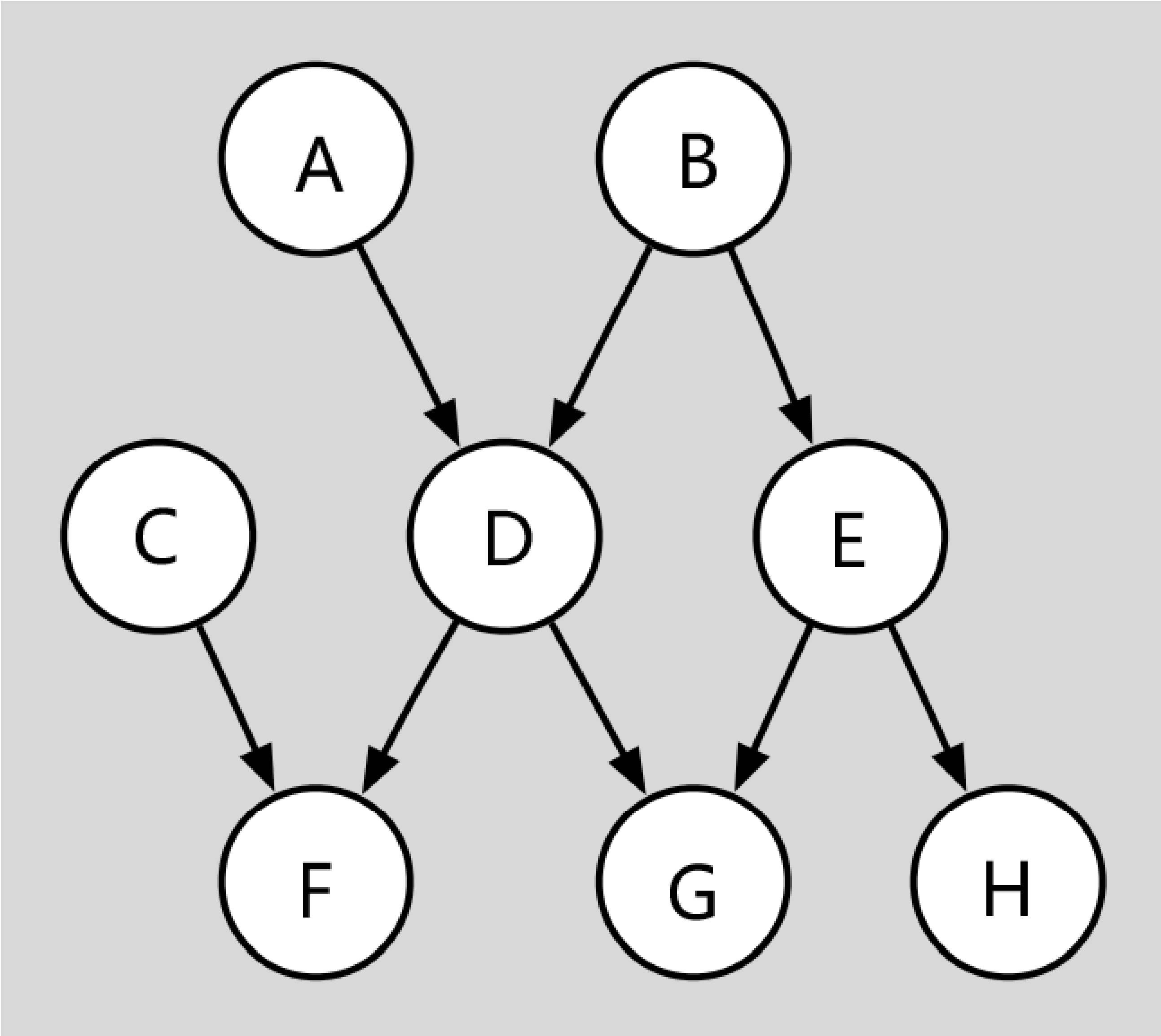
Cluster C, D into a single node F, then the DAG is a polytree.

2.



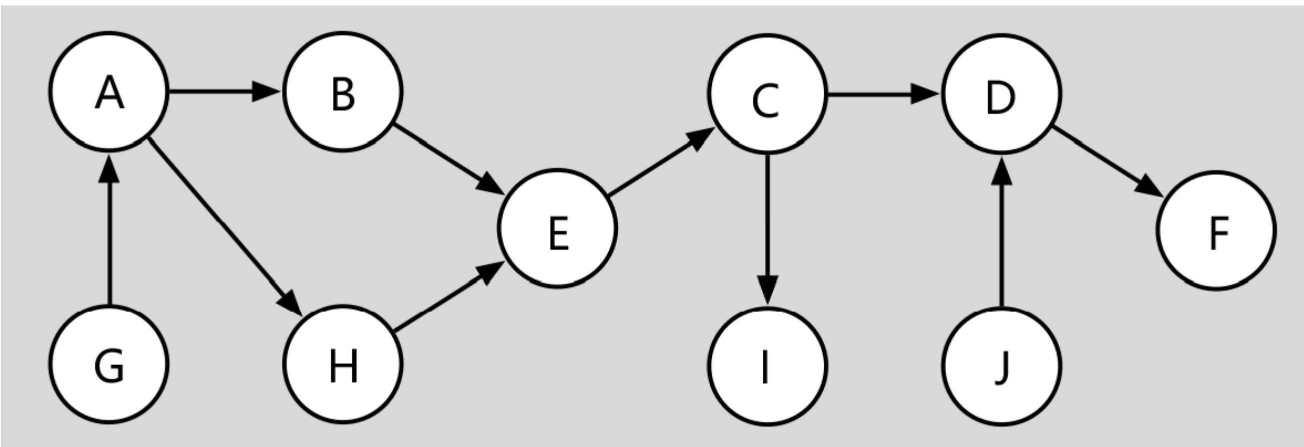
Cluster C, E into a single node H, then the DAG is a polytree.

3.



Cluster D, E into a single node I, then the DAG is a polytree.

4.



Cluster B, H into a single node K, then the DAG is a polytree.

3.3 Node clustering

Y_1	Y_2	Y_3	Y	$P(Y\|X=0)$	$P(Y\|X=1)$	$P(Z_1=1\|Y)$	$P(Z_2=1\|Y)$
0	0	0	1	0.0525	0.14625	0.8	0.2
1	0	0	2	0.2975	0.04875	0.7	0.3
0	1	0	3	0.0225	0.07875	0.6	0.4
0	0	1	4	0.0525	0.34125	0.5	0.5
1	1	0	5	0.1275	0.02625	0.4	0.6
1	0	1	6	0.2975	0.11375	0.3	0.7
0	1	1	7	0.0225	0.18375	0.2	0.8
1	1	1	8	0.1275	0.06125	0.1	0.9

3.4 Maximum likelihood estimation for an n-sided die

(a)

$$\begin{aligned}
\mathcal{L}(p) &= \log P(\text{data}) \\
&= \log P(X_1 = x^{(1)}, X_2 = x^{(2)}, \ldots, X_k = x^{(T)}) \\
&\overset{i.i.d.}{=} \log \left[P(X = x^{(1)})P(X = x^{(2)}) \ldots P(X = x^{(T)})\right] \\
&= \log \left[P(X = 1)^{C_1} P(X = 2)^{C_2} \ldots P(X = n)^{C_n}\right] \\
&= C_1 \log P(X = 1) + C_2 \log P(X = 2) + \ldots + C_n \log P(X = n) \\
&= \sum_{k=1}^{n} C_k \log p_k
\end{aligned}$$

(b)

$$\begin{aligned}
\sum_{k=1}^{n} q_k \log q_k - \frac{\mathcal{L}(p)}{T} &\overset{(a)}{=} \sum_{k=1}^{n} (q_k \log q_k - \frac{C_k}{T} \log p_k) \\
&= \sum_{k=1}^{n} (q_k \log q_k - q_k \log p_k) \\
&= \sum_{k=1}^{n} q_k \log \frac{q_k}{p_k} \\
&\overset{1.6}{=} \mathrm{KL}(q, p)
\end{aligned}$$

(c)

Using the conclusion of question 1.6, we can find that when $\frac{q_1}{p_1} = \frac{q_2}{p_2} = \ldots = \frac{q_n}{p_n}$, then $\mathrm{KL}(q, p)$ get the minimum value, which means that $\mathcal{L}(p)$ get the maximum number.

In conclusion,

$$p_k = q_k = \frac{C_k}{T}$$

is the maximum-likelihood estimate.