Chapter 10. Directed graphical models (Bayes nets)

10.1.
$C, A, B \to E,\ A, B, D \to F$

10.2.
a. None.
b. C, F.

10.3.

$$p(X_i|X_{-i}) = \frac{p(X)}{p(X_{-i})} = \frac{p(X)}{\sum_{X_i} p(X)} = \frac{\prod_j p(Y_j|\text{pa}(Y_j))}{\sum_{X_i}\prod_j p(Y_j|\text{pa}(Y_j))}$$
$$= \frac{p(X_i|\text{par}(X_i))\prod_{Y_j\in \text{ch}(X_i)} p(Y_j|\text{par}(Y_j)}{\sum_{X_i} p(X_i|\text{par}(X_i))\prod_{Y_j\in \text{ch}(X_i)} p(Y_j|\text{par}(Y_j)}.$$

10.4.
a. $p(X_{1:6}) = \sum_n (\prod_{i=1}^3 p(X_i))(\prod_{j=4}^6 p(X_j|H=h))p(H=h|X_{1:3})$.
Number of free parameters: $p(X_1) : 1, p(X_2) : 1, p(X_3) : 1, p(X_4|H=h) : 2, p(X_5|H=h) : 2, p(X_6|H=h) : 2, p(H=h|X_{1:3}) : 8$.
Total : 17
b. $p(X_{1:6}) = \sum_n (\prod_{i=1}^3 p(X_i))p(X_4|X_{1:3})p(X_5|X_{1:4})p(X_6|X_{1:5})$.
Number of free parameters: $p(X_1) : 1, p(X_2) : 1, p(X_3) : 1, p(X_4|X_{1:3}) : 8, p(X_5|X_{1:4}) : 16, p(X_6|X_{1:5}) = 32$. Total : 59
c. For binary features doing so becomes easier, because we have less number of free parameters.

10.5.

a. $P(S=1|V=1) = P(S=1) = P(S=1|G=0)P(G=0) + P(S=1|G=1)P(G=1) = \alpha(1-\gamma) + (1-\alpha)(1-\beta)$.
b. $P(S=1|V=0) = P(S=1) = P(S=1|V=1)$.
c. $\hat{\delta} = 0, \hat{\alpha} = 1/3, \hat{\gamma} = 1, \hat{\beta} = 0$.

10.6.
a. $p(x_2 = S) = 0.85, p(x_2 = B) = 0.15$.
$p(x_2 = S|x_4 = T) \simeq 0.872, p(x_2 = B|x_4 = T) \simeq 0.128$, therefore, the fish is much more likely to be a salmon.
b. $p(x_1|x_3 = M, x_4 = T) = p(x_1|x_2 = S)p(x_2 = S|x_3 = M, x_4 = T) + p(x_1|x_2 = B)p(x_2 = B|x_3 = M, x_4 = T)$.

$$p(x_2|x_3 = M, x_4 = T) = \frac{p(x_3 = M|x_2)p(x_4 = T|x_2)p(x_2)}{\sum_f p(x_3 = M|x_2 = F)p(x_4 = T|x_2 = F)p(x_2 = F)}.$$

$p(x_2) = \sum_s p(x_2|x_1 = s)p(x_1 = s) = [0.6, 0.4]$
$\Rightarrow p(x_2|x_3 = M, x_4 = T) \simeq [0.99832, , 0.00168]$
$p(x_1|x_2 = S) \simeq [0.375, 0.125, 0.167, 0.333]$
$p(x_1|x_2 = B) \simeq [0.0625, 0.4375, 0.375, 0.125]$
$\Rightarrow p(x_1|x_3 = M, x_4 = T) \simeq [0.3745, 0.1256, 0.167, 0.333]$
Therefore, the season is most likely to be spring.

10.7.
a. $p(\mathbf{z}_{1:3}|x_1, x_2, x_4) = p(z_1|x_1, x_2, x_4)p(z_2|x_2)p(z_3|x_4)$.
b. By Noisy-OR assumption, if child node is off, all parents' link to that node fails, so it becomes removable, just marginalizing out from the prior.

10.8.
$p(\mathbf{d}|\mathbf{f}^-) \propto p(\mathbf{d})p(\mathbf{f}^-|\mathbf{d})$
$p(\mathbf{d}) = \prod_i p(d_i)$, so this can be done in $O(|\mathbf{d}|)$ times.
$p(\mathbf{f}^-|\mathbf{d}) = \prod_j p(f_j^-|\mathbf{d}) = \prod_j \mathrm{sigm}(\mathbf{w}_j^T \mathbf{d}_{\mathrm{parent}(f_j^-)})$, so this can be done in $O(|\mathbf{d}||\mathbf{f}^-|)$ times.
Hence the total time complexity is $O(|\mathbf{d}||\mathbf{f}^-|)$.

10.9.
$X \to Y \leftarrow Z$ moralizes to cycle of $X - Y - Z - X$. There is no addition in conditional independence statement sets by this operation.