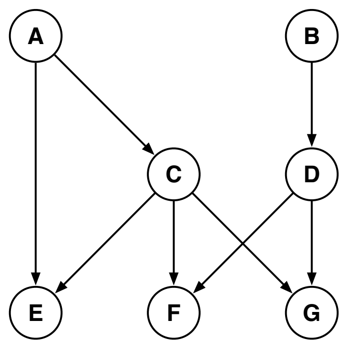
CSE 250A FINAL EXAM SOLUTIONS
FALL QUARTER 2020

1. d-separation (10 pts)



(a) $P(B|F) \stackrel{?}{=} P(B|E,F)$

False. For instance, the path $E \leftarrow C \rightarrow F \leftarrow D \leftarrow B$ is not blocked.

(b) $P(B,E) \stackrel{?}{=} P(B)\,P(E)$

True. There are four paths from node B to node E:
$E \leftarrow C \rightarrow G \leftarrow D \leftarrow B$,
$E \leftarrow A \rightarrow C \rightarrow G \leftarrow D \leftarrow B$,
$E \leftarrow C \rightarrow F \leftarrow D \leftarrow B$
$E \rightarrow A \leftarrow C \rightarrow F \leftarrow D \rightarrow B$.
The first two paths are blocked by node G, and the second two paths are blocked by node F.

(c) $P(A|B,F) \stackrel{?}{=} P(A|F)$

False. The path $A \rightarrow C \rightarrow F \leftarrow D \leftarrow B$ is not blocked.

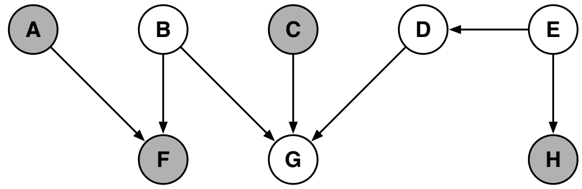
(d) $P(A,B|E) \stackrel{?}{=} P(A|E)\,P(B|E)$

True. There are four paths from node A to node B:
$A \rightarrow C \rightarrow G \leftarrow D \leftarrow B$,
$A \rightarrow E \rightarrow C \rightarrow G \leftarrow D \leftarrow B$,
$A \rightarrow C \rightarrow F \leftarrow D \leftarrow B$
$A \rightarrow E \leftarrow C \rightarrow F \leftarrow D \rightarrow B$.
The first two paths are blocked by node G, and the second two paths are blocked by node F.

(e) $P(C|A,E,F,G) \stackrel{?}{=} P(C|A,B,E,F,G)$

False. For instance, the path $C \rightarrow F \leftarrow D \leftarrow B$ is not blocked.

2. Polytree inference (10 pts)



(a) (2 pts)

$$\begin{aligned} P(E|H) &= \frac{P(E)\,P(H|E)}{P(H)} && \boxed{\textbf{Bayes rule}} \\ &= \frac{P(E)\,P(H|E)}{\sum_e P(E\!=\!e)\,P(H|E\!=\!e)} && \boxed{\textbf{normalization}} \end{aligned}$$

(b) (2 pts)

$$\begin{aligned} P(D|H) &= \sum_e P(D, E\!=\!e|H) && \boxed{\textbf{marginalization}} \\ &= \sum_e P(E\!=\!e|H)\,P(D|E\!=\!e, H) && \boxed{\textbf{product rule}} \\ &= \sum_e P(E\!=\!e|H)\,P(D|E\!=\!e) && \boxed{\textbf{conditional independence}} \end{aligned}$$

(c) (3 pts)

$$\begin{aligned} P(B|A,F) &= \frac{P(F|B,A)\,P(B|A)}{P(F|A)} && \boxed{\textbf{Bayes rule}} \\ &= \frac{P(F|B,A)\,P(B)}{P(F|A)} && \boxed{\textbf{marginal independence}} \\ &= \frac{P(F|B,A)\,P(B)}{\sum_b P(F|B\!=\!b, A)\,P(B\!=\!b)} && \boxed{\textbf{normalization}} \end{aligned}$$

(d) (3 pts)

$$\begin{aligned} &P(G|A,C,F,H) \\ &= \sum_{b,d} P(G, B\!=\!b, D\!=\!d|A,C,F,H) && \boxed{\textbf{marginalization}} \\ &= \sum_{b,d} P(B\!=\!b|A,C,F,H)\,P(D\!=\!d|A,B\!=\!b,C,F,H)\,P(G|B\!=\!b,D\!=\!d,A,C,F,H) && \boxed{\textbf{product rule}} \\ &= \sum_{b,d} P(B\!=\!b|A,F)\,P(D\!=\!d|H)\,P(G|B\!=\!b,C,D\!=\!d) && \boxed{\textbf{conditional independence}} \end{aligned}$$

3. Naive Bayes versus logistic regression (10 pts)

(a) Naive Bayes model (3 pts)

$$\begin{aligned} P(y|x_1,x_2,\ldots,x_d) &= \frac{P(y)\,P(x_1,x_2,\ldots,x_d|y)}{P(x_1,x_2,\ldots,x_d)} && \boxed{\textbf{Bayes rule}} \\ &= \frac{P(y)\,\prod_{i=1}^d P(x_i|y)}{P(x_1,x_2,\ldots,x_d)} && \boxed{\textbf{conditional independence}} \\ &= \frac{P(y)\,\prod_{i=1}^d P(x_i|y)}{\sum_{y'} P(Y\!=\!y')\,\prod_{j=1}^d P(x_j|Y\!=\!y')} && \boxed{\textbf{normalization}} \end{aligned}$$

(b) Log-odds (2 pts)

$$\begin{aligned} \log\frac{P(Y\!=\!1|x_1,\ldots,x_d)}{P(Y\!=\!0|x_1,\ldots,x_d)} &= \log\frac{P(Y\!=\!1)\,\prod_{i=1}^d P(x_i|Y\!=\!1)}{P(Y\!=\!0)\,\prod_{i=1}^d P(x_i|Y\!=\!0)} \\ &= \log\frac{P(Y\!=\!1)}{P(Y\!=\!0)} + \sum_{i=1}^d \log\frac{P(x_i|Y\!=\!1)}{P(x_i|Y\!=\!0)} \end{aligned}$$

(c) Linear decision boundary (3 pts)

A linear expression takes the form:

$$\log\frac{P(Y\!=\!1|x_1,x_2,\ldots,x_d)}{P(Y\!=\!0|x_1,x_2,\ldots,x_d)} = a_0 + \sum_{i=1}^d a_i x_i$$

From the hint:

$$\begin{aligned} \log\frac{P(x_i|Y\!=\!1)}{P(x_i|Y\!=\!0)} &= x_i \log\frac{P(X_i\!=\!1|Y\!=\!1)}{P(X_i\!=\!1|Y\!=\!0)} + (1-x_i)\log\frac{P(X_i\!=\!0|Y\!=\!1)}{P(X_i\!=\!0|Y\!=\!0)} \\ &= x_i\left[\log\frac{P(X_i\!=\!1|Y\!=\!1)}{P(X_i\!=\!0|Y\!=\!1)} - \log\frac{P(X_i\!=\!1|Y\!=\!0)}{P(X_i\!=\!0|Y\!=\!0)}\right] + \log\frac{P(X_i\!=\!0|Y\!=\!1)}{P(X_i\!=\!0|Y\!=\!0)} \end{aligned}$$

Matching terms to part (b):

$$\begin{aligned} a_0 &= \log\frac{P(Y\!=\!1)}{P(Y\!=\!0)} + \sum_{i=1}^d \log\frac{P(X_i\!=\!0|Y\!=\!1)}{P(X_i\!=\!0|Y\!=\!0)} \\ a_i &= \log\frac{P(X_i\!=\!1|Y\!=\!1)}{P(X_i\!=\!0|Y\!=\!1)} - \log\frac{P(X_i\!=\!1|Y\!=\!0)}{P(X_i\!=\!0|Y\!=\!0)} \qquad \text{for} \quad i = 1,2\ldots,d. \end{aligned}$$

(d) Logistic regression (2 pts)

The inverse of the sigmoid function is the log-odds. Therefore we can immediately identify $(w_0, w_1, \ldots, w_d) = (a_0, a_1, \ldots, a_d)$ in terms of the coefficients from part (c).

4. Nonnegative random variables (5 pts)

In this problem you will derive some elementary but useful properties of the exponential distribution $P(z) = (1/\mu)\, e^{-z/\mu}$ for a continuous, nonegative random variable with mean $\mu > 0$.

(a) Log-likelihood of i.i.d. data (1 pt)

$$\mathcal{L}(\mu) \;=\; \sum_{t=1}^{T} \log P(z_t) \;=\; \sum_{t=1}^{T}\left[-\log\mu - \frac{z_t}{\mu}\right] \;=\; -T\log\mu - \frac{1}{\mu}\sum_{t=1}^{T} z_t.$$

(b) Maximum likelihood estimation (1 pt)

$$0 \;=\; \frac{d\mathcal{L}}{d\mu} \;=\; -\frac{T}{\mu} + \frac{1}{\mu^2}\sum_{t=1}^{T} z_t \qquad\Longrightarrow\qquad \mu \;=\; \frac{1}{T}\sum_{t=1}^{T} z_t.$$

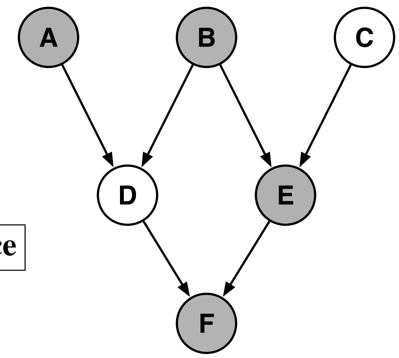
(c) Cumulative distribution (1 pt)

$$P(Z\!<\!a) \;=\; \int_0^a dz\, P(z) \;=\; \frac{1}{\mu}\int_0^a dz\, e^{-z/\mu} \;=\; 1 - e^{-a/\mu}.$$

(d) Comparison (2 pts)

$$\begin{aligned}
P(Z_1\!>\!Z_2) \;&=\; \int_0^\infty da\, P(Z_1\!=\!a)\, P(Z_2\!<\!a) \\
&=\; \frac{1}{\mu_1}\int_0^\infty da\, e^{-a/\mu_1}(1 - e^{-a/\mu_2}) \\
&=\; 1 \;-\; \frac{1}{\mu_1}\int_0^\infty da\, \exp\left(-a\left[\frac{1}{\mu_1} + \frac{1}{\mu_2}\right]\right) \\
&=\; 1 \;-\; \frac{1}{\mu_1}\left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right)^{-1} \\
&=\; 1 \;-\; \frac{1}{\mu_1}\left(\frac{\mu_1\mu_2}{\mu_1 + \mu_2}\right) \\
&=\; \frac{\mu_1}{\mu_1 + \mu_2}
\end{aligned}$$

5. EM algorithm (10 pts)



(a) Hidden node C (2 pts)

$$
\begin{aligned}
P(C|A,B,E,F) &= P(C|B,E) && \boxed{\textbf{conditional independence}}\\
&= \frac{P(E|C,B)\,P(C|B)}{P(E|B)} && \boxed{\textbf{Bayes rule}}\\
&= \frac{P(E|C,B)\,P(C)}{P(E|B)} && \boxed{\textbf{marginal independence}}\\
&= \frac{P(E|C,B)\,P(C)}{\sum_c P(E|C\!=\!c,B)\,P(C\!=\!c)} && \boxed{\textbf{normalization}}
\end{aligned}
$$

(b) Hidden node D (2 pts)

$$
\begin{aligned}
P(D|A,B,E,F) &= \frac{P(F|D,A,B,E)\,P(D|A,B,E)}{P(F|A,B,E)} && \boxed{\textbf{Bayes rule}}\\
&= \frac{P(F|D,E)\,P(D|A,B)}{P(F|A,B,E)} && \boxed{\textbf{conditional independence}}\\
&= \frac{P(F|D,E)\,P(D|A,B)}{\sum_d P(F|D\!=\!d,E)\,P(D\!=\!d|A,B)} && \boxed{\textbf{normalization}}
\end{aligned}
$$

(c) Both hidden nodes (1 pt)

$$
P(C,D|A,B,E,F) = P(C|A,B,E,F)\,P(D|A,B,E,F) \qquad \boxed{\textbf{conditional independence}}
$$

(d) **Log-likelihood** (2 pts)

$$
\begin{aligned}
\mathcal{L} &= \sum_t \log P(a_t,b_t,e_t,f_t)\\
&= \sum_t \log \sum_{cd} P(a_t,b_t,c,d,e_t,f_t) && \boxed{\textbf{marginalization}}\\
&= \sum_t \log \sum_{cd} P(a_t)\,P(b_t|a_t)\,P(c|a_t,b_t)\,P(d|a_t,b_t,c)\,P(e_t,|a_t,b_t,c,d)\,P(f_t|a_t,b_t,c,d,e_t) && \boxed{\textbf{PR}}\\
&= \sum_t \log \sum_{cd} P(a_t)\,P(b_t)\,P(c)\,P(d|a_t,b_t)\,P(e_t,|b_t,c)\,P(f_t|d,e_t) && \boxed{\textbf{conditional independence}}
\end{aligned}
$$

(e) EM algorithm (3 pts)

(i)
$$P(c) \leftarrow \frac{1}{T}\sum_t P(c|a_t, b_t, e_t, f_t)$$

or

$$P(c) \leftarrow \frac{1}{T}\sum_t P(c|b_t, e_t)$$

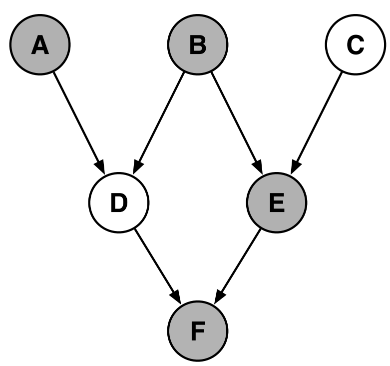
(ii)
$$P(d|a,b) \leftarrow \frac{\sum_t I(a,a_t)\, I(b,b_t)\, P(d|a_t,b_t,e_t,f_t)}{\sum_t I(a,a_t)\, I(b,b_t)}$$

(iii)
$$P(e|b,c) \leftarrow \frac{\sum_t I(b,b_t)\, P(c|a_t,b_t,e_t,f_t)\, I(e,e_t)}{\sum_t I(b,b_t)\, P(c|a_t,b_t,e_t,f_t)}$$

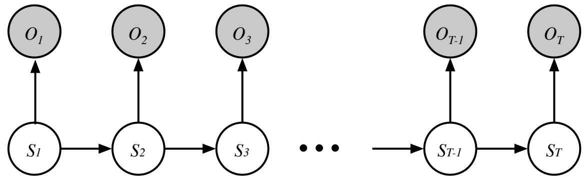
or

$$P(e|b,c) \leftarrow \frac{\sum_t I(b,b_t)\, P(c|b_t,e_t)\, I(e,e_t)}{\sum_t I(b,b_t)\, P(c|b_t,e_t)}$$

As noted in part (a), the red terms can be dropped due to conditional independence. However, either solution (with or without the red terms) is acceptable.



6. Inference in HMMs (8 pts)



(a) **Inference** (4 pts)

The claim is easiest to prove by induction. The base case $(t\!=\!1)$ is straightforward:

$$P(S_1\!=\!i) \;=\; \pi_i \;=\; \sum_{k=1} \pi_k I(k,i) \;=\; \sum_{k=1}^{n} \pi_k (a^0)_{ki}.$$

Then by induction we have:

$$\begin{aligned}
P(S_{t+1}\!=\!i) &= \sum_j P(S_t\!=\!j, S_{t+1}\!=\!i) && \boxed{\textbf{marginalization}} \\
&= \sum_j P(S_t\!=\!j)\, P(S_{t+1}\!=\!i|S_t\!=\!j) && \boxed{\textbf{product rule}} \\
&= \sum_{kj} \pi_k (a^{t-1})_{kj} a_{ji} && \boxed{\textbf{induction and substitution}} \\
&= \sum_k \pi_k (a^t)_{ki} && \boxed{\textbf{matrix multiplication}}
\end{aligned}$$

(b) **More inference** (4 pts)

$$P(o_1, o_2, \ldots, o_T | S_t\!=\!i, S_{t+1}\!=\!j) \;=\; \frac{P(o_1, o_2, \ldots, o_T, S_t\!=\!i, S_{t+1}\!=\!j)}{P(S_t\!=\!i, S_{t+1}\!=\!j)} \qquad \boxed{\textbf{product rule}}$$

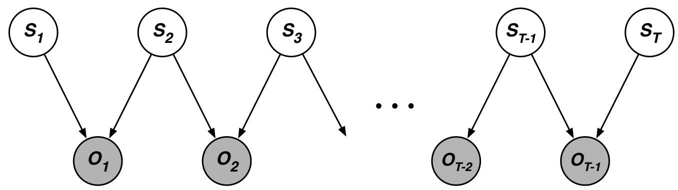
The denominator was essentially found in part (a):

$$P(S_t\!=\!i, S_{t+1}\!=\!j) \;=\; P(S_t\!=\!i)\, P(S_{t+1}\!=\!j|S_t\!=\!i) \;=\; \sum_k \pi_k (a^{t-1})_{ki} a_{ij}.$$

The numerator was computed in lecture:

$$\begin{aligned}
&P(o_1, o_2, \ldots, o_T, S_t\!=\!i, S_{t+1}\!=\!j) \\
&= P(o_1, \ldots, o_t, S_t\!=\!i) \times P(S_{t+1}\!=\!j|S_t\!=\!i, o_1, \ldots, o_t) \\
&\quad \times P(o_{t+1}|S_{t+1}\!=\!j, S_t\!=\!i, o_1, \ldots, o_t) \\
&\quad \times P(o_{t+2}, \ldots, o_T|S_{t+1}\!=\!j, S_t\!=\!i, o_1, \ldots, o_{t+1}) \qquad \boxed{\textbf{product rule}} \\
&= P(o_1, \ldots, o_t, S_t\!=\!i) \times P(S_{t+1}\!=\!j|S_t\!=\!i) \\
&\quad \times P(o_{t+1}|S_{t+1}\!=\!j) \times P(o_{t+2}, \ldots, o_T|S_{t+1}\!=\!j) \qquad \boxed{\textbf{conditional independence}} \\
&= \alpha_{it}\, a_{ij}\, b_j(o_{t+1})\, \beta_{j,t+1} \qquad \boxed{\textbf{substitution}}
\end{aligned}$$

7. Most likely hidden states (7 pts)



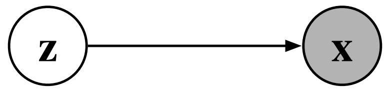
(a) Forward pass (4 pts)

$$
\begin{aligned}
(t\!=\!1) \qquad \ell^*_{i1} &= \log P(S_1\!=\!i) \\[1em]
(t\!>\!1) \qquad \ell^*_{j,t+1} &= \max_{s_1,\ldots,s_t} \log P(s_1,\ldots,s_t,S_{t+1}\!=\!j,o_1,\ldots,o_t) \\
&= \max_{s_1,\ldots,s_{t-1}} \max_i \log P(s_1,\ldots,S_t\!=\!i,S_{t+1}\!=\!j,o_1,\ldots,o_t) \\
&= \max_{s_1,\ldots,s_{t-1}} \max_i \log \Big[P(s_1,\ldots,S_t\!=\!i,o_1,\ldots,o_{t-1}) \times \\
&\qquad\qquad P(S_{t+1}\!=\!j|S_t\!=\!i,o_1,\ldots,o_{t-1}) \times \\
&\qquad\qquad P(o_t|S_t\!=\!i,S_{t+1}\!=\!j,o_1,\ldots,o_{t-1}) \Big] \quad \boxed{\textbf{product rule}} \\
&= \max_{s_1,\ldots,s_{t-1}} \max_i \log \Big[P(s_1,\ldots,S_t\!=\!i,o_1,\ldots,o_{t-1}) \times \\
&\qquad\qquad P(S_{t+1}\!=\!j) \times P(o_t|S_t\!=\!i,S_{t+1}\!=\!j) \Big] \quad \boxed{\textbf{marginal and conditional independence}} \\
&= \max_i \Big[\ell^*_{it} + \log P(o_t|S_t\!=\!i,S_{t+1}\!=\!j) \Big] + \log P(S_{t+1}\!=\!j)
\end{aligned}
$$

(b) Backward pass (3 pts)

$$
\begin{aligned}
(t\!=\!T) \qquad s^*_T &= \operatorname*{argmax}_i \left[\ell^*_{iT}\right] \\
(t\!<\!T) \qquad s^*_t &= \operatorname*{argmax}_i \left[\ell^*_{it} + \log P(o_t|S_t\!=\!i,S_{t+1}\!=\!s^*_{t+1})\right]
\end{aligned}
$$

8. Gaussian random variables (6 pts)



$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$ $\qquad\qquad$ $\mathbf{x} \sim \mathcal{N}(\mathbf{\Lambda z}, I)$

(a) Posterior mode (2 pts)

$$\begin{aligned}
\mathbf{z}^* &= \mathrm{argmax}_{\mathbf{z}} \Big[P(\mathbf{z}|\mathbf{x}) \Big] \\
&= \mathrm{argmax}_{\mathbf{z}} \left[\frac{P(\mathbf{z}), P(\mathbf{x}|\mathbf{z})}{P(\mathbf{x})} \right] \qquad \boxed{\textbf{Bayes rule}} \\
&= \mathrm{argmax}_{\mathbf{z}} \Big[P(\mathbf{z}, \mathbf{x}) \Big] \qquad \boxed{\textbf{denominator is independent of } \mathrm{z}} \\
&= \mathrm{argmax}_{\mathbf{z}} \log \Big[P(\mathbf{z})\, P(\mathbf{x}|\mathbf{z}) \Big] \qquad \boxed{\textbf{log is a monotonic function}}
\end{aligned}$$

(b) Maximization (3 pts)

Substituting the distributions:

$$\log \Big[P(\mathbf{z})\, P(\mathbf{x}|\mathbf{z}) \Big] = -\tfrac{1}{2}(d+D)\log(2\pi) - \tfrac{1}{2}\, \mathbf{z}^\top \mathbf{z} - \tfrac{1}{2}(\mathbf{x} - \mathbf{\Lambda z})^\top (\mathbf{x} - \mathbf{\Lambda z}).$$

Computing the gradient:

$$\frac{\partial}{\partial \mathbf{z}} \big[\log P(\mathbf{z}, \mathbf{x}) \big] \quad = \quad -\mathbf{z} \,+\, \mathbf{\Lambda}^\top (\mathbf{x} - \mathbf{\Lambda z}) \quad = \quad -(I \,+\, \mathbf{\Lambda}^\top \mathbf{\Lambda})\mathbf{z} \,+\, \mathbf{\Lambda}^\top \mathbf{x}.$$

Locating the maximum:

$$\frac{\partial}{\partial \mathbf{z}} \big[\log P(\mathbf{z}, \mathbf{x}) \big] \Big|_{\mathbf{z}^*} = 0 \quad \Longrightarrow \quad \boxed{\mathbf{z}^* \;=\; (I \,+\, \mathbf{\Lambda}^\top \mathbf{\Lambda})^{-1} \mathbf{\Lambda}^\top \mathbf{x}}$$

(c) Posterior mean (1 pt)

The posterior distribution $P(\mathbf{z}|\mathbf{x})$ is Gaussian because the joint distribution $P(\mathbf{z}, \mathbf{x})$ is Gaussian, and in *every* Gaussian distribution, the mean is equal to the mode.

Thus the statement is **true**: namely, $\mathrm{E}[\mathbf{z}|\mathbf{x}] = \arg\max_{\mathbf{z}} \big[P(\mathbf{z}|\mathbf{x}) \big]$.

9. Policy improvement (7 pts)

Consider the Markov decision process (MDP) with two states $s \in \{0, 1\}$, two actions $a \in \{\downarrow, \uparrow\}$, discount factor $\gamma = \frac{2}{3}$, and the reward function and transition matrices as shown below:

s	$R(s)$
0	-4
1	8

s	s'	$P(s' \vert s, a{=}\downarrow)$
0	0	$\frac{3}{4}$
0	1	$\frac{1}{4}$
1	0	$\frac{1}{4}$
1	1	$\frac{3}{4}$

s	s'	$P(s' \vert s, a{=}\uparrow)$
0	0	$\frac{1}{2}$
0	1	$\frac{1}{2}$
1	0	$\frac{1}{2}$
1	1	$\frac{1}{2}$

(a) State value function (4 pts)

From the Bellman equation:

$$\begin{aligned}
V_0 &= -4 + \tfrac{2}{3}\left(\tfrac{3}{4}V_0 + \tfrac{1}{4}V_1\right) \implies \tfrac{1}{2}V_0 - \tfrac{1}{6}V_1 = -4 \implies V_0 - \tfrac{1}{3}V_1 = -8 \\
V_1 &= 8 + \tfrac{2}{3}\left(\tfrac{1}{4}V_0 + \tfrac{3}{4}V_1\right) \implies \tfrac{1}{6}V_0 - \tfrac{1}{2}V_1 = -8 \implies V_0 - 3V_1 = -48
\end{aligned}$$

Solving this system gives $V_1 = 15$ and $V_0 = -3$.

(b) **Action value function** (2 pts)

$$\begin{aligned}
Q^\pi(0, \downarrow) &= V^\pi(0) = -3 \\
Q^\pi(0, \uparrow) &= -4 + \tfrac{2}{3}\left(\tfrac{1}{2}(-3) + \tfrac{1}{2}(15)\right) = 0
\end{aligned}$$

$$\begin{aligned}
Q^\pi(1, \downarrow) &= V^\pi(0) = 15 \\
Q^\pi(1, \uparrow) &= 8 + \tfrac{2}{3}\left(\tfrac{1}{2}(-3) + \tfrac{1}{2}(15)\right) = 12
\end{aligned}$$

(c) **Greedy policy** (1 pt)

$$\begin{aligned}
\pi'(0) &= \operatorname*{argmax}_a \left[Q^\pi(0, a)\right] = \uparrow \\
\pi'(1) &= \operatorname*{argmax}_a \left[Q^\pi(1, a)\right] = \downarrow
\end{aligned}$$

10. The simplest MDP (3 pts)

Theorem: If a Markov decision process (MDP) has a constant reward function (i.e., $R(s) = r$ for every state s of the state space $\mathcal{S}$), then every policy is optimal.

Proof: Consider the state value function $V^\pi(s)$ for any policy π of the MDP.

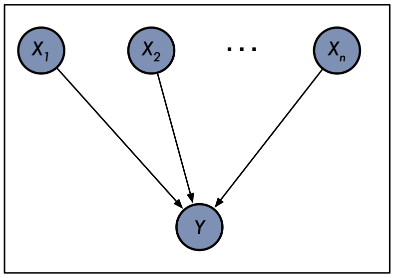
$$
\begin{aligned}
V^\pi(s) &= \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t) \,\middle|\, s_0 = s\right] \\
&= \mathrm{E}\left[\sum_{t=0}^{\infty} \gamma^t r \,\middle|\, s_0 = s\right] \qquad \boxed{\textbf{since the reward function is constant}} \\
&= \sum_{t=0}^{\infty} \gamma^t r \\
&= \frac{r}{1-\gamma}
\end{aligned}
$$

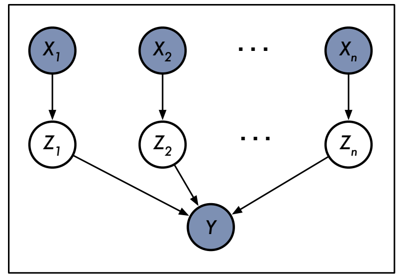
A policy π is optimal if for any policy $\tilde{\pi}$, we have $V^\pi(s) \geq V^{\tilde{\pi}}(s)$ for all $s \in \mathcal{S}$. In this MDP, all policies have the same (constant) state value function, so that

$$V^\pi(s) = V^{\tilde{\pi}}(s) = \frac{r}{1-\gamma}$$

for all $s \in \mathcal{S}$. Since π was an arbitrary policy, it follows that all policies are optimal.

11. Noisy parity model (9 pts)





(a) Noisy parity (2 pts)

$$\begin{aligned}
P(Y=1|x_1,x_2,\ldots,x_n) &= \frac{1}{2}\left[1-\prod_{i=1}^{n}(1-2p_i)^{x_i}\right] && \boxed{\textbf{general case}}\\
&= \frac{1}{2}\left[1-\prod_{i=1}^{n}(-1)^{x_i}\right] && \boxed{\textbf{setting all } p_i=1}\\
&= \frac{1}{2}\left[1-(-1)^{\sum_i x_i}\right]\\
&= \begin{cases} 1 & \text{if } \sum_i x_i \text{ is odd,}\\ 0 & \text{if } \sum_i x_i \text{ is even.}\end{cases} && \boxed{\textbf{parity}}
\end{aligned}$$

(b) Noisy copy (1 pt)

Suppose that $P(Z_i=0|X_i=0)=1$. Then we find:

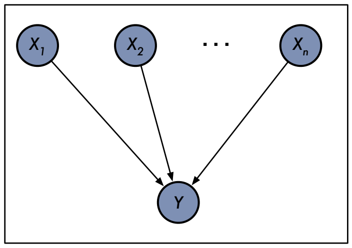
$$\begin{aligned}
(1-2p_i)^0 &= 1,\\
&= 1-0,\\
&= P(Z_i=0|X_i=0)-P(Z_i=1|X_i=0).
\end{aligned}$$

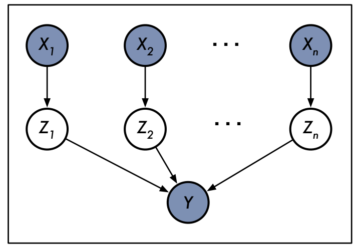
Likewise, suppose that $P(Z_i=1|X_i=1)=p_i$. Then we find:

$$\begin{aligned}
(1-2p_i)^1 &= 1-2p_i,\\
&= (1-p_i)-p_i,\\
&= P(Z_i=0|X_i=1)-P(Z_i=1|X_i=1).
\end{aligned}$$

Finally, combining these expressions, we find:

$$(1-2p_i)^{x_i} = P(Z_i=0|X_i=x_i)-P(Z_i=1|X_i=x_i).$$





(c*) Latent variable model (6 pts)

$$P(Y\!=\!1|x_1, x_2, \ldots, x_n)$$

$$= \sum_{z_1,\ldots,z_n} P(Y\!=\!1, z_1, z_2, \ldots, z_n|x_1, x_2, \ldots, x_n) \qquad \boxed{\textbf{marginalization}}$$

$$= \sum_{z_1,\ldots,z_n} P(Y\!=\!1|z_1, z_2, \ldots, z_n, x_1, x_2, \ldots, x_n)\, P(z_1, z_2, \ldots, z_n|x_1, x_2, \ldots, x_n) \quad \boxed{\textbf{product rule}}$$

$$= \sum_{z_1,\ldots,z_n} P(Y\!=\!1|z_1, z_2, \ldots, z_n)\, \prod_{i=1}^{n} P(z_1|x_i) \qquad \boxed{\textbf{conditional independence}}$$

$$= \sum_{z_1,\ldots,z_n} \frac{1}{2}\left[1-(-1)^{\sum_i z_i}\right] \prod_{i=1}^{n} P(z_1|x_i) \qquad \boxed{\textbf{parity}}$$

$$= \frac{1}{2} \sum_{z_1,\ldots,z_n} \prod_{i=1}^{n} P(z_1|x_i) \;-\; \frac{1}{2} \sum_{z_1,\ldots,z_n} \prod_{i=1}^{n} (-1)^{z_i} P(z_1|x_i)$$

$$= \frac{1}{2} \prod_{i=1}^{n} \sum_{z_i} P(z_1|x_i) \;-\; \frac{1}{2} \prod_{i=1}^{n} \sum_{z_i} (-1)^{z_i} P(z_1|x_i)$$

$$= \frac{1}{2} \prod_{i=1}^{n} (1) \;-\; \frac{1}{2} \prod_{i=1}^{n} \Big[P(Z_i\!=\!0|x_i) - P(Z_i\!=\!1|x_i)\Big]$$

$$= \frac{1}{2}\left[1 - \prod_{i=1}^{n} (1-2p_i)^{x_i}\right] \qquad \boxed{\textbf{from part (b)}}$$

12. Gamer rating engine (15 pts)

(a) Gradient ascent (3 pts)

Recall that $\frac{d\sigma}{dz} = \sigma(z)\sigma(-z)$, from which it follows that $\frac{d(\log\sigma)}{dz} = \sigma(-z)$. Thus we have:

$$\begin{aligned}
\mathcal{L} &= \sum_{i,j=1}^{n} G_{ij} \log\sigma(r_i - r_j) \\
\frac{\partial\mathcal{L}}{\partial r_k} &= \sum_{j=1}^{n} G_{kj}\sigma(r_j - r_k) \;-\; \sum_{i=1}^{n} G_{ik}\sigma(r_k - r_i)
\end{aligned}$$

Combining terms and renaming the indices:

$$\frac{\partial\mathcal{L}}{\partial r_k} = \sum_{j=1}^{n} \Big[G_{kj}\sigma(r_j - r_k) \;-\; G_{jk}\sigma(r_k - r_j) \Big].$$

Note how the positive terms in this partial derivative arise from games that the kth player wins, while the negative terms arise from games that the kth player loses.

(b) Self-check (2 pts)

Recall that $\sigma(z) + \sigma(-z) = 1$. Thus we can rewrite the partial derivative from part (a) as

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial r_k} &= \sum_{j=1}^{n} \Big[G_{kj}(1 - \sigma(r_k - r_j)) \;-\; G_{jk}\sigma(r_k - r_j) \Big] \\
&= \sum_{j=1}^{n} \Big[G_{kj} - (G_{kj} + G_{jk})\sigma(r_k - r_j) \Big].
\end{aligned}$$

This partial derivative vanishes when

$$\sum_{j=1}^{n} (G_{kj} + G_{jk})\sigma(r_k - r_j) \;=\; \sum_{j=1}^{n} G_{kj},$$

which is precisely the condition that the number of expected wins equals the number of observed wins.

(c) Nonnegative ratings (1 pt)

$$\sigma(r_i - r_j) \;=\; \frac{1}{1 + e^{r_j - r_i}} \;=\; \frac{e^{r_i}}{e^{r_i} + e^{r_j}} \;=\; \frac{\mu_i}{\mu_i + \mu_j}$$

(d) M-step (3 pts)

$$\begin{aligned}
Q(\boldsymbol{\mu}, \boldsymbol{\mu}_{\text{old}}) &= -\sum_{i,j=1}^{n} \left\{ G_{ij}\left(\log \mu_i + \frac{\mathrm{E}[Z_i|V_{ij}\!=\!1]}{\mu_i}\right) + G_{ji}\left(\log \mu_i + \frac{\mathrm{E}[Z_i|V_{ij}\!=\!0]}{\mu_i}\right)\right\} \\
\frac{\partial Q}{\partial \mu_i} &= -\sum_{j=1}^{n} \left\{ G_{ij}\left(\frac{1}{\mu_i} - \frac{\mathrm{E}[Z_i|V_{ij}\!=\!1]}{\mu_i^2}\right) + G_{ji}\left(\frac{1}{\mu_i} - \frac{\mathrm{E}[Z_i|V_{ij}\!=\!0]}{\mu_i^2}\right)\right\} \\
&= -\frac{1}{\mu_i}\sum_{j=1}^{n}(G_{ij}+G_{ji}) + \frac{1}{\mu_i^2}\sum_{j=1}^{n}\big(G_{ij}\mathrm{E}[Z_i|V_{ij}\!=\!1] + G_{ji}\mathrm{E}[Z_i|V_{ij}\!=\!0]\big).
\end{aligned}$$

The maximum occurs where the partial derivative vanishes. This yields the update

$$\mu_i \longleftarrow \frac{\sum_{j=1}^{n}\big(G_{ij}\mathrm{E}[Z_i|V_{ij}\!=\!1] + G_{ji}\mathrm{E}[Z_i|V_{ij}\!=\!0]\big)}{\sum_{j=1}^{n}(G_{ij}+G_{ji})}.$$

This update can be understood as an intuitive generalization of the ML estimate from problem 4(b). In particular, note how expected values take the place of observed values; in addition, these expected values are weighted (as appropriate) by the number of wins or losses.

(e) Posterior mean (when the player wins) (3 pts)

$$\begin{aligned}
\mathrm{E}[Z_i|V_{ij}\!=\!1] &= \frac{1}{P(Z_i\!>\!Z_j)}\int_0^\infty dz\; P(Z_i\!=\!z)\,P(Z_j\!<\!z)\,z \\
&= \frac{\mu_i+\mu_j}{\mu_i}\int_0^\infty dz\;\frac{1}{\mu_i}e^{-z/\mu_i}\left(1-e^{-z/\mu_j}\right)z \\
&= \frac{\mu_i+\mu_j}{\mu_i}\cdot\frac{1}{\mu_i}\cdot\left[\mu_i^2 - \left(\frac{1}{\mu_i}+\frac{1}{\mu_j}\right)^{-2}\right] \\
&= \frac{\mu_i+\mu_j}{\mu_i}\cdot\frac{1}{\mu_i}\cdot\left[\mu_i^2 - \left(\frac{\mu_i\mu_j}{\mu_i+\mu_j}\right)^{2}\right] \\
&= \mu_i+\mu_j - \frac{\mu_j^2}{\mu_i+\mu_j} \\
&= \mu_i\left[1+\frac{\mu_j}{\mu_i+\mu_j}\right]
\end{aligned}$$

Note that $\mathrm{E}[Z_i|V_{ij}\!=\!1] > \mathrm{E}[Z_i]$, as expected, since the posterior mean reflects a win.

(f) Posterior mean (when the player loses) (3 pts)

$$
\begin{aligned}
&P(V_{ij}\!=\!0)\,\mathrm{E}[Z_i|V_{ij}\!=\!0] \;+\; P(V_{ij}\!=\!1)\,\mathrm{E}[Z_i|V_{ij}\!=\!1] \\
&\quad= \; P(V_{ij}\!=\!0)\int_0^\infty dz\, z\, P(Z_i\!=\!z|V_{ij}\!=\!0) \;+\; P(V_{ij}\!=\!1)\int_0^\infty dz\, z\, P(Z_i\!=\!z|V_{ij}\!=\!1) \\
&\quad= \; \int_0^\infty dz\, z\, P(Z_i\!=\!z, V_{ij}\!=\!0) \;+\; \int_0^\infty dz\, z\, P(Z_i\!=\!z, V_{ij}\!=\!1) \qquad \boxed{\textbf{product rule}} \\
&\quad= \; \int_0^\infty dz\, z\left[P(Z_i\!=\!z, V_{ij}\!=\!0) \;+\; P(Z_i\!=\!z, V_{ij}\!=\!1)\right] \\
&\quad= \; \int_0^\infty dz\, z\, P(Z_i\!=\!z) \qquad \boxed{\textbf{marginalization}} \\
&\quad= \; \mathrm{E}[Z_i]
\end{aligned}
$$

Therefore we can easily compute $\mathrm{E}[Z_i|V_{ij}\!=\!0]$ in terms of our previous results.