10 pts each for a total of 30 pts

9.9

(Proof). We have to show that the following equation is true:

\[ P(X = x'|e) = \sum_{x \in \text{Val}(X)} P(X = x|e)T(x \rightarrow x') \]

where \( P \) factorizes according to \( G \), \( X = \chi - E = \{X_1, X_2, ..., X_k\} \) and \( T \) is a multiple transition chain with a local transition model \( T_i \) for each \( X_i \in X \) defined as \( T_i((u_i, x_i), (u_i, x_i')) = P(x_i'|u_i) \) where \( u_i \) denotes a instantiation of \( U_i = \chi - \{X_i\} \) and \( X_1, X_2, ..., X_k \) is an ordering of \( G - E \). Thus we can write \( T(x \rightarrow x') \) as:

\[
T(x \rightarrow x') = T_1((x_1, x_2, ..., x_k, e), (x_1', x_2, ..., x_k, e)) * \\
T_2((x_1', x_2, ..., x_k, e), (x_1', x_2', ..., x_k, e)) * \\
\vdots \\
T_k((x_1', x_2', ..., x_k, e), (x_1', x_2', ..., x_k', e)) = P(x_1'|x_2, ..., x_k, e)P(x_2'|x_1', x_3, ..., x_k, e) * \\
P(x_k'|x_1', ..., x_k'-1, e) = P(x_1'|e)P(x_2'|x_1'|e)P(x_3'|x_2'|x_1'|e) * \\
P(x_k'|x_1', ..., x_k'-1, e) = P(X = x'|e)
\]

where the second to last equality follows from the fact that the variables \( X_i \) are ordered according to \( G \). Finally plugging this into equation ?? we get:

\[
\sum_{x \in \text{Val}(X)} P(X = x|e)T(x \rightarrow x') = \\
= \sum_{x \in \text{Val}(X)} P(X = x|e)P(X = x'|e) \\
= P(X = x'|e) \sum_{x \in \text{Val}(X)} P(X = x|e)
\]
= P(X = x'|e)

10.1

\[
\frac{\partial}{\partial Q(x)} H_Q(X) = \frac{\partial}{\partial Q(x)} - E_Q[\text{lg}(Q)] \\
= - \frac{\partial}{\partial Q(x)} \sum_{y \in X} Q(y) \text{ lg}(Q(y)) \\
= - \frac{\partial}{\partial Q(x)} Q(x) \text{ lg}(Q(x)) \\
= - \frac{Q(x)}{Q(x)} - \text{lg}(Q(x)) \\
= -1 - \text{lg}(Q(x))
\]
\[ l(\theta|D) = \prod_{k}^{m} \theta_{k}^{M_{k}} \]
\[ L(\theta|D) = \log \prod_{k}^{m} \theta_{k}^{M_{k}} \]
\[ = \sum_{k}^{m} M_{k} \log \theta_{k} \]

We want to find the parameters that maximize the previous likelihood.

\[ \theta = \arg \max_{\theta'} L(\theta'|D) \]

subject to the condition that \( \sum_{k} M_{k} \theta_{k} = 1 \). We solve this using a Lagrangian \( \lambda \):

\[ \nabla_{\theta} \sum_{k}^{m} M_{k} \log \theta_{k} - \lambda (\sum_{k}^{m} \theta_{k} - 1) = 0 \]
\[ \nabla_{\theta} \sum_{k}^{m} M_{k} \log \theta_{k} = \lambda \nabla_{\theta} (\sum_{k}^{m} \theta_{k} - 1) \]
\[ (M_{1}/\theta_{1}, ..., M_{m}/\theta_{m}) = \lambda (1, ..., 1) \]
\[ \theta_{k} = M_{k}/\lambda \]

where \( \lambda = \sum M_{k} = N \) because

\[ \sum_{k} \theta_{k} = 1 \]
\[ \sum_{k} M_{k}/\lambda = 1 \]
\[ \sum M_{k} = \lambda \]

\[ \theta_{k} = M_{k}/N \]