1. The *conditional entropy* $H(X \mid Y)$ is defined as the average entropy in the conditional distribution

$$H(X \mid Y) = \mathbb{E}[\mathbb{E}\{-\log f(X \mid Y) \mid Y\}] = \sum_y \left\{ \sum_x -f(x \mid y) \log f(x \mid y) \right\} f(y).$$

   (a) Use the information inequality to show that

$$H(X \mid Y) \leq H(X),$$

   i.e. the *entropy is always reduced by conditoning*

   (b) Show that

$$H(X, Y) = H(X \mid Y) + H(Y).$$

   (c) For three discrete random variables, show that

$$H(X, Y, Z) + H(Z) \leq H(X, Z) + H(Y, Z).$$

   (d) Show further that

$$X \perp\!\!\!\perp Y \mid Z \iff H(X, Y, Z) + H(Z) = H(X, Z) + H(Y, Z).$$

2. Consider a directed graph $\mathcal{D} = (V, E)$ and assume given $k_v, v \in V$ with $k_v \geq 0$ and $\sum_{x_v \in \mathcal{X}_v} k_v(x_v \mid x_{\mathrm{pa}(v)}) = 1$. Define

$$p(x) = \prod_{v \in V} k_v(x_v \mid x_{\mathrm{pa}(v)}).$$

   (a) Show that when $\mathcal{D}$ is acyclic, i.e. a DAG, this yields a well-defined probability distribution;

   (b) Show that it holds that

$$p(x_v \mid x_{\mathrm{pa}(v)}) = k_v(x_v \mid x_{\mathrm{pa}(v)}). \tag{1}$$

   (c) Give a counterexample in the case where $\mathcal{D}$ has directed cycles.

   Hint: Use induction for (a) and (b), exploiting that a DAG always has a terminal vertex v_0, i.e. a vertex with no children.

3. Consider a DAG $\mathcal{D}$ with arrows $1 \to 2, 2 \to 5, 2 \to 3, 5 \to 6, 4 \to 5, 4 \to 7, 5 \to 7$.

   (a) Draw the DAG;

   (b) List all conditional independence relations corresponding to the local, directed Markov property;

   (c) List all conditional independence relations corresponding to the ordered Markov property for the well-ordering induced by the given numbering;

   (d) Find the ancestral sets generated by the following subsets:

      i. $\{5\}$;

      ii. $\{2, 7\}$;

      iii. $\{4, 6\}$;

   (e) Which of the following separation statements are true? For those that are not true, identify an active trail.

      i. $2 \perp_{\mathcal{D}} 4 \mid 5$;

      ii. $2 \perp_{\mathcal{D}} 7 \mid 5$,

      iii. $1 \perp_{\mathcal{D}} 7 \mid 5, 6$;

      iv. $1 \perp_{\mathcal{D}} 4 \mid 6$;

4. Let P be a distribution which factorizes over the DAG $\mathcal{D}$ and let $G(P)$ be its dependence graph. Show that $G(P) \subseteq \mathcal{D}^m$, where $\mathcal{D}^m$ is the moral graph of $\mathcal{D}$.

5. A DAG $\mathcal{D}$ is said to be *perfect* if all parents are married, i.e. if it holds that

$$\alpha, \beta \in \mathrm{pa}(\gamma) \Rightarrow \alpha \to \beta \text{ or } \beta \to \alpha.$$

   (a) Show that a perfect DAG $\mathcal{D}$ is Markov equivalent to its *skeleton* i.e. the undirected graph obtained by ignoring directions on all arrows;

   (b) Show the converse, i.e. that if $\mathcal{D}$ is Markov equivalent to its skeleton, the $\mathcal{D}$ is perfect.

   (c) Show that the skeleton of a perfect $\mathcal{D}$ is chordal.

   (d) Show that the edges of a chordal graph $\mathcal{G}$ can be directed to create a Markov equivalent DAG.

   Hint: Exploit the existence of a perfect numbering of a chordal graph.