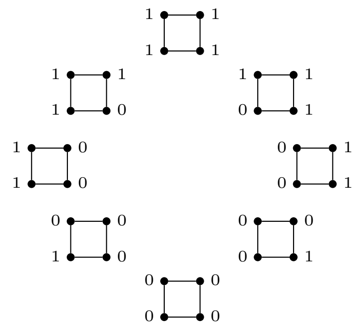
1. Consider the uniform distribution on the 8 configurations displayed in the figure below:



Show that this distribution satisfies (G) but the distribution does not factorize, i.e., it does not satisfy (F).

Conditioning on any pair of opposite corners renders one corner deterministic and therefore the global Markov property is satisfied.

However, the density does not factorize. To see this we assume the density factorizes. Then e.g.

$$0 \neq 1/8 = f(0,0,0,0) = \psi_{12}(0,0)\psi_{23}(0,0)\psi_{34}(0,0)\psi_{41}(0,0)$$

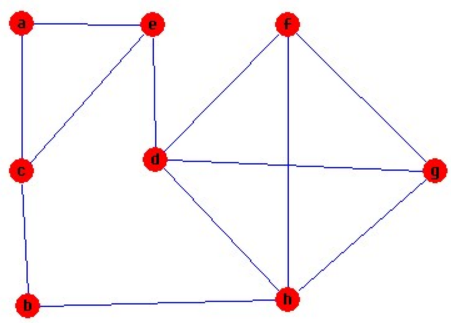
so these factors are all positive. Continuing for all possible 8 configurations yields that all factors $\psi_a(x)$ are strictly positive, since all four possible configurations are possible for every clique.

But this contradicts the fact that only 8 out of 16 possible configurations have positive probability.

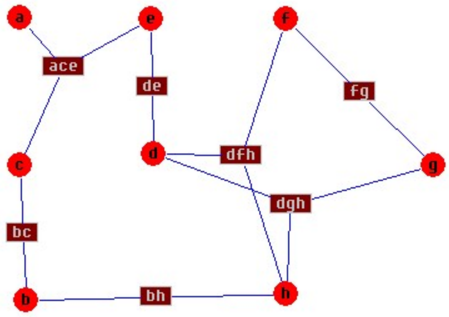
2. Consider the following generating class over $V = \{a, b, c, d, e, f, g, h\}$:

$$\mathcal{A} = \{\{a, c, e\}, \{b, c\}, \{b, h\}, \{d, e\}, \{d, f, h\}, \{d, g, h\}, \{f, g\}\}.$$

(a) Find the dependence graph $G(\mathcal{A})$;



(b) Find the factor graph $F(\mathcal{A})$;



(c) Find the cliques of $G(\mathcal{A})$;

The cliques of $G(\mathcal{A})$ are

$$\mathcal{C} = \{\{a, c, e\}, \{b, c\}, \{b, h\}, \{d, e\}, \{d, f, g, h\}\}.$$

(d) Show $\mathcal{A}$ is not conformal;

$\mathcal{A}$ is not conformal because $\{d, f, g, h\}$ is a clique which is not in $\mathcal{A}$.

(e) which of the following three statements are implied by $\mathcal{A}$?

$$\{a, b, c\} \perp\!\!\!\perp \{d, f, g\} \,|\, \{e, h\}, \quad \{a, b, c\} \perp\!\!\!\perp \{d, f, g, h\} \,|\, \{e\}, \quad d \perp\!\!\!\perp h \,|\, \{f, g\}.$$

The global Markov property yields that the conditional independence statement

$$\{a, b, c\} \perp\!\!\!\perp \{d, f, g\} \,|\, \{e, h\}$$

is implied by $\mathcal{A}$, but not any of the other two.

3. Prove the *information inequality:* For non-negative numbers $a(x)_{x\in\mathcal{X}}$ and $b(x)_{x\in\mathcal{X}}$ with $\sum_x a(x) = \sum_x b(x)$ it holds that

$$\sum_{x\in\mathcal{X}} a(x) \log b(x) \leq \sum_{x\in\mathcal{X}} a(x) \log a(x)$$

where the inequality is strict unless $a(x) = b(x)$ for all $x \in \mathcal{X}$. For the expressions to make sense we use the convention that $0 \log 0 = 0$.

We first show $\log y \leq y - 1$. Taylor's theorem yields

$$f(y) = f(1) + f'(1)(y-1) + f''(y^*)(y-1)^2/2$$

with y^* between y and 1,

For $f(y) = \log y$ we then have

$$\log y = 0 + (y-1) - \frac{(y-1)^2}{(y^*)^2} \leq y - 1$$

with equality if and only if $y = 1$.

Thus

$$\sum_{x\in\mathcal{X}} a(x) \log b(x) - \sum_{x\in\mathcal{X}} a(x) \log a(x) \quad = \quad \sum_{x:a(x)>0} a(x) \log b(x) - \sum_{x:a(x)>0} a(x) \log a(x)$$

$$
\begin{aligned}
&= \sum_{x:a(x)>0} a(x) \log \frac{b(x)}{a(x)} \\
&\leq \sum_{x:a(x)>0} a(x) \left\{ \frac{b(x)}{a(x)} - 1 \right\} \\
&= \sum_{x:a(x)>0} b(x) - \sum_{x:a(x)>0} a(x) \\
&\leq \sum_{x\in\mathcal{X}} b(x) - \sum_{x\in\mathcal{X}} a(x) = 0.
\end{aligned}
$$

The inequality is strict unless $a(x)/b(x) = 1$ for all x, i.e. $a(x) = b(x)$.

4. The *entropy* $H(X)$ of a discrete random variable X is

$$H(X) = \mathbf{E}\{-\log f(X)\} = \sum_{x\in\mathcal{X}} -f(x)\log f(x).$$

   (a) Show that $H(X) \geq 0$;
   Since $f(x) \leq 1$, $-\log f(x) \geq 0$, hence also its expectation;

   (b) Show that the entropy of the uniform distribution with $f(x) = 1/|\mathcal{X}|$ is $H(X) = \log|\mathcal{X}|$.

$$H(X) = \sum_{x\in\mathcal{X}} -\frac{1}{|\mathcal{X}|}\log\frac{1}{|\mathcal{X}|} = \sum_{x\in\mathcal{X}} \frac{1}{|\mathcal{X}|}\log|\mathcal{X}| = \log|\mathcal{X}|.$$

   (c) Show that the uniform distribution has maximal entropy, i.e. that

$$H(X) \leq \log|\mathcal{X}|.$$

   Using the information inequality with $a(x) = f(x)$ and $b(x) = 1/|\mathcal{X}|$ we get

$$-H(X) = \sum_{x\in\mathcal{X}} f(x)\log f(x) \geq \sum_{x\in\mathcal{X}} f(x)\log\frac{1}{|\mathcal{X}|} = -\log|\mathcal{X}|,$$

   and hence $H(X) \leq \log|\mathcal{X}|$.

5. Consider a three-way contingency table with variables A, B, C and cell counts

		C		
A	B	0	1	total
1	1	3	5	8
1	0	12	10	22
0	1	10	10	20
0	0	1	4	5

Perform one cycle of the IPS algorithm for the hierarchical log-linear model with generator $\{\{A,B\},\{B,C\},\{A,C\}\}$.

Initial uniform distribution $np_{abc} = 55/8 = 6.875$.

		C		
A	B	0	1	total
1	1	6.875	6.875	13.75
1	0	6.875	6.875	13.75
0	1	6.875	6.875	13.75
0	0	6.875	6.875	13.75

First scaling, with AB-marginal:

A	B	C = 0	C = 1	total
1	1	4	4	8
1	0	11	11	22
0	1	10	10	20
0	0	2.5	2.5	5

To scale with the BC-marginal, we need the observed and fitted BC-marginals:

B	C = 0	C = 1
1	13	15
0	13	14

B	C = 0	C = 1
1	14	14
0	13.5	13.5

Second scaling, with BC-marginal, thus yields:

A	B	C = 0	C = 1	total
1	1	3.71	4.29	8
1	0	10.59	11.41	22
0	1	9.29	10.71	20
0	0	2.41	2.59	5

Note that the AB-marginal still fits.

To scale with the AC-marginal, we need the observed and fitted BC-marginals:

A	C = 0	C = 1
1	15	15
0	11	14

A	C = 0	C = 1
1	14.30	15.70
0	11.70	13.30

And we are ready for scaling with the AC-marginal:

A	B	C = 0	C = 1	total
1	1	3.892	4.099	7.991
1	0	11.108	10.901	22.009
0	1	8.734	11.274	20.008
0	0	2.266	2.726	4.992

The AB-marginal does not fit anymore, but almost so.

If iteration is continued until convergence, the final fitted values are (rounded):

A	B	C = 0	C = 1	total
1	1	4.12	3.88	8
1	0	10.89	11.11	22
0	1	8.87	11.13	20
0	0	2.15	2.85	5

6. Consider a generating class $\mathcal{A} = \{a, b\}$ with only two elements. Show that the IPS algorithm converges after a single cycle.

   After fitting the a-marginal we have

$$p(x) = \frac{1}{|\mathcal{X}|} \frac{n(x_a)}{n|\mathcal{X}_a|/|\mathcal{X}|} = \frac{n(x_a)}{n|\mathcal{X}_a|}.$$

   Fitting now the b-marginal we get

$$p(x) = p(x) \frac{n(x_b)}{np(x_b)} = \frac{n(x_a)}{n|\mathcal{X}_a|} \frac{n(x_b)}{n(x_{a\cap b})/|\mathcal{X}_a|} = \frac{n(x_a)\, n(x_b)}{n\, n(x_{a\cap b})}.$$

   The marginals now fit. See overheads for Lecture 4.