3.1.4 Another Hint: Reverse KL minimization

You are instructed to use $q(x, y) = q(x)q(y)$ approximation. Therefore,

$$\text{KL}(q(x, y)||p(x, y)) = \sum_{x,y} q(x, y) \log \frac{q(x, y)}{p(x, y)} = \sum_{x,y} q(x)q(y) \log \frac{q(x)q(y)}{p(x, y)}$$

$$\text{KL}(q(x, y)||p(x, y)) = \mathbb{E}_{qx}[\log q(x)] + \mathbb{E}_{qy}[\log q(y)] - \mathbb{E}_{qx}\mathbb{E}_{qy}[\log p(x, y)]$$

where we have used $\sum_y q(y) = \sum_x q(x) = 1$.

For the KLD in above equation to be minimized ($= 0$), the following equality should hold:

$$\mathbb{E}_{qx}[\log q(x)] + \mathbb{E}_{qy}[\log q(y)] = \mathbb{E}_{qx}\mathbb{E}_{qy}[\log p(x, y)]$$

Using a standard factorisation of the joint distribution $p(x, y) = p(x|y)p(y)$ we can rewrite the above equation as:

$$\mathbb{E}_{qx}[\log q(x)] + \mathbb{E}_{qy}[\log q(y)] = \mathbb{E}_{qx}[\log p(x|y)] + \mathbb{E}_{qy}[\log p(y)]$$

By inspecting the above equation, we observe that the terms inside the expectations must correspond to each other, i.e., one valid solution can be: $q(y) = p(y)$ and $q(x) = p(x|y)$.

Of course, one could claim this at the outset, by noticing that $q(y) = p(y)$ and $q(x) = p(x|y)$ results in KLD to be zero as the ratio inside the log will be 1.

From the joint probability table:

$p(y) = [1/4, 1/4, 1/4, 1/4]$

We can work out conditional probability distributions for each value of y. For example:

$p(x|y = 1) = [1/2, 1/2, 0, 0]$

For these to be valid factorisations, we can initialize $q(y) = p(y)$ and $q(x) = p(x|y = 1)$ and optimize them such that $q(x)q(y) = 0$ wherever $p(x, y) = 0$.

	1	2	3	4
1	1/8	1/8	1/8	1/8
2	1/8	1/8	1/8	1/8
3	0	0	0	0
4	0	0	0	0

Table 2: Joint distribution for $q(x)q(y)$ without constraint.

By forcing entries in $q(x)q(y)$ to zero where $p(x, y) = 0$ and renormalizing it to obtain a valid joint distribution we obtain:

	1	2	3	4
1	1/4	1/4	0	0
2	1/4	1/4	0	0
3	0	0	0	0
4	0	0	0	0

Table 3: Joint distribution for $q(x)q(y)$ constrained to be zero wherever p(x,y) is zero.

The marginal densities can be obtained from it as $q(x) = [1/2, 1/2, 0, 0]$ and $q(y) = [1/2, 1/2, 0, 0]$.

Using intializations from the conditional distributions for other cases will result in the other minima.