Exercise: Forwards vs reverse KL divergence

(Source: Exercise 33.7 of (**?**).) Consider a factored approximation $q(x, y) = q(x)q(y)$ to a joint distribution $p(x, y)$. Show that to minimize the forwards KL $\mathbb{KL}(p||q)$ we should set $q(x) = p(x)$ and $q(y) = p(y)$, i.e., the optimal approximation is a product of marginals

Now consider the following joint distribution, where the rows represent y and the columns x.

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

Show that the reverse KL $\mathbb{KL}(q||p)$ for this p has three distinct minima. Identify those minima and evaluate $\mathbb{KL}(q||p)$ at each of them. What is the value of $\mathbb{KL}(q||p)$ if we set $q(x, y) = p(x)p(y)$?