The definition of conditional relative entropy is:

$$D(p(y/x) \| q(y/x)) \equiv \sum_x p(x) \sum_y p(y/x) \log\left(\frac{p(y/x)}{q(y/x)}\right)$$

Note that you might consider this the average of the relative entropy between $p(y/x)$ and $q(y/x)$. Since it is an average, we must weight the measure by the probability distribution of x (hence, infrequently occuring events will drive the log negative, but also scale a contribution towards zero). The measure will be dominated by values of x that occur frequently and for which $p(y/x) > q(y/x)$.

In practice, when computing any relative entropy, one must be extremely concerned about pdf's that achieve one or more zero values. This commonly occurs when dealing with small data sets. There are a variety of interpolation techniques designed to deal with this situation.

The relative entropy between two joint distributions is defined as:

$$D(p(x, y) \| q(x, y)) = \sum_x \sum_y p(x, y) \log\left(\frac{p(x, y)}{q(x, y)}\right)$$

This is just the expectation of the log, as we saw before for the one variable case. Noting that $p(y/x) = \frac{p(x, y)}{p(x)}$, we can show that the relative entropy between two joint distributions can be expressed as:

$$D(p(x, y) \| q(x, y)) = D(p(x) \| q(x)) + D(p(y/x) \| q(y/x))$$

Hence, the distance between the joint distributions is larger than the distance between the marginals. Only when x is independent of y are they equal.