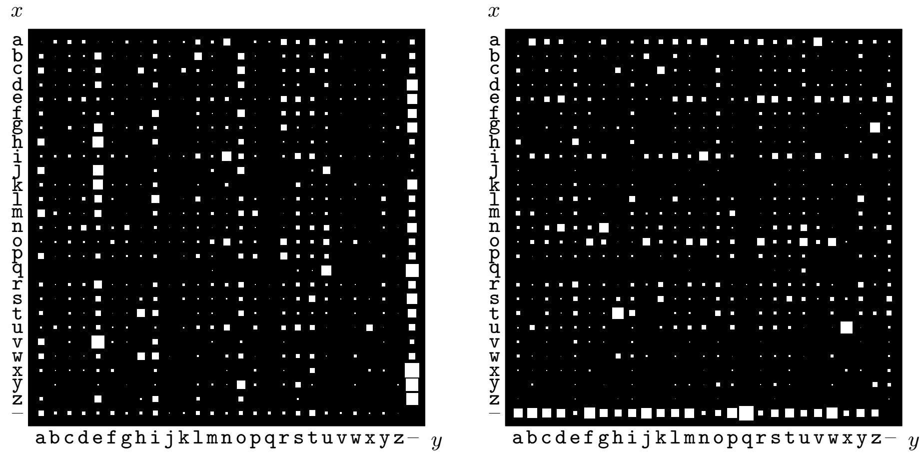

(a) $P(y \mid x)$ (b) $P(x \mid y)$

Figure 2.3. Conditional probability distributions. (a) $P(y \mid x)$: Each *row* shows the conditional distribution of the second letter, y, given the first letter, x, in a bigram xy. (b) $P(x \mid y)$: Each *column* shows the conditional distribution of the first letter, x, given the second letter, y.

that the first letter x is `q` are `u` and `-`. (The space is common after `q` because the source document makes heavy use of the word `FAQ`.)

The probability $P(x \mid y = \mathtt{u})$ is the probability distribution of the first letter x given that the second letter y is a `u`. As you can see in figure 2.3b the two most probable values for x given $y = \mathtt{u}$ are `n` and `o`.

Rather than writing down the joint probability directly, we often define an ensemble in terms of a collection of conditional probabilities. The following rules of probability theory will be useful. ($\mathcal{H}$ denotes assumptions on which the probabilities are based.)

Product rule – obtained from the definition of conditional probability:

$$P(x, y \mid \mathcal{H}) = P(x \mid y, \mathcal{H})P(y \mid \mathcal{H}) = P(y \mid x, \mathcal{H})P(x \mid \mathcal{H}). \tag{2.6}$$

This rule is also known as the chain rule.

Sum rule – a rewriting of the marginal probability definition:

$$\begin{aligned} P(x \mid \mathcal{H}) &= \sum_y P(x, y \mid \mathcal{H}) && (2.7) \\ &= \sum_y P(x \mid y, \mathcal{H})P(y \mid \mathcal{H}). && (2.8) \end{aligned}$$

Bayes' theorem – obtained from the product rule:

$$\begin{aligned} P(y \mid x, \mathcal{H}) &= \frac{P(x \mid y, \mathcal{H})P(y \mid \mathcal{H})}{P(x \mid \mathcal{H})} && (2.9) \\ &= \frac{P(x \mid y, \mathcal{H})P(y \mid \mathcal{H})}{\sum_{y'} P(x \mid y', \mathcal{H})P(y' \mid \mathcal{H})}. && (2.10) \end{aligned}$$

Independence. Two random variables X and Y are *independent* (sometimes written $X \perp Y$) if and only if

$$P(x, y) = P(x)P(y). \tag{2.11}$$

Exercise 2.2.[*1*, p.40] Are the random variables X and Y in the joint ensemble of figure 2.2 independent?