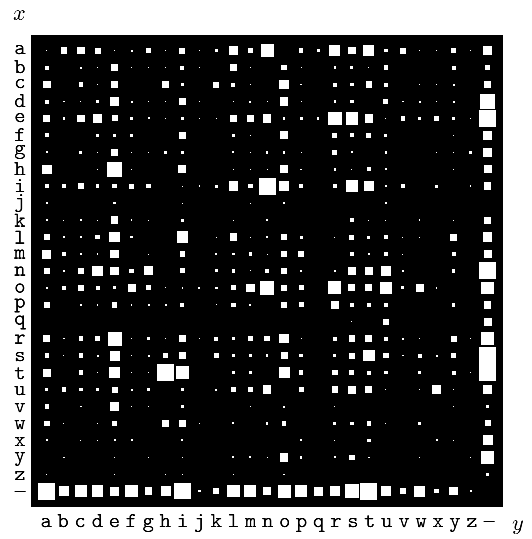

Figure 2.2. The probability distribution over the 27×27 possible bigrams xy in an English language document, *The Frequently Asked Questions Manual for Linux.*

Marginal probability. We can obtain the marginal probability $P(x)$ from the joint probability $P(x, y)$ by summation:

$$P(x = a_i) \equiv \sum_{y \in \mathcal{A}_Y} P(x = a_i, y). \tag{2.3}$$

Similarly, using briefer notation, the marginal probability of y is:

$$P(y) \equiv \sum_{x \in \mathcal{A}_X} P(x, y). \tag{2.4}$$

Conditional probability

$$P(x = a_i \,|\, y = b_j) \equiv \frac{P(x = a_i, y = b_j)}{P(y = b_j)} \;\text{ if }\; P(y = b_j) \neq 0. \tag{2.5}$$

[If $P(y = b_j) = 0$ then $P(x = a_i \,|\, y = b_j)$ is undefined.]

We pronounce $P(x = a_i \,|\, y = b_j)$ 'the probability that x equals a_i, given y equals b_j'.

Example 2.1. An example of a joint ensemble is the ordered pair XY consisting of two successive letters in an English document. The possible outcomes are ordered pairs such as `aa`, `ab`, `ac`, and `zz`; of these, we might expect `ab` and `ac` to be more probable than `aa` and `zz`. An estimate of the joint probability distribution for two neighbouring characters is shown graphically in figure 2.2.

This joint ensemble has the special property that its two marginal distributions, $P(x)$ and $P(y)$, are identical. They are both equal to the monogram distribution shown in figure 2.1.

From this joint ensemble $P(x, y)$ we can obtain conditional distributions, $P(y \,|\, x)$ and $P(x \,|\, y)$, by normalizing the rows and columns, respectively (figure 2.3). The probability $P(y \,|\, x = \texttt{q})$ is the probability distribution of the second letter given that the first letter is a `q`. As you can see in figure 2.3a, the two most probable values for the second letter y given