As an example, $P_\infty(p = 0.59) = 140/154$ for the single configuration shown in Figure 12.3b. An accurate estimate of P_∞ involves an average over many configurations for a given value of p. For an infinite lattice, $P_\infty(p) = 0$ for $p < p_c$ and $P_\infty(p) = 1$ for $p = 1$. Between p_c and 1, $P_\infty(p)$ increases monotonically.

More information can be obtained from the *cluster size distribution* $n_s(p)$ defined as

$$n_s(p) = \frac{\text{average number of clusters of size } s}{\text{total number of lattice sites}}. \tag{12.3}$$

For $p \geq p_c$, the spanning cluster is excluded from n_s. (For historical reasons, the *size* of a cluster refers to the *number* of sites in the cluster rather than to its spatial extent.) As an example, we see from Figure 12.3a that $n_s(1) = 20/256$, $n_s(2) = 4/256$, $n_s(3) = 5/256$, and $n_s(7) = 1/256$ for $p = 0.2$ and is zero otherwise.

Because $N \sum_s sn_s$ is the total number of occupied sites (N is the total number of lattice sites), and Nsn_s is the number of occupied sites in clusters of size s, the quantity

$$w_s = \frac{sn_s}{\sum_s sn_s} \tag{12.4}$$

is the probability that an occupied site chosen at random is part of an s-site cluster. The *mean cluster size* S is defined as

$$S(p) = \sum_s sw_s = \frac{\sum_s s^2 n_s}{\sum_s sn_s}. \tag{12.5}$$

The sum in (12.5) is over finite clusters only. As an example, the weights corresponding to the clusters in Figure 12.3a are $w_s(1) = 20/50$, $w_s(2) = 8/50$, $w_s(3) = 15/50$, and $w_s(7) = 7/50$, and hence $S = 130/50$.

Problem 12.5 Qualitative behavior of $n_s(p)$, $S(p)$, and $P_\infty(p)$

(a) Use `PercolationApp` to visually determine the cluster size distribution $n_s(p)$ for a square lattice with $L = 16$ and $p = 0.4$, $p = p_c$, and $p = 0.8$. Take $p_c = 0.5927$. Consider at least ten configurations for each value of p and average $n_s(p)$ over the configurations. For each value of p, plot n_s as a function of s and describe the observed s-dependence. Does n_s decrease more rapidly with increasing s for $p = p_c$ or for $p \neq p_c$? Plot $\ln n_s$ versus s and versus $\ln s$. Does either of these plots suggest the form of the s-dependence of n_s? Is there a qualitative change near p_c? You probably will not be able to obtain definitive answers to these questions at this point, but we will discuss a more quantitative approach later. Better results for n_s can also be found if periodic boundary conditions are used (see Project 12.17).

(b) Use the same configurations considered in part (a) to compute the mean cluster size S as a function of p. Remember that for $p > p_c$, the spanning cluster is excluded.

(c) Similarly, compute $P_\infty(p)$ for various values of $p \geq p_c$ and plot $P(p)$ as a function of p and discuss its qualitative behavior.

(d) Verify that $\sum_s sn_s(p) = p$ for $p < p_c$ and explain this relation. How is this relation modified for $p \geq p_c$? ■

It is useful to associate a characteristic linear dimension or *connectedness length* $\xi(p)$ with the clusters. One way to do so is to define the *radius of gyration* R_s of a single cluster of s particles as

$$R_s^2 = \frac{1}{s} \sum_{i=1}^{s} (\mathbf{r}_i - \overline{\mathbf{r}})^2, \tag{12.6}$$

where

$$\overline{\mathbf{r}} = \frac{1}{s} \sum_{i=1}^{s} \mathbf{r}_i, \tag{12.7}$$

and $\mathbf{r}_i$ is the position of the ith site in the same cluster. The quantity $\overline{\mathbf{r}}$ is the familiar definition of the center of mass of the cluster. From (12.6) we see that R_s is the root mean square radius of the cluster measured from its center of mass.

The connectedness length ξ can be defined as an average over the radii of gyration of all the finite clusters. To find the appropriate average for ξ, consider a site in a cluster of s sites. The site is connected to $s - 1$ other sites, and the mean square distance to these sites is the order of R_s^2. The probability that a site belongs to a cluster of site s is $w_s = sn_s$. These considerations suggest that a reasonable definition of ξ is

$$\xi^2 = \frac{\sum_s (s-1) w_s \langle R_s^2 \rangle}{\sum_s (s-1) w_s}, \tag{12.8}$$

where $\langle R_s^2 \rangle$ is the average of R_s^2 over all clusters of s sites. To simplify the expression for ξ, we write s instead of $s - 1$ and let $w_s = sn_s$:

$$\xi^2 = \frac{\sum_s s^2 n_s \langle R_s^2 \rangle}{\sum_s s^2 n_s}. \tag{12.9}$$

As before, the sum in (12.9) is over only nonspanning clusters.

Problem 12.6 Simple calculation of the connectedness length

To obtain a feel for how to compute the connectedness length ξ, calculate it for the configuration shown in Figure 12.3a. ■

12.3 ■ FINDING CLUSTERS

So far we we have visually determined the clusters for a given configuration at a particular value of p. We now discuss an algorithm due to Newman and Ziff for finding clusters at many values of p. This algorithm is based on one that is well known in computer science in the context of the union-find problem. In the Newman–Ziff algorithm, we begin with an empty lattice and keep track of the clusters as we randomly occupy new lattice sites. As each site is occupied, we determine whether it becomes a new cluster or whether it is a neighbor of an existing cluster (or clusters). Because $p = n/L^2$, where n is the number of occupied sites, p increases by $1/L^2$ each time we occupy a new site. The algorithm can be summarized as follows (see class `Clusters` in Listing 12.2).