20 Probability

20.1 The Percolation Model (7 units)

This project does not presuppose attendance at any particular Part IB or Part II course.

1 Introduction

The percolation process is a standard model for a random medium. Such a process possesses a singularity about which it is hard to prove much rigorous mathematics. The purpose of this project is to explore such a singularity by numerical methods.

Take as (directed) graph G the first quadrant of the square lattice, with northerly and easterly orientations. The vertices are the points $x = (x_1, x_2)$ with $x_1, x_2 \in \{0, 1, 2, \ldots\}$. We set $|x - y| = |x_1 - y_1| + |x_2 - y_2|$ for two such x, y, and we join x to y by an edge $\langle x, y\rangle$ if $|x - y| = 1$; this edge is directed upwards or rightwards as appropriate.

Let p satisfy $0 \leqslant p \leqslant 1$. Each edge of G is designated *open* with probability p, different edges having independent designations. Edges not designated open are called *closed*. Water is supplied at the origin (0,0), and is allowed to flow along open edges in the directions given. The problem is to study the geometry of the random set C containing all wetted points. In particular, for what values of p is there strictly positive probability that C is infinite?

Writing 'P_p' for the probability function when p is the parameter given above, we define $\theta(p) = P_p(|C| = \infty)$. It may be shown that θ is a non-decreasing function, and the *critical probability* is defined by

$$p_c = \sup\{p : \theta(p) = 0\}.$$

It may be shown that $0 < p_c < 1$ (and better bounds are known), but the true value of p_c is unknown.

You are required to investigate this percolation numerically and to comment on various aspects of the behaviour as described below. You are only expected to comment on results that you can obtain using a reasonable amount of computer time, and sensible discussions of the limitations of your methods will receive more credit than reports of excessive computations. You should, however, give some thought to how you design your programs in order to achieve larger values of n and m (defined below) than otherwise might be the case. You should comment on any such special features of your programs in *each section* of your write-up.

2 Estimating $\theta(p)$

One method for estimating $\theta(p)$ is as follows. Let Q_n be the set of all points $x = (x_1, x_2)$ with $x_1 + x_2 = n$. Define the sequence $C_0, C_1, C_2, \ldots$ of sets in the following inductive manner. First, $C_0 = \{(0, 0)\}$. Having found $C_0, C_1, \ldots, C_K$, we next define C_{K+1}. For $y = (y_1, y_2) \in Q_{K+1}$, we place y in C_{K+1} if and only if

$$\begin{aligned} \text{either:} \quad & y' = (y_1 - 1, y_2) \in C_K, \quad \text{and } \langle y', y\rangle \text{ is open,} \\ \text{and/or:} \quad & y'' = (y_1, y_2 - 1) \in C_K, \quad \text{and } \langle y'', y\rangle \text{ is open.} \end{aligned} \tag{1}$$

By generating pseudo-random numbers, we may obtain a realization of the model, and an associated sequence $C_0, C_1, \ldots$; for each such realization, define I_n to be 0 or 1 depending on whether C_n is empty or non-empty (respectively). If $I_n(1), I_n(2), \ldots, I_n(m)$ are the values of I_n obtained in m independent realizations of the model, then

$$\widehat{\theta}_{m,n}(p) = \frac{1}{m}\sum_{j=1}^{m} I_n(j)$$

may be used to estimate $\theta_n(p) = P_p(C_n \neq \emptyset)$. If n is sufficiently large, then $\widehat{\theta}_{m,n}(p)$ may be used to estimate $\theta(p) = \lim_{n\to\infty}\theta_n(p)$.

Question 1 Give an explanation of why θ is non-decreasing in p, that is $\theta(p_1) \leqslant \theta(p_2)$ if $p_1 \leqslant p_2$. Show also that $\theta_n(p)$ is decreasing in n. Give an estimate for the likely size of the error $\widehat{\theta}_{m,n}(p) - \theta_n(p)$.

Question 2 Use the scheme described above (but see the notes below) to plot $\widehat{\theta}_{m,n}(p)$ for $p \in [0.5, 0.75]$ for suitable n and m. How would you expect a graph of the true value $\theta(p)$ to look like in relation to your graph?

3 Estimating p_c

For fixed n and m an estimate of p_c may be obtained by finding $\sup\{p : \widehat{\theta}_{m,n}(p) = 0\}$. Denote this estimate by $\widehat{p_c} = \widehat{p_c}(m,n)$.

Question 3 Investigate the dependence of the estimate $\widehat{p_c}$ on the values of n and m. For fixed m, describe how the estimate varies with n, and explain why this should be so. How does the estimate vary with m for fixed n?

4 Subcritical behaviour

As above, let C_n be the set of points $x = (x_1, x_2)$ which satisfy $x \in C$, $x_1 + x_2 = n$. When $p < p_c$, it may be shown that there is a constant $\gamma > 0$ (depending on p) for which $P_p(C_n \neq \emptyset) \leqslant \exp(-\gamma n)$ and

$$\frac{1}{n}\log P_p(C_n \neq \emptyset) \to -\gamma \qquad \text{as} \qquad n \to \infty.$$

Question 4 Estimate γ for $p = 0.3$, 0.4, 0.5 and 0.6, choosing appropriate values of n and m for each case. Describe briefly how you chose n and m in each case and why you did so. What behaviour do you expect as $p \uparrow p_c$?

5 Notes

(i) You may find it helpful to know that we believe $p_c \approx 0.644$ (though you may obtain a different estimate).

(ii) At first sight, the scheme described above appears to require m realizations of the model *for each value of* p. The following construction enables the same realizations to be used *for all values of* p *simultaneously*. For each edge e, we choose a pseudo-random number R_e which is uniformly distributed on $[0,1]$; different edges receive *independent* numbers.

To each vertex x we assign a real number $Z(x)$ defined as follows. First, $Z(\mathbf{0}) = 0$, where $\mathbf{0} = (0,0)$ is the origin. Having calculated $Z(x)$ for $x \in Q_0 \cup Q_1 \cup \ldots \cup Q_K$, we define $Z(y)$ for $y \in Q_{K+1}$ by

$$Z(y) = \min\{A', A''\}$$

where

$$A' = \max\{Z(y'), R_{\langle y', y\rangle}\}, \quad A'' = \max\{Z(y''), R_{\langle y'', y\rangle}\},$$

where y' and y'' are given in (1). For given p, we may obtain a percolation realization as follows. Call an edge e *open* if $R_e \leqslant p$ (an event having probability p). It may now be seen that the set $\{y : Z(y) \leqslant p\}$ has the same distribution as the set C of wetted points given above. Much computational time may be saved by this device.

(iii) In practice, you may find that much of the computation time is spent in generating the pseudo-random numbers. Time may be saved if they are not calculated to excessive precision.

(iv) There are many interesting features of the super-critical behaviour ($p > p_c$); in particular the 'shape' of the infinite cluster when it exists. You are **not** asked to comment on these features for this project.

(v) Further details about percolation in general may be found in: *Percolation*, G R Grimmett, Springer, Berlin 1999. Further details about the particular percolation used in this project (two-dimensional oriented bond percolation) can be found in, for example, R. Durrett, *The Annals of Probability* **12**:999-1040 (1984). It is **not** necessary (nor even particularly desirable) to consult either of these references before attempting this project.

