20   Probability

20.5   Percolation and the Invasion Process   (9 units)

This project requires general knowledge of probability theory, at the level of IA Probability. It also requires competency in programming.

1   Introduction

This project concerns certain probability models for bond percolation. The book by Grimmett [1] is a good source to learn more about percolation.

We work on a connected graph $G = (V, E)$, that is, a collection of nodes V connected by edges E. To each edge $e \in E$, we assign, independently, a uniform random variable $U_e \sim U[0,1]$. We decide on a value $p \in [0,1]$; we declare the edge e to be *p-open* if $U_e < p$, and we declare it to be *p-closed* otherwise.

When p is very small, very few edges are open; but as we increase p, there appear open clusters, i.e. sets of nodes connected by open edges.

Percolation theory is the study of the geometry of the open clusters. In particular, important questions are whether or not there exists an infinite cluster of open edges; and if one does exist, how many infinite clusters there are. Clearly if $p = 0$ there is none and if $p = 1$ there is one open cluster, namely the graph G itself.

2   The binary tree

Let V, the set of nodes of the graph, consist of finite strings, as follows: V contains the empty string '' (also known as *Eve*), and the three strings '1', '2' and '3' (also known as *Eve's daughters*), and also every string that is one of Eve's daughters followed by a finite sequence of '1's and '2's. Two nodes are connected by an edge if one can be obtained by appending one digit to the other. For example, '3221' is connected to '322' (its mother) and to '32211' and '32212' (its two daughters). As before, each edge e is assigned a random variable $U_e \sim U[0,1]$.

(We can use this as a crude model to describe the propagation of a defective gene in a population.)

> **Question 1**   Let ϕ_p be the probability that Eve's daughter '1' is in an infinite open cluster consisting of her own descendents. Show that
>
> $$\phi_p = 2p(1-p)\phi_p + p^2\big(\phi_p^2 + 2\phi_p(1-\phi_p)\big).$$
>
> It can be shown that ϕ_p is the maximal solution to this equation. Find ϕ_p. (One way to obtain a 'merit' mark in this project, though not the only way, is to show that ϕ_p *is* the maximal solution.)
>
> Now let θ_p be the probability that Eve is in an infinite open cluster. Find θ_p and draw its graph as a function of p.

> **Question 2**   Show that, for $p \leqslant \frac{1}{2}$, there are almost surely no infinite clusters.
> How many infinite clusters are there if $\frac{1}{2} < p < 1$? Justify your answer.

This model exhibits a property which is general to percolation models: there exists a critical probability p_c such that for $p < p_c$ there is no infinite cluster and for $p > p_c$ we find at least one infinite cluster.

The region $p > p_c$ is called the *supercritical* region. We can ask questions like: what is θ_p, the probability that a node chosen arbitrarily lies in an infinite open cluster? (In Question 1, in calculating θ_p, we could of course have designated any node to be Eve.)

The region $p < p_c$ is called the *subcritical* region. We ask questions like: how likely are we to observe an open cluster of size n? if there is an open cluster of size n, what shape is it?

Most of the interesting (unsolved) problems relate to the geometry of open clusters when p is near p_c. For example, there are many graphs (e.g. $\mathbb{Z}^3$) where it is not known (but strongly conjectured) that there is no infinite p_c-open cluster.

3   The square lattice

The square lattice in two dimensions $\mathbb{L}^2$ is a graph with $V = \mathbb{Z}^2 = \{(m, n) : m, n \in \mathbb{Z}\}$. If the distance function is
$$d\big((k, l), (m, n)\big) = |k - m| + |l - n|,$$
the edges of the graph are straight lines connecting nodes which are distance 1 apart.

Let us look at two techniques which will help us estimate the critical probability p_c above which there is an infinite cluster and below which there is none.

Lower bound

Let us start at the origin. Let σ_n be the number of self-avoiding paths (i.e., paths which traverse each edge at most once) of length n leading away from the origin.

Question 3   Let $\lambda = \limsup_{n \to \infty} \sigma_n^{1/n}$. Show that $\lambda \leqslant 3$. Show further that $p_c \geqslant \lambda^{-1}$.

The actual value of λ is an open problem.

Upper bound

The general behaviour of large clusters in the subcritical region $p < p_c$ is described in the following result. Some notation first: We say $x \leftrightarrow y$ if there exists an open connected path between x and y. Define the open sphere S_n to be
$$S_n = \big\{x \in \mathbb{Z}^2 : d(x, 0) \leqslant n\big\}.$$

The boundary ∂S_n consists of the nodes where $d(x, 0) = n$. Let $P_p(0 \leftrightarrow \partial S_n)$ be the probability that there exists a p-open path connecting the origin to some node in ∂S_n. It can be shown that, for $p < p_c$, there exists $\psi_p > 0$ such that
$$P_p(0 \leftrightarrow \partial S_n) < e^{-n\psi_p} \quad \text{for all } n. \tag{1}$$

The proof is beyond the scope of this project but can be found in [1, Sections 5.2 and 6.1].

Armed with the above result, we aim to show that $p_c \leqslant \frac{1}{2}$. We do this as follows. Consider the following subgraph $G_n = (V_n, E_n)$ of the square lattice.

$$V_n = \big\{ (k,l) \,:\, 0 \leqslant k \leqslant n, \ 0 \leqslant l \leqslant n-1 \big\}.$$

Let E_n be all the edges in E connecting these nodes. We call $\{(k,l) \in V_n : k = 0\}$ the left boundary, and $\{(k,l) \in V_n : k = n\}$ the right boundary. Let A be the event that some node in the left boundary is connected to the right boundary via a path consisting of open edges.

Question 4   Show that $P_{1/2}(A) = \frac{1}{2}$. *Hint. You may find it useful to consider the dual graph $\bar{G}_n$, which has nodes at*

$$V_n' = \big\{ (k + \tfrac{1}{2}, l - \tfrac{1}{2}) \,:\, 0 \leqslant k \leqslant n-1, \ 0 \leqslant l \leqslant n \big\},$$

and edges joining those nodes which are distance 1 apart, and whose edges are open or closed depending on whether the edges of G are open or closed, in a manner which you should specify.

Question 5   By constructing n events $\{A_i\}$ such that $A = \cup A_i$, and using (1), prove that $p_c \leqslant \frac{1}{2}$.

4   The Invasion Process

We can try and use computing power to estimate p_c and θ_p. Suppose that at time $n = 0$ you are an invading force standing at the origin. We will call I_n the set of nodes you have invaded by time n. At time $n + 1$ you invade another node by looking at the edge-boundary of your territory and walking along the edge with the least value of U_e attached to it. Formally, we define

$$\partial I_n = \big\{ e \in E : I_n \overset{e}{\leftrightarrow} \mathbb{Z}^2 \setminus I_n \big\}.$$

We use the notation $x \overset{e}{\leftrightarrow} y$ to mean that the edge e connects x and y. You walk from I_n along the edge $e_n \in \partial I_n$ which satisfies

$$U_{e_n} = \min\{U_f : f \in \partial I_n\},$$

so that I_{n+1} is I_n with the node at the other side of e_n added. It can be shown that, almost surely,

$$\limsup_{n \to \infty} U_{e_n} = p_c. \tag{2}$$

(The proof is not hard, and is outlined at the end of this project.) The advantage of using the sequence U_{e_n} to estimate p_c is that the amount of memory required to store U_{e_n} and to calculate $U_{e_{n+1}}$ is $O(n)$. This is true whether we are working in $\mathbb{L}^2$ or $\mathbb{L}^{47}$.

Question 6   Implement the invasion process. Describe your algorithm.
Explain in particular why it only requires $O(n)$ storage space to calculate the first n values of U_{e_n}, and what it does to ensure it never revisits a vertex.
Comment also on the complexity of the algorithm (the number of time steps needed).

Question 7   Use your program to estimate p_c for $\mathbb{L}^2$, and explain your method. Estimate p_c for $\mathbb{L}^3$ (which is defined like $\mathbb{L}^2$, but using $\mathbb{Z}^3$ rather than $\mathbb{Z}^2$). Include in your report any appropriate plots.

It is desired to plot θ_p, the density of the infinite cluster.

> **Question 8**   Explain how the invasion process can be used to estimate θ_p simultaneously for all p. Produce a plot of θ_p against p for $\mathbb{L}^2$ by running the simulation for large n several times (at least $n = 5000$, at least 500 times). For what values of p do you expect your plot to be inaccurate? Why?

Appendix

Here is an outline of the proof of (2).

Let $p > p_c$. Then there exists an infinite p-open cluster. Let T_p be the first time that the invasion process hits the cluster (i.e. the first time that the vertex $I_{T_p} \setminus I_{T_p-1}$ is in the infinite p-open cluster). It can be shown that $P(T_p < \infty) = 1$. For all $n \geqslant T_p$, there will always be an edge in ∂I_n with $U_e \leqslant p$. It follows that

$$\limsup_{n \to \infty} U_{e_n} \leqslant p.$$

Since $p > p_c$ was arbitrary, $\limsup_{n \to \infty} U_{e_n} \leqslant p_c$.

Now let $p < p_c$. Suppose that with some probability $\alpha > 0$,

$$\limsup_{n \to \infty} U_{e_n} \leqslant p < p_c.$$

Then we have (with a positive probability) an infinite cluster, with all but finitely many of its edges p-open. It follows that (with a positive probability) there exists an infinite p-open cluster. This contradicts the definition of p_c as the critical probability.

References

[1] G.R. Grimmett. *Percolation.* Springer-Verlag, Berlin 1989 and 1999