P o C S What's The Story?

Principles of Complex Systems, CSYS/MATH 300
University of Vermont, Fall 2015
Assignment 6 • code name: The River Styx

Dispersed: Saturday, October 3, 2015.
Due: 5 pm, Friday, October 16, 2015.
Some useful reminders:
Deliverator: Peter Dodds
Office: Farrell Hall, second floor, Trinity Campus
E-mail: peter.dodds@uvm.edu
Office hours: 10:30 am to 11:30 am on Mondays, and 2:00 pm to 3:30 pm Wednesdays at Farrell Hall
Course website: http://www.uvm.edu/~pdodds/teaching/courses/2015-08UVM-300

All parts are worth 3 points unless marked otherwise. Please show all your workingses clearly and list the names of others with whom you collaborated.

Graduate students are requested to use LaTeX (or related TeX variant).

Email submission: PDF only! Please name your file as follows (where the number is to be padded by a 0 if less than 10 and names are all lowercase):
`CSYS300assignment%02d$firstname-$lastname.pdf` as in
`CSYS300assignment06michael-palin.pdf`

Please submit your project's current draft in pdf format via email. Please use this file name format (all lowercase after CSYS):
`CSYS300project-$firstname-$lastname-YYYY-MM-DD.pdf` as in
`CSYS300project-lisa-simpson-1989-12-17.pdf`

---

1. The 1-d theoretical percolation problem:

   Consider an infinite 1-d lattice forest with a tree present at any site with probability p.

   (a) Find the distribution of forest sizes as a function of p. Do this by moving along the 1-d world and figuring out the probability that any forest you enter will extend for a total length ℓ.

   (b) Find p_c, the critical probability for which a giant component exists.

   Hint: One way to find critical points is to determine when certain average quantities explode. Compute $\langle l \rangle$ and find p such that this expression goes boom (if it does).

2. Show analytically that the critical probability for site percolation on a triangular lattice is $p_c = 1/2$.

   Hint—Real-space renormalization gets it done.:
   `http://www.youtube.com/v/J1kbU5U7QqU?rel=0`

3. (3 + 3)

   Coding, it's what's for breakfast:

   (a) Percolation in two dimensions (2-d) provides a classic, nutritious example of a phase transition.

   Your mission, whether or not you choose to accept it, is to code up and analyse the L by L square lattice percolation model for varying L.

   Take L = 20, 50, 100, 200, 500, and 1000.

   (Go higher if you feel L = 1000 is for mere mortals.)

   (Go lower if your code explodes.)

   Let's continue with the tree obsession. A site has a tree with probability p, and a sheep grazing on what's left of a tree with probability $1 - p$.

   Forests are defined as any connected component of trees bordered by sheep, where connections are possible with a site's four nearest neighbors on a lattice.

   Each square lattice is to be considered as a landscape on which forests and sheep co-exist.

   Do not bagelize (or doughnutize) the landscape (no periodic boundary conditions—boundaries are boundaries).

   (Note: this set up is called site percolation. Bond percolation is the alternate case when all links between neighboring sites exist with probability p.)

   Steps:

   i. For each L, run N_{tests}=100 tests for occupation probability p moving from 0 to 1 in increments of 10^{-2}. (As for L, you may use a smaller or larger increment depending on how things go.)

   ii. Determine the fractional size of the largest connected forest for each of the N_{tests}, and find the average of these, S_{avg}.

   iii. On a single figure, for each L, plot the average S_{avg} as a function of p.

   (b) Comment on how $S_{\text{avg}}(p; N)$ changes as a function of L and estimate the critical probability p_c (the percolation threshold).

Helpful reuse of code (intended for black and white image analysis): You can use Matlab's bwconncomp to find the sizes of components. Very nice.

4. (3 + 3)

   (a) Using your model from the previous question and your estimate of p_c, plot the distribution of forest sizes (meaning cluster sizes) for $p \simeq p_c$ for the largest L your code and psychological makeup can withstand. (You can average the distribution over separate simulations.)

   Comment on what kind of distribution you find.

   (b) Repeat the above for $p = p_c/2$ and $p = p_c + (1 - p_c)/2$, i.e., well below and well above p_c.

   Produce plots for both cases, and again, comment on what you find.

5. (3 + 3 + 3)

   A courageous coding festival:

   Code up the discrete HOT model in 2-d. Let's see if we find any of these super-duper power laws everyone keeps talking about. We'll follow the same approach as the $N = L \times L$ 2-d forest discussed in lectures.

   Main goal: extract yield curves as a function of the design D parameter as described below.

   Suggested simulations elements:

   - Take $L = 32$ as a start. Once your code is running, see if $L = 64$, 128, or more might be possible. (The original sets of papers used all three of these values.) Use a value of L that's sufficiently large to produced useful statistics but not prohibitively time consuming for simulations.
   - Start with no trees.
   - Probability of a spark at the (i, j)th site: $P(i, j) \propto e^{-i/\ell} e^{-j/\ell}$ where (i, j) is tree position with the indices starting in the top left corner ($i, j = 1$ to L). (You will need to normalize this properly.) The quantity ℓ is the characteristic scale for this distribution. Try out $\ell = L/10$.
   - Consider a design problem of $D = 1$, 2, L, and L^2. (If L and L^2 are too much, you can drop them. Perhaps sneak out to $D = 3$.) Recall that the design problem is to test D randomly chosen placements of the next tree against the spark distribution.
   - For each test tree, compute the average forest fire size over the full spark distribution:

     $$\sum_{i,j} P(i, j) S(i, j),$$

     where $S(i, j)$ is the size of the forest component at (i, j). Select the tree location with the highest average yield and plant a tree there.

- Add trees until the 2-d forest is full, measuring average yield as a function of trees added.
- Only trees within the cluster surrounding the ignited tree burn (trees are connected through four nearest neighbors).

(a) Plot the forests.

(b) Plot the yield curves for each value of D, and identify (approximately) the peak yield and the density for which peak yield occurs for each value of D.

(c) Plot distributions of connected tree interval sizes at peak yield. Note: You will have to rebuild forests and stop at the peak yield value of D to find these distributions. By recording the sequence of optimal tree planting, this can be done without running the simulation again.

Hint: Working on un-treed locations will make choosing the next location easier.

6. The discrete version of HOT theory:

   From lectures, we had the following.

   Cost: Expected size of 'fire' in a d-dimensional lattice:

$$C_{\text{fire}} \propto \sum_{i=1}^{N_{\text{sites}}} p_i a_i^2$$

   where $a_i =$ area of ith site's region, and $p_i =$ avg. prob. of fire at site i over a given time period.

   The constraint for building and maintaining $(d-1)$-dimensional firewalls in d-dimensions is

$$C_{\text{firewalls}} \propto \sum_{i=1}^{N_{\text{sites}}} a_i^{(d-1)/d} a_i^{-1},$$

   where we are assuming isometry.

   Using Lagrange Multipliers, safety goggles, rubber gloves, a pair of tongs, and a maniacal laugh, determine that:

$$p_i \propto a_i^{-\gamma} = a_i^{-(2+1/d)}.$$

   and therefore that

$$\mathbf{Pr}(a_i) \propto a_i^{-\gamma}.$$

7. (3 + 3 + 3) *Highly Optimized Tolerance:*

   This question is based on Carlson and Doyle's 1999 paper "Highly optimized tolerance: A mechanism for power laws in design systems" [1]. In class, we made our way through a discrete version of a toy HOT model of forest fires. This paper revolves around the equivalent continuous model's derivation. You do not have to perform the derivation but rather carry out some manipulations of probability distributions using their main formula.

Our interest is in Table I on p. 1415:

$p(x)$	$p_{\text{cum}}(x)$	$P_{\text{cum}}(A)$
$x^{-(q+1)}$	x^{-q}	$A^{-\gamma(1-1/q)}$
e^{-x}	e^{-x}	$A^{-\gamma}$
e^{-x^2}	$x^{-1}e^{-x^2}$	$A^{-\gamma}[\log(A)]^{-1/2}$

and Equation 8 on the same page:

$$P_{\geq}(A) = \int_{p^{-1}(A^{-\gamma})}^{\infty} p(\mathbf{x})\mathrm{d}\mathbf{x} = p_{\geq}\left(p^{-1}\left(A^{-\gamma}\right)\right),$$

where $\gamma = \alpha + 1/\beta$ and we'll write $P_{\geq}$ for P_{cum}.

Please note that $P_{\geq}(A)$ for $x^{-(q+1)}$ is not correct. Find the right one!

Here, $A(\mathbf{x})$ is the area connected to the point $\mathbf{x}$ (think connected patch of trees for forest fires). The cost of a 'failure' (e.g., lightning) beginning at $\mathbf{x}$ scales as $A(\mathbf{x})^{\alpha}$ which in turn occurs with probability $p(\mathbf{x})$. The function p^{-1} is the inverse function of p.

Resources associated with point $\mathbf{x}$ are denoted as $R(\mathbf{x})$ and area is assumed to scale with resource as $A(\mathbf{x}) \sim R^{-\beta}(\mathbf{x})$.

Finally, $p_{\geq}$ is the complementary cumulative distribution function for p.

As per the table, determine $p_{\geq}(x)$ and $P_{\geq}(A)$ for the following (3 pts each):

(a) $p(x) = cx^{-(q+1)}$,

(b) $p(x) = ce^{-x}$, and

(c) $p(x) = ce^{-x^2}$.

Note that these forms are for the tails of p only, and you should incorporate a constant of proportionality c, which is not shown in the paper.

8. A spectacularly optional extra.

Warning:

- Only attempt if using registered safety equipment including welding goggles and a lead apron.

- Make sure to back up your brain in at least two geographically distant places beforehand (e.g., on different planets).

Dangerous feature:

- If you make it out, you will be very happy.

In lectures on lognormals and other heavy-tailed distributions, we came across a super fun and interesting integral when considering organization size distributions arising from growth processes with variable lifespans.

Show that

$$P(x) = \int_{t=0}^{\infty} \lambda e^{-\lambda t} \frac{1}{x\sqrt{2\pi t}} \exp\left(-\frac{(\ln \frac{x}{m})^2}{2t}\right) \mathrm{d}t$$

leads to:

$$P(x) \propto x^{-1} e^{-\sqrt{2\lambda(\ln \frac{x}{m})^2}},$$

and therefore, surprisingly, two different scaling regimes. Enjoyable suffering may be involved. Really enjoyable suffering. But many monks have found a way so you should follow their path laid out below.

Hints and steps:

- Make the substitution $t = u^2$ to find an integral of the form (excluding a constant of proportionality)

  $$I_1(a, b) = \int_0^{\infty} \exp\left(-au^2 - b/u^2\right) \mathrm{d}u$$

  where in our case $a = \lambda$ and $b = (\ln \frac{x}{m})^2/2$.

- Substitute $au^2 = t^2$ into the above to find

  $$I_1(a, b) = \frac{1}{\sqrt{a}} \int_0^{\infty} \exp\left(-t^2 - ab/t^2\right) \mathrm{d}t$$

- Now work on this integral:

  $$I_2(r) = \int_0^{\infty} \exp\left(-t^2 - r/t^2\right) \mathrm{d}t$$

  where $r = ab$.

- Differentiate I_2 with respect to r to create a simple differential equation for I_2. You will need to use the substitution $u = \sqrt{r}/t$ and your differential equation should be of the (very simple) form

  $$\frac{\mathrm{d}I_2(r)}{\mathrm{d}r} = -(\text{something}) I_2(r).$$

- Solve the differential equation you find. To find the constant of integration, you can evaluate $I_2(0)$ separately:

$$I_2(0) = \int_0^\infty \exp(-t^2)\mathsf{d}t,$$

where our friend $\Gamma(frac12)$ comes into play.

References

[1] J. M. Carlson and J. Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Phys. Rev. E*, 60(2):1412–1427, 1999. pdf