Econ 1033	**Problem set 3**
Prof. Ben Golub	Due: **Wed., Sep 23, 2020 11:50pm**

This problem set is worth 100 points. You should attempt problems totaling at least 100 points, and submit solutions to all parts of those problems. For any subset $\mathcal{P}$ of problems totaling at least 100 points, your $\mathcal{P}$-score is defined to be the fraction of points you receive for problems in $\mathcal{P}$ (i.e., points earned on problems in $\mathcal{P}$ divided by possible point total for problems $\mathcal{P}$, which may be more than 100). We will make your score equal to your $\mathcal{P}$-score where $\mathcal{P}$ (the set of problems that count) is chosen optimally for you.

1. (40 points) In this problem we will study a generalization of the branching process. The random variable X_n denotes the number of (new) infected individuals in wave n. Initialize $X_0 = 1$, which means there is one "patient zero" in wave 0. Now, for any $n \geq 1$, if we are given X_{n-1}, the number of infected individuals in wave $n-1$, here is how we generate X_n, the number of infected individuals in wave n. For each $i = 1, 2, \ldots, X_{n-1}$ (i.e, for each infected individual in wave $n-1$) draw a random variable $Z_{n-1,i}$ which is the number of "children" that i has. These $Z_{n-1,i}$ are independent and identically distributed with probability mass function P.[1] Let X_n be the sum of the random variables $Z_{n-1,i}$ as i ranges from 1 to X_{n-1}, i.e.

$$X_n = Z_{n-1,1} + Z_{n-1,2} + \cdots + Z_{n-1,X_{n-1}}.$$

Now define the function f on the domain [0,1] by

$$f(x) = 1 - \sum_{k=0}^{\infty} (1-x)^k P(k).$$

Let q_n be the probability that $X_n \geq 1$.[2] Assume that $0 < P(0) < 1$ and $0 < P(1) < 1$ to eliminate the trivial cases.

Throughout the problem, please interpret $0^0 = 1$.

- **a. (8 points)** What is q_0 (you should give an exact, numerical answer)? Now express q_1 in terms of $P(0)$, $P(1)$, $P(2)$, etc.
- **b. (8 points)** Is q_n increasing in n, decreasing in n, or neither? Argue from the definition of X_n and the basic properties of the process, without doing any calculations.
- **c. (8 points)** Give the sign of the first derivative and the sign of the second derivative of f. (Here P is arbitrary, subject to the assumptions given in the statement.) Justify your claims.
- **d. (8 points)** Show that for each $n > 0$, $q_n = f(q_{n-1})$ (Hint: use the same idea as in the lecture. Hint: first imagine that you knew the root has k children. How can you use this information even though you don't know k?)
- **e. (8 points)** Suppose that P is the Binomial distribution with k draws and probability of success p on each draw. Is that special case equivalent to the case studied in Easley and Kleinberg 21.2 and 21.8.A? Explain why or why not.

[1]This means that $\mathbb{P}(Z_{n-1,i} = k) = P(k)$.

[2]This probability is evaluated from the perspective of an observer who knows only that $X_0 = 1$.

2. (30 points) Maintain the assumptions and notations in Problem 1.

a. (10 points) Let $P(0) = 1/6$, $P(1) = 1/3$, and $P(2) = 1/2$, with $P(\ell) = 0$ for $\ell > 2$. Plot f. Illustrate using a "staircase plot" the values of q_n for $n = 0, 1, 2, 3$, as in Figure 21.18 of [EK]. Compute these values of q_n for $n = 4, 5, 6, 7$ as well (no need to plot this).

b. (5 points) Now let $P(0) = 1/2$, $P(1) = 1/3$, and $P(2) = 1/6$, with $P(\ell) = 0$ for $\ell > 2$. Plot f. Illustrate using a "staircase plot" the values of q_n for $n = 0, 1, 2, 3$. Compute these values for $n = 4, 5, 6, 7$ as well (no need to plot these).

c. (5 points) Define $q_\infty = \lim_{n\to\infty} q_n$ when this limit exists. Describe how you can find q_∞ by looking at a plot such as the ones you made in parts (a) and (b). Give a verbal statement of the meaning of q_∞ (this is required to get full points on this part).

d. (10 points) Give a necessary and sufficient condition for $q_\infty > 0$. (Your condition will, of course, be in terms of something about the distribution P—in fact, it will be in terms of a moment of this distribution.) Justify your answer clearly.

3. (35 points)

a. (10 points) Consider the following variant of the birthday problem. On Planet Kleinberg, there are two classes, with the people in one class comprising a set A while the people in another class comprise a set B, disjoint from A. The classes are of equal size, so that $|A| = k$ and $|B| = k$, for some integer $k \geq 1$. Each person $i \in A \cup B$ has a birthday $m_i \in \{1, 2, \ldots, M\}$ drawn uniformly at random, where M is the number of days in a year on Planet Kleinberg; all the birthdays are independent. Fix any constant $c > 0$ and assume $k \geq cM$. Show that for large enough M, the probability that the birthdays in class A are disjoint from the birthdays in class B (the event D) is arbitrarily small. More formally, show that for any $\epsilon > 0$, there is an $\underline{M}$ so that for all $M \geq \underline{M}$, the probability of D is less than ϵ. (You do *not* need to write a formula for $\underline{M}$; just show it is finite.)

b. (10 points) Now consider the random graph model $G(n, p)$, where there is a set N of n nodes, and each pair of distinct nodes is linked with probability p, independently. Fix $p > 0$. Let $d(i)$ be the (random) degree of node i in n. Let

$$E = \left\{ \frac{1}{2}p(n-1) \leq d(i) \leq \frac{3}{2}p(n-1) \text{ for all } i \in N \right\}$$

be the event that simultaneously, every node's degree is close enough to its expectation (within 50% of its expectation). Show that for large enough n, the probability of E is arbitrarily close to 1. More formally, show that for any $\epsilon > 0$, there is an $\underline{n}$ so that for all $n \geq \underline{n}$ the probability of E is at least $1 - \epsilon$. (You do *not* need to write a formula for $\underline{n}$; just show it is finite.)

c. (10 points) Show that for large enough n, with probability arbitrarily close to 1, the diameter of $G(n, p)$ is exactly 2. More formally, show that for any $\epsilon > 0$, there is an $\underline{n}$ so that for all $n \geq \underline{n}$, the probability that $\text{diam}(G(n, p)) = 2$ is at least $1 - \epsilon$. To do this, first use the information you know from (b) about the likely size of all neighborhoods. Then use an argument very similar to the one you used in (a). (Again, you do *not* need to write a formula for $\underline{n}$; just show it is finite.)

4. (40 points)

a. (5 points) Let G be the square lattice graph in 2D on a torus. The set of nodes V is the set of ordered pairs of integers (i, j) where $1 \leq i \leq n$ and $1 \leq j \leq n$. There is an edge between (i, j) and (i', j') if and only if (a) $i' \in \{i-1, i+1\}$ and $j' = j$ or (b) $j' = \{j-1, j+1\}$ and $i' = i$. Addition and subtraction in the adjacency definition are understood modulo n; this means that the nodes (n, j) are linked to $(1, j)$, for example. Sketch this graph.

b. (5 points) Take any node (i, j) in G. How many nodes are at distance k or less of (i, j)?

c. (5 points) Choose two distinct nodes uniformly at random. Prove that, if n is large enough, then with probability arbitrarily close to 1 the distance between them in G is more than $n^{0.99}$.

d. (5 points) Now we will create a new (random) graph G' whose edges are a superset of those of G. The construction is as follows: for each node, with probability 0.99 leave it alone and with probability 0.01 create one new edge from (i, j) to a different node in the graph, drawn uniformly at random. What is the expected degree of a node in this modified graph (before we realize the random draws)?

e. (10 points) Choose two distinct nodes uniformly at random. Prove that, if n is large enough, then with probability arbitrarily close to 1 the distance between them in G' is less than $n^{1/3}$. *Hint: chop up the graph into disjoint, connected, square subgraphs of size $k(n)$-by-$k(n)$, where $k(n)$ is wisely chosen so that you can prove any pair of subgraphs is very likely to be linked by a random edge.* (This will not quite prove the statement you want, but should give you an idea for how to proceed.)
Also, explain why your proof could be extended to show the same statement replacing $1/3$ with any positive number α.

f. (10 points) We can actually show that the average distance in G is bounded below by cn for some $c > 0$, and that in G' it is bounded above, with high probability, by $C \log(n)$ (for some constant C, where C depends on the probability you want to achieve). Feel free to convince yourself of these things, but don't feel obligated. However, do explain intuitively why the average distances are so different. Explain what this has to do with the plague reading in [HN, Ch. 3]. Feel free to refer to the paper "The small-world effect is a modern phenomenon" by Marvel et al. (arXiv:1310.263v1 [physics.soc-ph]).