Econ 1033	**Problem set 1**
Prof. Ben Golub	Due: **Wed., Sep 16, 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.

1. (20 points) Consider an "infection" process that evolves according to a more general rule: Start with a single infected individual in wave 0. In every wave, each infected individual produces a random number of children (nodes directly infected by it) in the next wave, independently of other individuals. The distribution of number of children, P, is a probability mass function with finite mean (a.k.a. expectation) denoted by R_0, and P is the same across individuals. Denote by X_n the number of infected individuals at wave n.

- **a. (5 points)** Describe the relationship between X_{n+1} and X_n by expressing X_{n+1} as a sum of X_n-many random variables. Your description can be in words as long as it is clear and correct.
- **b. (5 points)** Given X_n, compute the conditional expectation $\mathbb{E}[X_{n+1} \mid X_n]$.
- **c. (5 points)** Explain why the ratio $\mathbb{E}[X_{n+1}]/\,\mathbb{E}[X_n]$ is a deterministic constant, and say what this ratio is. Write $\mathbb{E}[X_n]$ in a simple way using this insight.
- **d. (5 points)** What is C_n, the *cumulative* number of infections up to and including wave n? What is the expectation of this random variable? Write your answer as simply as possible, and justify it.

3. (20 points) This exercise is about thinking through how the basic (k, p) branching process model applies to reality.

- **a. (10 points)** Describe three ways in which the model is not a realistic description of the transmission of disease on a social network. Think about both the process and the underlying network.
- **b. (10 points)** Imagine again that the (k, p) model is valid, with one tweak: because the individuals involved are reacting to the epidemic in real time, p actually depends on the cumulative number of cases to date. Recalling the notation of Problem 2(d), assume that the probability that individuals in wave $n+1$ get infected, p_{n+1}, is equal to $\mathcal{P}(C_n)$, where $\mathcal{P}$ is a decreasing function of its argument. You can think of this as a simple or "reduced-form" way of capturing rational agents' response to the epidemic: they protect themselves from exposure to infection when the infection is more widespread. Give reasonable conditions on $\mathcal{P}$ ensuring that the epidemic is eventually stopped.

4. (30 points) Recall the function

$$f(x) = 1 - (1 - px)^k, \quad x \in [0, 1]$$

from Lecture 1.[1] If you need to make (reasonable) assumptions to establish the statements below, make the assumptions clear.

[1] It also appears in Easley and Kleinberg 21.8.A. toward the end of the section.

a. (5 points) Show that $f(0) = 0$ and $f(1) < 1$, and compute $f'(0)$ in terms of k and p.

b. (5 points) Show that f is increasing.

c. (5 points) Show that f is strictly concave.

d. (5 points) When is there is a *positive* q^* such that $q^* = f(q^*)$? Explain your answer. Again, the condition you give will involve k and p.

e. (5 points) What happens to q^* if you hold p fixed and increase k to some other k'? Justify your answer. Also give a clear intuitive explanation of why this happens.

f. (5 points) What happens to q^* if you hold k fixed and increase p to some other p'? Justify your answer. Also give a clear intuitive explanation of why this happens.

5. (30 points) In this problem we will study a generalization of the branching process.[2] The random variable X_n will denote 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.[3] 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 by

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

Let q_n be the probability that $X_n \geq 1$.[4] Assume that $0 < P(0) < 1$ and $0 < P(1) < 1$ to eliminate the trivial cases.

- **a. (5 points)** What is q_0 as a number? Now express q_1 in terms of $P(0)$, $P(1)$, $P(2)$, etc.
- **b. (5 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. (5 points)** Give the sign of the first-order derivative and the sign of the second-order derivative of f. (Here P is arbitrary, subject to the assumptions given in the statement.) Justify your claims.
- **d. (7 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.

[2]This problem and the next are good if you have a bit of background in probability, rigorous statistics, stochastic processes, or something similar; if the notation and language seem a bit alien, consider one of the other problems.

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

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

6. (30 points) Maintain the assumptions and notations in Problem 5.

- **a. (5 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 $n = 0, 1, 2, 3$ and compute them for $n = 4, 5, 6, 7$ as well (no need to plot the latter).

- **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 $n = 0, 1, 2, 3$, and compute them for $n = 4, 5, 6, 7$ as well.

- **c. (10 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.) Justify your answer clearly.

7. (30 points) This exercise is for those who know some basic programming (in Python, MATLAB, or any other reasonable tool for quantitative simulation) and enjoy understanding things via simulations. Attach your code to your solution. Consider the branching process studied in Easley and Kleinberg 21.2/21.8.A. Recall that X_n is the number of newly infected individuals at wave n.

- **a. (5 points)** The analysis from Lecture 1 tells us that $\mathbb{E}[X_n] = (R_0)^n$. This gives us a prediction of the mean number of infected, but nothing about its variability. Choose five combinations (k, p) with $kp = 1.5$ and compute X_{20}. Do at least 10,000 simulations for each combination and report the empirical standard deviation of X_{20} in your simulations.

- **b. (5 points)** Imagine that you see one realization of a branching process, and observe that in this realization $X_{20} = 30,000$. Discuss how you could use your work in (a) to statistically test (and possibly reject) the null hypothesis that $k = 3$ and $p = 0.5$.

- **c. (5 points)** Do the simulations of (a) again but keep only the simulations where $X_{10} \geq 30$. For those simulations, what is the fraction of cases in which $X_{20} \geq 3000$? How does this compare to the unconditional frequency of the same event (i.e. the frequency of $X_{20} \geq 3000$ without throwing away any simulations)?

- **d. (5 points)** Explain what is going on in (c).

- **e. (10 points)** Do one other simulation that teaches you something interesting about the behavior of the model. Describe your question, the simulation you ran, and how the simulation sheds light on your question.