MATH 210, PROBLEM SET 6

DUE BY E-MAIL TO HAO ZHANG BY 5 P.M. APRIL 28.

Email of Hao Zhang: zhangphy@sas.upenn.edu

1. Random walks

Suppose that a virus particle makes a random walk on the real line by starting at some integer z and by flipping a fair coin at each time step to decide whether either to move one unit to the right or one unit to the left. Thus if Y_i is the random variable giving the position of the particle at time $i = 0, 1, 2, \ldots$, we have $Prob(Y_0 = z) = 1$, and for $i \geq 0$, $Y_{i+1} - Y_i = X_{i+1}$ is the random variable with density function f_{X_i} defined by

$$f_{X_{i+1}}(1) = Prob(X_{i+1} = 1) = 1/2 \quad \text{and} \quad f_{X_{i+1}}(-1) = Prob(X_{i+1} = -1) = 1/2.$$

Suppose now that the initial position z lies in $\{0, 1, 2, \ldots, N\}$. Let T_z be the random variable with values in $\{0, 1, 2, 3, \ldots,\} \cup \{\infty\}$ that gives the first time that the virus reaches either 0 or N.

1. Explain why we can take as a sample space the set S of all infinite sequences

$$\{s = (s_1, s_2, \ldots,) : s_i = \pm 1\}$$

with $X_i : S \to \{\pm 1\}$ sending $s \in S$ to its i^{th} component s_i. Describe $T_z(s)$ explicitly for $s \in S$. What should $T_z(s) = \infty$ mean?

2. Recall that $E(T_z)$ is the expected value of the random variable T_z. If positions 0 and N represent the front and the back of a bus, respectively, and a passenger at position z sneezes, what is the relevance of $E(T_z)$?

3. Show that $E(T_0) = E(T_N) = 0$. Then show

$$E(T_z) = 1 + \frac{1}{2}(E(T_{z+1} + E(T_{z-1})) \quad \text{for} \quad z \in \{1, \ldots, N-1\}.$$

Hints: If $0 < z < N$, then the virus must first take a step to the right or left in order to eventually reach 0 or N. Try showing that if $0 < z < N$, then for all $i \geq 0$,

$$\begin{aligned} Prob(T_z = i) &= Prob(T_z = i \text{ and } X_1 = 1) + Prob(T_z = i \text{ and } X_1 = -1) \\ &= Prob(T_z = i | X_1 = 1) \cdot \frac{1}{2} + Prob(T_z = i | X_1 = -1) \cdot \frac{1}{2} \\ &= Prob(T_{z+1} = i - 1) \cdot \frac{1}{2} + Prob(T_{z-1} = i - 1) \cdot \frac{1}{2} \end{aligned}$$

Then plug this into the definition of $E(T_z)$ and use the definitions of $E(T_{z+1})$ and $E(T_{z-1})$. You can take for granted that for all $z' \in \{0, \ldots, N\}$,

$$\sum_{i=0}^{\infty} Prob(T_{z'} = i) = 1. \tag{1.1}$$

For extra credit, try to prove (1.1) by using the central limit theorem to show that as $M \to \infty$, the set of M successive coin flips $(s_1, s_2, \ldots, s_M)$ such that

$$0 \leq z' + \sum_{i=1}^{M} s_i \leq N$$

has probability going to 0 as $M \to \infty$ among all possible sequences of M flips of a fair coin.

4. Show that the function $h(z) = z \cdot (N - z)$ for $0 \leq z \leq N$ and $h(z) = 0$ for all other z satisfies $h(0) = h(N) = 0$ and $h(z) = 1 + \frac{1}{2}(h(z+1) + h(z-1))$ if $z \in \{1, \ldots, N-1\}$. Suppose $\tilde{h}(z)$ is another function that satisfies these conditions. Show that the function $g(z) = \tilde{h}(z) - h(z)$ satisfies

$$g(z) = \frac{1}{2}(g(z+1) + g(z-1)) \quad \text{for} \quad z \in \{1, \ldots, N-1\} \quad \text{and} \quad g(0) = g(N) = 0. \tag{1.2}$$

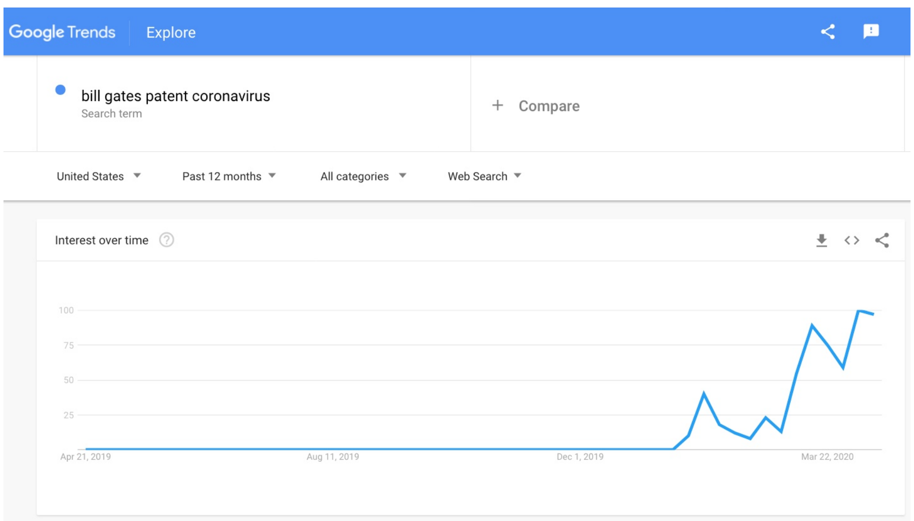
The first equality in (1.2) is sometimes called a discrete Laplace equation, and the last two equalities are called boundary conditions.

5. Show that the only function $g : \{0, \ldots, N\} \to \mathbb{R}$ satisfying (1.2) is the zero function. (Hint: Suppose the maximum of g over $\{0, \ldots, N\}$ occurs at z. Show that if $0 < z < N$ then (1.2) forces $g(z-1) = g(z) = g(z+1)$. Conclude from this that g has to attain its maximum at either 0 or N. Use the same argument for the minimum of g, and then use $g(0) = g(N) = 0$.)

6. Conclude from problems 4 and 5 that $E(T_z) = z \cdot (N - z)$. If $N = 2z$, corresponding to the passenger in problem 2 being exactly in the middle of the bus, what is the expected number of time steps till the virus particle reaches either the front or the back of the bus? By what factor does this change if the length of the bus is doubled?

2. Epidemic modeling of a conspiracy theory

Since he started criticizing the U.S. response to the covid-19 epidemic, Bill Gates has become a target for conspiracy theorists. One particular theory has been that Gates patented the covid-19 virus in order to profit from the epidemic. This theory is absurd. A British group, the Pirbright Institute, received funding from Gates and took out a patent on a potential vaccine for a different coronavirus which affects poultry. However, shortly after this patent was awarded, a conspiracy website, Infowars, claimed falsely that it was for "the deadly virus". (See "Bill Gates, at odds with Trump, becomes a right wing target," New York Times, April 17, 2020.) This has prompted an explosion of google searches for "Bill Gates patent coronavirus," as indicated by this graph from google trends:



This graph has the characteristic shape of a breakout epidemic. In this problem we are going to use the epidemic model discussed in class to model this particular conspiracy theory.

The population we will consider is the set of people who are prone to believing conspiracy theories. Let t be time, measured in weeks, and let $t = 0$ correspond to Jan. 19, 2020, which is the first date when searches for "Bill Gates patent coronavirus" appeared. We will set $S(0) = 1$ since at time 0, we are interested only in the population which is susceptible to conspiracy theories.

Let $I(t)$ be the proportion of the population of people prone to conspiracy theories who at time t consider the Gates conspiracy credible enough to be willing to conduct a google search for "Bill Gates patent coronavirus" at time t. In other words, $I(t)$ is the proportion infected by the Gates conspiracy theory. We will suppose that $I(t)$ is proportional to the actual number of such google searches being conducted at time t, by a proportionality constant that does not depend on t. Let $S(t)$ be the proportion of the population prone to conspiracy theories who are susceptible to making such a search about the Gates conspiracy theory, but who have not yet been infected by the Gates conspiracy theory in the above way. Let $R(t) = 1 - S(t) - I(t)$ be the proportion not in either of these groups. So $R(t)$

consists of people susceptible in general to conspiracy theories, but who have recovered from thinking the Gates conspiracy theory is credible.

We will suppose $S(t) + I(t) + R(t) = 1$ and there are differential equations

$$\frac{dS}{dt} = -\alpha SI \quad \text{and} \quad \frac{dI}{dt} = \alpha SI - \delta_2 I \tag{2.3}$$

for rate constants $\alpha, \delta_2 > 0$, as discussed in the course notes. We would like to know the largest value $I(t)$ attains. This represents the maximum proportion of the susceptible population who will consider the Gates conspiracy theory credible.

7. At time $t = 0$ we have $S(0) = 1$ since $S(t)$ is the proportion of the population of conspiracy minded people who are susceptible to believing the above conspiracy theory about Gates. We will suppose $I(0) = 0$. Writing

$$\frac{d(\ln I)}{dt} = \frac{1}{I}\frac{dI}{dt} = \alpha S - \delta_2 \tag{2.4}$$

we see that the limit

$$\lim_{t\to 0^+} \frac{d(\ln I)}{dt} = \alpha S(0) - \delta_2 = \alpha - \delta_2 \tag{2.5}$$

is the exponential growth rate of $I(t)$ near $t = 0$. One can approximate this as the slope of the graph of $y(t) = \ln I(t)$ for when t is near 0. On google trends, one can click the down arrow underneath a time graph of a search trend to get an excel spreadsheet of the data used to produce the graph. This leads to the following data describing the above graph when t is measured in weeks and $t = 0$ is January 19, 2020:

Time	1	2	3	4	5	6	7	8	9	10	11	12
Searches	25	14	7	7	24	3	54	67	58	38	100	97

Taking $I(t)$ to be proportional to the number $C(t)$ of searches, we get

$$\ln I(t) = \ln C(t) + a$$

for a constant a which is independent of time. So the slope of the graph of $y = \ln C(t)$ as a function of t should be the same as the slope of $y = \ln I(t)$, and for t near 0 this slope should be $\alpha - \delta_2$ by (2.5). Use the above table and the linear fit function in Wolfram alpha to estimate $\alpha - \delta_2$. You can do this by first googling Wolfram alpha, then entering "linear fit" into the search window and then using the example template which comes up.

8. We need now an estimate for δ_2. As discussed in the class notes, one such estimate is $\delta_2 = 1/T$ when T is the time needed for a member of the infected population to become removed. In this case, T would be the number of weeks on average which it takes a person believing the Gates conspiracy theory to drop this theory. This video from Bill Nye suggests 2 years is the minimal time it takes for someone believing in conspiracy theory to be persuaded it is not true:

   https://www.youtube.com/watch?v=MFLtTK13G2w

   What value of δ_2 does $T = 2$ years lead when we measure time in weeks? Using problem #7, what value do you get for α? Recall that α is the rate at time 0 at which one infected person creates new infections. This corresponds to the rate, in persons per week, that one person believing the Gates conspiracy brings onboard to thinking the Gates conspiracy theory is credible.

9. In the notes there is a derivation that at the time t^* when the proportion $I(t^*)$ of people believing the Gates conspiracy theory is largest, one will have

$$I(t^*) = 1 - \frac{\delta_2}{\alpha} \cdot (1 - \ln(\frac{\delta_2}{\alpha})) \tag{2.6}$$

What value do you get for $I(t^*)$?

10. Show that

$$\lim_{c \to 0^+} 1 - c(1 - \ln(c)) = 1.$$

What does this say about the maximal size $I(t^*)$ of $I(t)$ as the constant $c = \delta_2/\alpha$ tends toward 0? Note that if δ_2 is a fixed positive constant, then $c \to 0$ as $\alpha \to \infty$, where α measures the infectivity of a conspiracy theory.

3. Extra Credit Problems: Log normal distributions

These problems are strictly optional and can be turned in at any time before the date of the final exam.

Dave Eggers in his recent novel, "The Captain and the Glory," describes the captain of a ship who enjoys a random approach to steering:

"He nudged the wheel a bit left, and the entire ship listed leftward, which was both frightening and thrilling. He turned the wheel to the right, and the totality of the ship, and its uncountable passengers and their possessions, all were sent rightward. In the cafeteria, where the passengers were eating lunch, a thousand plates and glasses shattered. An elderly man was thrown from his chair, struck his head on the dessert cart and died later that night. High above, the Captain was elated by the riveting drama caused by the surprises of his steering."

Suppose the leader of a country adopts the same approach to containing an epidemic. Each day, the leader flips a fair coin and tries a new strategy. With equal likelihood, that strategy will either multiply the growth rate of the epidemic by a fixed constant $r > 1$ or multiply it by r^{-1}. Let Y_N be the random variable given by the net amount by which the growth rate is multiplied after N days.

11. Find an explicit expression involving binomial coefficients for the probability density function f_{Y_N} of Y_N. Here $f_{Y_N}(r^a)$ is the probability that Y_N equals r^a as a ranges over all integers.

12. Show $f_{Y_N}(z) = f_{Y_N}(z^{-1})$ for all non-zero real numbers z, and $f_{Y_N}(0) = 0$. Conclude that Y_N and Y_N^{-1} have the same density function.

13. Show that the expectation $E(Y_N)$ satisfies $E(Y_N) > 1$. (Hint: First show $z + z^{-1} \geq 2$ for all real $z > 0$, with equality if $z = 1$). What would you say about this if the leader says his approach has an equally likely chance of improving the situation or making it worse, so what have we got to lose?

14. We showed in class that as $N \to \infty$, the probability density function of Y_N approaches a log normal distribution with expectation

$$e^{N(E(X)+\sigma(X)^2/2)}$$

when X is the random variable with values in $\{\ln(r), -\ln(r)\}$ corresponding to a fair coin flip. Why is this consistent with problem 13?

15. Suppose that the fair coin flip represented by X in problem 14 is replaced by the flip of an unfair coin, corresponding to a random variable $\tilde{X}$ with density function given by $Prob(\tilde{X} = 1) = p$ and $Prob(\tilde{X} = -1) = 1 - p$. Suppose that if $\tilde{X} = 1$ on a given flip, the growth constant of the epidemic is multiplied by r and if $\tilde{X} = 1 - p$ it is multiplied by r^{-1}. This means that if $\tilde{X} = -1$, the growth rate is decreased. Find all p such that $\lim_{N \to \infty} E(Y_N)$ is not $+\infty$. What does this say about what is necessary for the leader's chaotic decision process to lead eventually to a lower growth rate?