CS584 Spring 2015, Problem Set 2
Due Thursday March 19, 11:59pm

Ymir Vigfusson

Problem 1

It's time for some game theory! In this problem, we will focus on 2-player games like we considered in class. The payoff matrix has two rows, corresponding to the strategies of the row player (player 1), and two columns, corresponding to the strategies of the column player (player 2). Suppose the two players represent two noble companies that are trying to improve the quality of people's lives by selling them "food that they want". The row player has a choice between manufacturing salty pretzels or even saltier potato chips. The column player is deciding whether to enter the market with a sugary "sports" drink, or an even sweeter variant of soda.

		Player 2	
		sports Drink	Soda
Player 1	Pretzels	$2, 15$	$4, 20$
	Chips	$6, 6$	$10, 8$

(a) What are the pure (*i.e.,* not mixed/randomized) Nash equilibria in the game above? Note that there could be multiple.

(b) Suppose there is massive competition on the drink and snacks markets. Consider changing the payoff matrix to the following. What are the pure (*i.e.,* not mixed/randomized) Nash equilibria in this game?

		Player 2	
		sports Drink	Soda
Player 1	Pretzels	$3, 5$	$4, 3$
	Chips	$2, 1$	$1, 6$

Note that there could be multiple.

(c) The repertoire of the two companies spans pretty much all the healthy foods. Now the two companies are contemplating going head-to-head on the same market: ice cream and candy. Consider the following payoff matrix.

		Player 2	
		Ice cream	Candy
Player 1	Ice cream	$1, 1$	$4, 2$
	Candy	$3, 3$	$2, 2$

Find *all* the Nash equilibria of this game, including mixed strategy equilibria.

Hint: Recall that to find a mixed strategy equilibrium, we define p to be the probability that Player 1 chooses to manufacture Ice cream (and Candy with probability $1 - p$), and q to be the probability that Player 2 does so. If a player is using a mixed strategy equilibrium, as opposed to a pure strategy (*e.g.*, $p = 0$ or $p = 1$), then the player must have been indifferent between the two strategies they have. In other words, the two strategies must have equal expected payoff. If we focus on Player 1, then their expected payoff for the Ice cream strategy, $q + 4(1 - q)$ must equal the expected payoff for the Candy strategy, $3q + 2(1 - q)$, when Player 2 is using strategy q.

Problem 2

Drivers between Atlanta and Athens in Georgia have effectively two routes to choose from: route 29 via Lawrenceville and route 78 via Conyers. Let us imagine some 100 drivers begin in Atlanta and wish to end up in Athens, and for simplicity assume that the roads are one-way. We'll measure the cost of travel in both the time wasted on the road, as well as the cost of fuel – none of the roads currently have any tolls. The road from Atlanta to Lawrenceville has a cost-of-travel of $0.5 + \frac{x}{200}$ where x is the number of travelers over the road. After Lawrenceville, highway 29 is pretty efficient and has a cost of 1 for everyone. The other route from Atlanta begins with the nicely wide Interstate 20 at a cost-of-travel 1 regardless of the number of drivers, but after Conyers the more narrow route 78 kicks in with a cost-of-travel of $0.5 + \frac{y}{200}$ where y is the number of concurrent drivers taking that road.

(a) Draw a picture of the network we described, labeling edges with the cost-of-travel. Note that the edges are directed arcs since we assumed the roads to be one-way.

(b) The 100 drivers simultaneously choose which of the two routes to use. Find the Nash equilibrium values of x and y on this network.

(c) To facilitate faster travel, Georgia – in a record first – contracted local wizards, shamans and Avengers from all creeds to come together and create a traffic portal: a wormhole in spacetime between Lawrenceville and Conyers. The wormhole is large enough to fit a wide highway and allow the drivers to incur negligible cost-of-travel. However, due to cosmic interference with negative energy flares from the Tesseract, the portal is one-way: accepting only traffic from Lawrenceville to Conyers.

Find the Nash equilibrium for this road network that includes the traffic portal.

What happens to the total cost-of-travel as a result of the wormhole?

(d) Georgia decides that despite the wormhole, traffic times between Atlanta and Athens are still not ideal. They decide to collect toll to the road between Atlanta and Lawrenceville, charging each driver 0.125 for using the road. The toll is collected automatically using the Georgia Peach Pass. The government also decided to further incentivize drivers to use the other route, and will give a subsidy of 0.125 to drivers who use the Atlanta-Conyers road. (Think of a subsidy as negative toll – it decreases the cost-of-travel for that player). Find the new Nash equilibrium in this setting.

(e) Notice how the toll and subsidies in part **(d)** balance out in the Nash equilibrium, so there is no effective cost (or revenue) to the state. This is a method governments sometime use to force a particular outcome in game-theoretic settings. But how does the total cost-of-travel change between **(c)** and **(d)**? What is a possible explanation for the difference? Can you think of similar break-even tolls between Lawrenceville and Conyers and from Conyers to Athens that would further reduce the total cost-of-travel?

Problem 3

A seller is auctioning off a very precious pair of vintage lederhosen in a sealed-bid second price auction. If a bidder who has value v secures an item after bidding p, the payoff is $v - p$, whereas a bidder who does not win the item has a payoff of 0. Two bidders, Tobias Fünke and Barry Zuckerkorn, are anxious to get their hands (or rather legs) on the item, and decide to collude, thinking they can at least share the item sometimes. Let's suppose their independent private valuations for the lederhosen are values uniformly chosen from $[0, 1]$. Their objective is to maximize the sum of their payoffs. The bidders can submit any bid in the range $[0, 1]$ (think of the numbers as having been normalized).

(a) Suppose Tobias and Barry are the only ones making a bid on the lederhosen. How much should each of them bid? Explain your answer.

(b) Now suppose a third bidder, Lucille Austero, enters the picture. She is not a part of the Fünke-Zuckerkorn collusion. Does the fact that she's also bidding on the item change the optimal bids for Tobias and Barry? How so? Explain.

Problem 4

Let's do some statistical analysis of the books *Les Misèrables* by Victor Hugo and *Dracula* by Bram Stoker from Project Gutenberg. Download the following versions that have one word on each line (UNIX format, UTF-8):

- Les Misèrables: http://www.mathcs.emory.edu/ cs584000/material/les-miserables.txt
- Dracula: http://www.mathcs.emory.edu/ cs584000/material/dracula.txt

(a) Draw a log-log plot of the word frequencies for each file. This means that the x-axis should show the frequency of a word (how often it appears in the text) and the y-axis shows the fraction of distinct words in the text that have that frequency. A particular point (x, y) on the plot then means that there are y distinct words in the book that have exactly x appearances in the text.

(b) Recall from lecture and the book that the probability density function (pdf) of the power-law distribution f where $f(x) \propto x^{-\alpha}$ is

$$\mathbb{P}[X = x] = \frac{\alpha - 1}{x_{\text{min}}} \left(\frac{x}{x_{\text{min}}} \right)^{-\alpha}. \tag{1}$$

Derive an expression for $\mathbb{P}[X \geq x]$, known as the complementary cumulative distribution function (CCDF), in terms of α and x_{min}.

(c) We now want to estimate the power-law exponent from the data. On a log-log plot, observe that a true power-law function $f(x) \propto x^{-\alpha}$ becomes a straight line: $\log(f(x)) \propto -\alpha \log(x)$. The slope of the line is then $-\alpha$, negative the exponent we are searching for! When we look at real data, we don't expect a fully straight line, but something that kind of looks like it over some orders of magnitude[1].

Let's first take a naïve approach. Create a histogram of frequencies of x. One way to do this is to round the values of x to the nearest integer (or perform some other division into bins), and tally up the number of values that end in the same bin. We then plot that histogram on a log-log scale, and try to assess the slope of the line that materializes. We can do this using *least squares linear regression*, which is a fancy way to say "find the best line that goes through the data on the plot", measured by square of the distance between a data point and the value of the line we found. The square here has the effect of giving lower penalty to *two* data points that are some distance away from the line we found than a *single* data point that's *twice* as far away from the line. Thus we're okay with lines that stay relatively close to all the points, but dislike lines where there are points that are far away from the line.

Include a plot showing the data, your line, and the power-law constant α that you found. (You can ignore x_{min} for now).

Hint: Gnuplot is very easy to use for this kind of data exploration. The following commands are useful. Suppose you have a TAB separated file called `filename` containing the values `x f(x)`.

```
plot "filename" using 1:2 title "Data"
a = 2    # Initial values for the regression
b = 1000
f(x) = b*x**-a
fit f(x) "filename" using 1:2 via a,b
plot "filename" u 1:2 t "Data", f(x) t sprintf("Regression a=%2.2f",a)
```

[1]Note that there are many distributions (such as log-normal) that kind of look like straight lines on a log-log plot. A big effort has been made recently to be more precise about what data really exhibits power-laws and which do not.

(d) Repeat (c) but using the empirical CCDF of the data.

Hint: You'll want to take each data point x in the data, and calculate how many data points $F(x)$ exceed or equal x. Then use the same script as in (c) on the file containing `x F(x)`. When you report the value of α, remember your result from part (b) !

(e) **(OPTIONAL)** We can go further and use maximum likelihood estimators (MLE) to estimate the power-law exponent from the data. The idea behind the statistical concept of likelihood is roughly "assuming that the data we saw was generated by a model with some particular parameters, what parameter values are the most likely to have produced it?" For example, if you observed a sequence of coin flips by a biased coin to produce 90% heads and 10% tails, then it is most likely that the bias of the coin is 90-10 heads to tails.

Suppose the data we collected follows the discrete probability density function from Equation (1), and let's refer to the data as $f(x_1), f(x_2), \ldots, f(x_n)$. (Recall that $f(x_i)$ is the fraction of the data that had x-value x_i). Our question is: "if the data actually came from a power-law distribution, what would be the most likely value of α?" (For convenience, let's fix $x_{\text{min}} = 1$ although it is technically another parameter).

We define the likelihood function $\mathcal{L}$ in terms of a joint density function (over all of the observed data), and assume that each data point is identically and independently drawn from the unknown power-law distribution with parameter α. Formally,

$$\mathcal{L}(\alpha; x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\alpha) = f(x_1|\alpha) \cdot f(x_2|\alpha) \cdots f(x_n|\alpha).$$

It is often easier to actually compute at the *logarithm* of the likelihood function, called log-likelihood, since it converts the products into more workable sums. Conveniently, the maximum of both the likelihood and log-likelihood functions are taken at the exact same value, so whatever parameters we derive from a log-likelihood estimate hold true for the original likelihood function.

Continuing the derivation above, the log-likelihood can now be computed:

$$\begin{aligned}
\ln \mathcal{L}(\alpha; x_1, \ldots, x_n) &= \ln\left(\prod_{i=1}^{n} f(x_i|\alpha)\right) \\
&= \sum_{i=1}^{n} \ln f(x_i|\alpha) \\
&= \sum_{i=1}^{n} \ln\left(\frac{\alpha - 1}{x_{\text{min}}}\left(\frac{x_i}{x_{\text{min}}}\right)^{-\alpha}\right) \\
&= \sum_{i=1}^{n} \ln(\alpha - 1) - \alpha \ln(x_i),
\end{aligned}$$

where we plugged in the density function for f, and $x_{\text{min}} = 1$.

Now we have derived a function of α that will tell us how likely it is that the observations we saw in our data $(x_1, \ldots, x_n)$ came from an actual power-law distribution with slope parameter α.[2] The next question is then: what is the *best* value for α to choose? We can answer that by maximizing the function, high-school style. Differentiate the function above with respect to α and set it to zero. This gives

$$0 = \frac{n}{\alpha - 1} - \sum_{i=1}^{n} \ln x_i$$

or by isolating α,

$$\alpha = 1 + \frac{n}{\sum_{i=1}^{n} \ln x_i}.$$

This value of α is known as the maximum likelihood estimator (MLE) of the data.

Use the MLE formula to estimate the α value from the data, showing your work. How does the value compare to the estimate derived from other methods?

[2] Actually, we computed the logarithm of that likelihood, but we also just argued that it doesn't matter