MA/CS 109 Lecture 18

Some Cautionary Tales...
Today: Three more case studies

1. (Last time) Admissions at Berkely and lurking variables.
2. Romney beats Obama in 2012
3. An example of (incredibly) poor journalism.
4. SIR Model of a flu epidemic
2. Romney beats Obama in 2012

Results of a Gallup poll, Nov 5, 2012

The Gallup organization is a well respected polling company. The poll predicts Romney will get 49% of the vote and Obama 48%. The margin of error was + or – 2%, so Gallup said it was too close to call...but Romney had the edge. The vote the next day was Romney 47%, Obama 51%
What happened?

The results of the election are outside the confidence interval at the 95% confidence level (the poll was of more than 3000 people).

Gallup could say “we were 95% confident...not 100% confident”
More Details

However, if you make your living predicting elections via polls, it is not good to miss an important prediction.

Soul searching at Gallup
http://www.huffingtonpost.com/2013/03/08/gallup-presidential-poll_n_2806361.html

Turns out, they do not use the Equally Likely Outcome Model to choose their samples. They call people on land lines (choosing numbers from directories ”at random”) and they call people
on cell phones by randomly dialing.

But some people have unlisted land lines... and lots of people don’t answer calls from strangers on their cell phones... and lots of people don’t like to talk to pollsters...

Also, lots of people don’t vote, but they have telephones—so to predict an election accurately you have to determine “likely voters”. But how do you determine likely voters?
Gallup addresses these issues using models. For example, they ask questions about voter registration and other information to try to determine likelihood a person will vote. And they “weight” some responses more heavily if they think it comes from a group that is underrepresented by the polling technique.

So they are on thin ice.....this is an occupational hazard for pollsters. They can’t force people to answer the phone or tell the truth (or have a land line!).
When you can’t use the Equally likely outcomes model to choose your sample, you must worry—is there a Central Limit Theorem that applies in this case??

(For the 2012 election, voter turn out was particularly high, so it could be the “likely voters” model that was at fault in the predictions.)
3. A Surprising Article in 2012

From the New York Times

http://www.nytimes.com/roomfordebate/2012/02/06/is-assads-time-running-out/syria-after-assad-could-be-even-worse

This refers to an article from the Guardian, a British news paper

http://www.guardian.co.uk/commentisfree/2012/jan/17/syrians-support-assad-western-propaganda

Results of a survey say 55% of Syrians wanted Assad to stay as president.
This would change my view of the whole situation in Syria....

What are the details? 
Survey done by “The Doha Debate”

http://www.thedohadebates.com/debates/item/index46d6.html

This page has disappeared....But a good discussion of the issues is at a BBC page

What it said: The details...

Poll of 1012 respondents...sample size 1012

\( \frac{1}{\sqrt{1012}} = 0.03 \) so margin of error is 3%...

Not bad.

But wait,
So only 97 respondents were in Syria.

So we should probably use a margin of error more like $1/\sqrt{97} = 0.10$ or 10%.

But wait, how was the sample determined?
• The research was conducted using YouGov’s regional online panel of 220,000+ respondents. Respondents from across the region (Arab world) were invited to participate in the survey.

• A total of 1,012 respondents completed the 5-7 minute survey. Fieldwork was conducted from the 14th to 19th of December 2011. (of whom 97 were in Syria)...

Oh man...this was “talk radio”. Participants in a particular club decided themselves if they wanted to participate.
Talk radio is the opposite of Equally Likely Outcomes model for choosing a sample.

On talk radio you only have a chance of hearing from those already listening to the station and then only those motivated to call in, wait in line, and willing to be publicly identified as a member of a certain group (not to mention the editing of the callers by the “host”).

Talk radio is evil...
The information from the survey is useless. It only shows that there are at least 55% of 97 people from Syria (or 53 people) who were not randomly selected thought Assad should remain president...

Completely the wrong impression was given by the NY Times article and The Guardian article The Doha Debates gave all the data it should on where its information came from... So this was bad journalism by NYT and Guardian.
Lesson’s Learned

1. Ask questions—How large was the study? How was it run? How was the sample chosen? How were the questions asked?

2. Assume there are things you don’t know… lurking variables—fight them with the Equally Likely Outcomes Model for choosing your sample..

3. Tools like the Central Limit Theorem are theorems inside a model universe. They are incredibly useful, but must be used carefully.
It is flu season...nobody wants to get the flu, particularly not the day before (or day of) an exam. What can mathematics say about the flu?

Well, remember the process...
Template for Doing Mathematics

Problem

| Model-----------------Repeat-----------------|
| Modify              |
| Examples/Conjectures|

Proof---Did we answer the question?---No

Yes—Fame $$$
Building the model of the spread of flu:

Divide the population into 3 groups:

Susceptibles (S): Those who can get the flu

Infecteds (I): Those with the flu who can spread it to others

Removed (or Recovereds (R)): Those who no longer spread the flu and can’t catch it again
Assumptions:

Assumption 1: People catch the flu at a rate proportional to the number of susceptible people and the number of infected people.

Assumption 2: People “recover” at a rate proportional to the number of infected.

Assumption 3: Recovered people are not susceptible or infectious ever again.
Equations

We need to relate the number of Susceptibles, Infecteds and Recovereds tomorrow with the number today—the process is the same as we have gone through before

\[ S_{\text{tomorrow}} = S_{\text{today}} - a \left( S_{\text{today}} \times I_{\text{today}} \right) \]
\[ I_{\text{tomorrow}} = I_{\text{today}} + a \left( S_{\text{today}} \times I_{\text{today}} \right) - b I_{\text{today}} \]
\[ R_{\text{tomorrow}} = R_{\text{today}} + b I_{\text{today}} \]
Parameters (constants) a

a: The proportionality constant a measures how likely it is for the flu to spread if a susceptible meets an infected. (How contagious is the flu.) We have some control over a—washing your hands, sneezing into your elbow, using the hand sanitizers all decrease the value of a.
Parameter b

b: This parameter measures how quickly you “recover” from the disease—or how quickly you stop being infectious.

We have some control of b too—people can stay away from others if they are infectious. In extreme cases governments can enforce quarantines. These all increase b.
What does the model predict?

We need to provide a starting condition and values of a and b.

Suppose $a=0.3$ and $b=0.1$.

We will let $S$, $I$ and $R$ represent fractions of the population. So all three are between 0 and 1 and $S+I+R$ always equals 1 (everybody is in a group).
\[ S_{\text{today}} = 0.999 \text{ (or } 99.9\% \text{ of the population)} \]
\[ I_{\text{today}} = 0.001 \text{ (or } 0.1\% \text{ of the population)} \]
\[ R_{\text{today}} = 0 \]

Then tomorrow

\[ S_{\text{tomorrow}} = S_{\text{today}} - 0.3 \left( S_{\text{today}} \times I_{\text{today}} \right) \\
= 0.999 - 0.3 \left( 0.999 \times 0.001 \right) = 0.9987 \]

\[ I_{\text{tomorrow}} = I_{\text{today}} + 0.3 \left( S_{\text{today}} \times I_{\text{today}} \right) - 0.1 I_{\text{today}} \\
= 0.002 + 0.3 \left( 0.999 \times 0.001 \right) - 0.1 \left( 0.001 \right) \\
= 0.0012 \]

\[ R_{\text{tomorrow}} = R_{\text{today}} + 0.1 I_{\text{today}} = 0 + 0.1 \left( 0.001 \right) = 0.0001 \]
Boring...let Excel do it \((a=0.3, \ b=0.1)\)

Blue=Susceptible,
Red=Infected,
Green=Recovered

Fraction of the population

![Graph showing the fraction of the population over days, with blue, red, and green lines representing susceptible, infected, and recovered states respectively. Days range from 0 to 100, and the fraction of the population ranges from 0 to 1.2.](image-url)
Pretty bad!

Predicts that ALMOST everybody (but not everybody) will get the flu before the epidemic dies out.

Also at about 40 days about 30% of the population will be sick. This would be a disaster!
React!!

Public health measures:
Wash hands, sneeze into your elbow, use hand sanitizer...Get a from 0.3 to 0.27

Stay home when you realize you have the flu.
Get b from 0.1 to 0.12
New $a=0.27$ and $b=0.12$

Blue=Susceptible,
Red=Infected,
Green=Recovered

Fraction of the population

Days

0 20 40 60 80 100 120 140

0 0.2 0.4 0.6 0.8 1 1.2
Better

The epidemic evolves more slowly, only 20% of the population is sick at the worst time and only about 85% of the population ever gets the disease (15% escapes...even though they are as susceptible as everyone else!!).
Vaccine!

Suppose we can get 40% of the population vaccinated before the flu arrives. Then S starts at only 0.599 and R starts at 0.4...

What effect will this have in each case?
40% vaccinated, $a=0.27$, $b=0.12$

Blue=Susceptible
Red=Infected
Green=Recovered

Max infected 3%. Total infected by flu over 6 months is only 30%
Important predictions

The details of the predictions can be very useful in public health planning. For example, knowing when that the peak can be delayed by public health measures...

For biology it is very interesting that the model predicts that the epidemic will end BEFORE everyone gets the disease (even though everybody was equally susceptible).
Hidden assumptions

This is the “SIR” model of epidemics. It does a good job predicting the spread but there are some hidden assumptions.

Most important is that the population is “well mixed”. Everybody has the same chance of running into an infected—e.g., a dorm, but probably not an entire college campus.
Morals

1. Learn from models (so you don’t have to learn from experience!!).
2. Use words carefully—they have precise meaning.
3. If you don’t follow the rules of your model, your results are not useful!
More Lessons

1. Talk radio is evil
2. Exponential growth can’t go on forever (change must happen).
3. Garbage in, Garbage out.
4. If you don’t get your vaccination, be sure to thank those that do for keeping you safer.