ISA 562	December 2, 2017

Homework 4

Question 1: In this problem, we will quantify how much is learned about an individual contributor to the "noisy poll" we computed in class. We will do this in two ways.

Approach 1 (Pen and paper): In this approach, we will compute the probability that a person voted for Trump, *conditioned* on the fact that they reported having done so in the poll. To understand conditional probability, it might help to think of an example. If you roll a die, the probability it comes up '6' is 1/6. However, if I told you that the roll came up even, then, conditioned on that information, the probability the value is '6' is now 1/3.

We'll let Trump denote the event that a student in fact voted for Trump, and we'll let $\widetilde{\mathsf{T}}$ denote the event that a student reported voting Trump. The conditional probability that we wish to compute can be written as follows. $\Pr[\mathsf{Trump} \mid \widetilde{\mathsf{T}}]$. Bayes's rule will be helpful here. It says that

$$\Pr[\mathsf{Trump} \mid \widetilde{\mathsf{T}}] = \frac{\Pr[\widetilde{\mathsf{T}} \mid \mathsf{Trump}] \cdot \Pr[\mathsf{Trump}]}{\Pr[\widetilde{\mathsf{T}}]}$$

We will treat $\Pr[\mathsf{Trump}]$ as a variable that captures our prior knowledge about the student in question. If we truly knew nothing about the student in question, and we knew nothing about how the general population voted, then we might set this variable as $\Pr[\mathsf{Trump}] = 1/2$. (This was effectively our assumption in class: we had $\Pr[\mathsf{Trump}] = 1/2$, since we flipped a coin to determine people's "true" votes.) On the other hand, if what we happen to know about the student in question is that they identify as male, then $\Pr[\mathsf{Trump}] = 62\%$. Answer the following questions.

(a) Calculate $\Pr[\widetilde{\mathsf{T}} \mid \mathsf{Trump}]$. (I'm looking for an exact number here.)

(b) Calculate $\Pr[\widetilde{\mathsf{T}}]$. Calculate this as a function of $\Pr[\mathsf{Trump}]$, leaving that term as a variable. Note that for any two probability events, A and B, $\Pr[A] = \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \overline{B}] \cdot \Pr[\overline{B}]$, where $\overline{B}$ is the probability that event B did *not* occur.

(c) Calculate exact values of $\Pr[\mathsf{Trump} \mid \widetilde{\mathsf{T}}]$ in the cases where $\Pr[\mathsf{Trump}] = 1$, $\Pr[\mathsf{Trump}] = .62$, $\Pr[\mathsf{Trump}] = .5$ and $\Pr[\mathsf{Trump}] = .25$.

(d) In the noisy poll we did in class, the probability that an individual gave a random response was 1/2. Suppose we modified the experiment, giving a true response with probability 1/4 and a noisy response with probability 3/4. Recalculate part (c) under this new experiment. What happens to user privacy in this case?

Approach 2 (Simulation): We will now verify the probabilities that we calculated in the previous approach by writing a program that estimates these probabilities experimentally. Write a program (in any language) that runs the same experiment we performed in class; however, instead of initializing the true votes using a random coin, explicitly set half of them to have voted for Trump, and half to have voted for Clinton. Assume there are 10,000 students responding to the poll. What fraction of the ones that respond that they voted for Trump actually did? Given that, compute $\Pr[\mathsf{Trump} \mid \widetilde{\mathsf{T}}]$. Then use/modify your program to answer (c) and (d) above. Submit the numerical answers, as well as (one version of) your code. (The only difference in these experiments will be the initialization of a few variables. We don't need to see all the ways you initialized these variables in the various experiments.)