COS 445 - PSet 3, Problem 2

Anakin Skywalker

March 31, 2021

Problem 2: Limited Information Scoring Rules

Part a

I don't know.

Part b

We wish to design a proper scoring rule which takes as input only the reported mean of X, and strictly incentivizes the predictor to report the true mean of X.

Let us define the scoring function $S(y, x) = -(y - x)^2$. To find the unique argmaximum, we compute the derivative of $\mathbb{E}[S(y, x)]$:

$$\begin{aligned} 0 &= \mathbb{E}[-2(y - x)] \\ 0 &= -2(y - \mathbb{E}[x]) \\ 0 &= y - \mathbb{E}[x] \\ y &= \mathbb{E}[x] \end{aligned} \tag{1}$$

Therefore, the unique argmaximum is $y = \mathbb{E}[x]$. We note that $\mathbb{E}[S]$ is taken by linearity of expectation, with y not being a random variable (and thus its expectation is always itself). If this is proper, than $y = \mathbb{E}[x]$ must be the global max. We know that S is strictly increasing on the left side, because its derivative $-2(y - \mathbb{E}[x])$ is positive for all $y < \mathbb{E}[x]$. Likewise, S is strictly decreasing on the right side because $-2(y - \mathbb{E}[x])$ is negative for all $y > \mathbb{E}[x]$. Therefore, $y = \mathbb{E}[x]$ is the global max and so S is a proper scoring rule.

Part c

We wish to prove that there does not exist any scoring rule S which takes only as input the reported variance of X that strictly incentivizes the predictor to report the true variance of X.

Let us assume for contradiction that such an S existed. We consider two random variables X and Y with the same variance, and a third distribution Z which is equal to X with probability $\frac{1}{2}$ and Y with probability $\frac{1}{2}$. We set the value of $X = 1$ with variance 0, and $Y = 2$ with variance 0. In other words, X and Y will be always be a constant value. We also assume that the scoring rule works for X and Y.

We then compute the unique argmaximum $\mathbb{E}[S(y, z)]$. By linearity of expectation, this can be related to the expectations of Y and Z. We also note that because the variance of X and Y is 0, taking the argmax for each one will also be 0 since the scoring function incentivizes the predictor to report the true variance. Since Z is equal to X with probability $\frac{1}{2}$ and Y with probability $\frac{1}{2}$, we compute the argmaximum as follows:

$$\mathbb{E}[S(y, z)] = 0.5 \cdot 0 + 0.5 \cdot 0 = 0 \tag{2}$$

However, this is a contradiction because the true variance of Z is in fact $\frac{1}{2}$. Therefore, it is not possible to create a scoring rule that incentivizes reporting the true variance when given only the reported variance as input.