Homework 2 — Data Science II
Decision Theory

Instructions

Please prepare a writeup following the general writeup instructions and submit on Brightspace. Please show all your work.

Problem 1. Consider the phrase, *"The map is not the territory"*. What does this mean in the context of decision analysis? Write an essay reflecting upon all the possible ways you can think of for a decision-theoretic analysis to *fail*. (Note this means only considering data-driven, rational analyses of problems, not failures due to humans making non-rational decisions.) Provide concrete examples of decision problems, including some from your own past experiences, as context for your reflections.

Problem 2. *Branching decision points*. Recall from class the example of the 95-yo man with a potentially malignant lung tumor. Doctors could recommend (R)adiotherapy, (S)urgery, or (N)o treatment. The medical literature gleans the following relevant statistics:

- There is a 90% chance the tumor is malignant.
- If the man does not have cancer, his life expectancy (LE) is 34.8 mo.
- If he does have cancer:
  - Choosing R, his LE = 16.7 mo.
  - Choosing S, there is a 35% chance of immediate death, but if he survives then his LE = 20.3 mo.
  - Choosing N, his LE = 5.6 mo.
- Either treatment will cause approximately one month of considerable discomfort. We decided to incorporate this into our utility function using Quality-Adjusted LE ($QALE$) = $LE - 1$ mo. for treatment options.

In class, we found that R was the preferred treatment for this man.

Now, suppose there is a fourth option, a ***biopsy***. This biopsy can test whether the tumor is malignant, and this information can inform a subsequent $R/S/N$ decision. Relevant facts:

- The test has a 75% chance of detecting cancer if the tumor is truly malignant.
- The test has a 4% chance of detecting cancer if the tumor is truly benign.
- In addition, there is an estimated 6% chance that the procedure to extract a tissue sample from the tumor will lead to fatal complications. Otherwise, the biopsy does not lead to patient discomfort.

Should the man choose to have a biopsy? Why or why not? Compute all the possible $QALE$s to justify your decision.

Hint: To answer this requires reasoning about the potential $R/S/N$ decision after (i) deciding whether or not to have the biopsy and (ii) if choosing the biopsy, whether the biopsy results were positive or negative. Bayes' Theorem is helpful in this regard as it allows us to estimate $\Pr(\text{cancer} \mid B)$, a new conditional probability for cancer given the biopsy result B. We can then use $\Pr(\text{cancer} \mid B)$ in our utility in place of the original prior cancer probability $\Pr(\text{cancer}) = 0.9$.

Problem 3. *Statistical Decision Theory*. Formulating statistical model selection as a decision problem (using "risk" as an equivalent to "utility") allowed us to derive new statistical estimators $\hat{\theta}$ for unknown parameter(s) θ. For example, in class, we sought as utility the minimization of Bayesian risk using a squared error loss function, and derived a Bayes estimator $\hat{\theta}(x) = E[\theta \mid X = x]$.

Derive the corresponding Bayes estimators under an *absolute error loss* and under a *zero-one loss*.

Hint: Recall that $|x|$ has a discontinuous first derivative. *Hint:* The *Dirac delta* $\delta(x)$ may be helpful (but not strictly necessary) due to the following useful property:

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a)$$

for at least the functions $f(x)$ we'll typically expect to encounter.