MATH 341 HW 3

Isaac Boaz

January 24, 2023

Problem 4

Measuring the effectiveness of treatment.

- RR: Relative Risk
- ARR: Absolute Risk Reduction
- T: Treatment Group $\implies \overline{T}$: Control Group
- B: 'Bad Outcome'

a) $RR = P(B|T)/P(B|\overline{T})$
RR should be > 1 if the treatment increases the chances of harm, and < 1 if it reduces the chances of harm. As such, the numerator represents people that got 'bad' despite being treated, and the denominator represents people that got 'bad' without being treated.

b) $ARR = P(B|\overline{T}) - P(B|T)$
The **A** in ARR stands for **A**bsolute, and as such, it is the difference between the two probabilities.

c) Is it possible to have a sutation where $RR \approx 0$, and$ARR \approx 0$?
Consider a "Rare disease", where

Yes, it is possible to have a low relative risk (RR) and a high absolute risk reduction (ARR) when the baseline risk of the outcome is low, and the treatment has a large effect on reducing the risk.

Example: Consider a treatment for a rare disease that affects only 1 in 10,000 people. If the treatment is given to 1000 people with the disease, it prevents one death. Without the treatment, one death would have occurred.

$$RR = \frac{(1/1000)}{1/10000} = 0.1$$
$$ARR = \frac{1}{1000} = 0.1\%$$

In this example, the relative risk is low (10% reduction) but the absolute risk reduction is high (0.1%).

d) Show that $P(B|\overline{T}) = \frac{ARR}{1-RR}$

$$\text{Let } a = P(B|\overline{T}) \text{ and } b = P(B|T)$$
$$\begin{aligned} \frac{ARR}{1-RR} &= \frac{a-b}{1-b/a} \\ &= \frac{a}{a} \cdot \frac{a-b}{1-b/a} \\ &= \frac{a(a-b)}{a-b} \\ &= a = P(B|\overline{T}) \end{aligned}$$

Problem 8

The following website compares different algorithms for predicting Australian credit approval results for 517 individuals. We focus on the confusion matrix (where each number is divided by the total) for the logistic regression model.

	Predicted Denied	Predicted Approved	Total
Actually Denied	249/517	38/517	287/517
Actually Approved	18/517	149/517	230/517
Total	330/517	187/517	1

a) Let A denote 'Actually Approved' and B denote 'Predicted Approved'. Show mathematically whether or not A and B are independent
We know A and B are independent if $P(A \cap B) = P(A) \cdot P(B)$.

$$
\begin{aligned}
P(A) &= \frac{230}{517} \\
P(B) &= \frac{187}{517} \\
P(A) \cdot P(B) &= \frac{230}{517} \cdot \frac{187}{517} \\
&= \frac{43010}{267289} = \frac{3910}{24299} \\
P(A \cap B) &= \frac{149}{517} \\
\frac{149}{517} &\neq \frac{3910}{24299}
\end{aligned}
$$

Since this is not the case, we can conclude that A and B are not independent.

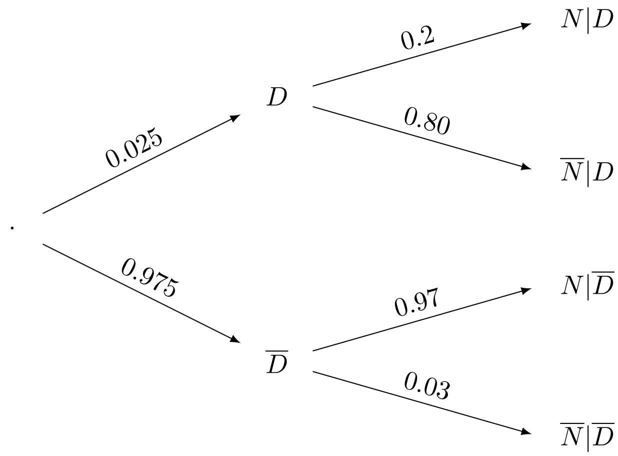
b) Does the result from (a) seem to indicate that the logistic regression model has some ability to predict the outcome correctly? In other words, discuss what must happen if the logistic regression model simply generates random predictions.
This model would not be able to accurately predict the outcome correctly, as the events are not independent.

c) Let C denote 'Actually Denied' and D denote 'Predicted Denied'. Without doing any calculation, discuss whether or not C and D are independent by referring to the result from (a) and one of the results mentioned in the lecture notes.
We can rewrite these as

$$
\begin{aligned}
C &= \text{'Actually Denied'} = \overline{A} \\
D &= \text{'Predicted Denied'} = \overline{B}
\end{aligned}
$$

Since we know that A and B are not independent, we can conclude that C and D are also not independent.

Problem 11



a) What is the probability that an individual tests positive?

$$\begin{aligned} P(\overline{N}) &= P(\overline{N}|D) + P(\overline{N}|\overline{D}) \\ &= 0.8 \cdot 0.025 + 0.03 \cdot 0.975 \\ &= 0.04905 \end{aligned}$$

b) Given an individual tests positive, what is the probability that they have the disease?

$$\begin{aligned} P(D \mid \overline{N}) &= \frac{P(D \cap \overline{N})}{P(\overline{N})} \\ &= \frac{P(\overline{N} \cap D)}{P(\overline{N})} \\ &= \frac{P(\overline{N} \mid D)P(D)}{P(\overline{N})} \\ &= \frac{0.8 \cdot 0.025}{0.04905} = 0.40774719673 \end{aligned}$$

c) Given an individual tests negative, what is the probability that they don't have the disease?

$$\begin{aligned} P(\overline{D}|N) &= \frac{P(\overline{D} \cap N)}{P(N)} \\ &= \frac{P(N \cap \overline{D})}{P(N)} \\ &= \frac{P(N|\overline{D})P(\overline{D})}{P(N)} \\ &= \frac{0.97 \cdot 0.975}{1 - P(\overline{N})} \\ &= \frac{0.97 \cdot 0.975}{1 - 0.04905} \approx 0.9945317 \end{aligned}$$

d) Comment on (b) and (c).
I found these results rather surprising, as it shows that WHO cares (statistically) more about false positives than false negatives. Considering the probability of having the disease given a positive test is significantly lower than the probability of not having the disease given a negative test, it goes to show how the multiplicative aspect probabilities can heavily impact raw statistics.