Admission probability distribution evaluation

This evaluation does not try to evaluate only the expected value of the result (no. of admissions) but instead evaluates how well the distribution yielded by the model correlates with reality.

Let denote the random variable of the number of admissions and be a threshold between and . corresponds to a certain number such that . The approximation is there because we are dealing with a discrete distribution. For example, suppose a cumulative distribution is given such that and , and let . We will make three approximations: a low, middle, and high approximation. In this example, a low approximation means corresponds to . A high approximation means corresponds to A middle approximation would mean that corresponds to the smaller if it is closer to and to the greater if it is closer to ; since 0.3 is closer to 0.35 in our example, corresponds to .

A probability distribution function is created each day based on the model’s output of that day. For a given , this pdf yields the value such that there is a chance that the number of admissions of that day is fewer than or equal . Using the data, we can answer whether the number of admissions that day is under or equal , or it is over . Over the period of time, we collect the number and separately collect a if the number of admissions that day is under or equal to and record otherwise. We then take the average of these collections, respectively, that is divide them by the number of predictions made in total.

What we have at the end is, for each , the pair of values:

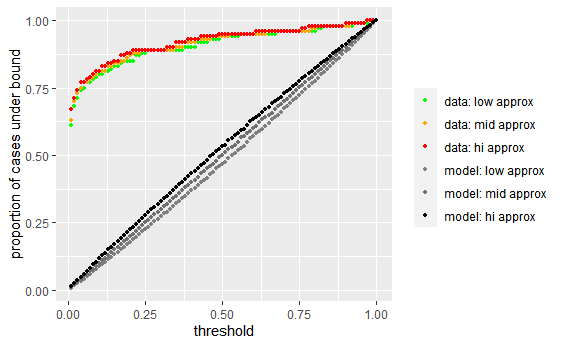
1. An average proportion of days at which the number of admissions would fall below the daily bound
2. The actual proportion of days at which the number of admissions fall below the daily bound

Because of the three ways we obtain the based on , the low, middle, and high approximation, we have three pairs of these values and we plot them against varying values of .

A summary of steps:

1. For each distribution, obtain the cumulative distribution
2. For each threshold , obtain the corresponding . ( Three ’s will be obtained for the low, middle, and high approximation )
3. Over the period of time in consideration, and for each collect the numbers and also record if the number of actual admissions is under or equal to and otherwise.
4. Take average by dividing all collection of numbers by the number of predictions made.

Using data of daily admissions for the first 100 days since the beginning of the data, we obtain the following plot.



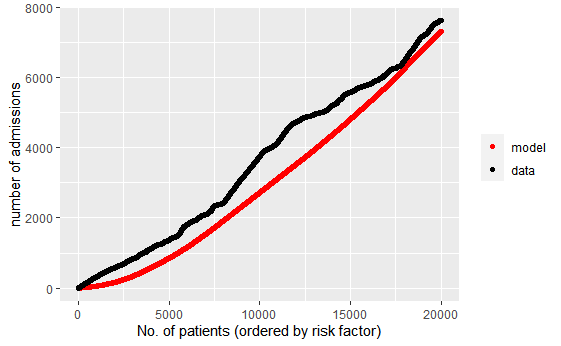
What can be concluded from this graph is that for any given and their corresponding given by the model, the proportion of days on which admission is less than is greater than expected, hence the model overestimates. I believe the reason for this large discrepancy is a difference in assessing the model’s prediction of admission on an individual level in the binary sense and assessing the quality of the probability of admission. The model can be good for the former but not the latter and when I use the probability to generate a distribution on an aggregate level, the poor quality will compound. Another note is that the model may be better when more days are taken into account; the reason I ran 100 days is a computational issue.

MADCAP

The model is evaluated at varying subset of patients based on the patients’ risk factor. In our case, a patient’s risk factor is the predicted probability of admission. The procedure in doing the evaluation is as follows:

1. Order patients by their risk factor
2. For a number , we look at first patients in increasing order of risk.
3. Using the model, we obtain the probability distribution of the number of admissions in this group
4. Take the expectation and count from the data how many admissions are there in reality
5. Plot this against varying values of

Below is the plot produced for Tim’s data using the first 20000 patients in order of probability of admission. The model starting off poorly indicates an error in prediction of patient admission at low risk. The model and data start to fit better when a patient is at higher risk.



Hosmer-Lemeshow Test

The Hosmer-Lemeshow Test determines whether the differences between the observed probability (from data) and the expected probability (from model distribution) is significant.

Suppose there is a dataset, and the model predicts the probability, given a set of inputs, of one of two outcomes (True or False). Then we can obtain an observed probability distribution by grouping similar inputs together and calculating the expected probability for each group. However, in most cases, the number of observations for each combination of inputs are few or none.

The Hosmer-Lemeshow statistic takes a different approach by grouping observations similar in expected probability, not in input. The statistic is given by

where is the number of groups, is the number of observations with an outcome of in group , is the expected number of observations with an outcome of in group . Usually, is picked to be around .

After calculating this statistic , the p-value is calculated using and compared to a predetermined level of significance, and the null hypothesis, that the expected and the observed is the same, is rejected when the p-value is lower than .

A summary of steps:

1. Split dataset into groups
2. For each group, find number of observed and expected admissions, and also find the number of observed and expected non-admissions.
3. Plug into equation above
4. Calculate p-value using chi-square distribution with degree of free = number of groups - 2

In our case, using Tim’s data and using 10 groups, the statistic value is ~ 80,000, yielding a p-value of essentially 0.