Conditional Probability: Continuous Test Values

BIOS 6611

CU Anschutz

Week 5

Background and Example

Motivation

We previously examined how to summarize a diagnostic test based on a confusion matrix. In that context, we assumed that the treadmill test had a known threshold to predict coronary heart disease that classified each person as either positive or negative.

However, what if we were given some continuous test or one with multiple categories instead of a dichotomous positive or negative outcome? How would we know whether someone should be classified as positive based on this test?

Receiver Operating Characteristic (ROC) Curves

One solution is to consider that there is a natural trade-off between sensitivity and specificity: as a test becomes more sensitive, the higher the false positive rate will be. Thus, specificity decreases.

Therefore, the cutoff or criterion value used for a diagnostic or screening test will determine the sensitivity and specificity of the test.

ROC curves are a visual way of showing the sensitivity/specificity trade-off as the cutoff varies across all possible values. They can be used to identify optimal cutoffs, and to compare one test to another, based on the area under the ROC curve. They can also be used to compare combinations of tests.

Example Dataset

We will use a subset of the baseline measures from the Framingham Heart Study ($n = 4434$), a long term prospective study of the etiology of cardiovascular disease among a population of free living participants in Framingham, MA. The data is available courtesy of BioLINCC from the NHLBI and can be found in our Canvas "Data Repository."

There are a variety of different measures collected in the dataset. For our purposes, we will focus on:

Outcome (Disease Status): `DIABETES` (is the participant diabetic)

Candidate Baseline Biomarkers: `GLUCOSE` (range: 40-394 mg/dL), `CIGPDAY` (range: 0-70 cigarettes per day), `BMI` (range: 15.5-56.8 kg/m^2)

ROC Curves In Practice

ROC Curves - the General Idea

To obtain ROC curves we take each possible cutoff value and create a 2x2 table showing the classification of positive (above the cutoff), or negative (below the cutoff) vs. the outcome of interest (e.g. gold standard: +/-, disease: yes/no, etc.).

From each table the corresponding sensitivity and specificity are obtained. These values are then plotted for several values of the cutoff.

The ROC curve specifically is plotted with the FPR (false positive rate = 1-specificity) along the x-axis and the true positive rate (TPR = sensitivity) on the y-axis.

Using the pROC Package

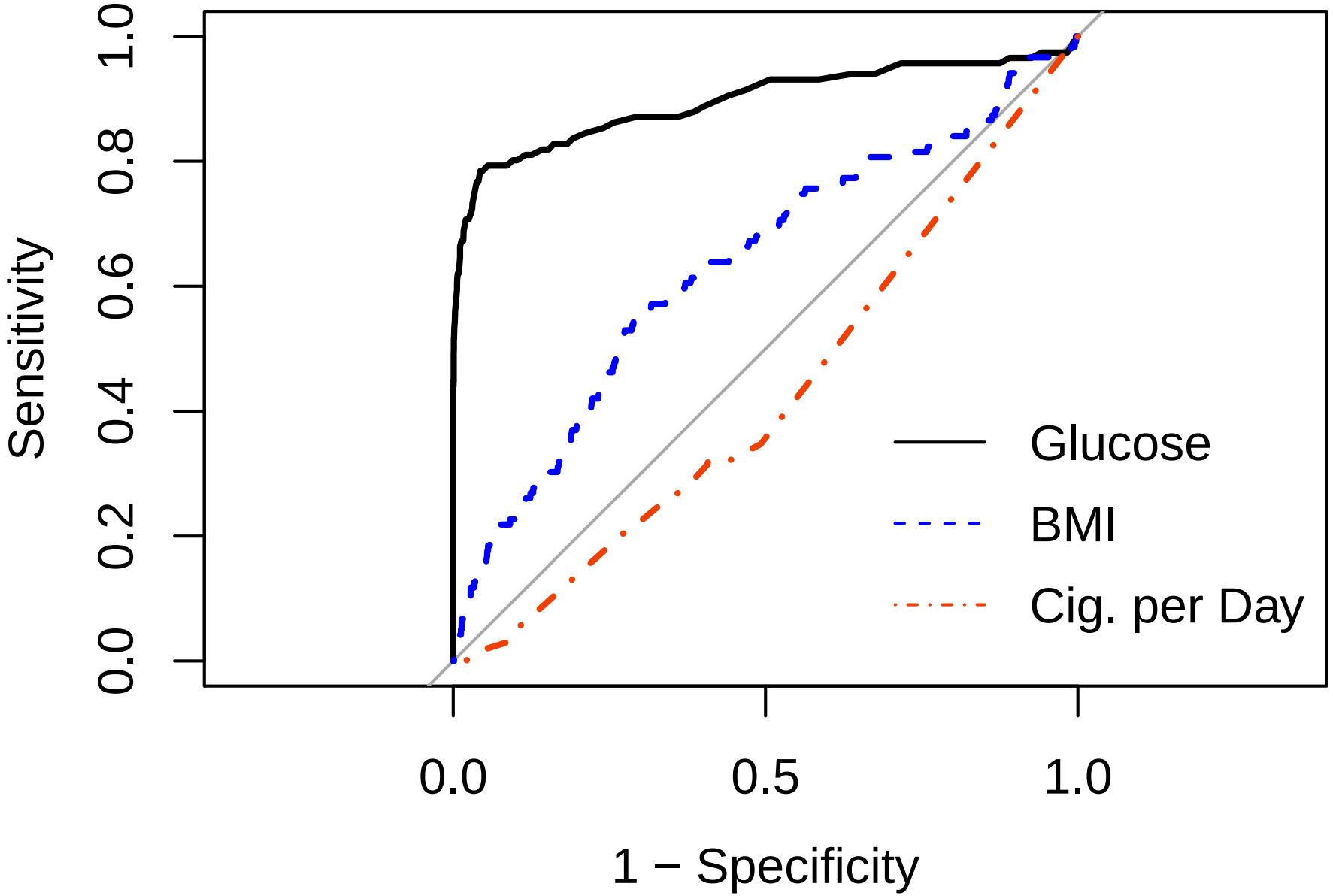
There are a host of packages that will generate and plot ROC curves (e.g., ROC, Epi, plotROC, ROCit, etc.), we will focus on the pROC package in this slide deck.

```
library(pROC)

# Read in our data for analysis
dat <- read.csv('frmgham2_baseline_subset.csv')

# Fit ROC curves
roc_glu <- roc(DIABETES ~ GLUCOSE, data=dat)
roc_bmi <- roc(DIABETES ~ BMI, data=dat)
roc_cig <- roc(DIABETES ~ CIGPDAY, data=dat)
```

ROC Curve Example



ROC Curve Example Code

```
# Note, `fig.height=3.5, fig.width=5` set in code chunk
# and \vspace{-15mm} to move plot up

# Create initial plot with glucose ROC
plot(roc_glu, legacy.axes=T)

# Add BMI ROC
plot(roc_bmi, add=T, lty=2, col='blue')

# Add Cig per Day ROC
plot(roc_cig, add=T, lty=4, col='orangered2')

# Add legend
legend('bottomright', lty=c(1,2,4),
       col=c('black','blue','orangered2'),
       legend=c('Glucose','BMI','Cig. per Day'),
       bty='n') # don't include black border around legend
```

Overall Performance: The Area Under the Curve (AUC)

The ability of a given predictor to discriminate between the outcome can summarize from the ROC curve using the **area under the curve** (AUC), which is the area that lies under the ROC curve.

The AUC can be interpreted as the probability that, for a randomly selected pair of a participant with and a participant without the outcome, the predictor of interest will rank higher for the participant with the outcome.

The AUC can range from 0 to 1:

- A test with an AUC of 1 has perfect accuracy
- A test with an AUC=0.5 is no better than random chance
- A test with an AUC<0.5 is actually worse than random chance

The AUC is also known as the c-index (concordance index) and is equivalent to a non-parametric (rank-based) statistic we will talk about later (the Mann-Whitney U-statistic).

AUC Example

```
auc(roc_glu) # AUC for the ROC curve for glucose
```

```
## Area under the curve: 0.8965
```

```
auc(roc_bmi) # AUC for the ROC curve for BMI
```

```
## Area under the curve: 0.6337
```

```
auc(roc_cig) # AUC for the ROC curve for cigs per day
```

```
## Area under the curve: 0.4279
```

These results suggest that glucose is an overall good predictor, followed by BMI. However, cigarettes per day is *worse than random chance*. Strictly speaking, `pROC` treats the order as important, so the increasing number of cigarettes per day does not match to an increasing likelihood of developing diabetes (based on our results).

Picking the "Best" Threshold

Oftentimes the ROC curve is a first step. To put the marker/predictor into practice, we need to choose a threshold that we believe is actionable.

In reality, there is rarely one "best" threshold. Rather, we should choose a threshold based on the given context and our acceptance of false positives or false negatives.

However, there are some mathematical approaches the are built into the `pROC` package that can give us a jump start:

- `best.method = 'youden'`: optimality is based on the coordinate(s) that achieve max(sensitivity + specificity), also called Youden's J statistic
- `best.method = 'closest.topleft'`: optimality is based on the coordinate(s) that achieve min((1-sensitivity)^2 + (1-specificity)^2)

"Best" Threshold Example

Using the `coords()` function we can extract the "best" threshold based on the two criteria on the previous slide. If we do not specify `x='best'` we will receive a data frame of all thresholds and the corresponding sensitivity and specificity.

```
coords(roc_glu, x='best',
       best.method=c("closest.topleft")) # closest top left
```

```
##   threshold specificity sensitivity
## 1     102.5    0.944657   0.7931034
```

```
coords(roc_glu, x='best',
       best.method=c("youden")) # Youden's J
```

```
##   threshold specificity sensitivity
## 1     104.5   0.9566437   0.7844828
```

These results suggest glucose has a higher specificity (true negative rate) than sensitivity (true positive rate) for both Youden's J and the closest top left coordinates.

Diagnostic Testing Summary

In our three sets of lectures we introduced how to evaluate the performance of a diagnostic test, how the probabilities of suspecting a disease changes with a test in a population with a given prevalence, and how to examine and evaluate the potential for continuous biomarkers/predictors.

In practice we may wish to optimize different parameters (sensitivity, specificity, NPV, PPV, etc.). So ultimately there may not be a perfect threshold or cut-off, but we can attempt to balance the trade-offs for different values or note a candidate marker is not promising for further study (e.g., AUC < 0.5).