Using Receiver Operating Characteristic (ROC) Analysis to Evaluate Information -Based Decision-Making

Nan Hu
University of Utah, USA

INTRODUCTION

Business operators and stakeholders often need to make decisions such as choosing between A and B, or between yes and no. These decisions include, but are not limited to, whether to invest in project A versus project B, or whether to continue running a company. These are often made by using a classification tool or a set of decision rules. For example, banks often use credit scoring systems to classify lending companies or individuals into a high or low risk of default, thus helping to decide whether to grant a loan. One important question businesses need to answer is how accurate the information based on these classification tools can help them make a correct decision, or how correctly they can be used to discriminate between two groups of subjects. In this chapter, we address this important issue by presenting accuracy parameters for assessing classification tools such as test modalities, scoring systems, and prediction models. Specifically, we introduce the receiver operating characteristics (ROC) curve as a statistical tool to evaluate these modalities. The ROC curve is widely used in business optimization analysis, health policy making, clinical studies, and health economics (Kampfrath & Levinson, 2013). In the Background section, we give updated examples of using the ROC related methods for assessing decision-makings based on our most current literature review. In the Main Focus section of this chapter, we provide mathematical definitions of the classification accuracy parameters, and describe the procedure to obtain an ROC curve. In addition, we present recent statistical developments in ROC curve methodologies and applications of ROC analysis in a diversity of research areas.

DOI: 10.4018/978-1-5225-2255-3.ch192

BACKGROUND

Business classification tools include scoring systems, predictive models, and quantitative test modalities. A classification tool is useful in business analytics only if it is shown to distinguish entities with a certain condition from those without that condition. For instance, a credit scoring system is a valuable classification tool for bankers when it can accurately classify between companies with default status (cases) and without default status (controls). A perfect test modality would categorize all default companies as cases and all non-default companies as controls. However, in practice, almost none of the testing modalities can make such a perfect classification. This implies that misclassifications can always exist and the correct classification rate may vary from one test to another. Thus, assessing classification performance among different test modalities is always a necessary step in making important business-related decisions.

MAIN FOCUS

We first define accuracy parameters of binary classification tools, and then extend the evaluation method to test modalities with continuous or discrete ordinal values. By applying accuracy parameters and ROC analysis, business analysts can easily examine the expected downstream harms and benefits of positive and negative test results based on these test modalities, and directly link the classification accuracy to important decision making (Cornell, Mulrow & Localio, 2008).

Accuracy Parameters for Classification and Decision Making

The accuracy of decision making should be measured by comparing the decision taken by a business to the choice that would be taken in order to maximize its benefit. In this section, we introduce two basic accuracy parameters, sensitivity and specificity, and two misclassification measures, the false positive rate and false negative rate. We define accuracy parameters in the context of classifying the default status of borrowers (companies that apply for a loan). Let S denote the dichotomous true default status such that $S = 0$ represents "no default," and $S = 1$ indicates "default." Let Y be the value of a test modality or scoring system. We suppose that Y is also binary such that $Y = 1$ denotes the test positive for default, and $Y = 0$ indicates the test negative. In reality, companies with a positive test result are often refused for a loan. The *sensitivity* of the binary test Y is defined as the probability of test positive among companies with default status ($S = 1$). Mathematically, this probability can be expressed as

$$Sensitivity = \Pr(Y = 1 \mid S = 1),$$

where the symbol | denotes the statistical concept of *conditioning*, the definition of which can be found in introductory statistics books such as Wasserman (2004), Chap. 1. The sensitivity of a test is also known as the *true positive rate* (*TPR*). Another important accuracy parameter is the *specificity* of Y, which is defined as the probability of test negative when the default status is absent. This probability is given by

$$Specificity = \Pr(Y = 0 \mid S = 0).$$

Specificity is often used interchangeably with the *true negative rate* (*TNR*) in the literature. Both sensitivity and specificity are correct classification rates of a test. Since such a test may also misclassify subjects, error rates are of interest as well.

There are also two types of misclassification rates. The first is the *false positive rate* (*FPR*), which is defined as the probability of test positive when the default status is absent. Mathematically,

$$FPR = \Pr(Y = 1 \mid S = 0).$$

A false positive occurs when a "refusal of loan" decision is made to companies that would never default. By examining the definitions of *FPR* and *specificity*, we note that *FPR* = 1- *specificity*.

Another misclassification rate is the *false negative rate* (*FNR*), which is the probability of test negative when the default status is present. This rate can be expressed by

$$FNR = \Pr(Y = 0 \mid S = 1).$$

A false negative occurs when a loan is granted to a company that later defaults on the loan. Also, we note that *FNR*=1-*sensitivity*.

Table 1 summarizes the aforementioned accuracy and misclassification parameters. The rows of this two-by-two table are split by the true default status ($S = 1$ versus $S = 0$), and columns are classified by test results ($Y = 1$ versus $Y = 0$). In each of the four cells defined by S and Y, the top row displays the frequency of the cell and the bottom row lists the mathematical equation for the accuracy parameter or misclassification rate corresponding to that cell.

Comparing Test Modalities With Binary Values

In the process of decision making, business analysts often have several candidate test modalities with binary values without knowing which modality has the best classification accuracy.