

Operating characteristic analysis of attribute ratings

Z. JOSEPH ULEHLA and ROBERT F. MARTIN
University of Denver, Denver, Colorado 80210

The analysis of receiver operating characteristics as employed in psychophysics is suggested as a way of obtaining several useful measures in the context of attribute ratings. These include the difference between two stimuli on the attribute, the tendency for Ss to favor one pole of the rating scale, width of rating categories, and equal interval properties of the rating scale. The underlying measurement model is described along with means of evaluating its basic assumptions.

In psychophysics, the analysis of receiver operating characteristics (ROCs) has been widely used as a method by which aspects of covert perceptual experience can be inferred from overt perceptual indicator responses. Such ROCs reflect the theoretical and mensurational relations between rating-scale responses to alternative stimuli, on the one hand, and an underlying continuum of covert reactions to the stimuli, on the other hand. These relations are basic to signal detection theory (SDT). The ROC analysis of this theoretical and mensurational relationship between a covert perceptual continuum and overt rating behavior may have general applicability in areas of the behavior sciences where rating responses are used as indicators of inherently covert axes such as attitudes, opinions, expectations, etc. Thurstonian scaling is based on the same theoretical assumptions as SDT (Lee, 1969), so that no assumptions new to Thurstonian scaling are required for ROC analysis. In this paper the theoretical basis of ROC analysis and possible applications to attribute ratings are discussed.

THE THEORETICAL BASIS FOR ROC ANALYSIS

In SDT, as well as in Thurstonian (1927) scaling, a key assumption is that the psychological values resulting from repeated presentation of a given stimulus are normally distributed along a covert decision axis. In the use of rating scales, stable decision criteria are assumed to partition the decision axis into several response regions, one for each category of the rating scale. The probability that a given response (R_a) will follow the presentation of a specified stimulus (S_a) is given by the proportion of the distribution of S_a psychological values that falls into the R_a response region. The unit for the hypothetical decision axis is derived from the distribution of the psychological values of the stimuli. Since this distribution is assumed to be

normal, normal deviate values provide an interval scale for the decision axis.

If more than one stimulus is presented, each stimulus yields a distribution of psychological values. All of these distributions will project onto the same decision axis. For every stimulus, the probability of response alternative R_a is given by the proportion of the psychological value distribution bounded by the criteria that define the response region for R_a .

If more than one stimulus alternative is presented, each stimulus yields an interval scale on the decision axis. These scales provide the coordinates for the type of ROC of interest here. The abscissa represents the normal deviate scale yielded by one stimulus (S_a), and the ordinate represents the other (S_b). Data points portray the location of the criteria on the respective normal deviate scales. Therefore, the response regions occupy on these scales the intervals between adjacent data points.

The data needed for plotting the ROC points are the conditional response proportions, reflecting the probability of each stimulus being assigned to each rating category. When each rating category is added to the cumulation, a "new" decision criterion is involved, and a "new" ROC point is provided. Because the coordinates are normal deviate scales, locating an ROC point requires determining the normal deviate of the cumulative response proportions for the response regions beyond that point. In other words, the pertinent response proportion, for a given point on the ROC plot, is the sum of the response proportions for all responses whose response regions fall to one (specified) side of the decision criterion. When normal deviates of the cumulative response proportions for one stimulus are plotted on the abscissa and equivalent normal deviates for the other stimulus are plotted on the ordinate, a graphic representation of the ROC is obtained.

In SDT, the decision axis provides the information for S's overt judgment

as to which stimulus alternative was presented. In the present application, as in some forms of Thurstonian scaling, the decision axis is associated with the attribute under consideration.¹ Although the attribute need not be "good" or "bad" for ROC analysis to be relevant, for simplicity in exposition attributes will be considered to range from an "unfavorable" pole to a "favorable" pole.

ROC analysis of attribute ratings has several uses. First, the two key measures drawn from SDT can be estimated directly from the ROC. In this application, the d' measure constitutes a ratio-scaled index of the difference between two stimulus alternatives in terms of the specified attribute (this is equivalent to their difference in Thurstonian scale value). The second measure drawn from SDT is an index of responsivity or criterion location which we can label "C," measured independently of d' and also on a scale with ratio properties. The degree to which S perceives the attribute as characterizing the stimulus set as a whole can be inferred from C, as will be shown below. Another important use of the ROC is to test the basic assumptions upon which these measures are based. Finally, the ROC can be employed to measure the category width, i.e., the range of psychological values included in the response region for each rating category. In connection with this, ROC analysis can be used to evaluate the extent to which the E-constructed response scale has equal-interval properties. Having discussed the measurement model in general terms, its specific applications will now be presented.

The d' measure represents the difference between the two interval scales on the decision axis yielded by the normal deviates of the two normal distributions of psychological values. This difference in psychological value underlies the differential evaluation the S(s) makes of the two alternative stimuli. Empirical estimates of the location of the criterion points on the normal deviate scales are provided by the ROC data points. The difference between the abscissa value and the ordinate value of a given ROC point estimates the distance between the normal deviate scales on the decision axis and thus provides an estimate of d' . On the ROC plot, this difference is manifest in the distance between the data point and the positive diagonal, a line with a slope of one passing through the coordinate (0,0) and connecting all of the points of zero difference. One direction of difference would plot above the positive diagonal and would bear positive sign; the

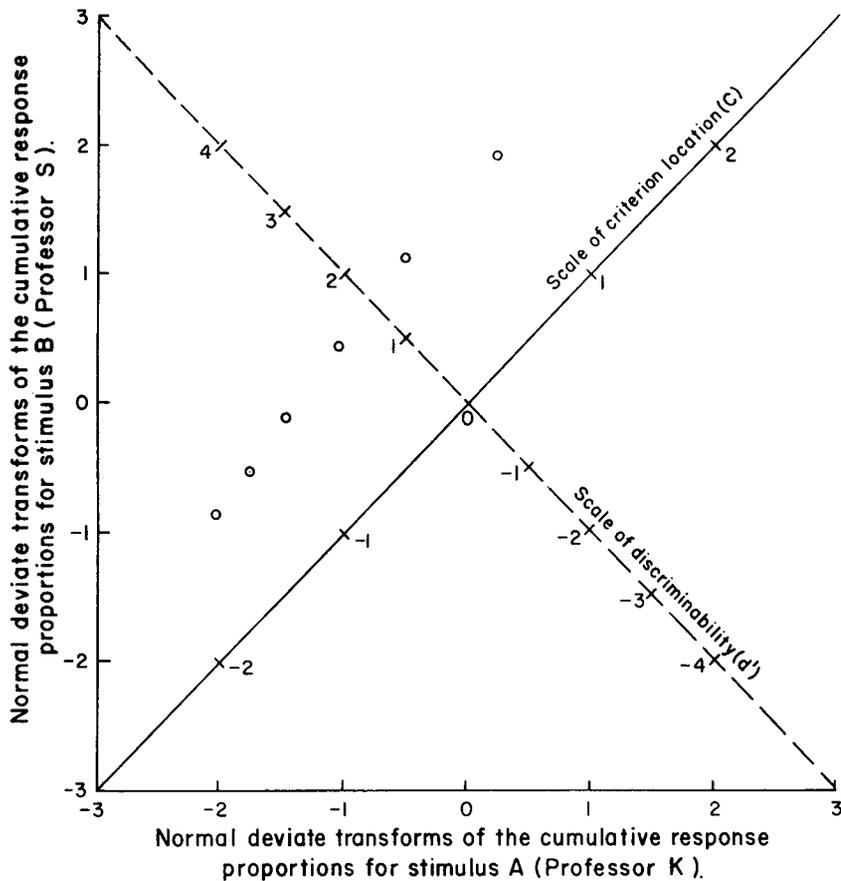


Fig. 1. ROC with data points obtained from ratings of an attribute of teachers.

opposite direction of difference would plot below the diagonal and take on negative sign.

The second key measure derived from the normal deviate scales is the mean of the two normal deviates yielded by the alternative stimuli, following Grice (1968). This measure locates the decision criterion in terms of its distance from the midpoint between the means of the psychological value distribution. In essence, this procedure also scales the decision axis in normal deviate units. Such a scale has a natural origin—the point on the axis where the two psychological value distributions are of equal density (where the two curves cross). On the ROC this origin is represented by the negative diagonal, a line with slope of -1 , passing through the coordinate $(0,0)$. Distance of a criterion from the origin is thus represented by the distance of the associated ROC point from the negative diagonal. Because criterion location is the mean of the two normal deviate values and d' is their difference, the unit for the graphic criterion scale is twice the linear size of the graphic d' scale unit. However, on the hypothetical decision axis the scale units for these two measures are equal.

If a given criterion—perhaps the median of the set of criteria—can be taken as partitioning the response scale into “favorable” and “unfavorable” halves, then the C associated with the criterion can serve as an index of the overall or average favorability of the stimuli in the eyes of the raters. For comparisons of *relative* favorability, between attributes, between groups of raters, or between stimuli sets, the C associated with any criterion could be used. For greater stability, the mean C over all of the criteria could be employed to index relative favorability.

There are also implications for the interaction of variance and criterion location indicated by the spread of data points. With the assumption of stable criterion location, if data points are widely spaced then low variance of the underlying distributions (discriminal dispersions, or DDs , in Thurstonian writing) is indicated. When data points are “bunched,” variance of the DDs is high. (Equivalently, bunched points imply narrow categories, spread-out points imply wide categories, if variance is assumed to be constant.)

Of final consideration in this section is the implication of equal or unequal

spacing of the obtained data points. Given the assumptions of the SDT model (normality and equal variance), if data points are equally spaced, then the E -defined scale, utilized by the respondent in making evaluations, is equal intervalled. If the category widths are not equal, relative widths are implied by relative spacing.

The ROC provides a test of the normality and equal variance assumptions previously noted. The ROC plots the locations of the decision criteria upon normal deviate scales derived from the two stimulus alternatives (i.e., the two rated objects under consideration). If both assumptions hold, the two normal deviate scales will be equal interval, with a common scale unit. The coordinates of a given data point on the vertical and horizontal axes represent the location on the respective normal deviate scales of a particular decision criterion. Thus, if the two normal deviate scales are equal interval with the same scale unit, the ROC data points will fall along a straight line with a slope of 1.0 . A ROC which yields a slope different from 1.0 provides negative evidence for the equal-variance assumption. A nonlinear configuration of the ROC data points constitutes negative evidence for the normality assumption.

Violation of the normality assumption would imply that the normal deviate transformation is inappropriate. If some other transformation yields a linear ROC, measures based on that transformation can be used in ROC analysis. For example, if the psychological value distributions were rectilinear, a linear transformation of cumulative response proportions would yield equal-interval scaling of the psychological value continuum. Such a transformation can be found in the standard score scale (e.g., Guilford, 1956, p. 490), which has its zero at the distribution midpoint and the standard deviation as its unit. Thus, the ROC axes would be scaled in these standard scores, the equivalent to d' would be the standard score difference, and the mean standard score would now provide a measure of criterion location.

Violation of the equal variance assumption means d' will not be the same for different criteria. The relative favorability of two rated objects is then a function of criterion location on the attribute scale. For example, Stimulus S_a may exceed Stimulus S_b at the favorable end of the scale, but not at the unfavorable end. In this case, both stimuli are equally prone to be seen as very bad, but S_a is more likely to be seen as very good. In a more extreme situation, the line of

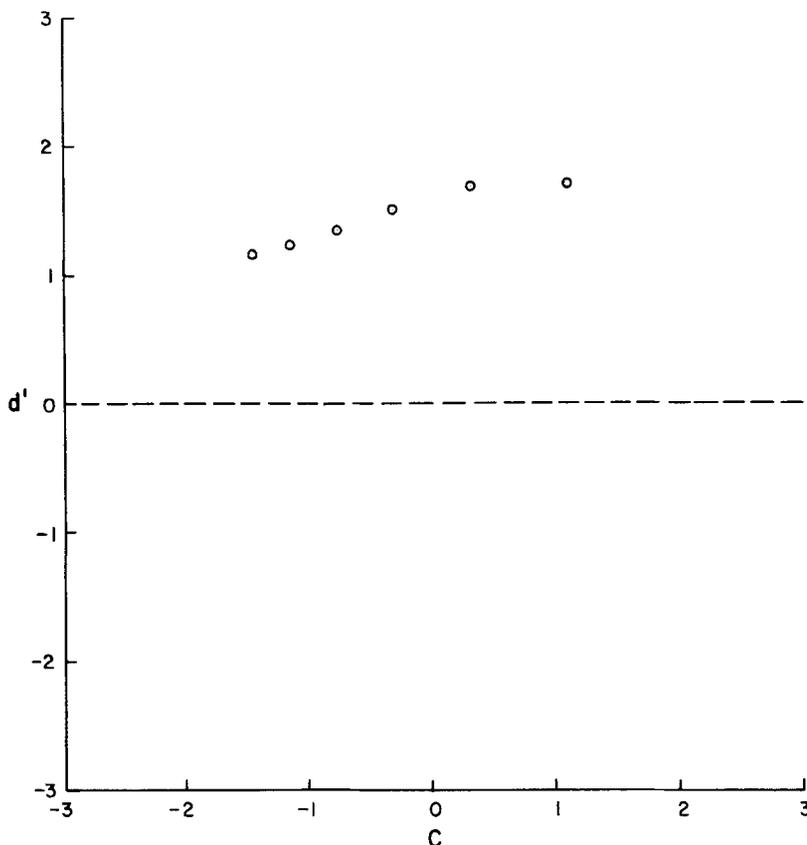


Fig. 2. Parameter plot of the same data as shown in Fig. 1.

ROC data points might intersect the line of chance performance; in this case, the stimulus most likely to be seen as very good would also be most likely to be seen as very bad.

In the ROC, transformed response proportions, d' , and criterion location are all graphically represented. The vertical-horizontal coordinate systems scale the normal deviate transforms of the cumulative response proportions to two stimulus alternatives, which permits data points to be plotted directly from transformed cumulative probabilities. Although redundant, a second coordinate system provides a direct reading of the values of the SDT measures; d' is scaled along the negative diagonal and criterion location is scaled along the positive diagonal. Thus, the projections of a data point upon the diagonals yields the SDT measures. It should be noted that the units are not equal for the two diagonal scales.

If the ROC is rotated 45 deg (following Grice, 1968), the criterion location and d' scales now constitute the abscissa and ordinate, respectively,

and the scale unit can be equalized by using the normal deviate for both. Abscissa points represent the mean normal deviate, ordinate points represent the algebraic difference between the normal deviates. This rotated ROC, which we may call a parameter plot because the axes represent the SDT measures, has an advantage over the unrotated ROC, in that the representation of the SDT measures is more straightforward.

Illustrative empirical data have been included in both the ROC and the parameter plot pictured at Figs. 1 and 2. The data represent evaluations by students in an introductory psychology class of two of the course instructors in the attribute, "intellectually stimulating." Ratings were on a 7-point numerical scale in which 0 was defined as "does poorly" and 6 was defined as "does well."

The linearity of the data points supports the normality assumption; thus, support is obtained for the measurement model with normal deviate scaling of the decision continuum. However, the assumption

of equal variance is contradicted because the slope departs from 1.0 in Fig. 1 and 0 in Fig. 2. Because of the slightly unequal variance, d' varies slightly with criterion location. In this example, Professor S is rated as more "intellectually stimulating" over the whole "does well" to "does poorly" range, but enjoys his greatest advantage at the favorable end of the range. This is most clearly shown in the parameter plot, where d' is seen to be highest for the criteria associated with the most favorable rating categories.

The set of data points is centered somewhat to the left of the zero criterion, indicating greater usage of the favorable rating categories than the unfavorable ones. The mean C of -0.39 reflects this overall favorability, and could be employed as a numerical index of favorability to be compared with similar indices drawn from other data sets.

The spacing of the data points increases as one moves from the unfavorable to the favorable end of the continuum. Coupled with the apparent good fit of the measurement model with the normality assumption, this suggests broader category widths for the more favorable rating categories.

In sum, ROC analysis provides graphic techniques for discovering several characteristics of the perceptions of attributes. These include difference between two stimuli on the attributes, overall favorability of judgment, width of rating categories and equal interval properties of the rating scale. Basic assumptions of the measurement model can be evaluated by reference to the form and slope of the ROC.

REFERENCES

- GRICE, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, 75, 359-373.
- GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956. P. 409.
- LEE, W. L. Relationships between Thurstone category scaling and signal detection theory. *Psychological Bulletin*, 1969, 71, 101-107.
- THURSTONE, L. L. A law of comparative judgement. *Psychological Review*, 1927, 34, 273-286.

NOTE

1. Because of this, the unidimensionality of the decision axis is implied by usage of a unidimensional rating scale. Even stimulus sets which may require a multidimensional space to be represented simultaneously on a variety of attributes or for other types of analysis can be assumed to be projected upon a unidimensional axis for ROC analysis of ratings of one attribute.