A Non-Parametric ROC-Based Measure of Sensitivity

Shane T. Mueller STMUELLE@INDIANA.EDU
Indiana University Department of Psychological and Brain Sciences, 1101 East 10th Street, Bloomington, IN 47404

Jun Zhang JUNZ@UMICH.EDU
University of Michigan Department of Psychology, 530 Church Street Ann Arbor, MI 48109-1043

Abstract

In the signal detection paradigm, the non-parametric index of sensitivity A', as first introduced by Pollack and Norman (1964), is a popular alternative to the more traditional d' measure of sensitivity. Smith (1995) clarified a confusion about the interpretation of A' in relation to the area beneath proper ROC (Receiver-Operation Characteristic) curves, and provided a formula (which he called A'') for this commonly held interpretation. However, he made an error in his calculations. Here we (i) rectify this error by providing the correct formula (which we call A) and compare the discrepancy that would have resulted, and (ii) prove a property of ROC curves parameterized by the likelihood-ratio.

1. Estimates of sensitivity based on ROC areas

Signal detection theory (Green & Swets, 1966) is commonly used to interpret data from tasks in which stimuli (e.g., tones, words, or medical images) are presented to an observer who must determine which one of two categories the stimulus belongs in. These tasks yield two measures of behavioral performance: the Hit Rate (H) and the False Alarm Rate (F). H and F are typically transformed into indices of sensitivity and bias based on assumptions about an underlying statistical model. However, other transformations are also possible, including ones not tied directly to an underlying statistical detection model, including ones based on the range of "proper" ROC curves[1] that could pass through any single point.

Appearing in *Proceedings of the ICML 2006 workshop on ROC Analysis in Machine Learning*, Pittsburgh, USA, 2006.

Figure 1. Proper ROC curves through point p must lie within or on the boundaries of the light shaded regions A_1 and A_2. The minimum-area proper ROC curve through p lies on the boundary of region I.

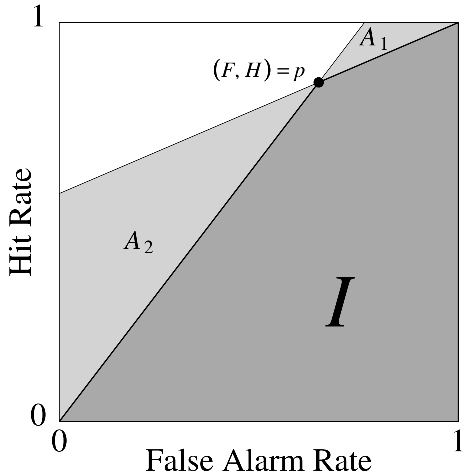


Figure 1, shows that for the point $p = (F, H)$, all proper ROC curves must fall within or on the bounds of light shaded regions A_1 and A_2. The proper ROC curve with the smallest area lies on the boundary between I and A_1 (to the right of p) and A_2 (to the left of p), whereas the proper ROC curve with the largest area lies within or on the boundaries of A_1 and A_2.

Pollack and Norman (1964) proposed such a measure of sensitivity (A') as the average of the ar-

[1]A proper (or admissible) ROC curve is a monotonically nondecreasing function with a non-increasing slope connecting the end points (0,0) and (1,1). It necessarily lies above the line $H = F$. Consequently, the path an ROC curve follows between (0,0) and (1,1) is convex.

eas $A_1 + I$ and $A_2 + I$, which turns out to equal $1/2+(H-F)(1+H-F)/(4H(1-F))$ (Grier, 1971). This alternative sensitivity measure is at least partially motivated and justified by Green's (1964) proof that the area under the ROC curve in a yes-no detection task is equal to the overall probability correct in the task. Since its introduction, A' has become a popular "non-parametric" alternative to d' for measuring sensitivity in detection and categorization tasks (Macmillan & Creelman, 1996). Given its widespread use, it is interesting to note that A' does not measure what many people believe it measures (Smith, 1995). Smith reported that A' is commonly (and incorrectly) believed to be an average between the minimum-area and maximum-area proper ROC curves that can pass through the observed point $p = (F, H)$.

Smith noted that A' is not the average of the maximum-area and minimum-area proper ROC curves, although it is easily shown to be the average of the area I and $I + A_1 + A_2$ from Figure 1. While the smallest area associated with a proper ROC curve passing through p is I, the A_{max} (the area of the largest proper ROC curve) is not $I + A_1 + A_2$, and its proper calculation is a much subtler issue. Smith (1995) claimed that, depending on whether p is to the left or right of the negative diagonal $H + F = 1$, A_{max} is the larger of $I + A_1$ and $I + A_2$, shown in Figure 1. Smith then defined a new measure called A'' which was the average of I and the $\max\{I + A_1, I + A_2\}$. However, Smith's (1995) claim about A_{max} was erroneous when p is in the upper left quadrant of ROC space (i.e., $F < .5$ and $H > .5$). In this region, neither $I + A_1$ nor $I + A_2$ is the region bounded by the maximum-area proper ROC curve passing through p.

2. The proper ROC curve with maximum area

Zhang and Mueller (2005) showed that the maximum area A_{max} of all proper ROC curves passing through a given point $p = (F, H)$ is actually:

$$A_{max} = \begin{cases} 1 - 2H(1-F) & \text{if} \quad F < 0.5 < H\,, \\ \frac{1-F}{2H} & \text{if} \quad F < H < 0.5\,, \\ 1 - \frac{(1-H)}{(2(1-F))} & \text{if} \quad 0.5 < F < H\,. \end{cases} \qquad (1)$$

Given this result, neither Pollack & Norman's (1964) A' nor Smith's (1995) A'' provide an accurate measure of the common interpretation of non-parametric sensitivity index as the averaged of the maximum and minimum proper ROC curves constrained by the experimental data (F, H). We provide here the correct formula (which we denote A) for the the average of A_{max} (Equation 2) and $A_{min} = (1 + H - F)/2$:

$$A = \begin{cases} \frac{3}{4} + \frac{H-F}{4} - F(1-H) & \text{if} \quad F \le 0.5 \le H\,; \\ \frac{3}{4} + \frac{H-F}{4} - \frac{F}{4H} & \text{if} \quad F < H < 0.5\,; \\ \frac{3}{4} + \frac{H-F}{4} - \frac{(1-H)}{(4(1-F))} & \text{if} \quad 0.5 < F < H\,. \end{cases} \qquad (2)$$

A is the correct formula for determining the average of the minimum-area proper ROC curve and the maximum-area proper ROC curve. However, this formulation ignores the fact that the minimum-area and maximum-area proper ROC curves might differ considerably for some points, whereas they might be very similar for others. Consequently, one might be reasonably confident about A in regions where the maximum and minimum areas are nearly the same, and be more hesitant about using A for data in regions where there is a large difference between the minimum-area and maximum-area proper ROC curves.

Figure 3 examines this issue, by plotting the difference between the minimum-area and maximum-area proper ROC curves that can pass through each point. The figure shows that the smallest differences occur along the positive and negative diagonals of ROC space, especially for points close to $(0, 1)$ and $(.5, .5)$. The points where there is the greatest difference between minimum and maximum-area proper ROC curves are near the lines $H = 0$ and $F = 1$. Thus, data observed near

Figure 2. Example of a proper ROC curve through p. The ROC curve $\mathcal{C}$, a piecewise linear curve denoted by the dark outline, is formed by following a path from $(0, 0)$ to $(0, H_0)$ to $(F_1, 1)$ (along a straight line that passes through $p = (F, H)$) and on to $(1, 1)$.

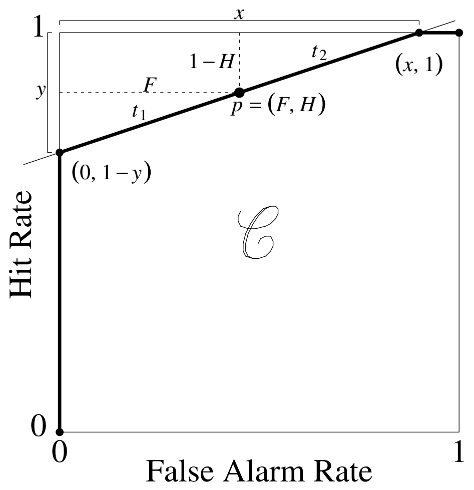
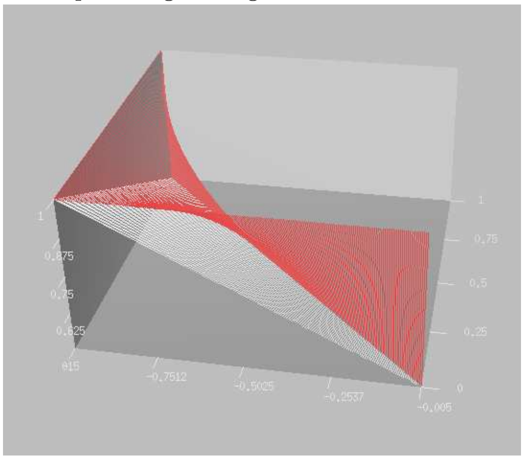


Figure 3. Difference between the minimum-area and maximum-area proper ROC curves through every point in ROC space. Lighter regions indicate smaller differences.



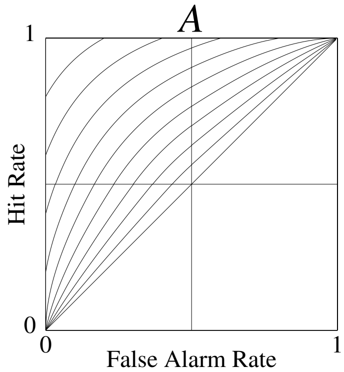
these edges of ROC space can be passed through by proper ROC curves with a large range of underlying areas. Consequently, care should be taken when interpreting results observed in these latter regions. This caution is not only applicable to area-based measures of sensitivity; these are also the regions where d' has the steepest slope with respect to H and F; and the regions where slightly different assumptions about distributions can have the largest impact.

Another way to examine A and related area-based measures of sensitivity is to examine the "iso-sensitivity" curve, i.e, the combinations of F and H will produce a given value of A. The topography of A in ROC space can be mapped by drawing isopleths for its different constant values. Figure 4 shows these topographic maps for A. The corresponding functions for A' and A'' were examined by Zhang and Mueller (2005).

The slope of any proper ROC curve is related to the likelihood ratio of the underlying distributions (for signal and for noise), and as such, it can be used as an index of decision bias. Zhang & Mueller (2005) showed that the slope of the constant-A curve (called b) can be found with the following formula:

$$b = \begin{cases} \frac{5-4H}{1+4F} & \text{if} \quad F \le 0.5 \le H\,; \\ \frac{(H^2+H)}{(H^2+F)} & \text{if} \quad F < H < 0.5\,; \\ \frac{(1-F)^2+(1-H)}{(1-F)^2+(1-F)} & \text{if} \quad 0.5 < F < H\,. \end{cases} \qquad (3)$$

Figure 4. Constant-sensitivity isopleths for A. A line maps out combinations of F and H that produce equal values of sensitivity, in increments of 0.05.



3. Slope of ROC curve and likelihood ratio

According to the traditional signal detection framework, an ROC curve $(F(t_c), H(t_c))$ is parameterized by the cutoff criteria value t_c along the measurement (sensory encoding) axis based on which a Yes/No decision is made. The signal and noise distributions may be generated by systematically varying the signal strength while recoding the subject response. Given underlying signal distribution $f_s(u)$ and noise distribution $f_n(u)$ of measurement (encoded) value u, a criterion-based response strategy will give rise to

$$\begin{aligned} H(u_c) &= \Pr(Yes|s) = \Pr(u > u_c|s) = \int_{u_c}^{\infty} f_s(u)du\,, \\ F(u_c) &= \Pr(No|s) = \Pr(u > u_c|n) = \int_{u_c}^{\infty} f_n(u)du\,. \end{aligned}$$

As u_c varies, so do H and F; they trace out the ROC curve. Its slope is

$$\left.\frac{dH}{dF}\right|_{F=F(u_c),H=H(u_c)} = \frac{H'(u_c)}{F'(u_c)} = \frac{f_s(u_c)}{f_n(u_c)} \equiv l(u_c)\ .$$

With an abuse of notation, we simply write

$$\frac{dH(u)}{dF(u)} = l(u)\,. \qquad (4)$$

It is well known (see Green and Swets, 1966) that a monotone transformation of measurement axis $u \mapsto v = g(u)$ does not change the the shape of the ROC curve (since it is just a re-parameterization of the curve). Furthermore, when $l(u)$ is a monotonically non-increasing function of u, the slope is non-increasing as one increases the cutoff criterion value u.

Complications arise when likelihood function $l(u)$ is a non-monotonic (still assumed to be continuous and differentiable) function of the measurement variable u. In this case, however, one can define the decision criterion by a cutoff likelihood-ratio value l_c, with the corresponding hit rate (H) and false-alarm rate (F) given by

$$\begin{aligned} H(l_c) &= \int_{\{u: l(u) > l_c\}} f_s(u) du \\ F(l_c) &= \int_{\{u: l(u) > l_c\}} f_n(u) du \,. \end{aligned}$$

Zhang and Mueller (2005) proved the following:

Lemma 2. The slope of a ROC curve (generated by a criterion-based decision mechanism on the likelihood-ratio axis) equals the likelihood-ratio value at the criterion l_c

$$\frac{dH(l_c)}{dF(l_c)} = l_c \ .$$

This conclusion also holds for multidimensional observations, i.e., when evidence consists multi-dimensional observation vector $\mathbf{u}$, with density of likelihood function $f_s(\mathbf{u})$, $f_n(\mathbf{u})$ (e.g., Ashby and Townsend, 1986). Using likelihood-ratio $l(\mathbf{u}) = \frac{f_s(\mathbf{u})}{f_n(\mathbf{u})}$ as the decision axis and l_c as cutoff criterion, the hit and false-alarm rates are

$$\begin{aligned} H(l_c) &= \int_{\{(\mathbf{u}): l(\mathbf{u}) \geq l_c\}} f_s(\mathbf{u}) d\mathbf{u} \,, \\ F(l_c) &= \int_{\{(\mathbf{u}): l(\mathbf{u}) \geq l_c\}} f_n(\mathbf{u}) d\mathbf{u} \,. \end{aligned}$$

It can be analogously proven that

Corollary 3. The slope of a ROC curve (generated by a criterion-based decision mechanism on multi-dimensional observations) equals the criterion likelihood-ratio

$$\frac{dH(l_c(\mathbf{u}))}{dF(l_c(\mathbf{u}))} = l_c(\mathbf{u}) \,.$$

4. Curvature of a likelihood-ratio based ROC curve

Lemma 2 shows that the slope of ROC curve is always equal the likelihood-ratio value regardless how it is parameterized, i.e., whether the likelihood-ratio is monotonically or non-monotonically related to the sensory dimension and whether the likelihood is based on uni- or multi-dimensional signals. The ROC curve is a signature of a criterion-based decision strategy, as captured by the unified expression

$$\frac{dH(l)}{dF(l)} = l \,. \tag{5}$$

Since $H(l)$ and $F(l)$ give the proportion of hits and false alarms when a decision-maker says "Yes" whenever the likelihood-ratio (of the data) exceeds l, then

$$\begin{aligned} \delta H &= H(l + \delta l) - H(l) \,, \\ \delta F &= F(l + \delta l) - F(l) \end{aligned}$$

are the amount of hits and false-alarms if he says "Yes" only when the likelihood-ratio falls within the interval $(l, l+\delta l)$. Their ratio is of course simply the likelihood-ratio.

The shape of the ROC curve is determined by $H(l)$ or $F(l)$. Differentiation both sides of (5) with respect to l

$$\frac{dF}{dl} \cdot \frac{d}{dF}\left(\frac{dH}{dF}\right) = 1$$

which implies that

$$\frac{d}{dF}\left(\frac{dH}{dF}\right) = \left(\frac{dF}{dl}\right)^{-1} < 0$$

since $dF/dl < 0$. This reflects the fact a proper ROC curve is always concave function bending downward. In fact, its curvature is

$$\kappa = \frac{d}{dl}\left(\frac{dH}{dF}\right) / \left(1 + \left(\frac{dH}{dF}\right)^2\right) = \frac{1}{1 + l^2} \,.$$

This likelihood-ratio based analysis is seen (Balakrishnan, 1998; 1999) as an alternative to the traditional stimulus-based paradigm. Using confidence rating as measurements, Balakrishnan showed that the likelihood ratio is a monotonic (though not identical) function of the confidence rating. Under manipulation of prior odds, the criterion value of confidence rating is unchanged in a classification task – decision criterion is uniformly unbiased. Under the likelihood-ratio parameterization, the signal distribution $f_s(l) = dH/dl$ and the noise distribution $f_n(l) = dF/dl$ can be shown to satisfy (i) $E_n l = \int l f_n(l) dl = 1$ and (ii) $E_s l = \int l f_s(l) dl \geq 1$, with the average separation satisfying $E\{|l - l'|\} = 4R - 2$ where $R = \int H(l) dF(l)$ is the ROC area. So indeed one can construct alternative measures for sensitivity based these distributions.

5. Conclusion

We described a correct formula for the maximal area of proper ROC curves passing through a data point and, through which, the non-parametric estimate of sensitivity as defined by the average of maximal and minimum ROC areas. We also showed that the relationship of ROC slope to likelihood-ratio is a fundamental relation in ROC analysis, as it is invariant with respect to any continuous reparameterization of the stimulus, including non-monotonic mapping of uni-dimensional signal and multi-dimensional signals in general.

References

Balakrishnan, J. D. (1998). Some more sensitive measures of sensitivity and response bias. *Psychological Methods*, *3*, 68–90.

Balakrishnan, J. D. (1999). Decision processes in discrimination: fundamental misrepresentations of signal detection thory. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1189–1206.

Green, D. M. (1964). General prediction relating yes-no and forced-choice results. *Journal of the Acoustical Society of America, A*, *36*, 1024.

Green, D. M., & Swets, J. A. (1964). *Signal detection theory and psychophysics*. New York: John Wiley & Sons.

Jr., W. P. T., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in roc space: History and theory of "non-parametric" measures of sensitivity and response bias. *Psychonomic Bulletin & Review*, *3*, 164–170.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory* (pp. 171–212).

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under roc curve and d'_e. *Psychological Bulletin*, *1*, 161–173.

Pollack, I., & Norman, D. A. (1964). Non-parametric analysis of recognition experiments. *Psychonomic Science*, *1*, 125–126.

Smith, W. D. (1995). Clarification of sensitivity measure a'. *Journal of Mathematical Psychology*, *39*, 82–89.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1964). Decision process in perception. *Psychological Review*, *68*, 301–340.

Zhang, J., & Mueller, S. T. (2005). A note on roc analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*, 145–154.