Medical Physics, vol. 34 (5), pp. 1556–1564, 2007

Nonparametric ROC and LROC analysis

Lucreţiu M. Popescu
University of Pennsylvania, Department of Radiology
423 Guardian Drive, 4-th floor Blockley Hall,
Philadelphia, PA 19104-6021 *

In this paper we review several results of the nonparametric receiver operating characteristic (ROC) analysis and present an extension to the nonparametric localization ROC (LROC) analysis. Equations for the estimation of the area under the characteristic curve and for the variance calculations are derived. Expressions for the choice of the optimal ratio between the number of signal absent and signal present image samples are also presented. The results can be applied both with continuous or discrete scoring scales. The simulation studies carried out validate the theoretical derivations and show that the LROC analysis is considerably more sensitive than the ROC analysis.

I. INTRODUCTION

The receiver (or relative) operating characteristic (ROC) analysis has become a standard procedure to evaluate the performance of observers in radiological tasks. One common approach of ROC analysis assumes a simple continuous binormal model of a hypothetic latent decision variable[1,2]. However, the mechanism that human observers use to analyze medical images (find suspicious features) is more complex and far from being completely understood. Theoretical considerations,[3,4] as well as eye movement experiments,[5] indicate that the detection process comprises two steps: first a scan process that identifies a number of suspicious locations, and second, a decision step that establishes whether these are abnormal or not. Although significant progress has been made in understanding the human visual system, the decision step proves to be more elusive to understand and model. Even within the more predictable realm of mathematical models and numerical procedures, the decision processes are not always that simple[6]. Multiple clues (size, contrast, or number of apparent features) can indicate an image as suspicious, and each clue can be considered with a different weight by different observers, depending on their prior assumptions.

Originally developed to characterize the operation of an apparatus, the binormal ROC model can be seen as a pragmatic data modeling approach when it is applied to human decision processes. An alternative pragmatic approach is the nonparametric data analysis of the observers' results that, as is the case with the nonparametric methods in general, does not rely on specific models or assumptions that are difficult to confirm experimentally. Instead of assuming the existence of a latent decision variable the nonparametric methods use directly the confidence scores reported by the observers. Bamber[7] has shown the relation that exists between the area under the ROC curve and the Mann-Whitney-Wilcoxon statistic[8,9] that can be used to show the difference between two populations, or to assess the accuracy of the discrimination performance of a test. The use of the nonparametric ROC analysis in radiology is discussed by Hanley and McNeil[10].

As an alternative to the ROC analysis, Swensson[11] proposed the localization ROC (LROC) analysis in which the locations of the suspected abnormalities in images are used to improve the statistical analysis[12]. The Swensson model is a significant improvement from the modeling realism point of view compared to the binormal model. It considers a search mechanism and elegantly solves the problem of improper or degenerate ROC curves[13], that is a symptom of the shortcomings of the simple binormal model. However, the parametric LROC model of Swensson does not account for the multiple clues situation mentioned above, and there is no guarantee that the latent variables involved are normally distributed. Therefore a nonparametric equivalent procedure would be a desirable addition to the LROC methodology.

In this paper we review several results of the nonparametric ROC analysis and present an extension to the nonparametric LROC analysis.

II. THEORY

A. Nonparametric ROC analysis

Consider an observer analyzing a total of $M+N$ images, M images with the signal present, and N images with the signal absent. The observer gives a subjective confidence score z to each image. We assume that the observer is stationary, so that its criteria and behavior do not change with time, and the confidence scores are independent of each other. In the next derivations for convenience we will assume that the confidence scores continuously cover an infinite range, but as we will show later on, the results obtained are also valid for discrete scoring schemes.

We denote by $\hat{f}(z)$ the nonparametric estimate of the distribution $f(z)$ of the confidence scores z for the case of the images with the signal present, and by $\hat{g}(z)$ the nonparametric estimate of the distribution $g(z)$ for the signal absent case. If we use the density kernel method[14] to estimate the distributions from the sample data, we have

$$\hat{f}(z) = \frac{1}{M}\sum_{m=1}^{M}\phi_f(z - x_m)\,, \tag{1}$$

$$\hat{g}(z) = \frac{1}{N}\sum_{n=1}^{N}\phi_g(z - y_n)\,, \tag{2}$$

where $\phi_f(z)$ and $\phi_g(z)$ are the density kernels $\int_{-\infty}^{+\infty}\phi_{f,g}(z)\mathrm{d}z = 1$, and $X \equiv \{x_m \mid m = 1,\ldots,M\}$ and $Y \equiv \{y_n \mid n = 1,\ldots,N\}$ are the sets of the scores received by the positive and the negative, respectively, images.

The estimates of the cumulative distributions of f, $F(z) = \int_{-\infty}^{z} f(z')\mathrm{d}z'$, and g, $G(z) = \int_{-\infty}^{z} g(z')\mathrm{d}z'$ are

$$\hat{F}(z) = \frac{1}{M}\sum_{m=1}^{M}\int_{-\infty}^{z}\phi_f(z' - x_m)\mathrm{d}z'\,, \tag{3}$$

$$\hat{G}(z) = \frac{1}{N}\sum_{n=1}^{N}\int_{-\infty}^{z}\phi_g(z' - y_n)\mathrm{d}z'\,. \tag{4}$$

With $\Phi_{f,g}(z) = \int_{-\infty}^{z}\phi_{f,g}(z')\mathrm{d}z'$ we have

$$\hat{F}(z) = \frac{1}{M}\sum_{m=1}^{M}\Phi_f(z - x_m)\,, \tag{5}$$

$$\hat{G}(z) = \frac{1}{N}\sum_{n=1}^{N}\Phi_g(z - y_n)\,. \tag{6}$$

The probability of a true positive result with the score greater than or equal to z is $P_1(z) = 1 - F(z)$ and the probability of a false positive result is $P_0(z) = 1 - G(z)$. The variation of P_1 as a function of P_0 represents the receiver (or relative) operating characteristic (ROC) curve.

In the case of human observers the rating is subjective, so it is difficult to define precise operating points. This is one reason why as performance indicators are used quantities integrating over a range of operating points. The area under the ROC curve is such an integral indicator, and it is the most widely accepted since for other performance indicators integrating only over a partial range, the selection of the appropriate operating range is still a matter of subjective judgment. The theoretical area under the ROC curve is

$$A = \int_0^1 P_1\mathrm{d}P_0 = \int_{-\infty}^{+\infty} f(z)G(z)\mathrm{d}z\,. \tag{7}$$

The estimated area under the ROC curve is

$$\begin{aligned}\hat{A} &= \int_{-\infty}^{+\infty}\left(\frac{1}{M}\sum_{m=1}^{M}\phi_f(z - x_m)\right)\left(\frac{1}{N}\sum_{n=1}^{M}\Phi_g(z - y_n)\right)\mathrm{d}z \\ &= \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\int_{-\infty}^{+\infty}\phi_f(z - x_m)\Phi_g(z - y_n)\mathrm{d}z\,.\end{aligned} \tag{8}$$

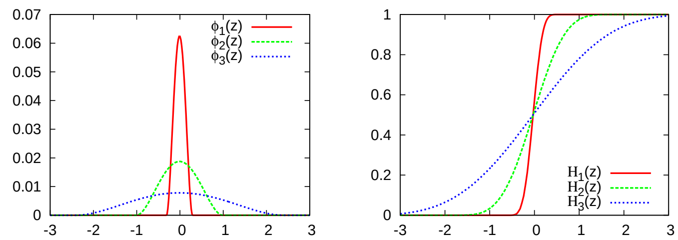

FIG. 1: Example of density estimation kernels $\phi_{f,g}(z) = \frac{1}{h}\frac{15}{16}\left[1 - \left(\frac{z}{h}\right)^2\right]^2$ and their corresponding rank function $H_\varepsilon(z)$ (equation (10)) for three different values of the smoothing parameter $h = 0.3, 1.0, 2.4$.

In the end the estimate of the area under the ROC curve is given by the following summation

$$\hat{A} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}H_\varepsilon(x_m - y_n)\,, \tag{9}$$

where

$$H_\varepsilon(y) = \int_{-\infty}^{+\infty}\phi_f(x)\Phi_g(x + y)\mathrm{d}x \tag{10}$$

is the rank function.

In Silverman[14] are given criteria for choosing the optimal shape and parameters of the density kernels. In general the density kernels can be expressed as $\phi_{f,g}(z) = \phi(\frac{z}{h_{f,g}})$, with h_f and h_g smoothing parameters (or kernel window widths) that depend on the shape of the original functions and the number of discrete events available. Typical shapes of such density kernels and of their corresponding $H_\varepsilon(z)$ rank functions are shown in FIG. 1. At the asymptotic limit the density kernels converge towards the Dirac δ function and H towards the Heaviside function

$$H(z) = \begin{cases} 0 & \text{if } z < 0 \\ \frac{1}{2} & \text{if } z = 0 \\ 1 & \text{if } z > 0 \end{cases}\,. \tag{11}$$

In this form H can also be used when the data take only discrete values. Choosing the optimal kernels can bring significant improvements[15,16], however, in this paper we will consider generic H_ε functions, and will present numerical examples obtained for H given by equation (11). The results and the conclusions will not lose their generality.

The variance of the area estimate $\hat{A}$ is

$$\sigma_{\hat{A}}^2 = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2\,, \tag{12}$$

where by $\langle\cdot\rangle$ we denote the expectation of the given statistic. We have

$$\hat{A}^2 = \left[\frac{1}{MN}\sum_{m_1=1}^{M}\sum_{n_1=1}^{N} H_\varepsilon(x_{m_1}-y_{n_1})\right]\left[\frac{1}{MN}\sum_{m_2=1}^{M}\sum_{n_2=1}^{N} H_\varepsilon(x_{m_2}-y_{n_2})\right] \tag{13}$$

$$= \frac{1}{M^2N^2}\Bigg[\sum_{m=1}^{M}\sum_{n=1}^{N} {H_\varepsilon}^2(x_m-y_n) + \sum_{m=1}^{M}\sum_{\substack{n_1=1\\ n_2=1\\ n_1\neq n_2}}^{N} H_\varepsilon(x_m-y_{n_1})H_\varepsilon(x_m-y_{n_2}) +$$

$$\sum_{\substack{m_1=1\\ m_2=1\\ m_1\neq m_2}}^{M}\sum_{n=1}^{N} H_\varepsilon(x_{m_1}-y_n)H_\varepsilon(x_{m_2}-y_n) + \sum_{\substack{m_1=1\\ m_2=1\\ m_1\neq m_2}}^{M}\sum_{\substack{n_1=1\\ n_2=1\\ n_1\neq n_2}}^{N} H_\varepsilon(x_{m_1}-y_{n_1})H_\varepsilon(x_{m_2}-y_{n_2})\Bigg],$$

where in order to estimate the average $\langle\hat{A}^2\rangle$ we have separated the overall sum into terms in which all appearing x_m and y_n data points are distinct random variables. With $\{x_m\}$ and $\{y_n\}$ independent, each of the above terms can be averaged separately. We obtain

$$\langle\hat{A}^2\rangle = \frac{1}{M^2N^2}\left[MNB_{11} + MN(N-1)B_{12} + M(M-1)NB_{21} + M(M-1)N(N-1)A^2\right], \tag{14}$$

where A, B_{11}, B_{12} and B_{21} denote the following expectations

$$A = \langle H_\varepsilon(x-y)\rangle_{f(x)g(y)} = \int_{-\infty}^{+\infty}\mathrm{d}x\, f(x)\int_{-\infty}^{+\infty}\mathrm{d}y\, g(y)H_\varepsilon(x-y) \tag{15}$$

$$B_{11} = \langle {H_\varepsilon}^2(x-y)\rangle_{f(x)g(y)} = \int_{-\infty}^{+\infty}\mathrm{d}x\, f(x)\int_{-\infty}^{+\infty}\mathrm{d}y\, g(y){H_\varepsilon}^2(x-y) \tag{16}$$

$$B_{12} = \langle H_\varepsilon(x-y_1)H_\varepsilon(x-y_2)\rangle_{f(x)g(y_1)g(y_2)}$$
$$= \int_{-\infty}^{+\infty}\mathrm{d}x\, f(x)\int_{-\infty}^{+\infty}\mathrm{d}y_1\, g(y_1)\int_{-\infty}^{+\infty}\mathrm{d}y_2\, g(y_2)H_\varepsilon(x-y_1)H_\varepsilon(x-y_2) \tag{17}$$

$$B_{21} = \langle H_\varepsilon(x_1-y)H_\varepsilon(x_2-y)\rangle_{f(x_1)f(x_2)g(y)}$$
$$= \int_{-\infty}^{+\infty}\mathrm{d}x_1\, f(x_1)\int_{-\infty}^{+\infty}\mathrm{d}x_2\, f(x_2)\int_{-\infty}^{+\infty}\mathrm{d}y\, g(y)H_\varepsilon(x_1-y)H_\varepsilon(x_2-y) \tag{18}$$

With $\langle\hat{A}\rangle = A$ and with

$$S_0 = B_{11} - A^2\,,\quad S_1 = B_{12} - A^2\,,\quad S_2 = B_{21} - A^2\,, \tag{19}$$

we have

$$\sigma^2_{\hat{A}} = \frac{S_0 + (N-1)S_1 + (M-1)S_2}{MN}\,. \tag{20}$$

The equation (20) was given in a slightly different form by Bamber[7] by applying the results of Noether[17].

B. Nonparametric LROC analysis

In the case of detection of a small localized signal in an image, a reader may mistakenly identify a suspicious background feature as a signal, while neglecting the actual true signal. Therefore an observer can correctly report that a signal is present but indicate a false signal location. To avoid this ambiguity, in the LROC experiments the readers are asked to report the most suspicious finding and indicate its location together with a confidence score. In the example given above, the reader in fact could notice both the false and the true signals, but be forced to make a wrong decision by being asked to report only the most apparently suspicious one. From this perspective the LROC analysis has a forced choice component unlike the free response analysis[18–21], where the readers are asked to score all suspicious locations they find in an image. The detection of a small signal at an unknown location in an image is a more complex task than the binary classification that is the object of the usual ROC analysis.

Let L be the number of images in which the signal was correctly localized by the observer, out of the total M abnormal images scored. We denote with $\{x_l\}$ the confidence scores assigned to those images. The distribution of the total positive results is the sum of the distribution of the correctly localized results $f_{\mathcal{L}}$, and the contamination distribution due to the background noise $f_{\mathcal{C}}$

$$f(x) = f_{\mathcal{L}}(x) + f_{\mathcal{C}}(x)\,. \tag{21}$$

The average ratio of the correctly localized positive results is

$$Q = \int_{-\infty}^{+\infty} f_{\mathcal{L}}(x)\mathrm{d}x\,, \tag{22}$$

and we have the normalized distribution of the correctly localized results

$$f^*_{\mathcal{L}}(x) = \frac{1}{Q} f_{\mathcal{L}}(x)\,. \tag{23}$$

The probability to report a correctly localized signal (a true positive) with the score greater than or equal to z is $P_{1\mathcal{L}}(z) = \int_z^{\infty} f_{\mathcal{L}}(z')\mathrm{d}z' = Q - F_{\mathcal{L}}(z)$, where $F_{\mathcal{L}}(z) = \int_{-\infty}^{z} f_{\mathcal{L}}(z')\mathrm{d}z'$. The variation of $P_{1\mathcal{L}}$ as a function of P_0 (the probability of a false positive result) represents the localization receiver operating characteristic (LROC) curve. An example of a typical LROC curve and how it differs from its ROC counterpart is shown in FIG. 2 (bottom).

The nonparametric estimate $\hat{f}_{\mathcal{L}}(z)$ of the correctly localized results distribution $f_{\mathcal{L}}(z)$ according to the confidence score z is

$$\hat{f}_{\mathcal{L}}(z) = \frac{1}{M} \sum_{l=1}^{L} \phi_f(z - x_l)\,. \tag{24}$$

Also, the estimated cumulative distribution is $\hat{F}_{\mathcal{L}}(z) = \frac{1}{M}\sum_{l=1}^{L} \Phi_f(z - x_l)$.

Following a similar derivation as in the ROC case, the area under the LROC curve is given by

$$\hat{A}_{\mathcal{L}} = \frac{1}{MN} \sum_{l=1}^{L}\sum_{n=1}^{N} H_\varepsilon(x_l - y_n)\,. \tag{25}$$

The derivation of the equation for the standard deviation ($\sigma_{\hat{A}_{\mathcal{L}}}$) of the $\hat{A}_{\mathcal{L}}$ estimate is presented in appendix A, and is obtained similarly to the ROC case with the difference that we now have L that is also a random variable.

A different version of equation (A12) without the $\langle L\rangle$ factors can be obtained if instead of the expectations defined by the equations (A5)–(A8) for the normalized distribution $f^*_{\mathcal{L}}$, we use the equivalent quantities for the non-normalized distribution $f_{\mathcal{L}}$. They are related by the following relations

$$\begin{aligned} A^*_{\mathcal{L}} &= \frac{A_{\mathcal{L}}}{Q}\,, & B^*_{\mathcal{L}11} &= \frac{B_{\mathcal{L}11}}{Q}\,,\\ B^*_{\mathcal{L}12} &= \frac{B_{\mathcal{L}12}}{Q}\,, & B^*_{\mathcal{L}21} &= \frac{B_{\mathcal{L}21}}{Q^2}\,, \end{aligned} \tag{26}$$

and with the notations

$$\begin{aligned} S_{\mathcal{L}0} &= B_{\mathcal{L}11} - \frac{1}{Q}A^2_{\mathcal{L}}\,, & S_{\mathcal{L}1} &= B_{\mathcal{L}12} - \frac{1}{Q}A^2_{\mathcal{L}}\,,\\ S_{\mathcal{L}2} &= B_{\mathcal{L}21} - A^2_{\mathcal{L}}\,, & S_{\mathcal{L}3} &= \frac{1-Q}{Q}A^2_{\mathcal{L}}\,, \end{aligned} \tag{27}$$

we obtain the more compact form

$$\sigma^2_{\hat{A}_{\mathcal{L}}} = \frac{S_{\mathcal{L}0} + (N-1)S_{\mathcal{L}1} + (M-1)S_{\mathcal{L}2} + NS_{\mathcal{L}3}}{MN}\,. \tag{28}$$

If the Heaviside function (11) is used as rank function and the data follows a continuous scale, as is often considered in the theoretical evaluations, then the terms B_{11} given by equation (16), and $B^*_{\mathcal{L}11}$ given by equation (A6) are equal to A and $A^*_{\mathcal{L}}$, respectively. Therefore equation (20) is sometimes presented in the literature[10] with A in place of the B_{11} term. However, if the data scale is discrete, which is almost always the case in practice, or a different rank function is used, the B_{11} and $B^*_{\mathcal{L}11}$ factors should be computed separately, otherwise unwanted discrepancies may result. In the discrete case the expectations given by equations (15)–(18) and (A5)–(A8) are obtained by using sums instead of integrals.

C. Estimation of the standard deviations from samples

In practice we do not know the distributions f, $f_{\mathcal{L}}$ or g, and all we have are the observed samples. In the case of ROC analysis we can obtain estimates of the variance $\hat{\sigma}^2_{\hat{A}}$ from equations (19) and (20) by replacing the theoretical A, B_{11}, B_{12}, B_{21} with the sample estimates $\hat{A}$, and respectively

$$\hat{B}_{11} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} {H_\varepsilon}^2(x_m - y_n) \tag{29}$$

$$\hat{B}_{12} = \frac{1}{MN^2}\sum_{m=1}^{M}\sum_{\substack{n_1=1\\ n_2=1}}^{N} H_\varepsilon(x_m - y_{n_1})H_\varepsilon(x_m - y_{n_2}) \tag{30}$$

$$\hat{B}_{21} = \frac{1}{M^2N}\sum_{\substack{m_1=1\\ m_2=1}}^{M}\sum_{n=1}^{N} H_\varepsilon(x_{m_1} - y_n)H_\varepsilon(x_{m_2} - y_n) \tag{31}$$

In the case of the LROC analysis we can obtain the variance estimate $\hat{\sigma}^2_{\hat{A}_{\mathcal{L}}}$ from equations (27) and (28) by replacing the theoretical values $A_{\mathcal{L}}$, $B_{\mathcal{L}11}$, $B_{\mathcal{L}12}$, $B_{\mathcal{L}21}$ and Q with the sample estimates $\hat{A}_{\mathcal{L}}$, and respectively

$$\hat{B}_{\mathcal{L}11} = \frac{1}{MN}\sum_{l=1}^{L}\sum_{n=1}^{N} {H_\varepsilon}^2(x_l - y_n) \tag{32}$$

$$\hat{B}_{\mathcal{L}12} = \frac{1}{MN^2}\sum_{l=1}^{L}\sum_{\substack{n_1=1\\ n_2=1}}^{N} H_\varepsilon(x_l - y_{n_1})H_\varepsilon(x_l - y_{n_2}) \tag{33}$$

$$\hat{B}_{\mathcal{L}21} = \frac{1}{M^2N}\sum_{\substack{l_1=1\\ l_2=1}}^{L}\sum_{n=1}^{N} H_\varepsilon(x_{l_1} - y_n)H_\varepsilon(x_{l_2} - y_n) \tag{34}$$

$$\hat{Q} = \frac{L}{M}\,. \tag{35}$$

The practical application of the ROC or LROC analysis is simple and can be done simultaneously for both cases if the ranking scores contain flags indicating the correct localizations. Once the ranking function H_ε has been specified, for instance in the form given by equation (11), the estimation of the area $\hat{A}$ or $\hat{A}_\mathcal{L}$ is done by the very simple evaluation of the equations (9) or (25), respectively. For the standard deviation calculations the procedure consists of evaluation of the equations (29)–(31) for the ROC case, and of the equations (32)–(34) and (35) for the LROC case. For the evaluation of the $\hat{B}_{12}$ and $\hat{B}_{21}$, and of the $\hat{B}_{\mathcal{L}\,12}$ and $\hat{B}_{\mathcal{L}\,21}$ respectively, the procedures involve just an extra exterior loop passing through the X or Y data sets. Computer coding of these procedures is very simple and can be done using simple scripting languages as we have initially done using the standard UNIX tool “awk”, or implemented as macro-commands in a spreadsheet program.

D. The optimal N/M ratios

We are interested in obtaining the desired level of precision for the estimates of the area under the curve by using a minimum number of image samples. The equations (20) and (28) can be rewritten as expressions of the variance σ^2 as a function of the number of signal present images M and the total number of images $T = M + N$. For a given T we want to determine the number M for which the variance $\sigma^2(M,T)$ is minimum . That is obtained for $\frac{\mathrm{d}\sigma^2(M,T)}{\mathrm{d}M} = 0$. In both ROC and LROC cases we obtain a second degree equation which has as positive solution the optimal M we are looking for. Eliminating the terms with negligible contributions we obtain the following relations

$$\left(\frac{N}{M}\right)_{\text{ROC}} = \sqrt{\frac{S_2}{S_1}}\,, \tag{36}$$

$$\left(\frac{N}{M}\right)_{\text{LROC}} = \sqrt{\frac{S_{\mathcal{L}2}}{S_{\mathcal{L}1}+S_{\mathcal{L}3}}}\,. \tag{37}$$

The optimal N/M ratios, as well as the total number of scored images T necessary to reach the desired precision of the estimations, can be determined from samples by performing a pilot study, as is usually recommended.

The simulation examples discussed in next section (see TABLE I and FIG. 6) show that there is an asymmetry between the required number of signal present images M and signal absent images N, and in the case of LROC this asymmetry is very pronounced.

III. SIMULATION STUDY

In order to test the estimators derived above and study their properties, we have applied a scan statistic model[6] to simulate the overall image abnormality test (no signal localization) and the signal detection test (with localization). The model considers that arbitrary noise clumps (noise nodules) appear in the image background which could be mistaken for true signals (features or targets) that appear also as small bumps of higher concentration than the surrounding background because of the limited detector resolution. The nodules' size, or the contrast or a combination of these, is determined by a scanning procedure (performed either numerically or by the human observers[5]). For each image, the maximum variable z obtained by scanning the entire interest region is used as an indicator of image normality/abnormality. For an ensemble of images, the distribution of the maximum scan value from background only regions of a given area is $g(z)$, while the distribution of the signals alone is $s(z)$. The cumulative distribution function of $s(z)$ is $S(z) = \int_{-\infty}^{z} s(z')\mathrm{d}z'$. If a signal is present in an image, the distribution of the max-scan value is[6,11]

$$f(z) = s(z)G(z) + g(z)S(z)\,, \tag{38}$$

while the distribution of the image max-scan that corresponds to an actual signal (correct localization) is

$$f_\mathcal{L}(z) = s(z)G(z)\,. \tag{39}$$

We have the cumulative distributions $F(z) = 1 - S(z)G(z)$, and $F_\mathcal{L}(z) = \int_{-\infty}^{z} s(z')G(z')\mathrm{d}z$, respectively.

The signal distribution can be modeled using a Gaussian $s(z) = C_s \exp\left[-\frac{(z-\mu_s)^2}{2\sigma_s^2}\right]$ with μ_s the average signal value and σ_s the standard deviation. The constant C_s is obtained so that $s(z)$ is normalized to unity on the interval of interest. According to our model[6] the background max-scan distribution is given by $g(z) = \nu_0 p(z)\exp\left[-\nu_0 q(z)\right]$, with $\nu_0 = n_0\Omega$ where n_0 is the density of the noise nodules with z greater than a certain threshold z_0, and Ω is the analyzed image area (or volume); $p(z)$ is the distribution of the noise nodules and $q(z) = \int_{z}^{\infty} p(z')\mathrm{d}z'$. The noise nodules distribution can be approximated by the tail of a Gaussian $p(z) = C_p \exp\left[-\frac{(z-\mu_p)^2}{2\sigma_p^2}\right]$, where μ_p, and σ_p are parameters and C_p is a normalization constant for the interval of interest.

The parameters for the $s(z)$ and $g(z)$ functions were chosen for two cases, so that the theoretical A and $A_\mathcal{L}$ are $A_1 = 0.79$, $A_{\mathcal{L}1} = 0.59$ in the first case and $A_2 = 0.86$, $A_{\mathcal{L}2} = 0.71$ in the second case. The underlying parameters are $\Omega = 100\,\mathrm{cm}^2$, $z_0 = 1.0$, and for the first case: $n_0 = 0.50$, $\mu_p = 1.0$, $\sigma_p = 0.23$, $\mu_s = 1.7$, $\sigma_s = 0.235$; and for the second case: $n_0 = 0.50$, $\mu_p = 1.0$, $\sigma_p = 0.26$, $\sigma_s = 1.88$, $\sigma_s = 0.272$. The two cases were selected from the examples used in a small nodule detectability study[6,22] considering 2D PET image reconstructions. The model functions and their corresponding ROC and LROC curves are shown in FIG. 2.

Random sets of test scores were generated using the Monte Carlo method. We sampled the the model distribution using the alias sampling method[23,24]. For the signal present images the test scores were sampled by randomly generating a variable x according to s, and a variable y according to g. The final score returned was $\max(x, y)$, and if the maximum corresponded to the x variable then a correct localization was recorded. In the discrete case, if a tie $x = y$ occurs, then a random assignment of the correct localization is made (50% chance). Also, the equations (38) and (39) should be adjusted

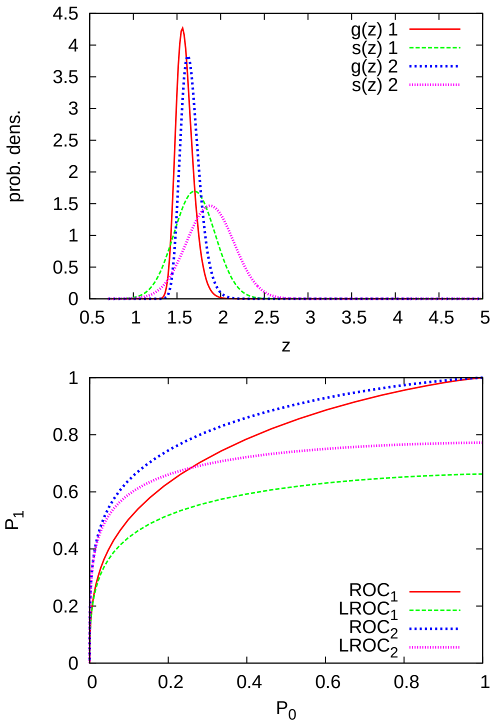

FIG. 2: Comparison between the model functions (top) and their corresponding ROC and LROC curves (bottom) for the two test cases considered.

for the non-negligible probability of a tie

$$
\begin{aligned}
f_{\mathcal{L}}(z) &= s(z)G(z) + \frac{1}{2}s(z)g(z)\,, &(40)\\
f(z) &= s(z)G(z) + g(z)S(z) + s(z)g(z)\,, &(41)
\end{aligned}
$$

where now $S(z) = \sum_{z'<z} s(z')$ and $G(z) = \sum_{z'<z} g(z')$.

From the randomly generated test score samples the $\hat{A}$, $\hat{A}_{\mathcal{L}}$ and their standard deviations $\hat{\sigma}_{\hat{A}}$, $\hat{\sigma}_{\hat{A}_{\mathcal{L}}}$, respectively, were computed and the results histogrammed. We also determined the averages $\langle\hat{A}\rangle$, $\langle\hat{A}_{\mathcal{L}}\rangle$, $\langle\hat{\sigma}_{\hat{A}}\rangle$, $\langle\hat{\sigma}_{\hat{A}_{\mathcal{L}}}\rangle$, and the sample standard deviation of the $\hat{\sigma}_{\hat{A}}$ and $\hat{\sigma}_{\hat{A}_{\mathcal{L}}}$, denoted here with $\sigma_{\hat{\sigma}_{\hat{A}}}$ and $\sigma_{\hat{\sigma}_{\hat{A}_{\mathcal{L}}}}$, respectively. In TABLE I are summarized the results obtained for the two test cases considered using sets of test scores with each set having $M = 120$ and $N = 90$ signal present and signal absent, respectively, samples. A few thousand of test score sets were used in the Monte Carlo simulations.

We have performed tests using various discretization levels, from hundreds to only a few discrete levels for the variable z over the main range of interest (the region where f and g vary significantly). The results from both theoretical calculations and simulation estimates showed consistency. The results presented below were obtained for several tens of discrete levels. A full study of the variation of the results with the discretization level of the scoring scale, which should also include the effect of changing the shape of the rank function H_{ε}, is beyond the scope of this paper.

TABLE I: Summary of the results obtained for the two test cases considered, with $M = 120$ and $N = 90$ signal present and signal absent, respectively, samples.

	1	2		1	2
A	0.795	0.857	$A_{\mathcal{L}}$	0.590	0.715
$\langle\hat{A}\rangle$	0.796	0.858	$\langle\hat{A}_{\mathcal{L}}\rangle$	0.591	0.714
$\sigma_{\hat{A}}$	0.030	0.025	$\sigma_{\hat{A}_{\mathcal{L}}}$	0.042	0.038
$\langle\hat{\sigma}_{\hat{A}}\rangle$	0.031	0.025	$\langle\hat{\sigma}_{\hat{A}_{\mathcal{L}}}\rangle$	0.044	0.037
$\frac{\sigma_{\hat{A}}}{A}$	3.9%	3.0%	$\frac{\sigma_{\hat{A}_{\mathcal{L}}}}{A_{\mathcal{L}}}$	7.4%	5.2%
$\frac{\sigma_{\hat{\sigma}_{\hat{A}}}}{\sigma_A}$	7.0%	9.5%	$\frac{\sigma_{\hat{\sigma}_{\hat{A}_{\mathcal{L}}}}}{\sigma_{A_{\mathcal{L}}}}$	2.1%	4.6%
$\left(\frac{N}{M}\right)_{\mathrm{ROC}}$	0.74	0.65	$\left(\frac{N}{M}\right)_{\mathrm{LROC}}$	0.22	0.21

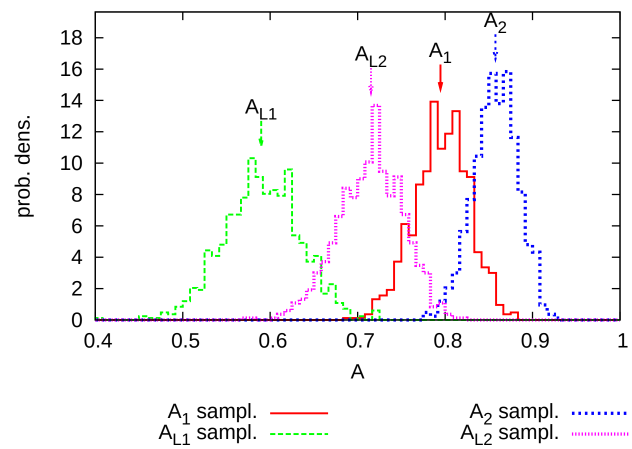

FIG. 3: Distributions of the $\hat{A}$ and $\hat{A}_{\mathcal{L}}$ estimates obtained by Monte Carlo simulation for the two test cases considered. The theoretical values are marked on the graph.

In FIG. 3 we compare the area estimates $\hat{A}$, $\hat{A}_{\mathcal{L}}$ distributions with the theoretical values, for the two cases considered. In FIG. 4 we compare the distributions of the standard deviation estimates $\hat{\sigma}_{\hat{A}}$, $\hat{\sigma}_{\hat{A}_{\mathcal{L}}}$. One can notice that for the LROC cases the standard deviation estimates are more precise (narrower and taller peaks).

The difference between two area estimates $\delta = \hat{A}_2 - \hat{A}_1$ is a random variable with the variance $\sigma_\delta^2 = \sigma_{\hat{A}_1}^2 + \sigma_{\hat{A}_2}^2$. The variable $\frac{\delta}{\sigma_\delta}$ is usually assumed to be standard normal distributed and can be used to determine the significance levels for the test comparing the two cases[10,25]. A larger $\frac{\delta}{\sigma_\delta}$ corresponds to better differentiations of the two imaging procedures. In FIG. 5 can be compared $\frac{\delta}{\sigma_\delta}$ for the ROC and LROC cases as a function of the number of samples M and N. The area estimates $\hat{A}_1$ and $\hat{A}_2$ represent the same two cases considered above.

In FIG. 6 (left) is shown the dependence of $\frac{\delta}{\sigma_\delta}$ on the total number of samples $M + N$ for the optimal N/M ratio point for the ROC and LROC, respectively, tests. FIG. 6 (right) shows the plots $N(M)$ that have as slopes the optimal N/M

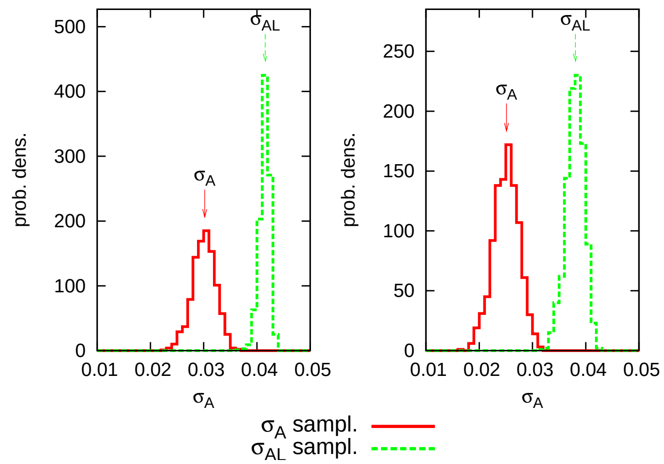

FIG. 4: The distribution of the standard deviation estimates $\hat{\sigma}_{\hat{A}}$, $\hat{\sigma}_{\hat{A}_{\mathcal{L}}}$ obtained by Monte Carlo simulation for the two test cases considered (left case 1, right case 2). The theoretical values are marked on the graphs.

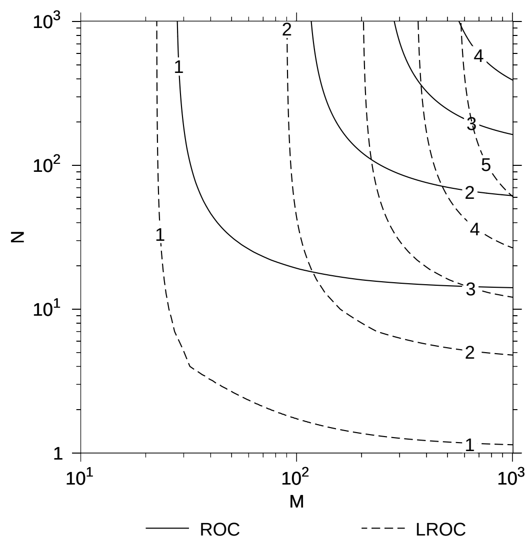

FIG. 5: Comparison between the ROC and LROC values of $\frac{\delta}{\sigma_\delta}$ as a function of the number of samples M and N, for the two test cases considered. The plotted curves are isolines on the 3-D surfaces defined by $\frac{\delta}{\sigma_\delta}$ as a function of M and N for several representative values marked on the plots.

ratio for the ROC and LROC tests. The optimal N/M ratio points shown in FIG. 6 were determined from the two-dimensional maps plotted in FIG. 5 and are specific to the comparison between the test case 1 and test case 2. These values are in between the optimal ratios of each distinct test case shown in TABLE I, which are optimal from the point of view of area under the curve estimation for that specific case.

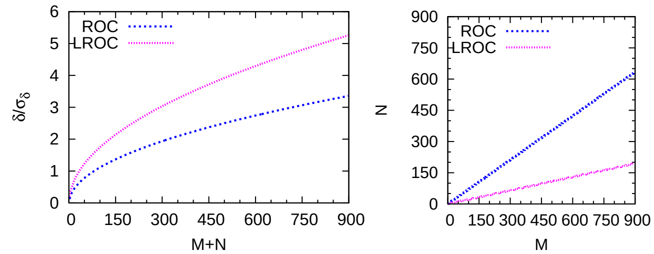

FIG. 6: The $\frac{\delta}{\sigma_\delta}$ ratio (left) as a function of the total number of samples $M + N$ for the optimal N/M ratio for the ROC and LROC tests, respectively. On the right are shown the plots of N as a function of M for the optimal N/M ratios for the comparison between the specific two test cases considered.

IV. DISCUSSION AND CONCLUSIONS

In this paper a nonparametric approach for LROC analysis is presented. The equations for estimation of the area under the LROC curve, and for the variance calculation, are derived. Equations are also derived for the optimal choice of the ratio between the number of signal absent and signal present image samples. These results can be applied with either continuous or discrete scoring scales.

The simulation studies carried out validate the theoretical derivations and show (see FIG. 5 and FIG. 6) that the LROC analysis is considerably more sensitive than the ROC analysis, as was previously demonstrated for the parametric case[12]. This is not surprising since the LROC approach takes advantage of the additional positional information. The area under the ROC curve (A) is bounded in the interval $(\frac{1}{2}, 1)$, which makes it a poor discriminator especially when the values to be compared are close to the extremities of the interval (very poor or very good tests). The area under the LROC curve ($A_{\mathcal{L}}$) is bounded in a larger interval $(0, 1)$ and the results of relatively good or poor tests do not cluster near the interval limits in the same manner as the ROC results.

The main advantages of a nonparametric method are that it does not rely on particular assumptions about the image score distributions, and its simplicity. The Swensson LROC model assumes a combination of two distributions: $g(z)$ for the image background, and $s(z)$ for the signal. In the applications of this model so far, the two distributions are assumed normal. There are theoretical reasons[6] to assume that the background distribution has an asymmetric shape instead of being a symmetric Gaussian. In the case of non-uniform images the asymmetry of the distribution tails can be even more pronounced since the distribution for the full image is the result of the distributions for different background regions with different noise levels. In more realistic conditions the signal distribution could also have asymmetric tails. The accuracy of the parametric fit depends on choosing an adequate parameterization. It must be emphasized that when the statistical model considered by the fit does not match the reality, the fit procedure could also report erroneous error estimates. The nonparametric approaches are more robust in this respect.

Similar to the ROC analysis, the LROC analysis can also be applied to more varied situations other than the radiological diagnostic case. We can consider the general case when we have signal present and signal absent results, and in the signal present case the results are partially contaminated with background noise. Provided that there is an additional test that can check for contamination (as we have the test for correct localization in the case of nodule detection) one can use a similar LROC type analysis.

The nonparametric LROC approach presented here can be extended in the same way as previously was done for the nonparametric ROC method to the cases of correlated data (e.g. same data multiple readers) or other situations documented in the literature[26–29]. Also, further improvements are possible by applying smoothing kernels as discussed here, and demonstrated previously for the ROC case[15,16]. By restricting the domain of integration (or summation) one can easily obtain the equations for the estimation of the partial areas A and $A_{\mathcal{L}}$, and their standard deviations.

The Monte Carlo simulation module we have used for this work can be easily modified and extended in order to be used as a tool for design of experiments and for analysis of data.

Acknowledgments

The author expresses his thanks to Dr. Robert Lewitt for his suggestions and critical reading of the manuscript.

This work was supported by the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health under grant R21-EB005434.

APPENDIX A: DERIVATION OF $\sigma^2_{\hat{A}_{\mathcal{L}}}$ EXPRESSION

From the variance definition we have

$$\sigma^2_{\hat{A}_{\mathcal{L}}} = \langle \hat{A}^2_{\mathcal{L}} \rangle - \langle \hat{A}_{\mathcal{L}} \rangle^2 \,. \tag{A1}$$

The average quantities $\langle \hat{A}_{\mathcal{L}} \rangle$ and $\langle \hat{A}^2_{\mathcal{L}} \rangle$ are given by

$$\langle \hat{A}_{\mathcal{L}} \rangle = \sum_{L=0}^{M} P(L|M) \langle \hat{A}_{\mathcal{L}} \rangle_L \,, \tag{A2}$$

$$\langle \hat{A}^2_{\mathcal{L}} \rangle = \sum_{L=0}^{M} P(L|M) \langle \hat{A}^2_{\mathcal{L}} \rangle_L \,, \tag{A3}$$

where $P(L|M)$ is the probability of correctly localizing L features out of M signal present images, $\langle \hat{A}_{\mathcal{L}} \rangle_L$ and $\langle \hat{A}^2_{\mathcal{L}} \rangle_L$ are the mean and the second moment of $\hat{A}_{\mathcal{L}}$ for a fixed value of L.

The distribution of the random variable L is given by the binomial distribution $P(L|M) = \binom{M}{L} Q^L (1-Q)^{M-L}$; its mean and variance are $\langle L \rangle = MQ$ and $\sigma^2_L = MQ(1-Q)$.

Proceeding as in the ROC case, we obtain

$$\langle \hat{A}^2_{\mathcal{L}} \rangle_L = L \frac{B^*_{\mathcal{L}11} - {A^*_{\mathcal{L}}}^2 + (N-1)(B^*_{\mathcal{L}12} - {A^*_{\mathcal{L}}}^2)}{M^2 N} + L(L-1) \frac{B^*_{\mathcal{L}21} - {A^*_{\mathcal{L}}}^2}{M^2 N} + L^2 \frac{{A^*_{\mathcal{L}}}^2}{M^2} \,, \tag{A4}$$

where $A^*_{\mathcal{L}}$, $B^*_{\mathcal{L}11}$, $B^*_{\mathcal{L}12}$ and $B^*_{\mathcal{L}21}$ denote the following expectations

$$A^*_{\mathcal{L}} = \langle H_\varepsilon(x-y) \rangle_{f^*_{\mathcal{L}}(x) g(y)} \tag{A5}$$

$$B^*_{\mathcal{L}11} = \langle {H_\varepsilon}^2(x-y) \rangle_{f^*_{\mathcal{L}}(x) g(y)} \tag{A6}$$

$$B^*_{\mathcal{L}12} = \langle H_\varepsilon(x-y_1) H_\varepsilon(x-y_2) \rangle_{f^*_{\mathcal{L}}(x) g(y_1) g(y_2)} \tag{A7}$$

$$B^*_{\mathcal{L}21} = \langle H_\varepsilon(x_1-y) H_\varepsilon(x_2-y) \rangle_{f^*_{\mathcal{L}}(x_1) f^*_{\mathcal{L}}(x_2) g(y)} \tag{A8}$$

We have used similar notations as defined in equations (15)–(18). From the equations (A3) and (A4) we obtain

$$\begin{aligned} \langle \hat{A}^2_{\mathcal{L}} \rangle &= \frac{B^*_{\mathcal{L}11} - {A^*_{\mathcal{L}}}^2 + (N-1)(B^*_{\mathcal{L}12} - {A^*_{\mathcal{L}}}^2)}{M^2 N} \sum_{L=1}^{M} P(L|M) + \\ &\quad \frac{B^*_{\mathcal{L}21} - {A^*_{\mathcal{L}}}^2}{M^2 N} \sum_{L=1}^{M} P(L|M) \cdot (L^2 - L) + \frac{{A^*_{\mathcal{L}}}^2}{M^2} \sum_{L=1}^{M} P(L|M) L^2 \\ &= \frac{B^*_{L11} - {A^*_{\mathcal{L}}}^2 + (N-1)(B^*_{\mathcal{L}12} - {A^*_{\mathcal{L}}}^2)}{M^2 N} \langle L \rangle + \frac{B^*_{\mathcal{L}21} - {A^*_{\mathcal{L}}}^2}{M^2 N} (\langle L^2 \rangle - \langle L \rangle) + \frac{{A^*_{\mathcal{L}}}^2}{M^2} \langle L^2 \rangle \end{aligned} \tag{A9}$$

We also have

$$\langle \hat{A}_{\mathcal{L}} \rangle_L = \frac{L}{M} A^*_{\mathcal{L}} \tag{A10}$$

and from equation (A2) results

$$\langle \hat{A}_{\mathcal{L}} \rangle = \frac{1}{M} A^*_{\mathcal{L}} \sum_{L=0}^{M} P(L|M) \cdot L = \frac{\langle L \rangle}{M} A^*_{\mathcal{L}} \tag{A11}$$

In the end the variance is given by the equation

$$\sigma^2_{\hat{A}_\mathcal{L}} = \frac{B^*_{\mathcal{L}11} - {A^*_\mathcal{L}}^2 + (N-1)(B^*_{\mathcal{L}12} - {A^*_\mathcal{L}}^2)}{M^2 N}\langle L\rangle + \frac{B^*_{\mathcal{L}21} - {A^*_\mathcal{L}}^2}{M^2 N}\left[\langle L\rangle(\langle L\rangle - 1) + \sigma_L^2\right] + \frac{{A^*_\mathcal{L}}^2}{M^2}\sigma_L^2$$

[*] e-mail: popescu@mipg.upenn.edu

[1] D. D. Dorfman and E. Alf, "Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals — rating-method data," *J. Math. Psychol.*, vol. 6, pp. 487–496, 1969.

[2] C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.*, vol. 21, pp. 720–733, 1986.

[3] R. G. Swensson, "A two-stage detection model applied to skilled visual search by radiologists," *Percep. and Psychophys.*, vol. 27, pp. 11–16, 1980.

[4] D. P. Chakraborty, "A search model and figure of merit for observer data acquired according to the free-response paradigm," *Phys. Med. Biol.*, vol. 51, pp. 3449–3462, 2006.

[5] H. L. Kundel, C. F. Nodine, and E. A. Krupinski, "Searching for lung nodules: visual dwell indicates location of false-positive and false-negative decisions," *Invest. Radiol.*, vol. 4, pp. 472–478, 1989.

[6] L. M. Popescu and R. M. Lewitt, "Small nodule detectability evaluation using a generalized scan statistic model," *Phys. Med. Biol.*, vol. 51, no. 23, pp. 6225–6244, 2006.

[7] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *J. Math. Psy.*, vol. 12, pp. 387–415, 1975.

[8] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[9] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistic*, vol. 18, no. 1, pp. 50–60, 1947.

[10] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, pp. 29–36, Apr. 1982.

[11] R. G. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Med. Phys.*, vol. 23, no. 10, pp. 1709–1725, 1996.

[12] R. G. Swensson, "Using localization data from image interpretations to improve estimates of performance accuracy," *Medical Decision Making*, vol. 20, no. 2, pp. 170–185, 2000.

[13] R. G. Swensson, J. L. King, and D. Gur, "A constrained formulation for the receiver operating characteristic (ROC) curve based on probability summation," *Med. Phys.*, vol. 28, pp. 1597–1609, Aug. 2001.

[14] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.

[15] C. J. Lloyd, "Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems," *J. Am. Stat. Assoc.*, vol. 93, no. 444, pp. 1356–1998, 1998.

[16] K. H. Zou, W. J. Hall, and D. E. Shapiro, "Smooth non-parametric receiver operating characteristic (ROC) for continuous diagnostic tests," *Statist. Med.*, vol. 16, pp. 2143–2156, 1997.

[17] G. E. Noether, *Elements of nonparametric statistics*. New York: John Wiley and Sons, 1967.

[18] P. C. Bunch, J. F. Hamilton, G. K. Sanderson, and A. H. Simmons, "Free-response approach to the measurement and characterization of radiographic-observer performance," *Journal of Applied Photographic Engineering*, vol. 4, no. 4, pp. 166–171, 1978.

[19] D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Phys. Med. Biol.*, vol. 16, no. 4, pp. 561–568, 1989.

[20] D. P. Chakraborty and L. H. L. Winter, "Free-response methodology: alternate analysis and a new observer-performance experiment," *Radiology*, vol. 174, pp. 873–881, 1990.

[21] D. C. Edwards, M. A. Kupinski, C. E. Metz, and R. M. Nishikawa, "Maximum likelihood fitting of FROC curves under an initial-detection-and-candidate-analysis model," *Med. Phys.*, vol. 29, no. 12, pp. 2861–2870, 2002.

[22] L. M. Popescu and R. M. Lewitt, "Comparison between TOF and non-TOF PET using a scan statistic numerical observer," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, (San Diego, California), 29 Oct. – 4 Nov. 2006. (M05-4).

[23] A. J. Walker, "An efficient method for generating discrete random variables with general distributions," *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 253–256, 1977.

[24] L. M. Popescu, "An extension of alias sampling method for parametrized probability distributions," *J. Comp. Phys.*, vol. 160, pp. 612–622, 2000.

[25] N. A. Obuchowski and M. L. Lieber, "Confidence intervals for the receiver operating characteristic area in studies with small samples," *Acad. Radiol.*, vol. 5, pp. 561–571, 1998.

[26] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, pp. 839–843, Sept. 1983.

[27] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, pp. 837–844, 1988.

[28] N. A. Obuchowski, "An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale," *Statist. Med.*, vol. 25, pp. 481–493, 2006.

[29] E. Clarkson, M. A. Kupinski, and H. H. Barrett, "A probabilistic model for the MRMC method, part 1: Theoretical development," *Acad. Radiol.*, vol. 13, no. 11, pp. 1410–1421, 2006.