Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

An R Package for (F)ROC Data Analysis: JAFROCwR

Xuetong Zhai
University of Pittsburgh

Dev Chakraborty
University of Pittsburgh

Abstract

TBD

Keywords: JAFROC, ROC, FROC, software, R.

1. Introduction

The evaluation of medical imaging systems is a major problem in diagnostic radiology. Receiver operating characteristic (ROC) analysis is widely used in the evaluation. The method is asking readers (observers) to give ratings to images for their confidence of lesion existence without knowing the truth. Higher ratings usually means higher confidence that there are lesions in the image. Then the ratings are used to calculate true positive fraction (TPF) and false positive fraction (FPF). The ROC curve is defined by the points of the sensitivity (aka TPF) versus specificity (aka 1 - FPF). There are several statistical methods available for multi-reader multi-case (MRMC) ROC analysis with different figures of merit. The Dorfman-Berbaum-Metz (DBM) algorithm is the first published and most widely used method for MRMC ROC analysis. It reports a p-value of the statistical test of the null hypothesis (NH) that the modalities are actually identical. Besides DBM method, there are at least five more approaches to the MRMC ROC study. They are (1) Obuchowski and Rockette's ANOVA approach; (2) Toledano and Gatsonis' ordinal regression methodology; (3) Song's analysis of correlated ROC areas in diagnostic testing; (4) Ishwaran and Gatsonis' hierarchical ordinal regression models; (5) the Beiden-Wagner-Campbell (BWC) multiple-bootstrap method.

Software implementation availability for these methods are given in table 1. Following is the brief descriptions of the available softwares.

- OR-DBM MRMC is designed by the Medical Image Perception Laboratory of the University of Iowa. OR and DBM method are implemented and unified for MRMC ROC

Table 1: Software availability of MRMC ROC analysis methods

Algorithm	Online Downloadable Softwares
Dorfman-Berbaum-and Metz (DBM)	OR-DBM MRMC, JAFROC, iMRMC
Obuchowski and Rockette (OR)	OR-DBM MRMC, JAFROC, iMRMC
Toledano and Gatsonis	None
Song	None
Ishwaran and Gatsonis	None
Beiden-Wagner-Campbell (BWC)	None

analysis. The area under ROC curve (AUC), partial area under ROC curve (pAUC), the sensitivity at specified specificity, and the specificity at specified sensitivity are used as figures of merit in the analysis. Four ROC curve estimation methods are implemented in the program, which are RSCORE, contaminated binormal method (CBM), PROPROC and the trapezoidal/Wilcoxon method. Trapezoidal/Wilcoxon method calculates the empirical estimation of AUC while the other methods fit the ROC curve using distributional assumptions with different parameters.

- JAFROC, an acronym for *j*ackknife *a*lternative *f*ree-response *r*eceiver *o*perating *c*haracteristic, is a free software applicable to the planning and analysis of FROC studies. The software is developed by Dr. Dev Chakraborty, Dr. Hong-Jun Yoon and Xuetong Zhai. For ROC data, JAFROC performs MRMC analysis using DBM method with AUC figure of method. Empirical estimation of AUC is computed using trapezoidal/Wilcoxon method, while no curve fitting method is included. JAFROC has the capability to analyze data accounting for localization and multiple lesions per image, which is described in section 4.

- iMRMC is a Java-based software. It is a open source project under the management of Dr. Brandon Gallas, Dr. Xin He and Rohan Pathare who are from US Food and Drug Administration. It estimates the variance components that are used in methods besides DBM and OR, including BDG, BCK and MS variance components. Hypothesis testing and confidence interval on the difference of two modalities are calculated using t-statistic with corresponding degrees of freedom. Similar with JAFROC, iMRMC also only selects AUC as the figure of merit, and estimates empirical AUC with trapezoidal/Wilcoxon method.

All of these three softwares are GUI (graphical user interface)-based, so it is difficult to call functions or grab the results from other program. But this is important for users with programming requirements. It is absolutely unreasonable to let users run softwares manually and copy the result to their own code every time. Considering R is widely used in statistic and data analysis, it is necessary to develop an R package with the capability to perform MRMC ROC analysis. Firstly, R users will be able to call functions in the package from their programs. Since all R packages are open source and R is a "higher" level language, users have the possibility to debug and modify the source code to satisfy their requirements without complicated programming skills. In addition, thanks to the amazing extensibility of R, packages will be available to be called from other languages, including C/C++, Python, Java and MATLAB, on all major platforms. Therefore, we developed the package **JAFROCwR**

(JAFROC within R) to make the analysis easy and convenient, which implements all features in JAFROC. The package is available to download from the Comprehensive R Archive Network at http://CRAN.R-project.org/package=JAFROCwR. Table 2 shows a brief comparison of the three softwares and the package **JAFROCwR**.

Table 2: Softwares comparison

Softwares	OR-DBM MRMC	JAFROC	iMRMC	JAFROCwR
Data Input	Plain text	Excel file	Plain text	Both
Open Source/Language	No/C++	No/C++	Yes/Java	Yes/R
Cross Platforms	No	No	Yes	Yes
Call from Other Languages	No	No	No	Yes
ROC Curve Fitting	Yes	No	No	No
Localization	No	Yes	No	Yes

This paper is organized as follows. The statistical model and method used in the package **JAFROCwR** are introduced in Section 2. Section 3 includes examples of the use of the package and analysis results of a same dataset from the package and other softwares. Localization issue and other additional features are shown in 4. Limitations and potential future updates is discussed in Section 5. Finally, the conclusions are described in Section 6.

2. Statistical Models and Methods

Let Z_{ijk} denote the rating given to the kth case by the jth reader using the ith modality with $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, where I is the number of modalities, J is the number of readers and K is the number of cases. Five or more integers are usually used with higher values meaning a higher confidence that the case contains lesions. The area under the ROC curve (AUC) can be used as a figure of merit, which is estimated by

$$\hat{\theta}_{ij} = \frac{1}{K_1 K_2} \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \psi(Z_{ijk_1}, Z_{ijk_2})$$

where k_1 and k_2 denote the indices for non-diseased and diseased cases with K_1 and K_2 being the total numbers of them respectively ($K = K_1 + K_2$). The Wilcoxon function ψ is defined as,

$$\psi(X, Y) = \begin{cases} 1 & X < Y \\ 0.5 & X = Y \\ 0 & X > Y \end{cases} \tag{1}$$

2.1. DBM Method

For the DBM method, the pseudovalues of AUC are calculated using jackknife for combination of each modality, reader and case. Let Y_{ijk} indicate the AUC pseudovalue of modality i, reader j and case k. Y_{ijk} is defined as $Y_{ijk} = K\hat{\theta}_{ij} - (K-1)\hat{\theta}_{ijk}$, where $\hat{\theta}_{ijk}$ is the AUC estimate of modality i, reader j with the rating of case k removed. The statistical model is given by

$$Y_{ijk} = \mu + \tau_i + R_j + C_k + (\tau R)_{ij} + (\tau C)_{ik} + (RC)_{jk} + (\tau RC)_{ijk} + \varepsilon_{ijk}$$

where τ_i denotes the fixed effect of the ith modality, R_j denotes the random effect of the jth reader, C_k denotes the random effect of the kth case, and ε_{ijk} is the error term. The interaction terms denote the mixed random effects of corresponding factors. All of the random terms are assumed to be independent and follow normal distributions with zero means and variances σ_R^2, σ_C^2, $\sigma_{\tau R}^2$, $\sigma_{\tau C}^2$, σ_{RC}^2, $\sigma_{\tau RC}^2$ and σ_ε^2.

The F statistic for testing the null hypothesis of equal modality effects is defined,

$$F_{DBM} = \frac{MS(T)}{MS(TR) + \max[MS(TC) - MS(TRC), 0]}$$

with the numerator degrees of freedom $t-1$ and denominator degrees of freedom

$$ddf_{DBM} = \frac{[MS(TR) + MS(TC) - MS(TRC)]^2}{\frac{MS(TR)^2}{(I-1)(J-1)} + \frac{MS(TC)^2}{(I-1)(K-1)} + \frac{MS(TRC)^2}{(I-1)(J-1)(K-1)}}$$

It is corrected by Hillis,

$$ddf_{DBM} = \frac{\{MS(TR) + \max[MS(TC) - MS(TRC), 0]\}^2}{MS(TR)^2/[(I-1)(J-1)]}$$

where $MS(T)$, $MS(TR)$, $MS(TC)$ and $MS(TRC)$ are the pseudovalue means squares of modality, modality by reader, modality by case, modality by reader by case correspondingly, by definition,

$$MS(T) = \frac{KJ\sum_{i=1}^{I}(Y_{i\cdot} - Y_{\cdot\cdot\cdot})}{t-1}$$

$$MS(TR) = \frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij\cdot} - Y_{i\cdot\cdot} - Y_{\cdot j\cdot} + Y_{\cdot\cdot\cdot})}{(I-1)(J-1)}$$

$$MS(TC) = \frac{J\sum_{i=1}^{I}\sum_{k=1}^{K}(Y_{i\cdot k} - Y_{i\cdot\cdot} - Y_{\cdot\cdot k} + Y_{\cdot\cdot\cdot})}{(I-1)(K-1)}$$

$$MS(TRC) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - Y_{ij\cdot} - Y_{i\cdot k} - Y_{\cdot jk} + Y_{i\cdot\cdot} + Y_{\cdot j\cdot} + Y_{\cdot\cdot k} - Y_{\cdot\cdot\cdot})}{(I-1)(J-1)(K-1)}$$

where Y with a "$\cdot$" subscript means the averaged pseudovalue over the replaced subscript.

2.2. OR Method

The statistical model of OR method is defined by

$$\hat{\theta}_{ij} = \mu + \tau_i + R_j + (\tau R)_{ij} + \varepsilon_{ij} \tag{2}$$

where $\hat{\theta}_{ij}$ denotes the AUC estimate for modality i and reader j. Other terms except for ε_{ij} are defined in the same way as DBM method. ε_{ij} is not assumed to be independent here, since each case is read by each reader with each modality. The covariances of error terms are defined by

$$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \begin{cases} \mathrm{Cov1} & i \neq i', j = j' \text{ (different modality, same reader)} \\ \mathrm{Cov2} & i = i', j \neq j' \text{ (same modality, different reader)} \\ \mathrm{Cov3} & i \neq i', j \neq j' \text{ (different modality, different reader)} \end{cases} \tag{3}$$

It is suggested that the covariances should be ordered as following sequence

$$\text{Cov1} \geq \text{Cov2} \geq \text{Cov3} \geq 0$$

In the condition of the model 2 (readers are regarded as fixed), covariances between error terms are also the corresponding covariances of the AUC estimates.

The OR variance and covariance components can be estimated using Jackknife, Bootstrap or DeLong's method.

- Jackknife method calculates AUCs for each modality-reader combination with one case deleted from the dataset. Let $\hat{\theta}_{ij\cdot}$ denote the the mean of $\hat{\theta}_{ijk}$ that is defined as same as in DBM model. The jackknife covariance estimate of the covariances between AUCs for readers j, j' and modality i, i' is given by

$$\text{Cov}_{jack}(\hat{\theta}_{ij}, \hat{\theta}_{i'j'}) = \frac{K-1}{K} \sum_{k=1}^{K} [(\hat{\theta}_{ijk} - \hat{\theta}_{ij\cdot})(\hat{\theta}_{i'j'k} - \hat{\theta}_{i'j'\cdot})]$$

- Let K_1 and K_2 denote the number of non-diseased and diseased cases respectively. Sample B groups of K cases indices with replacement from K_1 non-diseased cases and K_2 diseased cases. Let $\hat{\theta}_{ijb}$ be the AUC estimate for modality i and reader j calculated using the bth bootstrap of sample. The bootstrap covariance estimate of the covariances between AUCs for readers j, j' and modality i, i' is given by

$$\text{Cov}_{boot}(\hat{\theta}_{ij}, \hat{\theta}_{i'j'}) = \frac{1}{B-1} \sum_{b=1}^{B} [(\hat{\theta}_{ijb} - \hat{\theta}_{ij\cdot})(\hat{\theta}_{i'j'b} - \hat{\theta}_{i'j'\cdot})]$$

  where $\hat{\theta}_{ij\cdot}$ and $\hat{\theta}_{i'j'\cdot}$ denote the means of AUCs using all bootstrap samples for corresponding reader and modality.

- DeLong derived the variance and covariance for trapezoidal AUC estimates. This method can be only used to trapezoidal AUC.

The OR covariances components Cov1, Cov2 and Cov3 can be calculated by averaging the covariances following the condition in equation 3

The F statistic for testing the null hypothesis of no modality effects is given by

$$F_{OR} = \frac{MS(T)}{MS(TR) + \max[J(\widehat{\text{Cov}2} - \widehat{\text{Cov}3}), 0]}$$

where $MS(T) = [J/(I-1)] \sum_{i=1}^{I} (\hat{\theta}_{i\cdot} - \hat{\theta}_{\cdot\cdot})$, $MS(TR) = [1/((I-1)(J-1))] \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{\theta}_{ij} - \hat{\theta}_{i\cdot} - \hat{\theta}_{\cdot j} + \hat{\theta}_{\cdot\cdot})$, $\widehat{\text{Cov}2}$ and $\widehat{\text{Cov}3}$ denote the estimates of Cov2 and Cov3. $\hat{\theta}$ with a "$\cdot$" subscript means the averaged value over the replaced subscript. F_{OR} approximately follows a F distribution with degrees of freedom $I-1$ and $(I-1)(J-1)$. The denominator degrees of freedom ddf_{OR} is corrected by Hillis, where

$$ddf_{OR} = \frac{\left\{MS(TR) + \max[J(\widehat{\text{Cov}2} - \widehat{\text{Cov}3}), 0]\right\}^2}{MS(TR)^2/[(I-1)(J-1)]}$$

p-value can be calculated with the F value and degrees of freedom. Compare the p-value with the significance level α, we can determine if the null hypothesis should be rejected.

2.3. Sample Size Calculation

If a not significant result is obtained (i.e., $p > 0.05$) from a study, then the investigator may wish to estimate, from the observed variability of the data, how many cases and readers would be needed to achieve a given probability (typically chosen to be 80%) of detecting a true difference between the modalities in a subsequent study. This is termed sample size calculation. Method for statistical power calculation for ROC analysis is described in Ref. Briefly, we can calculate the statistical power with OR components via following steps. All OR components used are defined in the subsection 2.2.

1. *Specify the effect size.* Let d denote the effect size that is the absolute value of the difference of two AUCs. The effect size is defined by $d = |AUC_1 - AUC_2|$.

2. *Estimate OR parameter.* The variance of modality by reader interaction can be estimated by
$$\hat{\sigma}^2_{TR} = MS(TR) - \hat{\sigma}^2_{\varepsilon} + \widehat{\mathrm{Cov}}1 + \max\left(\widehat{\mathrm{Cov}}2 - \widehat{\mathrm{Cov}}3, 0\right)$$

3. *Estimate the noncentrality parameter and the degrees of freedom.* Let K^* denote the number of cases in current analysis, and let J and K be the numbers of readers and cases, which will be used to calculate the statistical power. The noncentrality parameter $\hat{\Delta}$ and the degrees of freedom of F_{OR}'s denominator $\widehat{ddf}_{OR}$ is estimated by
$$\hat{\Delta} = \frac{\frac{J}{2}d^2}{\hat{\sigma}^2_{TR} + \left(\frac{K^*}{K}\right)\left[\hat{\sigma}^2_{\varepsilon} - \widehat{\mathrm{Cov}}1 + (J-1)\max\left(\widehat{\mathrm{Cov}}2 - \widehat{\mathrm{Cov}}3, 0\right)\right]}$$
and
$$\widehat{ddf}_{OR} = \frac{\left\{\hat{\sigma}^2_{TR} + \left(\frac{K^*}{K}\right)\left[\hat{\sigma}^2_{\varepsilon} - \widehat{\mathrm{Cov}}1 + (J-1)\max\left(\widehat{\mathrm{Cov}}2 - \widehat{\mathrm{Cov}}3, 0\right)\right]\right\}^2}{\frac{\left\{\hat{\sigma}^2_{TR} + \left(\frac{K^*}{K}\right)\left[\hat{\sigma}^2_{\varepsilon} - \widehat{\mathrm{Cov}}1 - \max\left(\widehat{\mathrm{Cov}}2 - \widehat{\mathrm{Cov}}3, 0\right)\right]\right\}^2}{J-1}}$$

4. *Calculate the statistical power using the noncentrality parameter and degrees of freedom estimation.* The statistical power with significance level α can be calculated by
$$\text{Power} = \Pr\left(F_{1,\widehat{ddf}_{OR};\hat{\Delta}} > F_{1-\alpha;1,\widehat{ddf}_{OR}}\right)$$

To estimate the sample size for ROC analysis, we usually select a fixed number of readers(or cases) and calculate the statistical power by incrementing the number of cases(readers, vice versa) until the calculation result reaches the desired power. The number of readers and cases combination is a required sample size.

3. Examples and Results Comparisons

In this section, some examples for the usage of the package **JAFROCwR** are shown, and the analysis results of **JAFROCwR** and other softwares mentioned in Section 1 are provided and compared. The example dataset comes from Carolyn Van Dyke, MD. There are 45 diseased cases and 69 non-diseased cases in the dataset. Five radiologists (reader 0 to 4) read these images using two modalities (modality 0 and 1) and give their ratings for each case.

3.1. Analyze the Example Dataset Using JAFROCwR

The first steps are loading the package and read the data file.

```
> library(JAFROCwR)
> dataVanDyke <- readJAFROC("VanDyke.xlsx")
```

Then we analyze the data using both DBM and OR method.

```
> resultOR <- ORAnalysis(data = dataVanDyke, analysisFOM = "ROC", alpha = 0.05,
+                        covEstMethod = "Jackknife")
```

The result can be found in table

Following is an example of the plotting feature of this package. Figure 1 is the empirical ROC curves of each individual reader and their average performance using modality 0.

```
> rocCurveM1 <- plotROC(data = dataVanDyke, plottingModalities = 1,
+                       plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
+                       legendPosition = "bottom")
> rocCurveM1$ROCPlot
```

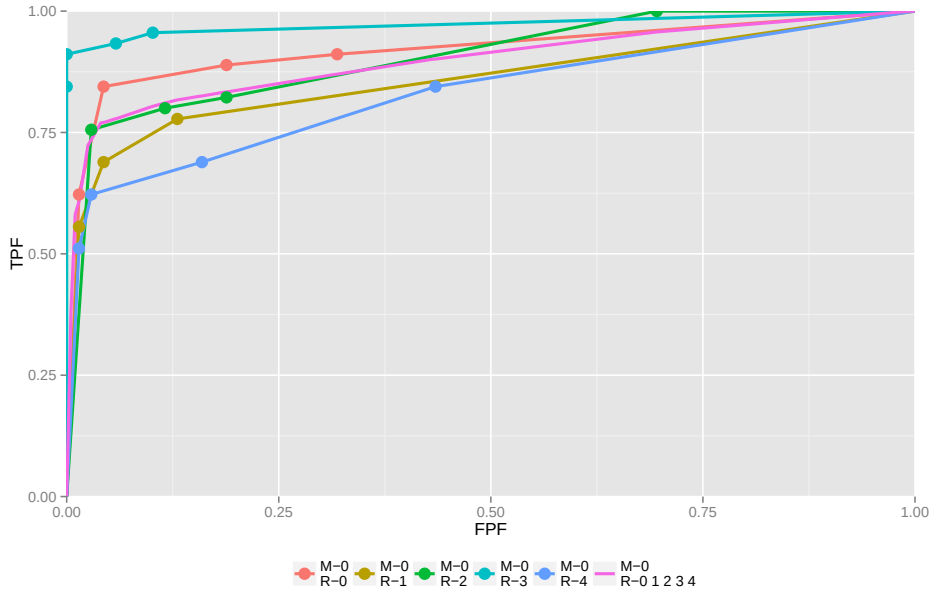


Figure 1: Plots of empirical ROC curves of each reader and their average performance using modality 0

The function for sample size calculation with effect size = 0.05 and desired power = 0.8 are as follows. Result can be found in table

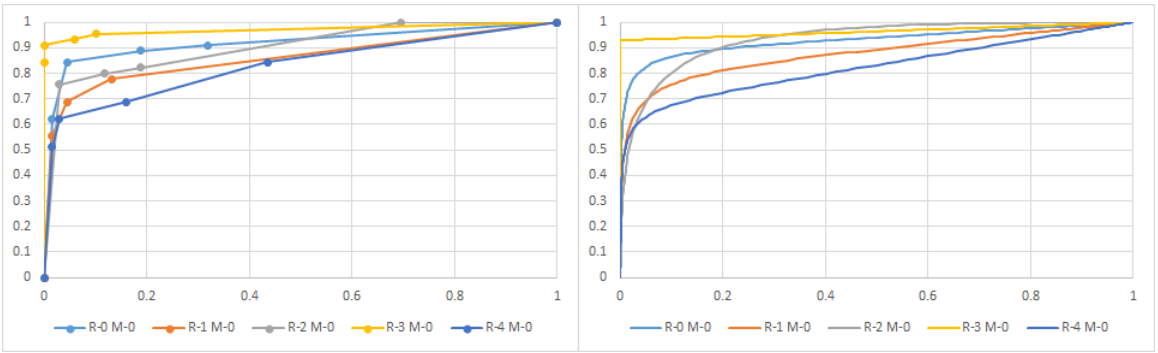
```
> calculateSampleSizeForData(data = dataVanDyke, alpha = 0.05,
+                            Effect_Size = 0.05, Desired_Power = 0.8)
```

3.2. Results of Other Softwares

Analysis results of the same dataset used in subsection 3.1 by other common ROC softwares are shown in this subsection. AUC is selected as the figure of merit. DBM method or OR method using jackknife estimation is used for the analysis. **JAFROCwR** includes functions that save JAFROC data using formats of other softwares. Users can use these functions to convert data between different formats.

OR-DBM MRMC

OR-DBM MRMC performs similar analysis as **JAFROCwR**. The example of how to use the software can be found in the user manual. Results is compared in table . Figure 2 show the empirical and fitted ROC curves for all readers using modalities 0.



(a) Empirical ROC curves

(b) CBM fitted ROC curves

(c) PROPROC fitted ROC curves

(d) RSCORE fitted ROC curves

Figure 2: OR-DBM MRMC ROC Curves Plots

iMRMC

iMRMC gives the test result of the difference between two modalities. Since we have only two modalities in the example datasets, we just choose them as the two modalities to be

tested. Testing results are reorganized and compared with other softwares in table . Variance components used in statistical models other than DBM and OR are provided and shown below.

Table 3: BDG variance components

	M1	M2	M3	M4
comp M0	8.6671E-01	8.4536E-01	8.1617E-01	8.0714E-01
coeff M0	6.4412E-05	4.3800E-03	2.8341E-03	1.9272E-01
comp M1	9.2320E-01	9.0951E-01	8.8986E-01	8.8622E-01
coeff M1	6.4412E-05	4.3800E-03	2.8341E-03	1.9272E-01
comp product	8.6164E-01	8.5978E-01	8.4613E-01	8.4528E-01
- coeff product	1.2882E-04	8.7601E-03	5.6683E-03	3.8544E-01
total	4.2921E-06	1.5470E-04	3.9001E-05	5.3964E-04
	M5	M6	M7	M8
comp M0	8.3454E-01	8.2855E-01	8.0738E-01	8.0468E-01
coeff M0	2.5765E-04	1.7520E-02	1.1337E-02	-2.2911E-01
comp M1	9.0132E-01	8.9642E-01	8.8668E-01	8.8517E-01
coeff M1	2.5765E-04	1.7520E-02	1.1337E-02	-2.2911E-01
comp product	8.5626E-01	8.5488E-01	8.4459E-01	8.4397E-01
- coeff product	5.1530E-04	3.5040E-02	2.2673E-02	-4.5823E-01
total	6.0143E-06	2.6665E-04	5.5381E-05	-4.3955E-04

Table 4: BCK variance components

	N	D	N ˜D	R
comp M0	2.7003E-03	2.3874E-02	3.2916E-03	2.4661E-03
coeff M0	1.4493E-02	2.2222E-02	3.2206E-04	2.0000E-01
comp M1	1.5063E-03	1.1249E-02	3.3939E-03	1.0437E-03
coeff M1	1.4493E-02	2.2222E-02	3.2206E-04	2.0000E-01
comp product	6.1991E-04	1.0910E-02	7.6453E-04	1.3141E-03
- coeff product	2.8986E-02	4.4444E-02	6.4412E-04	4.0000E-01
total	4.2996E-05	2.9559E-04	1.6607E-06	1.7633E-04
	N ˜R	D ˜R	R ˜N ˜D	
comp M0	6.3239E-03	1.4348E-02	9.0360E-03	
coeff M0	2.8986E-03	4.4444E-03	6.4412E-05	
comp M1	2.1399E-03	1.2039E-02	6.6583E-03	
coeff M1	2.8986E-03	4.4444E-03	6.4412E-05	
comp product	2.3462E-04	3.5850E-03	2.4796E-04	
- coeff product	5.7971E-03	8.8889E-03	1.2882E-04	
total	2.3172E-05	8.5409E-05	9.7896E-07	

Empirical ROC curve can be plotted using iMRMC. Figure 3 is the plot containing individual and averaged ROC curves for all readers using modality 0.

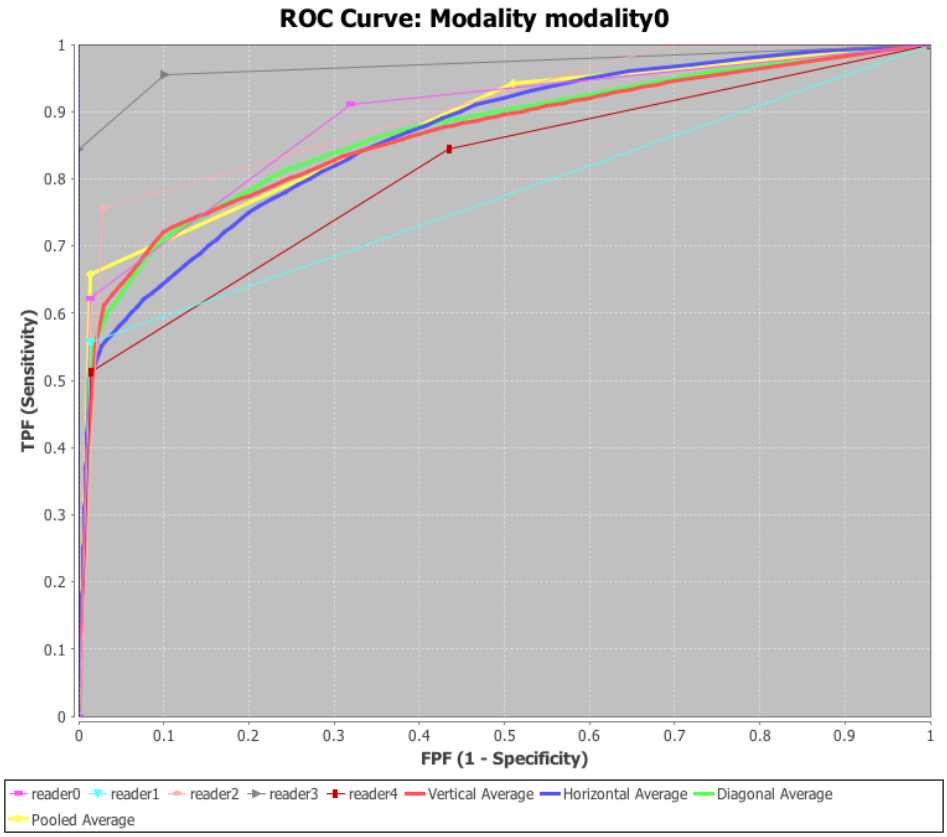


Figure 3: Individual and averaged ROC curves output by iMRMC

3.3. Results Comparison

In this section, we use some tables to show the results of the softwares. JAFROC is omitted since the package **JAFROCwR** produces identical results with it. Another dataset "franken" is also analyzed. Table 5

4. Additional Features of JAFROCwR

The main limitation of ROC analysis is that only one rating is used to describe each image. The radiologists may recognize a different region from where the true lesion is located, but a reasonable rating will still be given. The effect caused by this kind of errors is not revealed in the ROC analysis. The free-response receiver operating characteristic (FROC) analysis takes marks localization into account. In the free-response paradigm, the radiologist is free to mark and rate more than one suspicious regions that are considered. In this case, the number of

Table 5: AUC calculation comparison for the dataset "Van Dyke"

	OR-DBM MRMC		iMRMC		JAFROCwR	
	M - 0	M - 1	M - 0	M - 1	M - 0	M - 1
R - 0	0.9196	0.9478	0.9196	0.9478	0.9196	0.9478
R - 1	0.8588	0.9053	0.8588	0.9053	0.8588	0.9053
R - 2	0.9039	0.9217	0.9039	0.9217	0.9039	0.9217
R - 3	0.9731	0.9994	0.9731	0.9994	0.9731	0.9994
R - 4	0.8298	0.9300	0.8298	0.9300	0.8298	0.9300

marks is also regarded as a random variable. This source of randomness is the main difficulty of FROC analysis. Some figures of merit has been defined so DBM and OR method can be used for FROC analysis.

4.1. ROI and FROC Figures of Merit Used in JAFROCwR

Notation

For ROI and FROC study, lesion localization is taken into consideration. Hence more than one marks are allowed for each case. Let k_t denote the index of cases where t indicates the disease-status at the case (or patient) level, with $t = 1$ for non-diseased cases and $t = 2$ for diseased cases; k_1 ranges from 1 to K_1 for non-diseased cases and k_2 ranges from 1 to K_2 for diseased cases. Marks are indexed by l_s where s indicates the truth at the location level, with $s = 1$ for a non-lesion localization and $s = 2$ for a lesion localization; $l_1 = 1, 2, \ldots$, indexes marks of type $s = 1$ and $l_2 = 1, 2, \ldots, n_{k_2}$, indexes marks of type $s = 2$, where N_{k_2} is the number of lesions visible in image k_2. $Z_{k_t l_s}$ denotes the rating of mark l_s on case k_t. $Z_{k_2 l_2}$ is assigned $-\infty$ if the corresponding lesion localization is unmarked.

ROI Figure of Merit

In the region of interest (ROI) paradigm the investigator segments the image into a number of regions of interest (ROIs) and the radiologist rates each ROI for presence of at least one lesion somewhere within the ROI. The data consisted of a rating for each ROI. The ROI figure of merit θ_{ij}^{ROI} is defined by

$$\theta_{ROI} = \frac{1}{K_1^{ROI} K_2^{ROI}} \sum_{k_1=1}^{K_1^{ROI}} \sum_{k_2=1}^{K_2^{ROI}} \psi(Z_{k_1}, Z_{k_2})$$

where ψ is the function defined by equation 1, K_1^{ROI} is the number of non-diseased ROIs and K_2^{ROI} is the number of diseased ROIs, $Z_{k_1}(1 \leq k_1 \leq K_1^{ROI})$ and $Z_{k_2}(1 \leq k_2 \leq K_2^{ROI})$ are the ratings for corresponding ROIs.

FROC Figures of Merit

The weighted JAFROC figure of merit is defined by (the superscript c denotes case-based)

$$\theta^c_{JAFROC} = \frac{1}{K_1 K_2} \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \sum_{l_2=1}^{N_{k_2}} W_{k_2 l_2} \psi \left(\max_{l_1} \left(Z_{k_1 l_1} \right), Z_{k_2 l_2} \right)$$

$$\sum_{l_2=1}^{N_{k_2}} W_{k_2 l_2} = 1$$

where ψ is the function defined by equation 1, $\max_{l_s} \left(Z_{k_t l_s} \right)$ is the maximum over the ratings of all s localizations on the case k_t and $W_{k_2 l_2}$ is the weight of lesion l_2 in diseased case k_2.

The un-weighted JAFROC figure of merit is defined by (the superscript l denotes lesion-based)

$$\theta^l_{JAFROC} = \frac{1}{K_1 N_l} \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \sum_{l_2=1}^{N_{k_2}} \psi \left(\max_{l_1} \left(Z_{k_1 l_1} \right), Z_{k_2 l_2} \right)$$

where N_l is the total number of lesions in all diseased cases by definition

$$N_l = \sum_{k_2=1}^{K_2} N_{k_2}$$

The weighted JAFROC1 figure of merit is defined by

$$\theta^c_{JAFROC1} = \frac{1}{K_2 K_2} \sum_{k_2'=1}^{K_2} \sum_{k_2=1}^{K_2} \sum_{l_2=1}^{N_{k_2}} W_{k_2 l_2} \psi \left(\max_{l_1} \left(Z_{k_2' l_1} \right), Z_{k_2 l_2} \right)$$

The un-weighted JAFROC1 figure of merit is defined by

$$\theta^l_{JAFROC1} = \frac{1}{K_2 N_l} \sum_{k_2'=1}^{K_2} \sum_{k_2=1}^{K_2} \sum_{l_2=1}^{N_{k_2}} \psi \left(\max_{l_1} \left(Z_{k_2' l_1} \right), Z_{k_2 l_2} \right)$$

The inferred-ROC figure of merit is defined by (the superscript i denotes inferred)

$$\theta^i_{ROC} = \frac{1}{K_1 K_2} \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \psi \left(\max_{l_1} \left(Z_{k_1 l_1} \right), \max_{l_1 l_2} \left(Z_{k_2 l_2}, Z_{k_2 l_1} \right) \right)$$

The maximum lesion localization fraction figure of merit is defined by

$$\theta_{MaxLLF} = \frac{\sum_{k_2=1}^{K_2} \sum_{l_2=1}^{N_{k_2}} \phi \left(Z_{k_2 l_2} \right)}{N_l}$$

where ϕ is a function given by

$$\phi(Z) = \begin{cases} 0 & Z \text{ is } -\infty \\ 1 & \text{otherwise} \end{cases}$$

The maximum non-lesion localization fraction figure of merit is defined by

$$\theta_{MaxNLF} = \frac{\sum_{k_1=1}^{K_1} N_{k_1}}{K_1}$$

where N_{k_1} denotes the number of non-lesion localization marks in case k_1.

The maximum non-lesion localization fraction for all cases figure of merit is defined by

$$\theta_{MaxNLF}^{all} = \frac{\sum_{k=1}^{K} N_k}{K}$$

where N_k denotes the number of non-lesion localization marks in case k.

The exponential transformed specificity figure of merit is defined by

$$\theta_{ExpSP} = \exp(-\theta_{MaxNLF}) = \exp\left(-\frac{\sum_{k_1=1}^{K_1} N_{k_1}}{K_1}\right)$$

The Song figures of merit A0, A1 and A2 are also used in **JAFROCwR**. They are defined in Ref.

4.2. ROC/AFROC/FROC Plots

Empirical and parametric ROC/AFROC/FROC curves can be plotted using functions in **JAFROCwR**. These curves are defined as follows.

Empirical Plots

The ROC curve is a plot of true positive fraction (TPF) vs. false positive fraction (FPF), where TPF = number of positive decisions / number of diseased cases and FPF = number of positive decisions / number of non-diseased cases. For FROC data, TPF and FPF are calculated using the inferred ROC data by highest rating assumption. The AFROC curve is a plot of lesion localization fraction (LLF) vs. FPF, where LLF = number of lesions localizations / the total number of lesions. The FROC curve is a plot of LLF vs. non-lesion localization fraction (NLF), where NLF = number of non-lesion localizations / the number of cases.

Parametric Plots

Parametric ROC/AFROC/FROC curves are parametric predicted (fitted) curves using search model, which are defined in subsection 4.3. The steps that fit the curves are also given in that part.

4.3. Search-Model

The search-model is used to estimate sample size and fit ROC and (A)FROC curves for FROC data in **JAFROCwR**. According to Kundel and Nodine, image interpretation involves a search stage and a decision-making stage. During the search stage the radiologist quickly identifies suspicious regions. The search-stage is characterized by a random number of suspicious regions. During the decision making stage the observer calculates a rating at each suspicious region, and if it exceeds the minimum reporting threshold the region is marked.

Suspicious regions are termed *noise-sites* or *signal-sites* for benign tumors or lesions, respectively. The number of noise-sites on an image is assumed to be sampled from a Poisson distribution with mean λ. The number of signal-sites on a diseased image is assumed to be sampled from a binomial distribution with success probability ν and the total number of lesions on the diseased case k_2. The ratings form noise-site is assumed sampled from a normal distribution $N(0,1)$ and that from signal-site is assumed sampled from $N(\mu,1)$.

Search-model predicted ROC curves

The ROC curve of FROC data can be inferred using the highest rating assumption. The ROC curve predicted by the search-model is defined by

$$\begin{gathered} FPF(\zeta) = 1 - \exp\left(-\tfrac{\lambda}{2} + \tfrac{\lambda}{2}\text{erf}\left(\tfrac{\zeta}{\sqrt{2}}\right)\right) \\ TPF(\zeta) = \sum_{L=1}^{L^{MAX}} f_L \left[1 - \left(1 - \tfrac{\nu}{2} + \tfrac{\nu}{2}\text{erf}\left(\tfrac{\zeta-\mu}{\sqrt{2}}\right)\right)^L \exp\left(-\tfrac{\lambda}{2} + \tfrac{\lambda}{2}\text{erf}\left(\tfrac{\zeta}{\sqrt{2}}\right)\right)\right] \\ \sum_{L=1}^{L^{MAX}} f_L = 1 \end{gathered} \tag{4}$$

where ζ is the cutoff parameter determining an operating point on the ROC curve, L is the number of lesions in a diseased case, f_L is the fraction of diseased cases with L lesions, and $\text{erf}(x)$ is the error function. $TPF(\zeta)$ is a weighted average of true positive fraction for cases with $L = 1, 2, \ldots, L^{MAX}$ lesions.

Search-model predicted AFROC curves

The expression for the FPF predicted by the search-model has already been given in equation 4. That for the lesion localization fraction (LLF) is

$$LLF(\zeta) = \sum_{L=1}^{L^{MAX}} f_L \left[\nu\left(1 - \Phi(\zeta - \mu)\right)\right] = \nu\left(1 - \Phi(\zeta - \mu)\right) \tag{5}$$

where $\Phi(\zeta)$ is the cumulative distribution function (CDF) of standard normal distribution.

Search-model predicted FROC curves

The expression for the LLF predicted by the search-model has already been given in equation 5. That for the non-lesion localization fraction (NLF) is

$$NLF(\zeta) = \lambda\left(1 - \Phi(\zeta)\right) \tag{6}$$

ROC Likelihood function

Let (F_b, T_b) denote the number of false positive and true positives, respectively, in ratings bin b defined by neighboring cutoffs (ζ_b, ζ_{b+1}), where $b = 0, 1, \ldots, R$ where R is the number of bins, and $\zeta_0 = -\infty$ and $\zeta_{R+1} = +\infty$. For example, F_0 and T_0 represent the number of non-diseased and diseased cases with no marks respectively. The contribution of the ROC likelihood function from bin b is

$$\mathcal{L}_b^{ROC} = \left[FPF(\zeta_b) - FPF(\zeta_{b+1})\right]^{F_b}\left[TPF(\zeta_b) - TPF(\zeta_{b+1})\right]^{T_b} \tag{7}$$

The net likelihood $\mathcal{L}^{ROC}$ is the product of the ROC likelihood function from all $R+1$ bins,

$$\mathcal{L}^{ROC} = \prod_{b=0}^{R} \mathcal{L}_b^{ROC}$$

To fit the ROC curve, we need to estimate parameters λ, ν, $\overrightarrow{\zeta}$ and μ that maximize the logarithm of the likelihood 7. Following algorithm is used in **JAFROCwR**

1. For given λ, ν and cutoffs $\overrightarrow{\zeta}$, determine μ by minimizing the Chi-square goodness of fit statistic
2. Calculate the log-likelihood of $\mathcal{L}^{ROC}$
3. Repeat preceding steps using varied λ, ν and $\overrightarrow{\zeta}$ until reach the maximum of $\mathcal{L}^{ROC}$.

Search-model fitted ROC curve for (F)ROC data can be plotted using model 4 with optimized parameters.

AFROC Likelihood function

Let (F_b, L_b) denote the number of false positive and lesion localizations, respectively, in bin b between neighboring cutoffs (ζ_b, ζ_{b+1}). The contribution of the AFROC likelihood function from bin b is

$$\mathcal{L}_b^{AFROC} = [FPF(\zeta_b) - FPF(\zeta_{b+1})]^{F_b}[LLF(\zeta_b) - LLF(\zeta_{b+1})]^{L_b} \tag{8}$$

FROC Likelihood function

Let (N_b, L_b) denote the number of non-lesion localizations and lesion localizations, respectively, in bin b between neighboring cutoffs (ζ_b, ζ_{b+1}). The contribution of the FROC likelihood function from bin b is

$$\mathcal{L}_b^{FROC} = [NLF(\zeta_b) - NLF(\zeta_{b+1})]^{N_b}[LLF(\zeta_b) - LLF(\zeta_{b+1})]^{L_b} \tag{9}$$

Search-model fitted AFROC and FROC curves can be plotted using model 8 and 9 with optimized parameters.

4.4. FROC Example

ROI and FROC Analysis

ROI and FROC data file are able to be analyzed by **JAFROCwR** using figures of merit that were mentioned in subsection 4.1. The usage of functions for ROI/FROC data analysis are same as that for ROC analysis. For FROC data, empirical and search model parametric AFROC/FROC curves are available. Figure 4 and 5 are the empirical and parametric AFROC/FROC curves.

```
> frocData <- readJAFROC("frocDataFile.xls")
> afrocCurve <- plotAFROC(data = frocData, plottingModalities = 1,
+                         plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
```

```
+                          legendPosition = "bottom")
> fittedAfrocCurve <- plotFittedAFROCCurve(data = frocData, plottingModalities = 1,
+                                          plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
+                                          legendPosition = "bottom")
> frocCurve <- plotFROC(data = frocData, plottingModalities = 1,
+                       plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
+                       legendPosition = "bottom")
> fittedFrocCurve <- plotFittedFROCCurve(data = frocData, plottingModalities = 1,
+                                        plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
+                                        legendPosition = "bottom")
> multiplot(afrocCurve$AFROCPlot, fittedAfrocCurve$AFROCPlot, cols = 2)
```

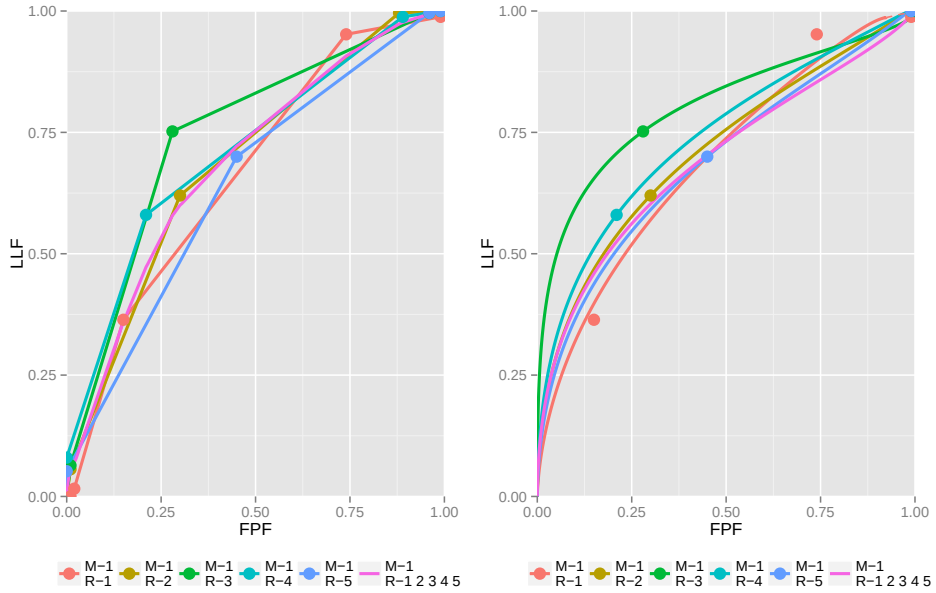


Figure 4: Plots of empirical (left) and fitted (right) AFROC curves of each reader and their average performance using modality 0

```
> frocCurve <- plotFROC(data = frocData, plottingModalities = 1,
+                       plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
+                       legendPosition = "bottom")
> fittedFrocCurve <- plotFittedFROCCurve(data = frocData, plottingModalities = 1,
+                                        plottingReaders = list(1, 2, 3, 4, 5, c(1:5)),
+                                        legendPosition = "bottom")
> multiplot(frocCurve$FROCPlot, fittedFrocCurve$FROCPlot, cols = 2)
```

Sample Size Estimation for FROC Data

For FROC data, sample size estimation can be performed using similar methods. Details are described in Ref. In short, estimate the search-model parameters of pilot study, then use them to calculate the area under ROC curve by numerical integration of the predicted ROC curve (equation 4). Add the desired ROC effect size, and new values of parameters can be determined. Calculate effect size in JAFROC unit with these new parameters. Finally,

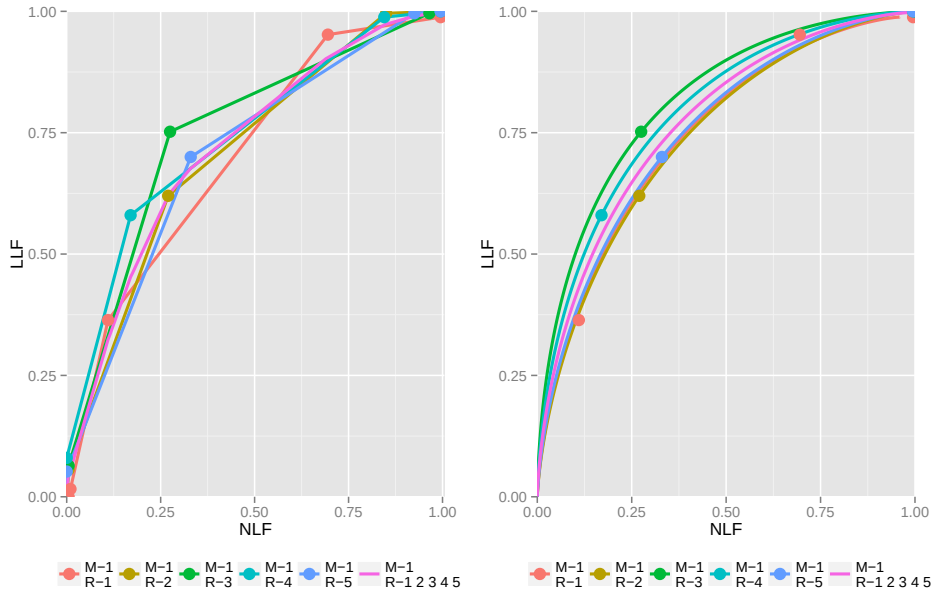


Figure 5: Plots of empirical (left) and fitted (right) FROC curves of each reader and their average performance using modality 0

required sample size to achieve the JAFROC unit effect size and statistical power can be estimated using same steps for ROC sample size estimation.

Operating Characteristics

Operating characteristics function provides the ability to input specific values for the search model parameters to visualize the effects on the predicted ROC, AFROC and FROC curves. It is a convenient tool to understand how changes in reader characteristics (parameters are defined in subsection 4.3) affect reader performance. The function accommodates one or more modalities and varying number of lesions per case and calculates ROC and AFROC areas for all modalities. Figure 6 shows example plots for $\mu = 2$, $\lambda = 1$, $\nu = 0.6$, 15 single-lesion-cases and 35 two-lesion-cases in the NH modality and $\mu = 3$, $\lambda = 0.5$, $\nu = 0.9$, 20 single-lesion-cases and 30 two-lesion-cases in the AH modality.

```
> lesionNumList <- list(rbind(c(1, 15), c(2, 35)),
+                       rbind(c(1, 20), c(2, 30)))
> oprtChrctResults <- operatingCharacteristics(mu = c(2, 3), lambda = c(1, 0.5),
+                                              nu = c(0.6, 0.9),
+                                              lesionNumTable = lesionNumList,
+                                              legendPosition = "bottom")
> oprtChrctResults$aucROC

[1] 0.8473118 0.9726777

> oprtChrctResults$aucJAFROC

[1] 0.6323901 0.9229876
```

```
> multiplot(oprtChrctResults$ROCPlot, oprtChrctResults$FROCPlot,
+           oprtChrctResults$AFROCPlot, oprtChrctResults$PDFPlot, cols = 2)
```

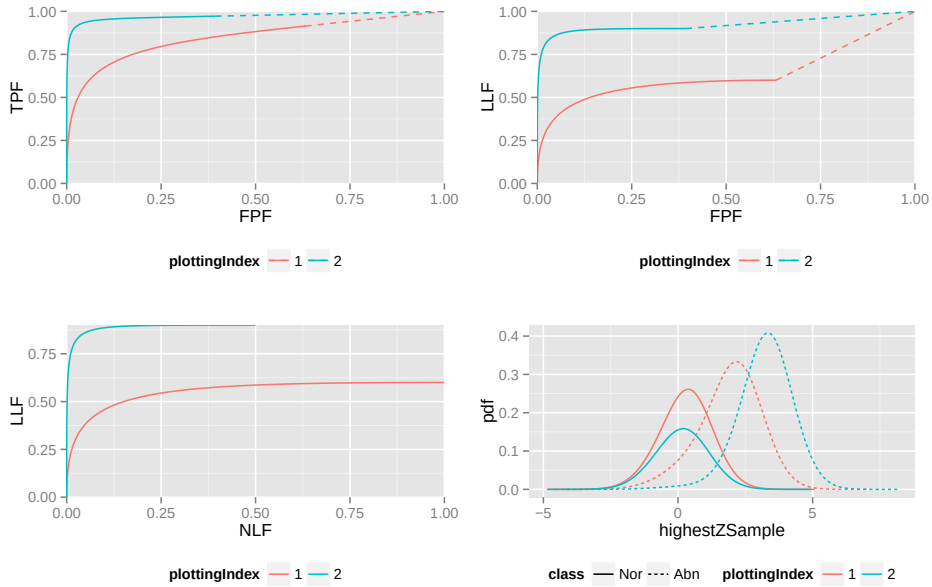


Figure 6: Operating Characteristics curves for $\mu = 2$, $\lambda = 1$, $\nu = 0.6$ (red) and $\mu = 3$, $\lambda = 0.5$, $\nu = 0.9$ (green)

5. Discussion

Compared with iMRMC and OR-DBM MRMC, advantages and limitations of **JAFROCwR** are summed up in following parts.

5.1. Advantages of JAFROCwR

In summary, **JAFROCwR** performs better than the others in following aspects.

1. *Programmabilty.* Both of OR-DBM MRMC and iMRMC are GUI-based softwares, so it is almost impossible to call them and catch the results from users' own program. **JAFROCwR** is distributed as a package of R. All functions in the package can be used in users' own R code. Considering the good interactivity between R and other programming languages, **JAFROCwR** can also be easily called from programming environments other than R including but not limited to C/C++, Java, Python and MATLAB. In additional, **JAFROCwR** is an open source package. Users can find the source code of the package online, and it is convenient to debug and modify the source code if the original version cannot meet users' specific requirements.

2. *Ability to Analyze FROC Data.* JAFROC and **JAFROCwR** are the only available software for free-response studies, and **JAFROCwR** implements OR method for FROC

data for the first time. Operating Characteristics is a useful tool to help users understand the effect of FROC parameters.

3. *Plotting Features* **JAFROCwR** provides empirical and search model fitted parametric ROC/AFROC/FROC curves. The R package **ggplot2** is used to plot all these curves. It gives fancy layouts, colors and legends, and **ggplot2** objects are also easy to be modified as the users requirements.

5.2. Limitations and Potential Future Updates

JAFROCwR has limitations need to be improved in the future. There are also some more features should be useful to be added in the package. General ideas are given as follows.

1. *Running Speed*. Since all the code of **JAFROCwR** are written using R and a lot of resampling operations (jackknife and bootstrap) are implemented by loops, the running time, especially for larger data file, is longer than that of the softwares written using "lower-level" languages such as C/C++. We are considering that convert some functions with many loop operations to C/C++ in the future. **Rcpp** will be used to combine them into the package. So the efficiency can be improved.

2. *Binormal Model*. Only search model is used to fit the curves in **JAFROCwR**. Binormal model estimation should also be included.

3. *Simulator*. Simulation plays an important role in ROC/FROC researches. It is always used to validate a new statistical method. Hence the simulator design is very significant. We are planning to provide a simulator in **JAFROCwR** to help users generate simulation data.

4. *Graphical User Interface*. GUI is necessary for some users, especially for the old users of JAFROC. An R based GUI package for **JAFROCwR** is in development. We will publish it when it is finished. JAFROC can be completely replaced at that time.

6. Conclusions

The **JAFROCwR** package is able to perform MRMC analysis of ROC/FROC/ROI data. The package provides OR and DBM methods for performing MRMC analysis, each of which can be used with different figures of merit. Empirical and parametric ROC/AFROC/FROC curves are also given by the package. Sample size calculation helps to estimate sample size for given effect size and statistical power in future study.

Search model is used to fit ROC/AFROC/FROC curves. Operating Characteristics shows the effect of search model parameters on the predicted ROC, AFROC and FROC curves. It is convenient to demonstrate how the parameters affect readers' performance.

Through the provided examples, **JAFROCwR** will be definitely helpful if you are using R as a main programming language. Future version of this package will attempt to reduce the running time, and provide more features.

Affiliation:

Xuetong Zhai
Department of Radiology
University of Pittsburgh
Pittsburgh, PA 15213
United States of America
E-mail: xuz19@pitt.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd