Communications for Statistical Applications and Methods
2015, Vol. 22, No. 4, 377–387

DOI: http://dx.doi.org/10.5351/CSAM.2015.22.4.377
Print ISSN 2287-7843 / Online ISSN 2383-4757

Test Statistics for Volume under the ROC Surface and Hypervolume under the ROC Manifold

Chong Sun Hong[1,a], Min Ho Cho[a]

[a]Department of Statistics, Sungkyunkwan University, Korea

Abstract

The area under the ROC curve can be represented by both Mann-Whitney and Wilcoxon rank sum statistics. Consider an ROC surface and manifold equal to three dimensions or more. This paper finds that the volume under the ROC surface (VUS) and the hypervolume under the ROC manifold (HUM) could be derived as functions of both conditional Mann-Whitney statistics and conditional Wilcoxon rank sum statistics. The nullhypothesis equal to three distribution functions or more are identical can be tested using VUS and HUM statistics based on the asymptotic large sample theory of Wilcoxon rank sum statistics. Illustrative examples with three and four random samples show that two approaches give the same VUS and HUM^4. The equivalence of several distribution functions is also tested with VUS and HUM^4 in terms of conditional Wilcoxon rank sum statistics.

Keywords: manifold, Mann-Whitney, nonparametric, ROC, surface, Wilcoxon

1. Introduction

The ROC curve for two dimensions is a popular method for biostatistics and credit evaluation study (Egan, 1975; Engelmann *et al.*, 2003; Fawcett, 2003; Hong, 2009; Hong *et al.*, 2010; Provost and Fawcett, 2001; Sobehart and Keenan, 2001; Swets, 1988; Swets *et al.*, 2000; Zou *et al.*, 2007). Let X_1 and X_2 be two random variables with their cumulative distribution functions $F_1(\cdot)$ and $F_2(\cdot)$, respectively. The area under the ROC curve (AUC) is defined as $P(X_1 \leq X_2)$. With an additional assumption $F_1(x) \geq F_2(x)$ for all x, the range of the AUC belongs to $[1/(2!), 1]$.

It is well known that the empirical AUC for sample data is represented by Mann-Whitney statistics (Bamber, 1975; Faraggi and Reiser, 2002; Hanley and McNeil, 1982; Mann and Whitney, 1947; Rosset, 2004) empirical AUC is also represented as Wilcoxon rank sum statistic since the Mann-Whitney statistic has a linear relationship with Wilcoxon rank sum statistic (Conover, 1980; Gibbons, 1971; Randles and Wolfe, 1979; Wilcoxon, 1945).

Next consider an ROC surface for three dimensions and an ROC manifold equal to four dimensions or more. The volume under the ROC surface (VUS) and hypervolume under the ROC manifold (HUM) are defined as $P(X_1 \leq X_2 \leq X_3)$ and $P(X_1 \leq X_2 \leq \cdots \leq X_k)$ (the HUM will be considered for only four dimensions in this paper). Hong and Cho (2015) showed that VUS and HUM are represented as functions of Mann-Whitney statistics by extension because AUC is derived as Mann-Whitney statistics. In this paper, we will propose that VUS and HUM could be represented as functions of Wilcoxon rank sum statistics as well because the Mann-Whitney statistic has a relationship with the Wilcoxon rank sum statistic.

[1] Corresponding author: Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea.
E-mail: cshong@skku.edu

Published 31 July 2015 / journal homepage: http://csam.or.kr

Joseph (2005) extended the method of Wilkie (2004) and proposed the standard criteria of the AUC for the probability of default based on Basel II under the assumption of homogeneous normal distribution functions. With similar arguments of Joseph (2005), Hong *et al.* (2013) and Hong and Jung (2014) suggested standard criteria for the VUS and HUM to discriminate three and four classification models, respectively. These works are only provided for 13 validation ranges based on VUS and HUM values. This paper proposes that the significance of the VUS and HUM could be determined using the Wilcoxon rank sum test method since VUS and HUM are functions of conditional Wilcoxon rank sum statistics.

In Section 2, the VUS and HUM are expressed by appropriate conditional probabilities derived as functions of conditional Mann-Whitney statistics and conditional Wilcoxon rank sum statistics as well. Therefore, VUS and HUM can also be derived as functions of conditional Wilcoxon rank sum statistics since VUS and HUM are represented as functions of conditional Mann-Whitney statistics. In Section 3, some statistical testing methods for the VUS and HUM are proposed based on the asymptotic large sample theory of conditional Wilcoxon rank sum statistics. We therefore suggest that null hypothesis equal to three distribution functions or more are identical and can tested with VUS and HUM values. Some illustrative examples for three and four random samples are provided in Section 4. The values of the VUS and HUM^4 using Mann-Whitney statistics are the same as those using Wilcoxon rank sum statistics. The significance of the VUS and HUM^4 can be tested with the asymptotic large sample theory of conditional Wilcoxon rank sum statistics with these examples. Section 5 provides the conclusion and future works.

2. Representation of VUS and HUM with Wilcoxon Rank Sum Statistics

For the discrete random variables X_1 and X_2, the AUC is derived as $P(X_1 \leq X_2) = P(X_1 < X_2) + P(X_1 = X_2)/2$. Suppose that $\{X_{1,1}, \ldots, X_{1,n_1}\}$ and $\{X_{2,1}, \ldots, X_{2,n_2}\}$ are two random samples of X_1 and X_2 with sizes n_1 and n_2, respectively. It is well known that the empirical AUC is obtained by using Mann-Whitney statistics such as $[U_{X_1<X_2} + U_{X_1=X_2}/2]/n_1n_2$, where $U_{X_1<X_2}$ and $U_{X_1=X_2}$ are defined as $\sum_{i,j} I(X_{1i} < X_{2j})$ and $\sum_{i,j} I(X_{1i} = X_{2j})$, respectively. With the relationship between Mann-Whitney and Wilcoxon rank sum statistics, the AUC is also obtained as $[\sum_j R_j^{X_2} - n_2(n_2 + 1)/2]/n_1n_2$, where $\sum_j R_j^{X_2}$ is denoted as Wilcoxon rank sum statistic of X_2 from the combined sample of X_1 and X_2.

Let us consider the VUS and HUM^4 for the ROC surface and manifold for three and four dimensions, respectively. Let $X_1, \ldots, X_4$ be four random variables with cumulative distribution functions $F_1(\cdot), \ldots, F_4(\cdot)$, respectively. With an assumption $F_1(x) \geq F_2(x) \geq F_3(x) \geq F_4(x)$ for all x, values of VUS and HUM^4 belong to $[1/(3!), 1]$ and $[1/(4!), 1]$. Hong and Cho (2015) expressed VUS with the following conditional probabilities.

$$\begin{aligned} P(X_1 \leq X_2 \leq X_3) = P(X_2 < X_3|X_1 < X_2)P(X_1 < X_2) + \frac{1}{2}P(X_2 = X_3|X_1 < X_2)P(X_1 < X_2) \\ + \frac{1}{2}P(X_2 < X_3|X_1 = X_2)P(X_1 = X_2) + \frac{1}{2^2}P(X_2 = X_3|X_1 = X_2)P(X_1 = X_2). \end{aligned} \tag{2.1}$$

Hong and Cho (2015) also showed that VUS can have a relationship with conditional Mann-Whitney statistics.

$$\mathrm{VUS}_{MW} = \frac{1}{n_1n_2n_3}\left[U_{X_2<X_3|X_1<X_2} + \frac{1}{2}U_{X_2=X_3|X_1<X_2} + \frac{1}{2}U_{X_2<X_3|X_1=X_2} + \frac{1}{2^2}U_{X_2=X_3|X_1=X_2}\right], \tag{2.2}$$

Table 1: Representation with Mann-Whitney or Wilcoxon rank sum statistics for VUS

$\left[P(X_2 < X_3 \vert X_1 < X_2) + \frac{1}{2}P(X_2 = X_3 \vert X_1 < X_2)\right]P(X_1 < X_2)$	$\frac{1}{n_1 n_2 n_3}\left[U_{X_2<X_3\vert X_1<X_2} + \frac{1}{2}U_{X_2=X_3\vert X_1<X_2}\right]$
	$\frac{1}{n_1 n_2 n_3}\left[\sum_k R_k^{X_3\vert X_1<X_2} - \frac{n_3(n_3+1)}{2}\right]$
$\left[P(X_2 < X_3 \vert X_1 = X_2) + \frac{1}{2}P(X_2 = X_3 \vert X_1 = X_2)\right]P(X_1 = X_2)$	$\frac{1}{n_1 n_2 n_3}\left[U_{X_2<X_3\vert X_1=X_2} + \frac{1}{2}U_{X_2=X_3\vert X_1=X_2}\right]$
	$\frac{1}{n_1 n_2 n_3}\left[\sum_k R_k^{X_3\vert X_1=X_2} - \frac{n_3(n_3+1)}{2}\right]$

VUS = volume under the ROC surface.

where conditional Mann-Whitney statistics are defined as $U_{X_2<X_3|X_1<X_2} = \sum_{j,k} I(X_{2j} < X_{3k}|X_{1i} < X_{2j})$, $U_{X_2=X_3|X_1<X_2} = \sum_{j,k} I(X_{2j} = X_{3k}|X_{1i} < X_{2j})$, $U_{X_2<X_3|X_1=X_2} = \sum_{j,k} I(X_{2j} < X_{3k}|X_{1i} = X_{2j})$, $U_{X_2=X_3|X_1=X_2} = \sum_{j,k} I(X_{2j} = X_{3k}|X_{1i} = X_{2j})$.

Note that if there is no tied sample of X_1 and X_2, then conditional Mann-Whitney statistics $U_{X_2<X_3|X_1=X_2}$ and $U_{X_2=X_3|X_1=X_2}$ have zero values. Mann-Whitney statistics can have a relationship with Wilcoxon rank sum statistics; therefore, VUS can have a relationship with the following conditional Wilcoxon rank sum statistics.

Theorem 1. *The VUS, $P(X_1 \le X_2 \le X_3)$, can also be represented as*

$$VUS_W = \frac{1}{n_1 n_2 n_3}\left\{\left[\sum_k R_k^{X_3|X_1<X_2} - \frac{n_3(n_3+1)}{2}\right]I(A) + \frac{1}{2}\left[\sum_k R_k^{X_3|X_1=X_2} - \frac{n_3(n_3+1)}{2}\right]I(B)\right\},$$

where $\sum_k R_k^{X_3|X_1<X_2}$ and $\sum_k R_k^{X_3|X_1=X_2}$ are conditional Wilcoxon rank sum statistics of X_3 from the combined sample of X_2 and X_3 given situations $X_1 < X_2$ and $X_1 = X_2$, respectively. The sets A and B mean that there exists at least one sample that satisfies the corresponding conditional states of $X_1 < X_2$ and $X_1 = X_2$, respectively.

Proof: Note that $U_{X_2<X_3|X_1<X_2} + U_{X_2=X_3|X_1<X_2}/2 = \sum_k R_k^{X_3|X_1<X_2} - n_3(n_3+1)/2$ and $U_{X_2<X_3|X_1=X_2} + U_{X_2=X_3|X_1=X_2}/2 = \sum_k R_k^{X_3|X_1=X_2} - n_3(n_3+1)/2$ implies that the VUS is represented as

$$\begin{aligned}&\frac{1}{n_1 n_2 n_3}\left\{\left[U_{X_2<X_3|X_1<X_2} + \frac{1}{2}U_{X_2=X_3|X_1<X_2}\right] + \frac{1}{2}\left[U_{X_2<X_3|X_1=X_2} + \frac{1}{2}U_{X_2=X_3|X_1=X_2}\right]\right\}\\ &= \frac{1}{n_1 n_2 n_3}\left\{\left[\sum_k R_k^{X_3|X_1<X_2} - \frac{n_3(n_3+1)}{2}\right] + \frac{1}{2}\left[\sum_k R_k^{X_3|X_1=X_2} - \frac{n_3(n_3+1)}{2}\right]\right\}.\end{aligned}$$

If there is no tied sample of X_1 and X_2, then the statistic $\sum_k R_k^{X_3|X_1=X_2}$ has a zero value, so that $[\sum_k R_k^{X_3|X_1=X_2} - n_3(n_3+1)/2] = 0$. Hence we obtain the Theorem 1 with two indicator functions. Note that we may conclude that the two terms in (2.1) can be obtained by either Mann-Whitney statistics in (2.2) or Wilcoxon rank sum statistic in Theorem 1 (Table 1).

Hong and Cho (2015) expressed the HUM^4 for four dimensions with following conditional probabilities.

$$\begin{aligned}&P(X_1 \le X_2 \le X_3 \le X_4) \qquad (2.3)\\ &= P(X_3 < X_4|X_1 < X_2 < X_3)P(X_1 < X_2 < X_3) + \frac{1}{2}P(X_3 = X_4|X_1 < X_2 < X_3)P(X_1 < X_2 < X_3)\\ &\quad + \frac{1}{2}P(X_3 < X_4|X_1 = X_2 < X_3)P(X_1 = X_2 < X_3) + \frac{1}{2^2}P(X_3 = X_4|X_1 = X_2 < X_3)P(X_1 = X_2 < X_3)\end{aligned}$$

$$+\frac{1}{2}P(X_3 < X_4|X_1 < X_2 = X_3)P(X_1 < X_2 = X_3) + \frac{1}{2^2}P(X_3 = X_4|X_1 < X_2 = X_3)P(X_1 < X_2 = X_3)$$
$$+\frac{1}{2^2}P(X_3 < X_4|X_1 = X_2 = X_3)P(X_1 = X_2 = X_3) + \frac{1}{2^3}P(X_3 = X_4|X_1 = X_2 = X_3)P(X_1 = X_2 = X_3).$$

Similarly, Hong and Cho (2015) showed that HUM^4 can have a relationship with conditional Mann-Whitney statistics.

$$\mathrm{HUM}^4_{MW} = \frac{1}{n_1n_2n_3n_4}\left\{\left[U_{X_3<X_4|X_1<X_2<X_3} + \frac{1}{2}U_{X_3=X_4|X_1<X_2<X_3}\right] + \frac{1}{2}\left[U_{X_3<X_4|X_1=X_2<X_3} + \frac{1}{2}U_{X_3=X_4|X_1=X_2<X_3}\right]\right.$$
$$\left. + \frac{1}{2}\left[U_{X_3<X_4|X_1<X_2=X_3} + \frac{1}{2}U_{X_3=X_4|X_1<X_2=X_3}\right] + \frac{1}{2^2}\left[U_{X_3<X_4|X_1=X_2=X_3} + \frac{1}{2}U_{X_3=X_4|X_1=X_2=X_3}\right]\right\}, \quad (2.4)$$

where $U_{X_3<X_4|X_1<X_2<X_3}$ is defined as $\sum_{k,l} I(X_{3k} < X_{4l}|X_{1i} < X_{2j} < X_{3k})$, and other $U_{X_3=X_4|X_1<X_2<X_3}$, $U_{X_3<X_4|X_1=X_2<X_3}$, $U_{X_3=X_4|X_1=X_2<X_3}$, $U_{X_3<X_4|X_1<X_2=X_3}$, $U_{X_3=X_4|X_1<X_2=X_3}$, $U_{X_3<X_4|X_1=X_2=X_3}$, $U_{X_3=X_4|X_1=X_2=X_3}$ can easily be defined using similar arguments with. □

Note that the four kinds of conditional Mann-Whitney statistics, $U_{\cdot|X_1<X_2<X_3}$, $U_{\cdot|X_1=X_2<X_3}$, $U_{\cdot|X_1<X_2=X_3}$, and $U_{\cdot|X_1=X_2=X_3}$, have non-zero values, if there exists at least one sample that satisfy corresponding conditional states of X_1, X_2 and X_3 such as $X_1 < X_2 < X_3$, $X_1 = X_2 < X_3$, $X_1 < X_2 = X_3$ and $X_1 = X_2 = X_3$, respectively. HUM^4 can have a relationship with the following modified Wilcoxon rank sum statistics in Theorem 2.

Theorem 2. *The HUM^4, $P(X_1 \le X_2 \le X_3 \le X_4)$, can also be derived with conditional Wilcoxon rank sum statistics as follows*

$$HUM^4_W = \frac{1}{n_1n_2n_3n_4}\left\{\left[\sum_l R_l^{X_4|X_1<X_2<X_3} - \frac{n_4(n_4+1)}{2}\right]I(A) + \frac{1}{2}\left[\sum_l R_l^{X_4|X_1=X_2<X_3} - \frac{n_4(n_4+1)}{2}\right]I(B)\right.$$
$$\left. + \frac{1}{2}\left[\sum_l R_l^{X_4|X_1<X_2=X_3} - \frac{n_4(n_4+1)}{2}\right]I(C) + \frac{1}{2^2}\left[\sum_l R_l^{X_4|X_1=X_2=X_3} - \frac{n_4(n_4+1)}{2}\right]I(D)\right\},$$

where $\sum_l R_l^{X_4|X_1<X_2<X_3}$, $\sum_l R_l^{X_4|X_1=X_2<X_3}$, $\sum_l R_l^{X_4|X_1<X_2=X_3}$ and $\sum_l R_l^{X_4|X_1=X_2=X_3}$ are the conditional Wilcoxon rank sum statistics of X_4 from the combined sample of X_3 and X_4 given situations $X_1 < X_2 < X_3$, $X_1 = X_2 < X_3$, $X_1 < X_2 = X_3$ and $X_1 = X_2 = X_3$, respectively. And the sets A, B, C, D in the indicator functions in Theorem 4 mean that there exists at least one sample satisfying corresponding conditional states of $X_1 < X_2 < X_3$, $X_1 = X_2 < X_3$, $X_1 < X_2 = X_3$ and $X_1 = X_2 = X_3$, respectively.

Proof: Since the following conditional Mann-Whitney statistics,

$$\left[U_{X_3<X_4|X_1<X_2<X_3} + \frac{1}{2}U_{X_3=X_4|X_1<X_2<X_3}\right], \quad \left[U_{X_3<X_4|X_1=X_2<X_3} + \frac{1}{2}U_{X_3=X_4|X_1=X_2<X_3}\right],$$
$$\left[U_{X_3<X_4|X_1<X_2=X_3} + \frac{1}{2}U_{X_3=X_4|X_1<X_2=X_3}\right], \quad \text{and} \quad \left[U_{X_3<X_4|X_1=X_2=X_3} + \frac{1}{2}U_{X_3=X_4|X_1=X_2=X_3}\right],$$

Table 2: Representation with Mann-Whitney or Wilcoxon rank sum statistics for HUM[4]

$\left[P(X_3<X_4\|X_1<X_2<X_3)+\frac{1}{2}P(X_3=X_4\|X_1<X_2<X_3)\right]P(X_1<X_2<X_3)$	$\frac{1}{n_1n_2n_3n_4}\left[U_{X_3<X_4\|X_1<X_2<X_3}+\frac{1}{2}U_{X_3=X_4\|X_1<X_2<X_3}\right]$
	$\frac{1}{n_1n_2n_3n_4}\left[\sum_l R_l^{X_4\|X_1<X_2<X_3}-\frac{n_4(n_4+1)}{2}\right]$
$\left[P(X_3<X_4\|X_1=X_2<X_3)+\frac{1}{2}P(X_3=X_4\|X_1=X_2<X_3)\right]P(X_1=X_2<X_3)$	$\frac{1}{n_1n_2n_3n_4}\left[U_{X_3<X_4\|X_1=X_2<X_3}+\frac{1}{2}U_{X_3=X_4\|X_1=X_2<X_3}\right]$
	$\frac{1}{n_1n_2n_3n_4}\left[\sum_l R_l^{X_4\|X_1=X_2<X_3}-\frac{n_4(n_4+1)}{2}\right]$
$\left[P(X_3<X_4\|X_1<X_2=X_3)+\frac{1}{2}P(X_3=X_4\|X_1<X_2=X_3)\right]P(X_1<X_2=X_3)$	$\frac{1}{n_1n_2n_3n_4}\left[U_{X_3<X_4\|X_1<X_2=X_3}+\frac{1}{2}U_{X_3=X_4\|X_1<X_2=X_3}\right]$
	$\frac{1}{n_1n_2n_3n_4}\left[\sum_l R_l^{X_4\|X_1<X_2=X_3}-\frac{n_4(n_4+1)}{2}\right]$
$\left[P(X_3<X_4\|X_1=X_2=X_3)+\frac{1}{2}P(X_3=X_4\|X_1=X_2=X_3)\right]P(X_1=X_2=X_3)$	$\frac{1}{n_1n_2n_3n_4}\left[U_{X_3<X_4\|X_1=X_2=X_3}+\frac{1}{2}U_{X_3=X_4\|X_1=X_2=X_3}\right]$
	$\frac{1}{n_1n_2n_3n_4}\left[\sum_l R_l^{X_4\|X_1=X_2=X_3}-\frac{n_4(n_4+1)}{2}\right]$

HUM = hypervolume under the ROC manifold.

are expressed as the following conditional Wilcoxon rank sum statistics,

$$\left[\sum_l R_l^{X_4|X_1<X_2<X_3}-\frac{n_4(n_4+1)}{2}\right],\quad \left[\sum_l R_l^{X_4|X_1=X_2<X_3}-\frac{n_4(n_4+1)}{2}\right],$$
$$\left[\sum_l R_l^{X_4|X_1<X_2=X_3}-\frac{n_4(n_4+1)}{2}\right],\quad \text{and}\quad \left[\sum_l R_l^{X_4|X_1=X_2=X_3}-\frac{n_4(n_4+1)}{2}\right],$$

respectively, the HUM^4 is represented as

$$\frac{1}{n_1n_2n_3n_4}\left\{\left[\sum_l R_l^{X_4|X_1<X_2<X_3}-\frac{n_4(n_4+1)}{2}\right]+\frac{1}{2}\left[\sum_l R_l^{X_4|X_1=X_2<X_3}-\frac{n_4(n_4+1)}{2}\right]\right.$$
$$\left.+\frac{1}{2}\left[\sum_l R_l^{X_4|X_1<X_2=X_3}-\frac{n_4(n_4+1)}{2}\right]+\frac{1}{2^2}\left[\sum_l R_l^{X_4|X_1=X_2=X_3}-\frac{n_4(n_4+1)}{2}\right]\right\}.$$

If there do not exist any sample points satisfying corresponding conditional states of X_1, X_2 and X_3 such as $X_1 < X_2 < X_3$ and $X_1 = X_2 < X_3$, for example, then both $\sum_l R_l^{X_4|X_1<X_2<X_3}$ and $\sum_l R_l^{X_4|X_1=X_2<X_3}$ have zero values, so that $[\sum_l R_l^{X_4|X_1<X_2<X_3} - n_4(n_4+1)/2] = 0$, $[\sum_l R_l^{X_4|X_1=X_2<X_3} - n_4(n_4+1)/2] = 0$. Hence we obtain Theorem 2 with appropriate four indicator functions. □

Note that we may conclude that the four term in the right hand side of (2.3) are represented with either a conditional Mann-Whitney or Wilcoxon rank sum statistic (Table 2). The HUM for more than four dimensions can be extended and represented with both conditional Mann-Whitney and Wilcoxon rank sum statistics with similar arguments to (2.4) and Theorem 2.

3. Hypotheses Test with VUS and HUM

The mean and variance of conditional Wilcoxon rank sum statistic could be derived based on the asymptotic large sample theory of a Wilcoxon rank sum statistic. VUS has two conditional Wilcoxon rank sum statistics of X_3, $\sum_k R_k^{X_3|X_1<X_2}$ and $\sum_k R_k^{X_3|X_1=X_2}$. The Wilcoxon rank sum statistic, $\sum_k R_k^{X_3|X_1<X_2}$,

has the mean $n_3(n_3 + U_{X_1<X_2} + 1)/2$ and variance $n_3 U_{X_1<X_2}(n_3 + U_{X_1<X_2} + 1)/12$, where $U_{X_1<X_2} = \sum_{ij} I(X_{1i} < X_{2j})$ is the sample size of X_2 satisfying states $X_1 < X_2$. The mean and variance of $\sum_k R_k^{X_3|X_1=X_2}$ are obtained similarly.

VUS distribution can then be derived with the properties of two conditional Wilcoxon rank sum statistics in Theorem 1. We can now suggest a hypothesis testing method.

Proposition 1. *Consider the hypotheses*

$$H_0 : F_1(x) = F_2(x) = F_3(x) \quad \textit{versus} \quad H_1 : F_i(x) > F_{i+1}(x), \quad \textit{for at least one } i. \tag{3.1}$$

Under the null hypothesis that all three distribution functions are the same, the p-value for a certain value c of the VUS, $P(VUS \geq c)$, could be defined as

$$P(VUS \geq c) = P\left(\sum_k R_k^{X_3|X_1<X_2} \geq c_1\right) I(A) + \frac{1}{2} P\left(\sum_k R_k^{X_3|X_1=X_2} \geq c_2\right) I(B), \tag{3.2}$$

where $\sum_k R_k^{X_3|X_1<X_2} = c_1$ and $\sum_k R_k^{X_3|X_1=X_2} = c_2$ when the VUS has a value c in Theorem 1.

With the assumption $F_1(x) \geq F_2(x) \geq F_3(x)$, we could conclude that the null hypothesis in (3.1) can be tested with the p-values of (3.2).

Let us extend to the HUM^4 for four dimensions. There are four conditional Wilcoxon rank sum statistics of X_4, $\sum_l R_l^{X_4|X_1<X_2<X_3}$, $\sum_l R_l^{X_4|X_1=X_2<X_3}$, $\sum_l R_l^{X_4|X_1<X_2=X_3}$ and $\sum_l R_l^{X_4|X_1=X_2=X_3}$ for the HUM^4. The Wilcoxon rank sum statistic, $\sum_l R_l^{X_4|X_1<X_2<X_3}$, has the mean $n_4(n_4 + U_{X_1<X_2<X_3} + 1)/2$ and variance $n_4 U_{X_1<X_2<X_3}(n_4 + U_{X_1<X_2<X_3} + 1)/12$. Other Wilcoxon rank sum statistics can easily be obtained their corresponding means and variances. The distribution of HUM^4 can then be derived with the properties of four conditional Wilcoxon rank sum statistics in Theorem 2. Hence, we suggest another hypothesis testing method.

Proposition 2. *For the hypotheses*

$$H_0 : F_1(x) = F_2(x) = F_3(x) = F_4(x) \quad \textit{versus} \quad H_1 : F_i(x) > F_{i+1}(x), \quad \textit{for at least one } i, \tag{3.3}$$

the p-value for a certain value c of the HUM^4, $P(HUM^4 \geq c)$, could be formulated as

$$\begin{aligned} P\left(HUM^4 \geq c\right) = {} & P\left(\sum_l R_l^{X_4|X_1<X_2<X_3} \geq c_1\right) I(A) + \frac{1}{2} P\left(\sum_l R_l^{X_4|X_1=X_2<X_3} \geq c_2\right) I(B) \\ & + \frac{1}{2} P\left(\sum_l R_l^{X_4|X_1<X_2=X_3} \geq c_3\right) I(C) + \frac{1}{4} P\left(\sum_l R_l^{X_4|X_1=X_2=X_3} \geq c_4\right) I(D), \end{aligned} \tag{3.4}$$

where $\sum_l R_l^{X_4|X_1<X_2<X_3} = c_1$, $\sum_l R_l^{X_4|X_1=X_2<X_3} = c_2$, $\sum_l R_l^{X_4|X_1<X_2=X_3} = c_3$ and $\sum_l R_l^{X_4|X_1=X_2=X_3} = c_4$ when the HUM^4 has a value c in Theorem 2.

We can therefore conclude that the null hypothesis in (3.3) can be tested with the p-value of (3.4). These findings about the VUS and HUM^4 are for only three and four distribution functions in this work, but we may extend to more than four distribution functions; therefore, HUM$^k = P(X_1 \leq X_2 \leq \cdots \leq X_k)$ can be represented with conditional Mann-Whitney and conditional Wilcoxon rank sum statistics, and HUM^k could also test the null hypothesis $H_0 : F_1(x) = F_2(x) = \cdots = F_k(x)$ with the asymptotic large sample theory of Wilcoxon rank sum statistics.

Table 3: Three random samples with some tied values

X_1	11	17		23	39	44							$n_1 = 5$
X_2		17	22		39		48	57		72			$n_2 = 6$
X_3					39			57	63		89	94	$n_3 = 5$

Table 4: The second stage data sets from Table 3

2^{nd} stage data - 1				2^{nd} stage data - 2			
$X_1 < X_2$	X_3	$R^{X_2\|X_1<X_2}$	R^{X_3}	$X_1 = X_2$	X_3	$R^{X_2\|X_1=X_2}$	R^{X_3}
(11, 17)		1		(17, 17)		1	
(11, 22)		2.5					
(17, 22)		2.5					
(11, 39)		5.5					
(17, 39)	39	5.5	5.5	(39, 39)	39	2.5	2.5
(23, 39)		5.5					
(11, 48)		10					
(17, 48)		10					
(23, 48)		10					
(39, 48)		10					
(44, 48)		10					
(11, 57)		15.5					
(17, 57)		15.5					
(23, 57)	57	15.5	15.5		57		4
(39, 57)		15.5					
(44, 57)		15.5					
	63		19		63		5
(11, 72)		22					
(17, 72)		22					
(23, 72)		22					
(39, 72)		22					
(44, 72)		22					
	89		25		89		6
	94		26		94		7
$U_{X_1<X_2} = 21$	$n_3 = 5$	260	91	$U_{X_1=X_2} = 2$	$n_3 = 5$	3.5	24.5

4. Some Illustrative Examples

4.1. Example with some tied values for VUS

Consider three random samples in Table 3. There are three tied values $X_1 = X_2 = 17$, $X_2 = X_3 = 57$, and $X_1 = X_2 = X_3 = 39$. The data set of X_1 and X_2 is divided into two data sets $\{(X_1, X_2)|X_1 < X_2)\}$ and $\{(X_1, X_2)|X_1 = X_2)\}$. These two data sets are called second stage data and are similar to data in Table 4. The two tied values $X_2 = X_3 = 57$, 39 are negligible at this moment since these two values will be considered when conditional Mann-Whitney and Wilcoxon rank sum statistics are obtained. Conditional Mann-Whitney statistics can be calculated from Table 4 by comparing X_3 and X_2, where X_2 is in the second stage data sets $\{(X_1, X_2)|X_1 < X_2)\}$ and $\{(X_1, X_2)|X_1 = X_2)\}$.

VUS using Mann-Whitney statistics presents

$$\begin{aligned}\mathrm{VUS}_{MW} &= \frac{1}{n_1 n_2 n_3}\left\{U_{X_2<X_3|X_1<X_2} + \frac{1}{2}U_{X_2<X_3|X_1=X_2} + \frac{1}{2}U_{X_2=X_3|X_1<X_2} + \frac{1}{2^2}U_{X_2=X_3|X_1=X_2}\right\}\\ &= \frac{1}{5\times 6\times 5}\left\{72 + \frac{9}{2} + \frac{8}{2} + \frac{1}{4}\right\} = \frac{80.75}{150} = 0.5383.\end{aligned}$$

Table 5: Three random samples with some tied values

X_1	11	17		23		45											$n_1 = 4$
X_2			22			45		61			77						$n_2 = 4$
X_3					29	45	54			72		83		90			$n_3 = 6$
X_4						45			69				88		95	100	$n_4 = 5$

Now put ranks on each value of X_3 and X_2 in two different data sets $\{(X_1, X_2)|X_1 < X_2)\}$ and $\{(X_1, X_2)|X_1 = X_2)\}$ that are similar to those in Table 4. The conditional Wilcoxon rank sum statistics, R^{X_3}, in Table 4 have 88.5 and 19.5 in the left and right table, respectively. VUS using Wilcoxon rank sum statistics then presents

$$\begin{aligned}\mathrm{VUS}_W &= \frac{1}{n_1 n_2 n_3}\left\{\left(\sum_k R_k^{X_3|X_1<X_2} - \frac{n_3(n_3+1)}{2}\right) I(A) + \frac{1}{2}\left(\sum_k R_k^{X_3|X_1=X_2} - \frac{n_3(n_3+1)}{2}\right) I(B)\right\} \\ &= \frac{1}{5\times 6\times 5}\left\{\left(91 - \frac{5\times 6}{2}\right) + \frac{1}{2}\left(24.5 - \frac{5\times 6}{2}\right)\right\} = 0.5383.\end{aligned}$$

VUS using Mann-Whitney statistics are shown to have the same value as those using Wilcoxon rank sum statistics.

In this Example 4.1, $\sum_k R_k^{X_3|X_1<X_2}$ has the mean $5(5 + 21 + 1)/2 = 67.5$ and variance $(5 \times 21)(5 + 21 + 1)/12 = 236.25$ with $n_3 = 5$ and $U_{X_1<X_2} = 21$. The mean and variance of $\sum_k R_k^{X_3|X_1=X_2}$ are 20 and 6.6667, respectively with $U_{X_1=X_2} = 2$. We have $\sum_k R_k^{X_3|X_1<X_2} = 91$ and $\sum_k R_k^{X_3|X_1=X_2} = 24.5$; therefore, the corresponding p-value of the VUS, $P(\mathrm{VUS} \geq 0.5383)$, can be obtained

$$p\text{-value} = \left[1 - \Phi\left(\frac{91 - 67.5}{\sqrt{236.25}}\right)\right] + \frac{1}{2}\left[1 - \Phi\left(\frac{24.5 - 20}{\sqrt{6.6667}}\right)\right] = 0.0835.$$

The null hypothesis in (3.1) cannot be rejected with the level of significance $\alpha = 0.05$ since its p-value is not small.

4.2. Example with one tied value for HUM^4

Consider four other random samples in Table 5. There is one tied value $X_1 = X_2 = X_3 = X_4 = 45$. Even though there is one tied value among X_1, X_2 and X_3, we may consider four different second stage data sets: $\{(X_1, X_2, X_3)|X_1 < X_2 < X_3)\}$, $\{(X_1, X_2, X_3)|X_1 = X_2 < X_3)\}$, $\{(X_1, X_2, X_3)|X_1 < X_2 = X_3)\}$ and $\{(X_1, X_2, X_3)|X_1 = X_2 = X_3)\}$ in Table 6. The two values of the HUM^4 using Mann-Whitney and Wilcoxon rank sum statistics are then shown to be the same.

$$\begin{aligned}\mathrm{HUM}^4_{MW} &= \frac{1}{n_1 n_2 n_3 n_4}\left\{\left(U_{X_3<X_4|X_1<X_2<X_3} + \frac{1}{2}U_{X_3=X_4|X_1<X_2<X_3}\right) + \frac{1}{2}\left(U_{X_3<X_4|X_1=X_2<X_3} + \frac{1}{2}U_{X_3=X_4|X_1=X_2<X_3}\right)\right. \\ &\quad \left. + \frac{1}{2}\left(U_{X_3<X_4|X_1<X_2=X_3} + \frac{1}{2}U_{X_3=X_4|X_1<X_2=X_3}\right) + \frac{1}{2^2}\left(U_{X_3<X_4|X_1=X_2=X_3} + \frac{1}{2}U_{X_3=X_4|X_1=X_2=X_3}\right)\right\} \\ &= \frac{1}{4\times 4\times 6\times 5}\left\{\left(130 + \frac{2}{2}\right) + \frac{1}{2}\left(12 + \frac{0}{2}\right) + \frac{1}{2}\left(12 + \frac{3}{2}\right) + \frac{1}{4}\left(4 + \frac{1}{2}\right)\right\} = 0.3018,\end{aligned}$$

Table 6: Second stage data sets from Table 5

2^{nd} stage data - 1			
$X_1 < X_2 < X_3$	X_4	$R^{X_3 \mid X_1<X_2<X_3}$	R^{X_4}
(11, 22, 29)		1.5	
(17, 22, 29)		1.5	
(11, 22, 45)	45	4	4
(17, 22, 45)		4	
(11, 22, 54)		8	
(17, 22, 54)		8	
(11, 45, 54)		8	
(17, 45, 54)		8	
(23, 45, 54)		8	
	69		11
(11, 22, 72)		16	
(17, 22, 72)		16	
(11, 45, 72)		16	
(17, 45, 72)		16	
(23, 45, 72)		16	
(11, 61, 72)		16	
(17, 61, 72)		16	
(23, 61, 72)		16	
(45, 61, 72)		16	
(11, 22, 83)		27	
(17, 22, 83)		27	
(11, 45, 83)		27	
(17, 45, 83)		27	
(23, 45, 83)		27	
(11, 61, 83)		27	
(17, 61, 83)		27	
(23, 61, 83)		27	
(45, 61, 83)		27	
(11, 77, 83)		27	
(17, 77, 83)		27	
(23, 77, 83)		27	
(45, 77, 83)		27	
	88		34
(11, 22, 90)		41	
(17, 22, 90)		41	
(11, 45, 90)		41	
(17, 45, 90)		41	
(23, 45, 90)		41	
(11, 61, 90)		41	
(17, 61, 90)		41	
(23, 61, 90)		41	
(45, 61, 90)		41	
(11, 77, 90)		41	
(17, 77, 90)		41	
(23, 77, 90)		41	
(45, 77, 90)		41	
	95		48
	100		49
$U_{X_1<X_2<X_3} = 44$	$n_4 = 5$	1079	146

2^{nd} stage data - 2			
$X_1 < X_2 = X_3$	X_4	$R^{X_3 \mid X_1<X_2=X_3}$	R^{X_4}
(11, 45, 45)		2.5	
(17, 45, 45)	45	2.5	2.5
(23, 45, 45)		2.5	
	69		5
	88		6
	95		7
	100		8
$U_{X_1<X_2=X_3} = 3$	$n_4 = 5$	7.5	28.5

2^{nd} stage data - 3			
$X_1 = X_2 < X_3$	X_4	$R^{X_3 \mid X_1=X_2<X_3}$	R^{X_4}
	45		1
(45, 45, 54)		2	
	69		3
(45, 45, 72)		4	
(45, 45, 83)		5	
	88		6
(45, 45, 72)		7	
	95		8
	100		9
$U_{X_1=X_2<X_3} = 4$	$n_4 = 5$	18	27

2^{nd} stage data - 4			
$X_1 = X_2 = X_3$	X_4	$R^{X_3 \mid X_1=X_2=X_3}$	R^{X_4}
(45, 45, 45)	45	1.5	1.5
	69		3
	88		4
	95		5
	100		6
$U_{X_1=X_2=X_3} = 1$	$n_4 = 5$	1.5	19.5

and

$$
\begin{aligned}
\mathrm{HUM}_W^4 &= \frac{1}{n_1 n_2 n_3 n_4}\left\{\left[\sum_l R_l^{X_4|X_1<X_2<X_3} - \frac{n_4(n_4+1)}{2}\right] I(A) + \frac{1}{2}\left[\sum_l R_l^{X_4|X_1=X_2<X_3} - \frac{n_4(n_4+1)}{2}\right] I(B)\right. \\
&\quad \left. + \frac{1}{2}\left[\sum_l R_l^{X_4|X_1<X_2=X_3} - \frac{n_4(n_4+1)}{2}\right] I(C) + \frac{1}{2^2}\left[\sum_l R_l^{X_4|X_1=X_2=X_3} - \frac{n_4(n_4+1)}{2}\right] I(D)\right\} \\
&= \frac{1}{4\times 4\times 6\times 5}\left\{\left(146 - \frac{5\times 6}{2}\right) + \frac{1}{2}\left(27 - \frac{5\times 6}{2}\right) + \frac{1}{2}\left(28.5 - \frac{5\times 6}{2}\right) + \frac{1}{4}\left(19.5 - \frac{5\times 6}{2}\right)\right\} \\
&= 0.3018.
\end{aligned}
$$

With $n_4 = 5$ and $U_{X_1<X_2<X_3} = 44$, $\sum_l R_l^{X_4|X_1<X_2<X_3}$ has the mean $5(5+44+1)/2 = 125$ and variance $(5\times 44)(5+44+1)/12 = 916.6667$. Since $U_{X_1=X_2<X_3} = 4$, $U_{X_1<X_2=X_3} = 3$, $\sum_l R_l^{X_4|X_1=X_2<X_3}$ and $\sum_l R_l^{X_4|X_1<X_2=X_3}$ have the mean 25, 22.5 and the variance 16.67, 11.25. The mean and variance of $\sum_l R_l^{X_4|X_1=X_2=X_3}$ are 17.5 and 2.9167, respectively with $U_{X_1=X_2=X_3} = 1$. We have $\sum_l R_l^{X_4|X_1<X_2<X_3} = 146$, $\sum_l R_l^{X_4|X_1=X_2<X_3} = 27$, $\sum_l R_l^{X_4|X_1<X_2=X_3} = 28.5$ and $\sum_l R_l^{X_4|X_1=X_2=X_3} = 19.5$; therefore, then the corresponding p-value of the HUM^4, $P(\mathrm{HUM}^4 \geq 0.3018)$ can be obtained

$$
\begin{aligned}
p\text{-value} &= \left[1-\Phi\left(\frac{146-125}{\sqrt{916.6667}}\right)\right] + \frac{1}{2}\left[1-\Phi\left(\frac{27-25}{\sqrt{16.67}}\right)\right] + \frac{1}{2}\left[1-\Phi\left(\frac{28.5-22.5}{\sqrt{11.25}}\right)\right] + \frac{1}{4}\left[1-\Phi\left(\frac{19.5-17.5}{\sqrt{2.9167}}\right)\right] \\
&= (0.2440) + \left(\frac{0.3121}{2}\right) + \left(\frac{0.0368}{2}\right) + \left(\frac{0.1208}{4}\right) = 0.4486.
\end{aligned}
$$

The null hypothesis in (3.3) cannot be rejected since its p-value is too big.

5. Conclusion

The AUC is represented in the ROC curve as Mann-Whitney statistics as well as a Wilcoxon rank sum statistic. In this paper, we extend this work to the ROC surface and manifold, so that we may conclude that the VUS and HUM are represented with conditional Mann-Whitney statistics as well as conditional Wilcoxon rank sum statistics. VUS and HUM^4 obtained by using the Mann-Whitney statistics for three and four random samples are the same as those obtained from Wilcoxon rank sum statistics.

The asymptotic large sample theory of Wilcoxon rank sum statistic allows us to find the asymptotic distribution of the conditional Wilcoxon rank sum statistic. Hence the distribution functions of the VUS and HUM for more than three dimensions could also derived with the conditional Wilcoxon rank sum statistics proposed in this paper. The corresponding p-value can be obtained with VUS and HUM distribution function when the VUS and HUM have a certain value. Therefore, the null hypothesis that all $k(\geq 3)$ distribution functions are identical, $H_0 : F_1(x) = F_2(x) = \cdots = F_k(x)$, could be tested with VUS and HUM that use the asymptotic large sample theory of Wilcoxon rank sum statistics.

References

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, *Journal of Mathematical Psychology*, **12**, 387–415.

Conover, W. J. (1980). *Practical Nonparametric Statistics*, John Wiley & Sons, New York.

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*, Academic Press, New York.

Engelmann, B., Hayden, E. and Tasche, D. (2003). Testing rating accuracy, *Risk*, **16**, 82–86.

Faraggi, D. and Reiser, B. (2002). Estimation of the area under the ROC curve, *Statistics in Medicine*, **21**, 3093–3106.

Fawcett, T. (2003). ROC graphs: Notes and practical considerations for data mining researchers, *HP Labs Tech Report HPL-2003-4*, Available from: http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf

Gibbons, J. D. (1971). *Nonparametric Statistical Inference*, McGraw-Hill, New York.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143**, 29–36.

Hong, C. S. (2009). Optimal threshold from ROC and CAP curves, *Communications in Statistics - Simulation and Computation*, **38**, 2060–2072.

Hong, C. S. and Cho, M. H. (2015). VUS and HUM represented with Mann-Whitney statistic, *Communications for Statistical Applications and Methods*, **22**, 223–232.

Hong, C. S., Joo, J. S. and Choi, J. S. (2010). Optimal thresholds from mixture distributions, *The Korean Journal of Applied Statistics*, **23**, 13–28.

Hong, C. S., Jung, E. S. and Jung, D. G. (2013). Standard criterion of VUS for ROC surface, *The Korean Journal of Applied Statistics*, **26**, 977–985.

Hong, C. S. and Jung, D. G. (2014). Standard criterion of hypervolume under the ROC manifold, *Journal of the Korean Data & Information Science Society*, **25**, 473–483.

Joseph, M. P. (2005). A PD validation framework for Basel II internal ratings-based systems, Available from: http://www.bu siness-school.ed.ac.uk/waf/crc_ archive/2005/papers/joseph-maurice.pdf

Mann, H. B. and Whitney, D. R. (1947). On a test whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics*, **18**, 50–60.

Provost, F. and Fawcett, T. (2001). Robust classification for imprecise environments, *Machine Learning*, **42**, 203–231.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*, John Wiley & Sons, New York.

Rosset, S. (2004). Model selection via the AUC, In *Proceedings of the 21st International Conference of Machine Learning*, Banff, Canada.

Sobehart, J. R. and Keenan, S. C. (2001). Measuring default accurately, *Risk: Credit Risk Special Report*, **14**, S31–S33.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems, *Science*, **240**, 1285–1293.

Swets, J. A., Dawes, R. M. and Monahan, J. (2000). Better decisions through science, *Scientific American*, **283**, 82–87.

Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin*, **1**, 80–83.

Wilkie, A. D. (2004). Measures for comparing scoring systems. In L. C. Thomas, D. B. Edelman, and J. N. Crook (Eds.), *Readings in Credit Scoring*, Oxford University Press, Oxford, 51–62.

Zou, K. H., O′Malley, A. J. and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models, *Circulation*, **115**, 654–657.

Received June 15, 2015; Revised July 13, 2015; Accepted July 13, 2015