**A new method to determine separated hyper-plane for non-parametric sign test in multivariate data**

Loc Nguyen

Vietnam Institute of Mathematics, Vietnam

Email: ng\_phloc@yahoo.com

Homepage: www.locnguyen.net

**Abstract**

Non-parametric testing is very necessary in case that the statistical sample does not conform normal distribution or we have no knowledge about sample distribution. Sign test is a popular and effective test for non-parametric model but it cannot be applied into multivariate data in which observations are vectors because the ordering and comparative operators are not defined in *n*-dimension vector space. So, this research proposes a new approach to perform sign test on multivariate sample by using a hyper-plane to separate multi-dimensional observations into two sides. Therefore, it is possible for the sign test to assign plus signs and minus signs to observations in each side. Moreover, this research introduces a new method to determine the separated hyper-plane. This method is a variant of support vector machine (SVM), thus, the optimized hyper-plane is the one that contains null hypothesis and splits observations as discriminatively as possible.

**Keywords:** separated hyper-plane, non-parametric sign test.

**1. Introduction to non-parametric sign test**

Nonparametric testing is used in case of without knowledge about sample distribution; concretely, there is no assumption of normality. The nonparametric testing begins with the test on sample *median* in univariate data. If distribution is symmetric, median is identical to mean. Given the median is the observation at which the left side data and the right side data are of equal accumulate probability.

*P*(*D <* ) *= P*(*D >* ) *=* 0.5

If data is not large and there is no assumption about normality, the median is approximate to population mean. Given null hypothesis *H*0: = and alternative hypothesis *H*1: ≠ , the so-called sign test (Walpole, Myers, Myers, & Ye, 2012, pp. 656-660) is performed as below steps:

* Assigning plus signs to sample observations whose values are greater than and minus signs to ones whose values are less than . Note that values which equal are not considered. Plus signs and minus signs represent the right side and left side of , respectively.
* If the number of plus signs is nearly equal to the number of minus signs, then null hypothesis *H*0 is true; otherwise *H*0 is false. In other words, that the proportion of plus signs is significantly different from 0.5 cause to rejecting *H*0 in flavor of *H*1.

The reason of *H*0 acceptance is that the probability that observations fall in both left side and right side of are of equal value 0.5 and of course, it is asserted that is a real median. Note that terms *data point*, *sample point*, *sample value* and *observation* are identical.

In the case that alternative hypothesis *H*1: < , if the proportion of plus signs is less than 0.5 then rejecting *H*0 in flavor of *H*1. In the case that alternative hypothesis *H*1: > , if the proportion of plus signs is greater than 0.5 then rejecting *H*0 in flavor of *H*1. Now let *X* be the discrete random variable representing the number of plus signs and suppose that *X* conforms binomial distribution *B*(*X*; *n*; *p*) where *n* and *p* are the total number of sample data points and the probability that plus sign is assigned to a data point, respectively. Because the proportion of plus signs gets 0.5 when *H*0: = is true, the parameter *p* is set to be 0.5. Given the distribution of plus signs is *B*(*X*; *n*; 0.5) and significant level *α* and let *x* be the instance of *X*, there are three following tests (Walpole, Myers, Myers, & Ye, 2012, pp. 657-660):

* *H*0: = and *H*1: ≠ : In case of *x < n*/2, if 2*P*(*X ≤ x*) *< α* then rejecting *H*0. In case of *x > n*/2, if 2*P*(*X ≥ x*) *< α* then rejecting *H*0. This test belongs to two-sided test family.
* *H*0: = and *H*1: < : if *P*(*X ≤ x*) *< α* then rejecting *H*0. This test belongs to one-sided test family.
* *H*0: = and *H*1: > : if *P*(*X ≥ x*) *< α* then rejecting *H*0. This test belongs to one-sided test family.

Note that *P*(…) is accumulated probability of binomial distribution *B*(*X*; *n*; 0.5), for example, *P*(*X ≤ x*) *=* . In case that *n* is large enough, for instance *n >* 10, *B*(*X*; *n*; 0.5) is approximate to standard normal distribution *N*(*Z*; 0; 1) where *Z* = . Let *z* be the instance of *Z*, there are three following tests:

* *H*0: = and *H*1: ≠ : if |*z*| > *zα*/2 then rejecting *H*0 where *zα*/2 is 100*α*/2 percentage point of standard normal distribution.
* *H*0: = and *H*1: < : if *z <* –*zα* then rejecting *H*0 where *zα* is 100*α* percentage point of standard normal distribution.
* *H*0: = and *H*1: > : if *z > zα* then rejecting *H*0.

In case of pair-test *H*0: – = *d*0 which we need to know how much median shifts from other one , sign test is applied in similar way with a little bit of change. If *d*0 = 0, *H*0 indicates whether equals . We compute all deviations between two samples *X* and *Y* where is sample median of *X* and is sample median of *Y*. Let *di = xi – yi* be the deviation between *x* ∈ *Y* and *y* ∈ *Y*. Plus signs (minus signs) are assigned to *di* (s) which are greater (less) than *d*0. Now signed test is applied into such plus signs and minus signs by discussed method.

**2. A proposal of non-parametric sign test in multivariate data**

If sample is extended from real number space to *n*-dimension space, observations become vectors composed of *n* partial values. Traditional sign test cannot be applied into multivariate data because null hypothesis and alternative hypothesis are vectors and it is not likely to compare hypothesis vector with observation vector with regard to assign plus signs. This paper proposes a new approach for sign test in case of *n*-dimension sample. The basic idea is to determine *a hyper-plane and its normal vector which is used to calculate signed-rank number*, namely, an observation is assigned plus sign if it is in the right side of hyper-plane. The right side, in turn, is defined by the direction of normal vector of hyper-plane. Because the hyper-plane is very important, research also proposes a method to determine it with proposition “the optimized hyper-plane is the one that contains null hypothesis and splits observations as discriminatively as possible”.

Suppose distribution is symmetric, median is identical to mean, we have null hypothesis *H*0: *μ* = *μ*0 and alternative hypothesis *H*1: *μ ≠ μ*0, the sign test in case of multivariate sample includes three following steps:

1. Determining separated hyper-plane that contains null hypothesis *μ*0 and splits observations as discriminatively as possible. The method to discover this hyper-plane is described particularly in next section. Another important thing is to specify the normal vector *W* of such hyper-plane. The normal vector *W* is used to make sign of observations.
2. Let Ω be n-dimension multivariate sample and let *Xi* = (*xi*1, *xi*2,…, *xin*) ∈ Ω be observation vector. The scalar product *WT*(*Xi – μ*0) between normal vector *W* and each standardized observation *Xi – μ*0 is calculated, where *T* denotes transposition operator and *W* and *Xi* are column vectors. If such scalar product is positive then *Xi* is assigned a plus sign; otherwise *Xi* is assigned a minus sign.
3. Let *b* is the number of plus signs, thus, *b* conform binomial distribution with proportion parameter 0.5. Let *z* be the *z*-score of *b*, we have *z* = where *m* is the number of observations in sample, *m* = |Ω|. If |*z*| *> zα*/2 then rejecting *H*0 in flavor of *H*1 where *zα*/2 is 100*α*/2 percentage point of standard normal distribution. It is easy to recognize that *z* is normal approximation of *b*.

The ideology behind this method is that if vector *μ*0 is the mean of symmetric multivariate distribution, then it exists a separated hyper-plane that divides the whole sample space into two half-spaces so that the cardinalities of these half-spaces are equal (or approximately equal). It is possible to exist many hyper-planes like that but the optimal hyper-plane is found out by the method described particularly in next section.

**3. A new method to determine separated hyper-plane**

Let Ω be *n-*dimension multivariate sample and let *Xi* = (*xi*1, *xi*2,…, *xin*) Ω be observation vector. Suppose Ω = {*X*1, *X*2,…, *Xm*} has *m* observations. The method is to find out an optimal separated hyper-plane is a variant of support vector machine (SVM) (Law, 2006). Hence, this hyper-plane satisfies two following propositions:

1. Containing specified null hypothesis *μ*0.
2. *S*plitting vector space Ω into two half-spaces as separately as possible.

The first proposition is by following equation that is the equation of hyper-plane.

*WT*(*Xi – μ*0) *=* 0

Where *W* is the normal vector of hyper-plane and *T* denotes transposition operator. The task of determining optimal hyper-plane is identical to find out its normal vector *W.* The second proposition is equivalent to maximizing the distance between two margins of two half-spaces. This proposition is similar to the methodology of support vector machine (SVM) method. These margins between two half-spaces are restricted by two parallel hyper-planes that are specified by following equations:

Two half-spaces are separated as much as possible if and only if following condition is satisfied:

Because the distance between two margins is , the optimal hyper-plane will maximize . It means that the optimal hyper-plane minimizes with constraint. According to Lagrange dual theorem (Boyd & Vandenberghe, 2009, p. 215), *W* is extreme point of following Lagrange function:

|  |  |
| --- | --- |
|  | (1) |

Where *λ =* (*λ*1, *λ*2,…, *λm*)*T* represents a set of Lagrange multipliers and . Suppose *W* minimizes then it is the solution of following equation when we set the derivative of *L*(*W, α*) with regard to *W* to 0.

|  |  |
| --- | --- |
|  | (2) |

Substituting equation 2 into equation 1, we have:

In case of , due to and , we have:

In case of , due to and , we have:

Due to

In any case we have:

It is easy to infer that

So *L*1(*λ*) and *L*2(*λ*) are inferior and supreme of *L*(*W*, *λ*), respectively.

*L*1(*λ*) *≤ L*(*W*, *λ*) *≤ L*2(*λ*)

Hence, *L*1(*λ*) and *L*2(*λ*) which are functions of Lagrange multipliers are re-written into matrix notation, according to equation 3:

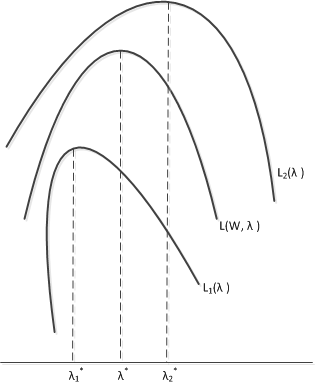
|  |  |
| --- | --- |
|  | (3) |

Where *I* = (1, 1,…, 1)*T* is one-column identity matrix and *S* is a symmetric *m*x*m* matrix with elements *sij =* . Minimizing is identical to maximizing *L*(*W*, *λ*) with regard to *λ*. Note that maximizing *L*(*W*, *λ*) is equivalent to maximizing *L*1(*λ*) and *L*2(*λ*) because *L*1(*λ*) and *L*2(*λ*) are inferior and supreme of *L*(*W, λ*), respectively. Suppose *λ*1*\** and *λ*2*\** is maximum points of *L*1(*λ*) and *L*2(*λ*), respectively, we have equation 4 for specifying them:

|  |  |
| --- | --- |
|  | (4) |

*L*1(*λ*) and *L*2(*λ*) are second-order functions and so the way to find *λ*1*\** and *λ*2*\** is quadratic programming (QP) problem. There are many approaches to solve this problem; for instance, the sequential minimal optimization is described in (Law, 2006, p. 15) and (Platt, 1998, pp. 8-9).

Although *λ*1*\** and *λ*2*\** is maximum points of *L*1(*λ*) and *L*2(*λ*), they are not asserted to be maximum points of *L*(*W*, *λ*) with regard to *λ*. Let *λ\** be the maximum point of *L*(*W*, *λ*) with regard to *λ*, figure 1 is the interpretation of *λ*1*\**, *λ*2*\**, *λ\**, *L*1(*λ*), *L*2(*λ*) and *L*(*W*, *λ*).



**Figure 1.** The interpretation of *λ*1*\**, *λ*2*\**, *λ\**, *L*1(*λ*), *L*2(*λ*) and *L*(*W*, *λ*)

So the maximum point *λ\** is approximated by the average of *λ*1*\** and *λ*2*\** according to equation 5.

|  |  |
| --- | --- |
|  | (5) |

Substituting maximum point *λ\** to equation 2, we have following equation:

|  |  |
| --- | --- |
|  | (6) |

Because equation 6 is the quadratic function, it is easy to find out its solutions. There are some methods to solve quadratic equation such as Newton-Raphson and bisection (Burden & Faires, 2011, pp. 48-74). If equation 6 has only one solution *W\**, then *W\** is normal vector of separated hyper-plane. Otherwise equation 6 has *k* > 1 solutions *W*1*\**, *W*2*\**,…, *Wk\**, then the normal vector *W\** of separated hyper-plane is the addition of these solutions.

So the best normal vector *W\** is considered as the intermediate one among possible normal vector (*W*1*\**, *W*2*\**,…, *Wk\**). As aforementioned, it is easy to determine the separated hyperplane with regard to *W\** according to equation 7.

|  |  |
| --- | --- |
|  | (7) |

Where *μ*0 is the null mean (median).

**4. A case study of non-parametric sign test based on separated hyperplane**

Given a sample including two data points *X*1={20, 55}, *X*2={35, 10} and a null median , it is necessary to perform a sign test with null hypothesis *H*0: = in flavor of alternative hypothesis *H*1: ≠ . Figure 2 shows points *X*1, *X*2 and median (Johansen, 2012).

A picture containing chart

Description automatically generated

**Figure 2.** Data sample and null median

According to equation 3, inferior and supreme of *L*(*W*, *λ*) are:

Figure 3 shows 3-dimension graphs of *L*1(*λ*) and *L*2(*λ*).

Chart, surface chart

Description automatically generated

**Figure 3.** Graphs of *L*1(*λ*) and *L*2(*λ*)

The maximum points of *L*1(*λ*) and *L*2(*λ*) are found by setting their partial derivatives to be zero. We have:

So *L*1(*λ*) gets maximal at . Similarly, we have:

So *L*2(*λ*) gets maximal at . According to equation 5, the maximum point *λ\** is the average as follows:

By substituting *λ\** into equation 6 where *W*=(*w*1, *w*2), we have (Wolfram):

The normal vector is solution of the above equation. The separated hyperplane is specified based on *W*\* according to equation 7 as follows:

Figure 4 shows the separated hyperplane *y* = 27 + 0.2*x* with regard to two variables.

A picture containing diagram

Description automatically generated

**Figure 4.** Separated hyperplane with regard to two variables

Due to

There are 1 positive sign and 1 negative sign. According to sign test, that the number of plus signs is equal to the number of minus signs leads to fail to reject the null hypothesis *H*0: *μ*=(15,30).

Moreover, if we substitute into equation 1, the Lagrange function *L*(*λ*) is the same to supreme *L*2(*λ*) as follows:

Figure 5 shows 3-dimension graph of *L*(*λ*).

Chart, surface chart

Description automatically generated

**Figure 5.** Graphs of *L*(*λ*)

**5. Conclusions**

The main idea of this paper is to find out a separated hyper-plane, which aims to count the number of observations which fall in left side or right side of such hyper-plane. The right side of hyper-plane is defined as the side that its direction is the same to normal vector of hyper-plane. The opposite to the right side is the left side. The method to find out hyper-plane is a variant of support vector machine (SVM) (Law, 2006) method except that hyper-plane contains the null hypothesis. SVM assumes that data points are labeled with classes {1 and –1} but observations in case of hypothesis testing are not classed. This research solves this problem by two-step process of separation:

* Firstly, the label of observations is specified indirectly via the scalar product between these observations and the normal vector of separated hyper-plane.
* Secondly, Lagrange multipliers are determined via arithmetic transformations when the goal of such transformations is to eliminate normal-vector-dependency from Lagrange function and find out the inferior and supreme of Lagrange function. After that the best-separated hyper-plane is the intermediate one among many possible hyper-planes.

**Acknowledgment**

This research is the place to acknowledge professor Ho, Phuc D. – Vietnam Institute of Mathematics who gave me valuable comments and advices. These comments help me to improve this research.

**References**

Boyd, S., & Vandenberghe, L. (2009). *Convex Optimization.* New York, NY, USA: Cambridge University Press.

Burden, R. L., & Faires, D. J. (2011). *Numerical Analysis* (9th Edition ed.). (M. Julet, Ed.) Brooks/Cole Cengage Learning.

Johansen, I. (2012, December 29). Graph software. *Graph version 4.4.2 build 543(4.4.2 build 543)*. GNU General Public License.

Law, M. (2006). *A Simple Introduction to Support Vector Machines.* Lecture Notes for CSE 802 course, Michigan State University, USA, Department of Computer Science and Engineering.

Platt, J. C. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.* Microsoft Research.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probability & Statistics for Engineers & Scientists* (9th ed.). (D. Lynch, Ed.) Boston, Massachusetts, USA: Pearson Education, Inc.

Wolfram. (n.d.). Mathematica. *Wolfram Mathematica, 10(4)*. Wolfram Research. Retrieved March 2016, from http://www.wolfram.com/mathematica